Redesign of the didactics of S(E)IR(D) → SI(EY)A(CD) models of infectious epidemics

Reference to exit strategies for the economic crisis. A nasty virus, asking for eradication. Could the virus be used as its own vaccine for some ? How immune will one be ? Didactics on modeling of the SARS-CoV-2 (Covid-19) pandemic of 2020, with the example case of Holland. Prevent or prepare for another pandemic, and not only for a second wave perhaps already this Fall or Winter.

Thomas Colignatus

June 15 & August 29 2020

https://zenodo.org/communities/re-engineering-math-ed/about/

Abstract

This notebook and package redesign the didactics of the classic SIR, SIRD and SEIR(D) models, using the example of the SARS-CoV-2 pandemic. (Issue 1) A first step is to relabel to SI(EY)A(CD). The acquitted $A = C + D$ are the cleared or deceased. This avoids the triple use of R for removed, recovered, and reproductive number. The infected $I = E + Y$ are the exposed and infectious. In SIA(CD) we have $A' = \gamma\, I$, with γ the acquittal rate from infectiousness, with I' incidence, I prevalence and A cumulated prevalence. The format of ordinary differential equations (ODE) should not distract. The basic structure is given by the Euler-Lotka renewal equation. The deceased are a fraction of the acquitted, $D = \varphi\, A$, with φ the infection fatality factor (IFF). The ODE format $D' = \mu\, I$ or $D' = \mu\, Y$ can be rejected since it turns the model into a course in differential equations, with the need to prove $D = \varphi\, A$ which can already be stated from the start. The ODE format also causes distracting questions what μ might be and whether there is a difference between a lethal acquittal period and a clearing acquittal period, and how parameters values must be adapted when the acquittal rate γ changes. (Issue 2) The term "herd resistance" is much more preferred above the term "herd immunity". The notion means that the herd survives and doesn't become extinct, even when the infection is lethal to perhaps most members. Policy makers and the general public associate "herd immunity" with protection for *all* members, which is not what the notion means (and which special outcome may only occur with vaccination in a well-mixed population before the onset of infection). This source of confusion between experts and non-experts can be prevented by using the term "herd resistance".

The common formula on herd resistance 1 - 1 / R_0 uses (i) a steady state or (ii) vaccination before the start of an epidemic. (ad i) SI(EY)A(CD) has only an *asymptotic* steady state, for which the formula does not apply. (ad ii) Vaccination during an ongoing epidemic means that there are still large numbers of (asymptomatic) infected persons, so that the infection continues, called "overshoot", which does not give the protection that is suggested by the term "herd immunity": there is only a herd resistance. For SI(EY)A(CD), a notion of "near herd immunity" might be 95% of the limit values. For SARS-CoV-2, RIVM (the Dutch CDC) has mentioned R_0 = 2.5 and "herd immunity" (i.e. herd resistance) of 60%, presumably using another type of model with a proper steady state. In SI(EY)A(CD) an infection with R_0 = 2.5 proceeds after 60% till the limit value of 89.3%, which, with IFF = 1.5%, would mean another 78,000 deceased in Holland, compared to 9,000 at the end of May. (Issue 3) The objective of Public Health is to balance medical and economic issues. A better understanding of the SI(EY)A(CD) family of models helps to gauge exit strategies for the pandemic and its economic crisis. A possible strategy is to eradicate the virus. With *test, test, and test* it would be possible to put positively tested persons in quarantine till they have cleared. Another possible strategy is that the vulnerable (elderly and comorbid younger) are put into quarantine while the less vulnerable are infected (in cohorts dictated by ICU capacity), effectively using the virus as its own vaccine, for a period of 12-16 months until there is a proper vaccine for the vulnerable compartment of society. If this is rejected because the virus is so nasty then this is an argument for eradication. It is remarkable that these scenario's are so little discussed in policy making circles, where there seems to be a preference for a lock-on - lock-off approach, that is risky and prolongs the economic crisis. The SI(EY)A(CD) model uses lives saved (lives extended) but other life gain measures are the (quality adjusted) life-years gained, fair innings, proportional shortfall and UnitSqrt. A life table computation that assumes annual loss of immunity shows that a 10% rise in annual cumulated prevalence implies about a 0.5 year drop in life expectancy. The effect is relatively small since the fatalities are mostly in the higher age groups. If cumulated prevalence would be 60% as RIVM suggested, then life expectancy would reduce with 3 years, but the overshoot to 90% would reduce life expectancy by some 4.5 years. The annual death toll is an acceleration of the mortality by comorbidity. (Issue 4) Public Health and epidemiology exist for longer than a century and there have been many warnings about the risk of pandemics. Lessons learnt at the level of cities and nations are now learnt at world level. There is something fundamentally wrong in the relation between society in general and science & learning. For the democratic setup of each nation it is advisable to have both an *Economic Supreme Court* and a *National Assembly of Science and Learning*. We want to save lives and livelihoods but let us not forget fundamental insights about democracy and science & learning.

Keywords and codes

Journal of Economic Literature (JEL) codes: P16 (political economy), E66 (macro-economics, general outlook), I18 (public health), Q56 (sustainability), D60 (welfare economics), A12 (general economics and teaching, relation to other disciplines)

Medical Subject Headings (MeSH) (IDs): D011634 (Public health), C000656484 (COVID-19), D004813 (Epidemiology), D000066949 (Biobehavioral sciences), D015233 (Models, Statistical), D000075082 (Proof of Concept Didactics), D017145 (Models, Educational), D004778 (Environment and public

health), D000072440 (Public health systems research), D040381 (Education, public health professional), D006285 (Health planning), D004989 (ethics)

MSC2010: 92D30 (Epidemiology), 92C60 (Medical epidemiology), 91B64 (Macroeconomic theory), 91B76 (Environmental economics), 00-01 (Introductory exposition)

Disclaimer and declaration of interest

Thomas Colignatus is the scientific name of Thomas Cool, econometrician (Groningen 1982) and teacher of mathematics (Leiden 2008), in Scheveningen, The Netherlands. He worked at Erasmus Medical Center (Rotterdam 2002-2004) on (Markov) modeling for screening on the Human Papilloma Virus (HPV) concerning cervical cancer. The article has used software of *The Economics Pack. Applications of Mathematica*, which has been developed by the author since 1993, which is proprietary software, available at a price, and not falling under the CC license for the PDF. The targeted publication venue is not a journal but a book publisher.

PM. This 2nd version includes vaccination, life gain measures and life table

Table of contents

Table of contents - detailed

8. SI(EY)A(CD)

1. Introduction

1.1. What this notebook and package do

1.1.1. Didactic redesign

This notebook and package redesign didactics for the S(E)IR(D) → SI(EY)A(CD) epidemic models.

- There is no new finding on content. There is no new analysis. There are no new parameter estimates.

- There is only an effort at *better didactics* of what already is known in the field.

- Readers of this notebook may be researchers from other fields than epidemiology or they may be first-year students from a general background. Readers will have at least a first-year student understanding of mathematics and they will be open to the use of *Mathematica.* This readership will *insist* upon the use of mathematics.

A readership with such competence should have little problem with the standard didactics of the S(E)IR(D) → SI(EY)A(CD) models. Thus, an effort at improving these didactics can only have a very marginal effect. However, advancement comes in small steps. The pandemic will cause an interest by new students and researchers from other areas than epidemiology itself. This new readership can be served by clearer exposition, with the benefit of computer algebra and a computable environment. One hope is that this notebook also reaches conventional teachers in S(E)IR(D) → SI(EY)A(CD) modeling: they may not adopt all suggestions but can find some alerts.

PM. A new element may be the distinction in quarantine classes with a colouring scheme in Chapter 4 that could contribute to better communication.

1.1.2. There already is an abundance

There exists an abundance in (empirical) modeling, and it is useful to mention some such modeling, and to again emphasize that there is no contribution on content here.

- The general modeling context is given by the *Global Burden of Disease* and in particular the *Dismod II* framework, see Barendregt et al. (2003), with incidence, prevalence, case fatality and recovery (remission). The latter modeling used Excel while a later development is *Dismod_mr* in Python, see Vos et al. (2012) and Flaxman & Collins (2019). For us, the duration of infection is so short that we neglect background mortality. The Dismod framework has a constant transmission hazard (rate of change) from the susceptibles to the incidence of disease (in variables given below: dLn[Sp] = -β), but for an infectious disease the hazard is proportional to the prevalence of the infection (dLn[Sp] = -β Ip). The *cumulated* prevalence (of who have had the disease) is important for the "herd resistance" to the infection. The distinction between *infection* (virus SARS-CoV-2) and *disease* (Covid-19) will be useful overall.

- There is an abundance of SI(EY)A(CD) modeling. Important is the Berger, Herkenhoff and Mongey (2020) (March 29) paper, that adapts the SEIRD → SEYCD model to test-dependent states of quarantine. Their code in Python is available. These authors are not epidemiologists but they present their

paper as a proof of concept for the exit scenario of *test, test, and test.* See Tabarrok (2020) for an independent appreciation that this exit scenario is cost-effective. Acemoglu et al. (2020) (May) employ a SIRD → SICD model with at least three age compartments, and plot GDP & Death policy frontiers. See another application in Python at Idmod.org, called "Emod". The basic SI(EY)A(CD) model has a constant basic reproduction factor R_0 but there is also an abundance of related models with time-varying parameters, for example (i) unrestrained development (with R_0), (ii) endogenous response (with $R_{0,\text{end}}$) when people notice that there is an epidemic, (iii) organised intervention on social distancing or vaccination (with $R_{0,\text{int}}$), see for example Silver (2020) with a spreadsheet and Jones (2020) with a historical perspective.

- Such abundance also exists for *Mathematica,* see the list of references. For the virus SARS-CoV-2 and its disease Covid-19, the Kaurov (2020) collection of *computational publications* must be mentioned. **Appendix D** shows that one does not need this present notebook and package of mine for epidemiological modeling with *Mathematica*. Who knows the S(E)IR(D) models and *Mathematica* will find it simple and straightforward to create such model in *Mathematica* and run a WhenEvent policy intervention. This present notebook and package only provide structure in use, though at the cost of a learning curve. We only assume now that (such) experienced readers are willing to look at the issue from the angle of didactics.

- Within *Mathematica* there are also many notebooks and demonstrations on actuarial computation, see for example Chandler (2011) with the Markov transition matrix with states of health, illness and death. See also the links provided from there. There is also the Wolfram Language & System Documentation Center (2020) discussion on actuarial computation. See Weisstein (2020b) (MathWorld) for a neat review of the life table and life expectation calculation, including an explanation how R_0 relates to reproduction over the life cycle. The short discussion by the Dutch KAG (2020) mentions (in Dutch) the key factors required for a proper assessment of the impact of Covid-19. **Appendix E** reviews the life table and the effect of SARS-CoV-2 upon life expectancy, with a summary in section 7.11.

- There are other approaches to epidemic modeling than the SI(EY)A(CD) models. A fine discussion has been given by 3Blue1Brown (2020). See the interview by Spinney (2020) with Karl Friston on models derived from physics; this kind of modeling is supported by Gill (2020). While SI(EY)A(CD) assumes a homogeneous and well-mixed population, SARS-CoV of 2003 and SARS-CoV-2 of 2019 appear to show clusters, so that more relevant models use heterogeneity, see Fox et al. (1971), Diekmann et al. (1990), Lloyd-Smith et al. (2005), Britton et al. (2020ab) and the editorial Thorp et al. (2020).

- There is an abundance of approaches and software for elements in SI(EY)A(CD). Obadia et al. (2012) present a package for R with five ways to estimate reproduction numbers for epidemic outbreaks, using the epidemic curves and the serial interval. Interesting are also the peer review reports available with the article. An estimate of R_0 requires a fully susceptible population, and the authors and some peer reviewers clarify that the software does not check upon this, so that it is upon the user to consider whether the estimate really can be interpreted as such. An application of this R-package for SARS-CoV-2 is by Michael Höhle (2020), with available source code. (I am not a user of R though.)

- A key assumption of SI(EY)A(CD) modeling (specifically) consists of lasting immunity after recovery from infection. Like the corona viruses for the common cold - Worrall (2011) - it seems that SARS-

CoV-2 does not generate lasting immunity, see Seow et al. (2020), Sample (2020b), Papachristodoulou et al. (2020), Kirkcaldy et al. (2020) and Irribando et al. (2020). This leads to (also abundant) models with (seasonal) waves of infection, with also births and background mortality. However, SI(EY)A(CD) modeling might still be useful for grasping the pandemic in the first half of 2020. Understanding of SI(EY)A(CD) modeling remains relevant as a base for understanding more complex approaches.

1.1.3. Overview of this *Introduction*

This *Introduction* proceeds with (i) the overview of didactic redesign, (ii) the context of the SARS-CoV-2 pandemic and exit strategies, (iii) the overview of the notebook itself.

The *context* can be summarised as follows. The virus is called SARS-CoV-2 and the disease Covid-19, see Enserink (2020) about these WHO choices. The name of the virus is no coincidence. The virus would be genetically very much like the original SARS-CoV from 2003, see Wilder-Smith et al. (2020): "The whole genome of SARS-CoV-2 has a 86% similarity with SARS-CoV" while referring to Chan et al. (2020), who actually state "Overall, the genome of 2019-nCoV has 89% nucleotide identity with bat SARS-like-CoVZXC21 and 82% with that of human SARS-CoV." SARS-CoV and SARS-CoV-2 have many similar properties, see De Wit et al. (2016), like (i) the effect on age and sex and comorbidity, (ii) asymptomatic transmission, see Wilder-Smith et al. (2005), and (iii) the relevance of superspreading events ("clusters") also related to aerosol transmission (see the overview in Colignatus (2020e)), which explains the success of source and contract tracing in Japan, see department director Saito (2020). The original SARS-CoV might have its peak viral shedding later in the symptomatic phase, see Cheng et al. (2004), which makes detection by symptoms and subsequent quarantine effective, while the current SARS-CoV-2 is already infectious in the asymptomatic phase and thus more difficult to detect. Molecular biologist Borger (2020) argues that we are basically dealing with the same virus of 2003 but a bit less deadly (as happens to viruses over time, though it may also be closer to the bat original), and he wonders why the world did not prepare for its remergence, and why it was claimed that the virus would be "new" and "unknown" - though admittedly the outbreak of 2003 did not reach the West (see Colignatus (2020a:265)). Borger (2020) has a higher similarity in the genome: "A study from March 2020 shows that the genetic material of the SARS-CoV2 virus is 96.11% identical to the SARS virus strain RaTG13", referring to Yu et al. (2020) who clarify that this concerns the bat virus. Of particular interest are pandemic experts Petersen et al. (2020) who compare influenza en the SARS 2003 and 2019 versions. They observe that the two versions are not identical, so that the English language allows that the 2019 version can be called "new", but it appears that the word "new" is overused here. By over-emphasizing the differences between the versions of the virus the authors seem to imply that there were no professional mistakes in their field of monitoring a pandemic. It would be better when pandemic experts state explicitly that they could have diagnosed the not-newness already in January 2020, with a subsequent faster and more effective response in the West, with a better balance between mortality, life-years gained, and costs to the economy. The response in Germany and Greece was faster and more effective than what happened in Holland.

PM. Overall, this notebook has a lot of repetition. Not only the models are much alike, but also the analytical properties and the explanations of the routines, and the discussion of the aspects of the

SARS-CoV-2 pandemic (e.g. in Holland and elsewhere). The assumption has been that readers might skip sections (e.g. jump to their model of interest), so that such repetition is beneficial.

1.2. The renaming S(E)IR(D) → SI(EY)A(CD)

The S(E)IR(D) abbreviation stands for the family of epidemic models that have these compartments or phases (switched on or off):

<p align="center">Susceptible - Exposed - Infectious - Cleared or Deceased</p>

The conventional S(E)IR(D) setup uses the label *"R"* in three different meanings: *Removed* in the SIR model, *Recovered* in SIRD and SEIRD, and overall for the (basic) *reproductive factor* R_0. Also, all models use the *"I"* for the infectious compartment, but some also for "Infected". SEIR(D) includes the *exposed,* as the infected but not infectious yet. Since all models claim to add up to the total population, and the exposed did not occur in SIR or SIRD, this implies that SEIR(D) would describe a different population.

This family of models has been presented in this manner for half a century. Within the computer algebra environment of *Mathematica* it appeared that there is a clear advantage of a better nomenclature. Unique names help the identification of variables, their parameters and the application of associated routines. When we discuss the same population then our analysis should not be cluttered with issues on the administration of the compartments. While humans are flexible, computers are a bit less so, and when humans work with computers then humans become rather inflexible too.

This notebook and package use:

- The Acquitted *A* instead of the *Removed.*
- The Cleared *C* instead of the *Recovered.*
- So that the *A* compartment can be split in *A = C + D* or the Cleared or Deceased.
- The Infected *I* can be *infectious* or *exposed but not infectious yet*.
- The Infectious *Y.*
- So that the *I* compartment can be split in *I = E + Y* or the Exposed and Infectious.
- So that the SIA and SICD models have *E* = 0 and *I = Y*, or that their *infected I* are also the *infectious Y.*

Thus we get the abbreviation SI(EY)A(CD). For all variables *V* we distinguish *Vq* for the levels and *Vp* for the proportions w.r.t. the population *N*[0]. For plotting we can also plot the current population *N* = *N*[0] - *D.*

The decision to choose these names in this manner was taken with quite some hesitation and consideration but would seem to be optimal. A practical argument for this compartmentalisation and nomenclature is that the limit properties of SI(EY)A(CD) are determined by *I* (i.e. the sum *I = E + Y*) just like in the SIA model, without the need for relabeling in SEYA(CD). The introduction of the exposed compartment has only effect for the short run behaviour. For the (crucial) limit property there really is a compartment *I = E + Y*, apart from the temporary distinction in the phases of *E* and *Y*. The package allows the user to adopt the conventional nomenclature, including the possibility of using *Y = E + I* instead. The underlying programming structure however is as above, and there is no reason to hide it and not take advantage of it.

With a standard SEYCD model, an unhindered epidemic in Holland would have given *roughly* the following path, using R_0, i.e. neglecting a change in parameters due to (i) endogenous reactions

when people hear that there is an epidemic ($R_{0,\text{end}}$), and (ii) the intervention by the government in March ($R_{0,\text{int}}$). It is unclear what an endogenous reaction would have been, and the intervention is discussed in Chapters 15 and 16. These figures for an unhindered epidemic have only a didactic meaning, see below for a comparison to RIVM figures (the Dutch CDC). The first death was reported on March 6 but according to the model there were actually already 6 that day. The unhindered eventual total death toll could have been around 250,000, with 95% reached by May 22, and the remaining 5% spreading over the remainder of the year. Only 2% of the population would not have been infected at one stage.

Table legend: S = Susceptibles (uninfected), I = Infected (exposed or infectious), D = Deceased, C = Cleared, N = N[0] - D remaining population, A = C + D (the acquitted). Percentages (p) are w.r.t. the onset, i.e. N[0]. (Table taken from below.)

	Date	Day	Sp	Ip	Dp	Cp	Np	Ap
Onset	2020–02–23	0	100.0	0.0	0.0	0.0	100	0.0
1st Death	2020–03–06	12	100.0	0.0	0.0	0.0	100.0	0.0
1st Intervention	2020–03–18	24	99.9	0.1	0.0	0.0	100.0	0.0
Top if free	2020–04–21	58	25.0	40.3	0.5	34.1	99.5	34.7
Ap = 50%	2020–04–25	62	13.5	36.5	0.7	49.2	99.2	50.0
95% Ap[∞]	2020–05–22	89	2.4	4.5	1.4	91.7	98.6	93.1
Limit if free	∞	∞	2.0	0.0	1.5	96.5	98.5	98.0

The levels (q) of the relevant compartments are as follows. (Table taken from below.)

	Date	Day	Sq	Iq	Cq	Dq
Onset	2020–02–23	0	17.4×10^6	100	0	0
1st Death	2020–03–06	12	17.4×10^6	1290	391	6
1st Intervention	2020–03–18	24	17.4×10^6	19438	6356	97
Top if free	2020–04–21	58	4.4×10^6	$7. \times 10^6$	5.9×10^6	90455
Ap = 50%	2020–04–25	62	2.4×10^6	6.3×10^6	8.6×10^6	130500
95% Ap[∞]	2020–05–22	89	419710	778037	$16. \times 10^6$	243034
Limit if free	∞	∞	344995	0	16.8×10^6	255825

ReadMeSIA["Symbol"]

N[0] = S + E + Y + C + D, with I = E + Y, is the startup population of units.

N = N[0] – D is the current population, mainly used in plotting.

S are the Susceptible.

E are the Exposed but not infectious yet. (They might become diseased.)

Y are the Infectious. The (currently) Infected are I = E + Y, while C + D are have–beens.

A = C + D are the Acquitted, or the Aftermath. (A avoids confusion

with R0. A is conventionally denoted as R that may cause that confusion.)

C are the Cleared. (Avoid the confusing use of R as "Recovered".)

D are the Deceased.

I = E + Y are the Infected. In SIA E = 0 and I = Y.

O = N[0] – S – I are "the other", so that these are all SIO models.

Variable V has level Vq and proportion Vp = Vq / N[0] (divided by N[0] and not N)

? ReadMeSIA

> Symbol
>
> ReadMeSIA[] gives an overall explanation of the SIA` package. See there for:
>
> ReadMeSIA["Terminology"]
>
> ReadMeSIA["Symbol"]
>
> ReadMeSIA["Period"]
>
> ReadMeSIA[model] for models SIA, SICD and SEYCD
>
> ReadMeSIA[SEYCD, Simplify]
>
> ˅

ReadMeSIA["Terminology"]

This terminology applies to the SICD model:

SICD improves on SIA by distinguishing A = C + D, for D deceased and C cleared.

> (–Sp') is the ∗incidence∗ of infections at moment t.
>
> Ip is the ∗prevalence∗ of infections at moment t.
>
> 1 – Sp = I + C + D is the ∗cumulated prevalence∗ at moment t (i.e. the haves and hads).
>
> 1 – Sp[∞] = Ap[∞] in both SIA and SICD is the ∗limit cum. prevalence∗, since Ip[∞] = 0.
>
> Cp is the cumulated survival factor at moment t (but the Sp also count as living).
>
> Dp is the ∗cumulated mortality factor∗ at moment t (mortality ~ population).
>
> Cp[∞] is the limit cumulated survival, and part of the burden of disease (infection).
>
> Dp[∞] is the ∗limit (specific) Infection Mortality Factor∗ (IMF) (prevalence of death).
>
> Dp[∞] / Ap[∞] = Dp[∞] / (Cp[∞] + Dp[∞]) is the ∗limit Infection Fatality Factor∗ (IFF).
>
> MuSICD[] is the ∗(instantaneous) Infection Fatality Rate∗ (IFR) (fatality ~ infection).
>
> MuSICD[] = GammaSIA[] ∗ PhiSICD[] in the model, with PhiSICD[] the IFF.
>
> In addition for SEYCD:
>
> Ip = Ep + Yp subdivides into Exposed (but not infectious) and Infectious.

1.3. Infection Fatality Factor (IFF) and symptomatic Case Fatality Factor (sCFF)

The conventional discussion of the SI(EY)A(CD) model family is oriented at the ordinary differential equations (ODE). In particular, the development of the deceased compartment is modeled as:

$D' = \mu\, I$ in SICD

$D' = \mu\, Y$ in SEYCD

This causes the question what μ stands for. What happens if the parameter γ for the infectious period $1/\gamma$ changes ? Would there be a distinction between a "lethal infectious period" and a "clearing infectious period" ? Maugeri et al. (2020) have the better format $D' = \gamma$ sCFF, for the symptomatic case fatality factor (sCFF). A change in γ will not affect the sCFF and the limit property of the model. The Maugeri et al. (2020) formulation still relies upon the ODE format.

It appears that the *C* and *D* compartments are mere proportions of the acquitted *A* department. The formulation as an ODE is only mathematical convenience of modelers who have adopted this

format but it is not the essential property of the model as such. *A discussion of these epidemic models should not be confused with a course on differential equations.* Assuming that didactics is targeted at the relevant property, we get this structure, with φ the infection fatality factor (IFF):

$A' = \gamma\, I$ in SIA(CD)

$A' = \gamma\, Y$ in SEYA(CD)

$D = \varphi\, A$

$C = (1 - \varphi)\, A = A - D$

This immediately clarifies that there is no distinction between a lethal infectious period and a clearing infectious period: both compartments have the common $1/\gamma$ infectious period. (Some readers might want to check Appendix A.)

The package allows the conventional format too, and such properties can then derived, even though it is also mathematically more elegant to directly specify what the model does.

This notebook and package use an Infection Fatality Factor (IFF) of 1.5%, based upon reweighing the Verity et al. (2020) and Ferguson et al. (2020) estimates, applying the Dutch age composition and inserting a special assumption on the age 70-79 group, see Colignatus (2020cd), part of which is reproduced below in section 2.7. In testimony for Dutch Parliament, RIVM & Van Dissel (2020e) mentions 9,000 deaths, and, using tests on antibodies that cause an expected value of the prevalence of infected (current and past) of 650,000 (3.74%), arrives at a crude IFF ≈ 1.38%, not far of the 1.5% used in this notebook (that is directed at didactics).

PM. Not irrelevant: these models for infections are also used for symptomatic phases, which is also the case for the Goh (2020) "epidemic calculator". Thus below we also discuss the distinction between the infection *generation interval* and the symptomatic *serial interval*. The sCFF has a higher value than the infection fatality factor (IFF) since sCFF excludes the *infections without symptoms*, a.k.a. *asymptomatic infections,* which is important in the SARS-CoV-2 case.

PM. For antibody testing, see Slot et al. (2020), who also mention a 9.5% prevalence in the hardest hit areas, and "that the protective effect of SARS-CoV-2-specific antibodies is not yet known".

1.4. The meaning of parameters R_0, α, β and γ

1.4.1. Euler 1767 and Lotka 1907

Leonhard Euler in 1767 already formulated a "*renewal equation*" for population dynamics, and the SI(EY)A(CD) model appears to be a reformulation in terms of (ordinary) differential equations (ODE). Relevant papers are Fine (2003), Svensson (2005, 2007), Wallinga & Lipsitch (2006, 2007), Breda et al. (2012) and Champredon, Dushoff & Earn (2018). The latter paper is quite recommendable, though they regrettably do not refer to Svensson while a comparison of results would have been informative. Delamater et al. (2019) clarify that George MacDonald who introduced the term of "reproductive number" (factor) in the epidemiological literature in the 1950s used Z_0. See Heesterbeek (2002) for an overall brief history.

The structure that seems best for didactics can be found in the Wallinga & Lipsitch (2006, 2007) paper. For understanding of the SI(EY)A(CD) model family we better now specify the latter main structure here at the outset. Let us mention two models that do not apply - the fixed period and the

normal distribution - and one model that does apply: the exponential distribution (to be distinguished from exponential growth). Park et al. (2019) can be recommended, also for didactic clarity, and I tend to suggest that one first reads Park et al. (2019), then Wallinga & Lipsitch (2006, 2007), and then proceeds here, since my summary might assume too much known. Alternatively, one proceeds here, and falls back to these two articles if something would be unclear.

The general structure is that R_0, β and γ are related to r, the rate of growth of the infected compartment, and the Euler-Lotka "renewal" format rather than the ODE. First consider growth. In general, a growth process $x[t]$ has an *average* factor of growth $f = f[t] = (x[t]/x[0])^{1/t}$ and then there are these formats:

- a "linear format": $f = 1 + \rho$ for rate $\rho = f - 1$. A stepwise format is $x[t] = x[t-1] (1 + \rho[t])$. For a constant ρ: $x[t] = x[0] (1 + \rho)^t$

- an "exponential format": $f = \text{Exp}[r]$ for $r = \text{Log}[f]$. A stepwise format is $x[t] = x[t-1] \text{Exp}[r[t]]$. For a constant r: $x[t] = x[0] \text{Exp}[r\, t]$.

While $r = \text{Log}[1 + \rho]$ or $1 + \rho = \text{Exp}[r]$, for small values $r \approx \text{Log}[1 + r]$ so that the formats differ but the numbers not so much.

A doubling period is $P_d = \text{Log}[2] / r$. The point for SI(EY)A(CD) is that exponential growth is impossible because the pool of susceptibles shrinks. Nevertheless, during the first growth phase the share of susceptibles is close to 1 for a relatively long period and we might e.g. estimate $\beta - \gamma \approx \text{Log}[2] / P_d$, see below.

For a population growth process, we count the numbers generated and measure the periods ("generation interval") over which this happens. (PM. This generation interval differs from the "serial interval" for symptomatic disease, see Chapter 10.) Two new variables are P_c and R_0. Wallinga & Lipsitch (2006, 2007) write P_c but it will be useful at times to write $\delta = P_c$. These variables basically restate what has already been stated above about the general growth process, but now the issue shifts to the context of an infection. The connection is that $r = \text{Log}[R_0] / P_c$ or $R_0 = \text{Exp}[r\, P_c]$, with P_c the "generation interval" (or "generation period"), with "c" standing for the *cohort* contribution to the 2nd generation. Some authors write "generation time" but it is better to call it a period.

- The first generation has index 0. Starting with infector 0, the multiplication factor for the secondary infections of the new generation with be denoted by R_0. Conventionally the factor is called a number, but it is in relation to the number count 1 of the first generation. The outcomes are in integer units. For generations with index $i = 0, 1, 2, \ldots$ outcomes can be like $\{1, 2, 4, 16, 32, \ldots\}$. The formula for the number of reproduced units is $R_i = (R_0)^{i+1}$, with the index shift because of the index and historical notation R_0.

- The generation interval starts when the infector got infected, and it ends when the infector infects another unit (and thus when a new generation starts with the infectee). When the interval is a constant then there is a simple exponential process and when the interval has a distribution then the complexity of epidemiological practice starts.

- The relation between index and time is $i + 1 = t / P_c$ or $t = P_c (i + 1)$. This gives the exponential process $R[t] = (R_0)^{t/P_c}$ that starts at the value $R[0] = 1$ for $t = 0$ for the first generation, and that has value $R[P_c] = R_0$ when $t = P_c$. After two generation intervals or $t = 2 P_c$ then $R[2 P_c] = (R_0)^2$.

- Observe the relationships between the present format $R[t] = (R_0)^{t/P_c} = \text{Exp}[t \, \text{Log}[R_0] / P_c]$ which has P_c in the denominator in the exponent, and the expression $R_0 = \text{Exp}[r \, P_c]$ that has P_c in the numerator of the exponent, and the standard growth format $x[t] = x[0] \, (1 + \rho)^t = x[0] \, \text{Exp}[r \, t]$ with rate r in the numerator of the exponent, with $r = \text{Log}[R_0] / P_c$. Park et al. (2019) call $C = 1 / r = P_c / \text{Log}[R_0]$ the "characteristic time" - though this is a period too.

Let us use $\delta = P_c$ now. The linear growth rate is $\rho = (R_0 - 1) / \delta$, so that $R_0 = 1 + \delta \, \rho$, in which $R_0 - 1$ is the increase of the next generation compared to the single originator, and $\delta = P_c$ is the duration over which this increase happens. Rewrite $\rho = (R_0 - 1) / \delta = (R_0 / \delta) - (1 / \delta) = \beta - \gamma$, with $\beta = R_0 / \delta$ as the contribution to the second generation over the generating / infectious period of the originator, and $\gamma = 1/\delta$ as the rate per time. Sometimes it may be convenient to write $\gamma = \delta^H$ or $\delta = \gamma^H$ with $H = -1$, thus in multiplicative format, see Colignatus (2018). SI(EY)A(CD) models are defined in terms of $\beta = R_0 \, \delta^H$ and $\gamma = \delta^H$ and not in terms of the direct R_0 and δ. At the start of the process, when R_0 applies and not subsequent R, the exponential rate $r \approx \rho = \beta - \gamma$ is close to the linear growth rate. Larger numerical values show important differences in outcomes of the linear or exponential format though. The question is how to proceed after the first generation. The standard approach uses r rather than ρ.

With the process now formulated as $R[t] = (R_0)^{t/P_c} = \text{Exp}[t \, \text{Log}[R_0] / P_c]$ the stage has been set for the question about the relationship between R_0, P_c and r when we assume a distribution for P_c rather than a constant value.

Wallinga & Lipsitch (2006, 2007) pose the question: "How should we choose the most appropriate equation for inferring the reproductive number from observed growth rates for a particular infection?" Their conclusion is: "To conclude, the variety of equations that relate observed growth rate to reproductive number can be understood within the Lotka-Euler [sic] framework which embraces both a description of the infection cycle and a description of the change in number of new case counts. The observed generation intervals and the observed epidemic growth, when taken together, specify the appropriate value of the reproductive number, and therefore, the required control effort to contain the epidemic. This means that infectious disease surveillance systems which have an objective to inform health policy makers on the required control effort should monitor the symptom onset date of new cases as well as their generation interval for new emerging infections."

These authors then consider the constant interval, the normal distribution, the exponential distribution, and the empirical distribution. Subsequently there is also the gamma distribution in general.

1.4.2. Constant generation interval

When the generation interval is no random number but a constant then all secondary infections arise exactly after the same period. This constant P_c then is also the *mean* generation interval. This particular distribution of the infections over time is called the "delta distribution". With the rate of growth r and the period over which this applies, the number of secondary infections will be, as already stated above:

$R_0 = \text{Exp}[r \, P_c] = e^{\, r \, P_c}$

The development over time would show $R[t] = R_0^{\, t/P_c} = \text{Exp}[t \, \text{Log}[R_0] / P_c] = \text{Exp}[r \, t]$.

Above, we had the rate $\rho = f - 1$ in linear format or $r = \text{Log}[f]$ in exponential format. Now we have $r = \text{Log}[R_0] / P_c$ so that the unit of account for r and R_0 is not 1 but the generation interval. A historical reason is that R_0 counted discrete units rather than being an average. The interval also has the interpretation for the cohorts or generations of pathogens.

1.4.3. Normal distribution

When infections arrive with a generation interval with mean P_c and a standard deviation σ then, with appeal to the law of large numbers, they might approximate the normal distribution. The Dublin-Lotka result of 1925 is:

$R_0 \text{ == R0DublinLotka}[r, P_c, \sigma]$

$$R_0 = e^{r\, P_c - \frac{r^2\, \sigma^2}{2}}$$

Note the constant value when $\sigma = 0$. Also, the more dispersion, the lower R_0, meaning that faster growing infection paths cannot compensate for the slower paths.

1.4.4. The SI(EY)A(CD) family

Epidemiology uses these notions from demographics for the SI(EY)A(CD) family of models. See the formulas and the plot of the growth rates in Section 5.7 on the SIA model. The demographic model assumes exponential growth of the population of pathogens. This will only work in the first and final phases of an epidemic. For a long period, the development is distinctly not exponential. However, the first and last phases can still be used for estimating the relevant parameters.

The relevant distinction within the SI(EY)A(CD) family is:

(1) $P_c = P_{\text{inf}}$, or the generation interval is also the *infectious period.* This assumes an exponential distribution.

(2) $P_c = P_{\text{exp}} + P_{\text{inf}}$, or the generation interval consists of the *infectious period* and a preceding *exposed* period during which the unit is not infectious. This assumes the sum of two exponential distributions, for both exposed and infectious.

Define the *infections over the infectious period* as $\beta = R_0 / P_{\text{inf}}$ (the 2nd generation considered over the infectious period of the originator). *This holds for both models.* Thus $R_0 = \beta\, P_{\text{inf}}$.

The rate of leaving the infectious period is denoted as γ, so that $\gamma = 1 / P_{\text{inf}}$ or $P_{\text{inf}} = 1 / \gamma$. This notebook takes $P_{\text{inf}} = 10$ days.

The rate of leaving the exposed period is denoted as α, so that $\alpha = 1 / P_{\text{exp}}$ or $P_{\text{exp}} = 1 / \alpha$. This notebook takes $P_{\text{exp}} = 1$ day.

Rewriting from periods to the parameters for the rates:

(ad 1) $P_c = 1 / \gamma$

(ad 2) $P_c = 1 / \alpha + 1 / \gamma$

Each case has its relationship between the reproductive number and the growth rate r.

(ad 1) $R_0 = 1 + r\, P_{\text{inf}} = 1 + r / \gamma$

(ad 2) $R_0 = (1 + r\, P_{\text{exp}})(1 + r\, P_{\text{inf}}) = (1 + r / \alpha)(1 + r / \gamma) = (\alpha + r)(\gamma + r) / (\alpha\, \gamma)$

PM. Roberts & Heesterbeek (2007:809) expand to $R_0 = 1 + r(1/\alpha + 1/\gamma) + \frac{r^2}{\alpha\gamma}$.

PM 1. We have done quite a roundabout to reproduce the original notion that $r = (R_0 - 1) / P_{\text{inf}} \approx \text{Log}[R_0] / P_{\text{inf}}$. However, while we can apply such transformation for any dynamic process at any particular moment in time, only the exponential distribution will generate constant parameter values, as assumed here.

PM 2. Using $R_0 = \beta / \gamma$:

(ad 1) $\beta / \gamma = 1 + r / \gamma$ or $r = \beta - \gamma$

For example, RIVM & Van Dissel (2020a)(the Dutch CDC) mention $R_0 = 2.7$ and a doubling period of 5-6 days. If the SIA model applies:

```
eqs = {r0 == 2.7, r0 == beta / gamma, r == beta - gamma} /. {r → Log[2] / 5.5}
```
$\left\{\text{r0} = 2.7, \text{r0} = \dfrac{\text{beta}}{\text{gamma}}, 0.126027 = \text{beta} - \text{gamma}\right\}$

```
sol = Solve[eqs, {r0, beta, gamma}] // Quiet
```
$\{\{\text{r0} \rightarrow 2.7, \text{beta} \rightarrow 0.20016, \text{gamma} \rightarrow 0.0741334\}\}$

```
1 / gamma /. sol[[1]]
```
13.4892

(ad 2) $\beta / \gamma = (1 + r / \alpha)(1 + r / \gamma)$

   (PM. ad 2) $\text{Log}[\beta / \gamma] = \text{Log}[(1 + r / \alpha)(1 + r / \gamma)]$

   (PM. ad 2) $\text{Log}[\beta] - \text{Log}[\gamma] \approx \beta - \gamma = \text{Log}[1 + r / \alpha] + \text{Log}[1 + r / \gamma] \approx r / \alpha + r / \gamma$

   (PM. ad 2) $r \approx (\beta - \gamma) / (1 / \alpha + 1 / \gamma)$

   (PM. ad 2) The latter expression reduces for $\alpha \rightarrow \infty$ to $r \approx \gamma (\beta - \gamma)$ which shows that the latter is a worse estimate than using the original non-approximated expression $(1 + r / \alpha)(1 + r / \gamma) \rightarrow (1 + r / \gamma)$. It seems advisable to use (2) rather than the linearisation.

For the RIVM figures, assuming the SEYCD model:

```
eqs = {r0 == 2.7, r0 == beta / gamma, r0 == (1 + r / alpha) (1 + r / gamma) } /.
   {r → Log[2] / 5.5, alpha → 1}
```
$\left\{\text{r0} = 2.7, \text{r0} = \dfrac{\text{beta}}{\text{gamma}}, \text{r0} = 1.12603\left(\dfrac{0.126027}{\text{gamma}} + 1\right)\right\}$

```
sol = Solve[eqs, {r0, beta, gamma}] // Quiet
```
$\{\{\text{r0} \rightarrow 2.7, \text{beta} \rightarrow 0.243432, \text{gamma} \rightarrow 0.09016\}\}$

```
1 / gamma /. sol[[1]]
```
11.0914

PM 3. The growth of the infected = infectious in SIA can be found in $r = \text{dLog}[I] = I' / I = i = \beta Sp - \gamma$, see Section 5.7. It will be useful to identify the *dynamic* reproductive factor $R_0[t] = \beta[t] / \gamma[t]$ and the *effective* reproductive factor $R[t] = R_0[t] Sp[t]$, where the latter corrects for the influence of the declining share of susceptibles. In SIA thus $r = (R_0 Sp - 1) / \delta$ for $\delta = 1/\gamma$, which gives an *effective* version of the above. Here Sp would be a dimensionless number, as units of susceptibles per units

of the base population.

PM 4. It is useful here to refer to the same result as in Chapter 10. The PDF of the random variable that is the sum of two exponentially distributed variables is (and observe the symmetry):

```
SumTwoExp[α, γ, t]
```

$$\frac{\alpha\,\gamma\left(e^{\alpha\,(-t)} - e^{\gamma\,(-t)}\right)}{\gamma - \alpha}$$

When we substitute this pdf in the Moment Generating Function as indicated by Wallinga & Lipsitch (2006), then (while remembering that this only holds for *part* of the process in SEYA(CD), namely the exponential phases):

```
R0 == 1 / Integrate[ Exp[-r t] SumTwoExp[γ, α, t], {t, 0, Infinity},
    Assumptions → {{r, γ, α} ∈ Reals, γ + r > 0, α + r > 0 }]
```

$$R0 == \frac{(\alpha + r)\,(\gamma + r)}{\alpha\,\gamma}$$

1.4.5. Empirical distributions

Wallinga & Lipsitch (2006, 2007) mention the important case when the $R[t]$ can be taken from the moment generating function of an empirically observed distribution of interarrival times. Wallinga has been affiliated with the RIVM, the Dutch counterpart of the US CDC, and has a special professorship in Leiden. He is also co-editor of the "Handbook of infectious disease data analysis". Our chapter 2 below will copy the RIVM graph of such $R[t]$. See section 5.2 for terminology, with $\beta[t]$ / $\gamma[t]$ for the "*dynamic* reproduction factor", while the "*effective* reproduction factor" $R[t]$ also corrects for the level of susceptibles Sp[t].

Aspects to consider are: (i) For a longer horizon, there are births and background deaths. (ii) When there is symptomatic disease, there may be asymptomatic transmission, see Chapter 10, and this may have complications for estimation and source and contact tracing. (iii) When one path systematically has short intervals and another path has systematically longer intervals, then the joint distribution is not stable, and one might better distinguish two compartments. A distinction may be that the elderly have faster symptoms and clear later. (iv) The RIVM $R[t]$ graph is estimated on hospital admissions but over time in Holland the empirical value of this series has not been constant. There was a learning-effect about the virus, with a shift away from hospitalisation, with deaths no longer tested and not officially recorded as "SARS-CoV-2 deaths" but showing up as "excess deaths" in the monthly statistics.

1.4.6. Gamma distribution

The extension with a gamma distribution was reported on by Wearing et al. (2005), was taken up by Roberts & Heesterbeek (2007), further developed by Nishiura et al. (2009) and reported on usefully by Park et al. (2019). The exponential distribution is a special case of the gamma distribution. While the exponential distribution fits the SI(EY)A(CD) model family, this also holds for the extended model with m latency periods and n infectious periods, which gives a Gamma[m, n] distribution. It is not clear however whether the gamma distribution (in general) without such assumption of subphases still fits the SI(EY)A(CD) model family. Nishiura et al. (2009) assume a gamma distribution

with mean μ and coefficient of variation κ, and use the following formula, for which Park et al. (2019) provide examples and plots.

$R_0 = (1 + \kappa^2\,\mu\,r)\wedge(1/\kappa^2)$

The suggestion is that one collects the follow-up data of the infected and their infectors, with the observed generation intervals, finds the gamma distribution that best fits those, and uses its parameters μ and κ. Potentially such parameters are taken from another study on the same disease. The rate of growth r would be taken from the development of incidence, for which Nishiura et al. (2009) state another formula.

1.4.7. Interpretation of γ, also for estimation

Above discussion with the generation interval has the didactically beneficial property that the emphasis lies on the periods and not the rates. In the conventional ODE format, the SI(EY)A(CD) models are formulated with β and γ, and subsequently the relevance of $R_0 = \beta / \gamma$ is deduced. Below we will also present the models with those rates. However, the above should provide some antidote, that underlying assumptions already have a close connection with the periods, the rate of growth and the reproductive factor, via the notion that $r = (R_0 - 1) / P_{inf} \approx \mathrm{Log}[R_0] / P_{inf}$.

The interpretations of "infectious period" and rate of acquittal are adequate for an application for e.g. such natural infectious period (a "real infectious period"). The issue may become more complicated when we change our assumptions and the model might not quite apply like we might suppose.

On occasion a better format for the model may be to replace $\beta \rightarrow R_0 / P_{inf}$. While R_0 gives the number of infections, those *given* infections are spread over said period. A change of the "infectious period" then does not affect the given number of infections. In this manner, β has been made "variable" in order to fit the given number of infections R_0. This perspective appears relevant when modeling and estimating an intervention. An estimation routine may be ill-conditioned for β and γ, and show no real differences in outcome, when their ratio R_0 does not change along the ray $\beta = R_0$ γ. In this perspective the P_{inf} can better be seen as a normalisation factor, that cannot be changed, and the estimation then concerns R_0 only.

Let us consider another case when both R_0 and β are reduced, with a reduction of the infections, accompanied by a reduction of the "infectious period". The rise in the rate of acquittals γ causes a surge of the acquittals, and thus also a rise in the death toll (in the short run). Let us first give a numerical example before discussing the issue further. The following has values relevant for SARS-CoV-2.

(base) There is a base case $\beta = 0.6$ and $\gamma = 0.1$. Thus $R_0 = 6$ and the infectious period is 10 days. The graph below depicts this base case with *dot-dashed* lines (and not just dashed).

(intervention) Subsequently there is an intervention on day 24 with $\beta = 0.4$ and $\gamma = 2$. Thus $R_0 = 0.2$ and the infectious period is 1/2 day (people are quickly put into quarantine). This is depicted with *drawn* lines (the deceased dashed though).

The intervention is dramatically effective. The death toll quickly tops at its limiting value of some 7000 persons (including those of before the intervention).

(The following graphs are copied from section 16.13.)

People & ICU capacity
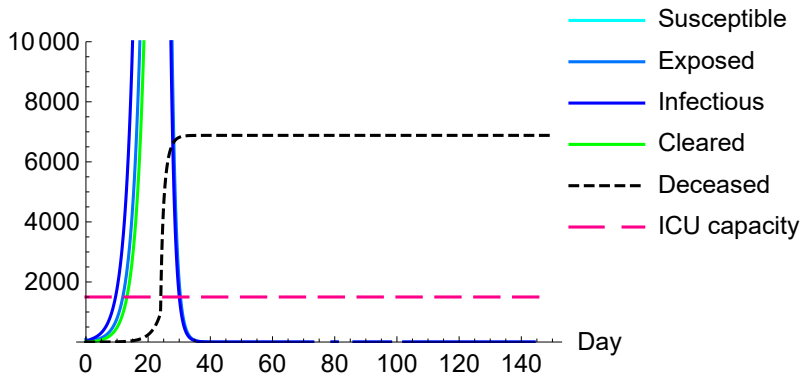
However, let us focus on the lower ranges. To the left of the intervention, the plotted lines overlap. At the intervention and some few days after it, we see a surge of clearings and deaths. With γ changing from 1/10 to 2, the model presumes that there is an acceleration in acquittals indeed.
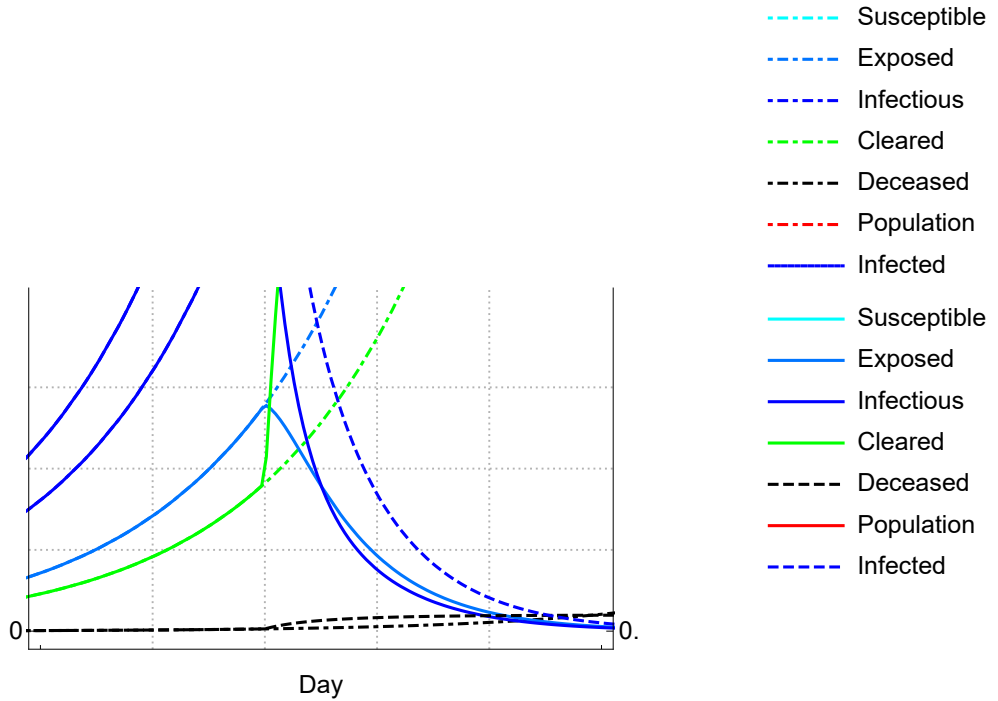
Now that the graphs have been shown it will be easier to discuss the issue. The point is that the SI(EY)A(CD) model concerns *compartments* and not necessarily *events.* The allocation to the compartment of *D* does not mean immediate death. It is true that the intervention with the quarantine reduced the "infectious period" from 10 days to 1/2 day. And it is true that the Infection Fatality Factor (IFF) proportion of those acquitted belongs to the compartment of the (soon to be) deceased. However, the intervention did not speed up the process of dying. Given the new "infectious period" from the intervention, the infected are properly allocated to their compartment but we should not assume that this "infectious period" still would be a "natural or real infectious period" (ending e.g. in death). (A consideration might be to use $D' = \mu\,I$ and not change μ, so that

the change in the "infectious period" is borne by $C' = \lambda I$. However, this implies changing the IFF. The intervention cannot change this. A given proportion of the infected belongs to the compartment of the deceased.) Thus, entry into the compartment of the deceased does not yet indicate the day of actual death. The estimation on the observed death toll requires that we take into account that the model speeds up the allocation to this compartment, so that we need a longer period from *entry into the compartment* to *actual death*.

The conclusion is unavoidable that the SI(EY)A(CD) model has a parameter γ that must be treated with some care. The reading of $1/\gamma$ as the "infectious period" would be unproblematic for the natural process. When this interpretation is changed, in the context of an intervention, then $1/\gamma$ is rather regarded as a normalising constant and the interpretation as "infectious period" should not be taken too literally with connotations from the other context.

PM. Also stated below: Yan & Chowell (2019:167) nicely show how the rescaling of time into $\tau = \gamma\, t$ changes the SEYCD model into one that is dominated by only two parameters: β/γ and α/γ. Thus we basically still need three parameters, the latter two, and the scaling of time. Their deduction however helps us to focus on R_0 and these ratios, and not on β/α.

That said, the notebook and package have the conventional formulation with β and γ and only consider this substitution $\beta \to R_0 / P_{\text{inf}}$ as a possibility for estimation.

1.5. Benefit of better didactics on SI(EY)A(CD) epidemic models. Understanding herd resistance ("immunity")

1.5.1. General understanding and communication

The 2020 pandemic (again) proves the relevance of the SI(EY)A(CD) models for the common understanding of infectious epidemics and their impact on the economy. The Goh (2020) "epidemic calculator" allowed perhaps millions of people a better understanding, and clarified for them the notion of "flattening the curve" to remain within the capacity of the health system. Pueyo (2020) provided clarity for many too. There is common use of the model in the literature (see the references). The model appears to be of key importance for a world at mercy of the pandemic, not only for education, but also for communication between research communities and for the media and general public. It stands to reason, but is not guaranteed, that better didactics would be beneficial for such understanding and communication.

The SI(EY)A(CD) model remains too simple as a foundation for practical policy making (when we have better models, see e.g. Diekmann et al. (1990) and Grant (2020)). However, the model remains relevant for general understanding, and such general understanding very much concerns issues of strategy too.

1.5.2. Interpreting official reports

RIVM - the Dutch equivalent of the US CDC - provides the national reservoir of experience in infectious diseases and must have employed more complex models than SI(EY)A(CD) for its substantiation to the Dutch government that there *was* a deadly pandemic. The international reports, the WHO PHEIC of January 30, the WHO March 11 actual declaration of a pandemic, and the growing

alarm in Dutch hospitals themselves were factors too of course. The point remains that, against this backdrop, the SI(EY)A(CD) family of models can still be regarded as a bedrock for our understanding of epidemics.

RIVM has had a policy of *not-mentioning the death-toll at stake* but *focusing on flattening the curve* with the ICU capacity as a target. The RIVM & Van Dissel (2020a-f) reports to Parliament have a speculation about "avoided ICU admissions" . However, "avoided deaths" are rather implied in the published parameters. RIVM has reported values of R_0 to Dutch Parliament starting with 2.7, down to 2.3 and up to 2.5 again. With R_0 = 2.5 and their IFF =1.38%, the death toll for a population of 17.4 million would have been, had there been no intervention:

```
Explain[SICD[Limit, BetaSIA[] → 2.5 GammaSIA[], PhiSICD[] → 0.0138], SICD]
```

{S → 0.107354, I → 0, C → 0.880327, D → 0.0123185, N → 0.987681, A → 0.892646}

```
17.4 × 10^6 * "D" /. TheFormer
```

214 342.

The RIVM policy of not mentioning an implied death toll seems rather a consequence of the attitude of avoiding speculation and "sticking to the facts". However, it is actually better to mention these implications, because death toll estimates make for a more transparant discussion about the "value of a statistical life". There may be a difference between the world of Public Health (including economics) and the world of Medicine (without economics), with RIVM too much influenced by the latter.

1.5.3. Remarkable statistical fits

This notebook took R_0 = 4 from the literature, and φ = 1.5 was found by using international age-specific factors for the Dutch population composition. The explanatory power of the SI(EY)A(CD) model remains baffling. The following plot has used SEYCD and shows the fit of the model to the official number of deceased in Holland till May 20 (week 21), to a total of 5748 deaths. This still excludes the untested non-hospitalised "excess deaths" of these months, so that the actual death toll in this period is likely over 9000, see CBS (2020). Jointly, CBS and PBL arrive at excess deaths including SARS-CoV-2 in week 11 to 21 of 10,164, see Husby et al. (2020). For all clarity: this estimation is only for didactic purposes. This uses only 15 data-points on the cumulative death toll, and includes a heuristic linear-proportional back-tracing of the recorded day of death to an estimated end of the "infectious period". (Graph copied from below.)
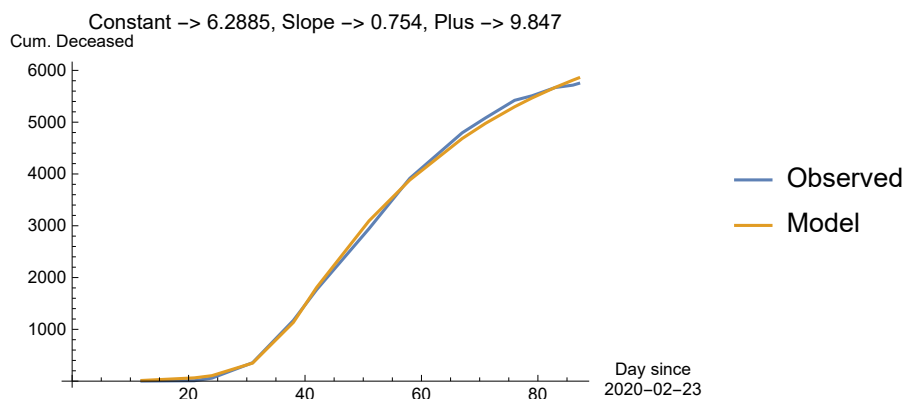
This "didactic estimate" also allows to locate the Dutch onset to between Valentine's Day (February 14) and Dutch Carnival (February 23). There are also reasonable estimates of the basic reproductive factor R_0 of the period before and after the intervention of the partial lockdown announced on March 12 and the closing of the schools on March 15. *The estimate indicates that there would have been some* 250,000 *deaths if we disregard (i) an endogenous response, (ii) the intervention by the government.* It is testimony of the quality of the model that such an estimate can be regarded as reasonable. The difference with the similarly implied outcome of the RIVM parameters is not overly large. It are the properties of exponential processes, the SI(EY)A(CD) modeling itself, and such statistical fitting that makes this inference about the avoided death toll so reasonable. The Dutch economy suffers a lot, but the alternative, this death toll, would have been stark - while it remains an open question how the economy would have reacted if these deaths had actually occurred. National disasters were the flood of 1953 with 1836 deaths and the 2014 MH17 plane with 298 deaths, and we are now considering disaster sizes of 130 floods or 800 planes.

Above graph and "didactic estimation" on the official death count figures awkwardly exclude the "excess deaths" while the parameter φ = IFF = 1.5% in the estimation includes them. It would seem that the estimate on the distribution of deaths is not affected by the proportional parameter. A reduction to φ = 1% within the estimation would basically mean a lower RMSE. The model is flexible and can be adapted. The following graph clarifies that above estimate has used such flexibility. This graph shows the Root Mean Squared Error (RMSE) as a function of both the R_0 before the intervention (R_0-pre) and the possibility that the onset of infections happened some days earlier than Carnival, February 23 2020. (Graph copied from below.)
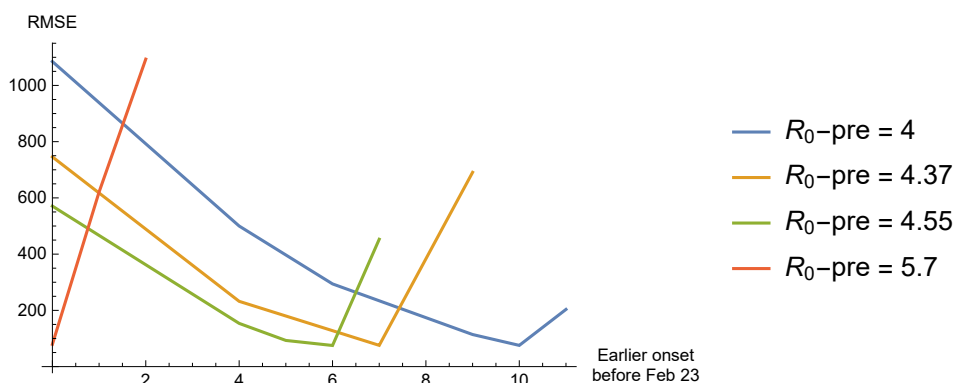
This notebook chose R_0 = 4 with reference to the literature, before doing this estimation on Dutch data. The value of 4 agrees with 9-10 days earlier than Carnival: Valentine's Day. With a higher R_0 then there is less need for an earlier onset to explain the cumulative death toll at the end of May. We can maintain the onset at Carnival, if R_0-pre would be 5.7, and then have the same RMSE level as in other cases - which RMSE is so uniform low because all fits are so well (or so worse since none reaches 0). In all these cases we maintained γ = 0.1. The range of 4 - 5.7 seems rather large. However, it so happens for the SI(EY)A(CD) model family that this range is not unreasonable.

1.5.4. Limit values

The following gives the contours of the proportion of the acquitted A = (1 - susceptible - infected), as a function of R_0. There is a close relation between R_0 and the limit value of those affected in society. Remember that the death toll is a fixed proportion of the acquitted as well. Based upon the age

composition of Dutch society, the IFF ≈ 1.5%. For our assumption of $R_0 = 4$, the top of infections would have been reached after 58 days. The level of 50% of the population would have been reached after 62 days. The limit value of some 98% would never have been reached because it is a limit value. However, 95% of this limit value would have been reached after 89 days, say three months. The point to observe is that the values of the limit shares are rather flat in the range of $4 \leq R_0 \leq 6$. This gives some leeway for above "didactic estimation" outcome. (Graph copied from below.)

A = 1 − S − I
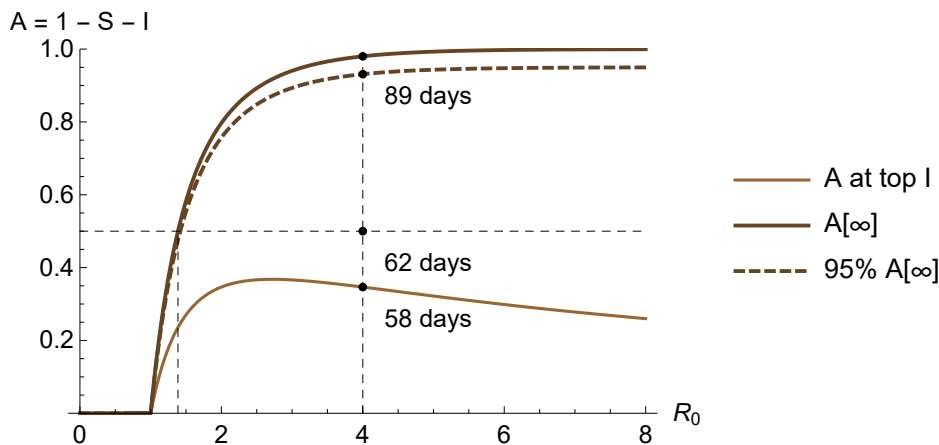
1.5.5. Vaccination

Let us assume that vaccination results into immediate and lasting immunity and that the susceptibles have been vaccinated with degree *v,* so that *v* is in the range [0, 1]. The first infectious unit (the seed) has an offspring of R_0 units, via direct infection of its contacts. With perfect mixing of units, this means that $v R_0$ would not be infected and $(1 - v) R_0$ would still be infected. The infection will remain constant when $(1 - v) R_0 = 1$; it will grow then $(1 - v) R_0 > 1$; and it will reduce when $(1 - v) R_0 < 1$. The latter can be rewritten into a condition for the degree of vaccination to warrant that the infection will be reduced: $v > 1 - 1 / R_0$. Normally there is a small number of infectious units (seed) at the start. If vaccination is done during a raging epidemic then we must take into account that there is a large number of infectious units present that will continue the infection.

The level of acquitted Ap = 1 - 1 / R_0 is sometimes called "herd immunity". Jaap Goudsmit (2020): "The idea is that you do not have to vaccinate everyone to still protect all against the consequences of an infection." (translated) However, there is a distinction between vaccination *before* an infection and vaccination *during* a raging epidemic: in the latter case there can be huge numbers of infected people so that the epidemic can proceed (called "the overshoot") before it subsides. Because of reasons like this, the better term is "herd resistance" than "herd immunity". The real meaning of "herd immunity" is that the herd does not become extinct, even when most members would succumb. Goudsmit (2020) is correct about the protection for all for a vaccination before an epidemic, but this is not the full and only meaning of the term.

It depends upon circumstances whether we can really compare *v* with *Ap,* see the discussion in the next subsection. At this point we can already plot the distinction between the final outcome in the limit and the vaccination requirement $v > 1 - 1 / R_0$, sometimes also referred to as "herd immunity". When a raging infection reaches level Ap = 1 - 1 / R_0 then the infection will continue till Ap[∞] is

reached, and then there is little protection by this level of so-called "herd immunity". For vaccination it is important to have it done *before* an outbreak, to prevent it.

ApByR0Plot["HerdImmunity"]

A_{HI} called Herd Immunity
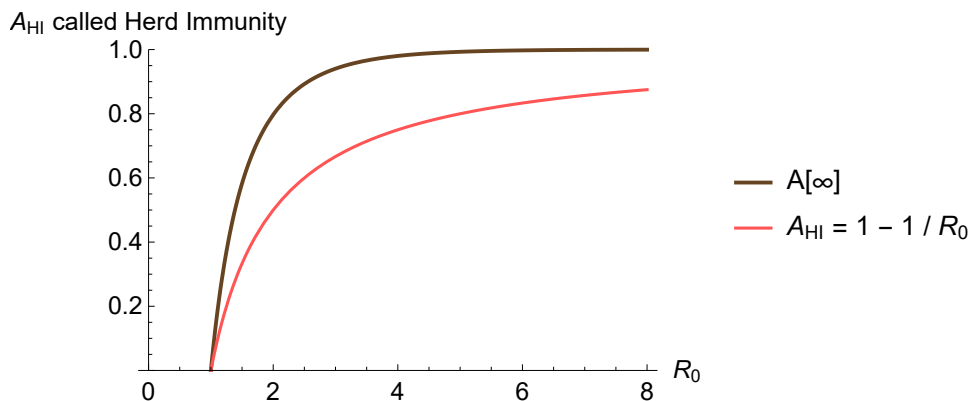
1.5.6. Understanding herd resistance (so-called "herd immunity")

RIVM informed the Dutch Prime Minister Mark Rutte about the notion of "herd immunity", which information the PM used in his television speech of March 16 2020. Rutte (2020) did not mention a specific number but RIVM circulated that herd immunity would be around 60% of the population, see the interview by journalist Tweebeeke and RIVM department head Van Dissel (2020). There are at least four possibilities how to arrive at such notion with such value, see **Section 6.12**. The former section has already clarified that we must distinguish between the start of an epidemic and a raging epidemic, and between the limit outcome and the level of her resistance (so-called "herd immunity").

(a) Potentially, 40% of society (7 million) are vulnerables who will remain in permanent quarantine, so that 60% (10.4 million) would be the less vulnerables who might be in mutual contact with a death toll of $10.4 \times 10^6 * 0.06\% = 6240$ persons of younger age. This would work, see below, but it is remarkable that RIVM does not discuss the issue in these terms.

(b) When $R_0 \approx 1.5$, then the *limit* value can be 60% (see above graph). It is less likely that RIVM used this notion, when they explicitly reported R_0 in the range of 2.3 to 2.7, and when the notion of herd immunity was indicated for the period *after* the lockdown. But conceivably RIVM might have the argument that $R_0 \approx 1.5$ might work for the new rules on social distancing, and then accept the limit value for the raging infection. RIVM suggested that the vulnerables would be protected from infection, potentially by quarantine but also that 60% herd immunity would protect them too. RIVM then seems to imply that *also the vulnerables would be exposed* (i.e. under the conditions of herd immunity) while the very idea is to protect them (rather with quarantine). If protection by means of quarantine fails then the implied death toll is 60% * 1.5% = 0.9% of the population, or 156,600 persons. Perhaps this issue clarifies to RIVM that it *is* useful to also mention the death toll, even while it is "speculative".

(c) Within the epidemiological literature there might be authors who are subject to confused reasoning with a formula $x = 1 - 1 / R_0$ for some (unstated) x, like potentially the level of vaccination v before the onset of an infection. When RIVM has $R_0 = 2.5$ and applies this formula then this gives 1 -

1 / 2.5 = 60% indeed. There is more than one way how this formula 1 - 1 / R_0 may arise though. Nevertheless, when also the vulnerables are exposed then the death toll would be said 156,000 persons.

In general, the effective reproduction $R = Sp\, R_0 = (1 - Ip - Ap)\, R_0 < 1$, iff $Ip + Ap > 1 - 1 / R_0$. With $Ip > 0$ then *infections continue,* even while they are slowed down. The effect can be large in a raging epidemic with $Ip \gg 0$. The literature calls this "overshoot", which doesn't provide the protection that RIVM refers to. With overshoot the death toll would be higher than 156,000.

(c1) If the formula for vaccination $v > 1 - 1 / R_0$ is used and the substitution $v \to Ap$ is done (as a "natural form of vaccination"), then the proper substitution actually has $Ip + Ap > 1 - 1 / R_0$, and one cannot overlook that the epidemic is already raging.

(c2) If $R_0 > 1$ then the infections have a top when the susceptibles are at Sp = $1/R_0$. Perhaps this invites attention for 1 - Sp = 1 - $1/R_0$ again. At this top Ip $\gg$ 0 and not a steady state or limit value. After the top the infections decline, but the infections continue nevertheless. When $R_0 = 2.5$, then the limit value Ap[∞] ≈ 90% (see above graph). When the Acquitted are at 60%, the infection may be reducing, but is not over yet, and still continues with 30% * 1.5% = 0.45% of the population or 78000 deaths.

```
Explain[SIA[Ip, Top, R0], SIA]  (* the real top of Ip *)
```

$$\left\{ S \to \frac{1}{R0},\ I \to \frac{R0 - \log(R0) - 1}{R0},\ A \to \frac{\log(R0)}{R0} \right\}$$

```
Explain[SICD[Limit, BetaSIA[] → 2.5 GammaSIA[]], SICD]
(* the limit values for R0 = 2.5 *)
```

{S → 0.107354, I → 0, C → 0.879256, D → 0.0133897, N → 0.98661, A → 0.892646}

In above plot, the difference between the curves (for $A[∞]$ and A_{HI}) *times* the population *times* the Infection Fatality Factor (IFF) is the overshoot in deaths, that is not accounted for by the claim that $Ap_{HI} = 1 - 1 / R_0$ would give "protection" ("herd immunity").

(d) The formula $Ap \geq 1 - 1 / R_0$ can also be found with a model with a steady state. The SI(EY)A(CD) family of models does not have a steady state but an "*asymptotical* steady state". Perhaps the model used by RIVM has a steady state. This was not explained however.

(abcd) In sum, the possibility cannot be excluded that Dutch RIVM was itself the cause for confusion here. RIVM was explicitly reticent in discussing the implied death counts for such policy scenario's. It would have been better, and for the future it will be better, when the death toll is explicitly mentioned so that it becomes clear what scenario is discussed. For journalists, it might be that it requires knowledge of mathematical modeling to know when those models are *not* needed when IFF values can applied directly.

PM. The three (inadequate) scenario's as proposed by RIVM are at the Dutch government archive *https://archief09.archiefweb.eu/archives/archiefweb/20200605115253/https://www.rivm.nl/coronavirus-covid-19/aanpak-bestrijding* or in English at *https://web.archive.org/web/20200527143138/https://www.rivm.nl/en/novel-coronavirus-covid-19/dutch-response-to-coronavirus.*
(Current strategy *https://www.rivm.nl/en/novel-coronavirus-covid-19/dutch-response-to-coronavirus*)

1.5.7. Relevance of better didactics

A presumption of better didactics is that it would be beneficial for the general discussion, at least within the wider community of researchers and policy makers, when such notions, as just discussed, would be well-understood.

PM. In a paper that got through peer-review, Ioannidis (2020) (my italics): "Focusing on protecting *susceptible* individuals may be preferable to maintaining countrywide lockdowns longterm." Likely he means "protecting *vulnerable* individuals". With the terminology of epidemiological models, protection of all *susceptibles* requires a national lockdown.

1.6. A nasty virus, asking for eradication. A challenge anyway

1.6.1. A world at risk but also badly managed

John Snow (1813-1858) is considered one of the founders of modern epidemiology, when he deliberately searched for and identified a Soho water pump in the London cholera epidemic of 1854. We are more than 150 years later. It is not for nought that the world has set up the WHO and created the notion of an official warning of a *public health emergency of international concern* (PHEIC). Recent more general warnings were e.g. from HIV / AIDS 1981-present, SARS 2003, MERS 2012-present, the Mexican "swine flu" 2009, Ebola 2013-2016 and 2018-2020 and still risky, and Zika 2015-2016. Microsoft founder Bill Gates is credited for a 2015 TED talk warning for "the next outbreak". Piot (2018) is more recent and essentially more qualified. The UK government provided for a study how to respond to pandemics, which study was duly used by Singapore for the current SARS-CoV-2 outbreak. However, it was not used by the UK itself. The USA response team was dismantled from the US White House a few years ago. Something has gone badly wrong in the last decade. The world in the last few years shows a huge *pandemic mismanagement,* likely caused by fundamental misunderstandings, while the issue was already clearly spelled out in the literature decades ago.

Observations are: (1) The health community has a hierarchical setup and mindset while the economic community is more liberally minded. When top health officials do not strongly criticise policy makers - because they have to work with them on a daily basis - then the health community apparently also subdues criticism. But economics is not perfect either. (2) What are severely lacking in the democratic setup of each nation are both an *Economic Supreme Court* and a *National Assembly of Science and Learning,* see Colignatus (2020a).

1.6.2. Prevent future pandemics

Future pandemics are better prevented. The BBC Horizon (2020) broadcast can be much recommended for its overview of the issue and essential information, including the discussion of the WHO 2018 warning on pandemics and the identification of risk "hotspots". At roots, the lessons already learnt on epidemics ages ago are rediscovered for the world-scale without world-governance, and with scientists already warning for decades. Mankind is advised to make some fundamental changes, towards much lower world population and environmentally sustainability, including less

contact with wildlife and their pathogens, but also in the information and decision making structure in society, see Colignatus (2020ab).

There is a perspective that SARS-CoV-2 might only mean a reduction of life-expectancy of 1%, see Richardson & Spiegelhalter (2020). For Dutch readers there is Jaspers (2020). From this perspective, there is only a limited effect, and there might be no need for more involved changes. However, the other view is that we got only lucky now. And perhaps we are not so lucky actually, since the virus appears to be a nasty one.

1.6.3. A nasty virus, asking for eradication

SARS-CoV-2 is a nasty virus that attacks many parts of the body, see the column by Collins (2020), the director of the US National Institutes of Health (NIH), and the interview with Longley by BBC Horizon (2020) (May 26). For Dutch readers there is De Visser (2020). A possibility is also that the body is cleared of infection but that the immune system is out of control and starts to attack the own body.

Draulans & Piot (2020), is an interview with Piot, who suffered the virus, and who says:

"Many people think COVID-19 kills 1% of patients, and the rest get away with some flu-like symptoms. But the story gets more complicated. Many people will be left with chronic kidney and heart problems. Even their neural system is disrupted. There will be hundreds of thousands of people worldwide, possibly more, who will need treatments such as renal dialysis for the rest of their lives. The more we learn about the coronavirus, the more questions arise. We are learning while we are sailing. That's why I get so annoyed by the many commentators on the sidelines who, without much insight, criticize the scientists and policymakers trying hard to get the epidemic under control. That's very unfair."

Viruses mutate fast, and are apt in exchanging parts of their RNA, and we would not want that this "multi-system attack property" is spread amongst viruses. Perhaps it is still feasible to choose for eradication. The example of smallpox, eradicated, and polio, though still with problems, is encouraging, see WHO GPEI & Heymann (2020). This notebook will not arrive at a choice, and only mention the options: either eradication or find a way to live with the virus till there is a vaccine. There is an overlap in measures: a discussion about "herd immunity" might also fit a scenario of eradication.

Piot also states: "Let's be clear: Without a coronavirus vaccine, we will never be able to live normally again. The only real exit strategy from this crisis is a vaccine that can be rolled out worldwide. That means producing billions of doses of it, which, in itself, is a huge challenge in terms of manufacturing logistics. And despite the efforts, it is still not even certain that developing a COVID-19 vaccine is possible."

1.6.4. Under normal conditions the virus can no longer be contained

If we want to eradicate SARS-CoV-2 then special conditions must apply. Under normal conditions it can no longer be contained in the world. The reasons for the latter are standard from an introductory course in infectious disease - see e.g. BCM (2020).

- getting a vaccine takes time (perhaps 1.5 years) … and then give it to 7.5 billion people on the planet

- health care is under strain, will perform less well, thus there will be breaches and ever newer infections
- the virus already shows mutations and is likely to continue to do so: every new mutation would require a swift reaction for new containment – but the health care system already is under strain
- special for SARS-CoV-2: the pandemic cannot be suppressed because of the *asymptomatic infectious*.

Kissler et al. (2020) at Harvard mention already two strains and suggest that intermediate periods of national lockdowns might be needed ("lock-on-off"). The logic is impeccable, and worthy of more than a century of research in epidemiology. The logic however leads to the absurd conclusion that we must destroy our economy in order to save lives, while the destruction of the economy also costs lives - not included in their study. Epidemiologists have lived too long in a world that accepted models without consequences for the economy.

Less developed nations lack the medical systems of the developed world, will be open to the virus, and will form breeding grounds for *new waves* and *mutations*. Developed nations might think that they might insulate themselves by containment and suppression, but it is dubious that they can insulate themselves forever. We can expect waves of new infections indeed, like with the flu, but then 10 times more infectious / deadlier than the flu, with the risk of shorter intervals because of faster mutations. The Northern hemisphere now benefits from the Summer, but in Autumn the reduced health because of the common cold and flu will combine with SARS-CoV-2, causing increased joint mortality. The Summer of 2020 should rather not be wasted.

There are more than 7.5 billion people on the planet. Not all will comply with quarantine or get the eventual vaccine in time. Viruses mutate. This virus derives from the common cold strain and people may lose immunity after a period. The economic price of national lockdowns is high and it seems that the virus is "not terrible or deadly enough" for people to bear that price, causing much indecision and delay perhaps till it is too late again.

1.6.5. Return to some normalcy, or consider what urgently requires improvement

With the pandemic still continuing, the medical world faces the challenge to restore care to the level of before the pandemic. Gupta (2020) reports that some 40% of normal care has fallen away because of the focus on SARS-CoV-2 and the fear amongst patients of getting infected in the hospital. De Rek & Goudsmit (2020) is an interview with virologist Jaap Goudsmit (involved in finding HIV) who comments that the emergency brake of the national lockdown contained the problem but that normal care ought to be restored, and: (i) there is a lack of data, (ii) reliance on models means also a reliance upon assumptions, (iii) there is too little attention for scenario's (the three scenario's presented by RIVM are not enough). See https://jaapgoudsmit-over-corona.nl/

The economy faces the similar challenge, see Baldwin (2020ab), Baldwin & Weder di Mauro (2020), Berger et al. (2020) and Acemoglu et al. (2020).

A core question is: if some return to normalcy would be possible today, in May through September 2020, then *why wasn't it considered in January ?* Was it really necessary for Western Europe to first have this wave and national lockdowns from February to May, before its CDCs, governments and

citizens would accept the idea that "normalcy" would require some adaptation ? If CDCs have only an advisory role, what happened at the *government* Departments of Health ? Indeed, it is reported that Taiwan, Hong Kong, Singapore and South Korea learned their lessons by the SARS-CoV-1 epidemic in 2003. *Would this mean that countries can only learn from disasters happening on their own territory ?* And if this would be the sad truth indeed, then *what have countries learned from the present episode ?* (It is said, however, that Singapore used the pandemic scenario developed in London, that London itself did not use. Thus Singapore can learn, and what is so special about Singapore ?)

RIVM, the Dutch counterpart of the US CDC, issued a health warning in January. RIVM (2020a) advised the Ministry to give alert status A to SARS-CoV-2, which e.g. allows mandatory quarantine. This status was assigned. (https://www.rivm.nl/meldingsplicht-infectieziekten/welke-infectieziekten-zijn-meldingsplichtig) However, when a general practitioner doctor (GP) reports a case, the case is only a "suspected case", and there is no legal obligation for microbiological testing, and somehow this then isn't done. Thus, Dutch statistics, with the official death toll of 5748 at the end of May 2020, are based upon hospital admissions. RIVM's modelers estimated the dynamic reproductive number $R[t]$ as getting below 1 by March 16, but based upon such hospital admissions where such testing is done. At the end of May, CBS Statistics Netherlands reported a total of 9000 "excess death", i.e. above the normal level expected from earlier years. Thus some 3000 cases (likely mostly reported) were left untested. *Why doesn't RIVM make sure that all suspected cases are tested and included in the official statistics ?* If making sure that statistics are collected is not in *their* job description, then in whose job description is it ?

Originally there was source and contact tracing by GGD but this was later repealed when the number of cases rose, instead of hiring more people to do the job. In May, the GGD director stated that they would be capable to take up source and contact tracing again. The media and eventually Parliament found it problematic that he mentioned some 2-3 hours per case, and that he intended to use of a letter of warning. Germany uses some 16 hours per case. A letter of warning might be used for a sexually transmitted infection while SARS-CoV-2 requires direct telephone calls.

If Holland had clammed down the infections from January like Taiwan did, and e.g. forbidden Carnival, and quarantined the provinces of Brabant and Limburg where they had the Carnival festivities anyway, then the economy would not have needed to lock down, and 9000 deaths by the end of May could have been prevented. Comparison with Malaysia Airlines MH17 with 298 deaths in 2014 might be seen as somewhat dramatic, but we are speaking about *thirty* of such disasters, caused by home-grown *failure of the protective system*.

So, if RIVM is only advisory, what prevented the Dutch *Ministry of Health* from seeing the notion of a *pandemic* in proper perspective, even while there had been ample warnings about pandemics before, and while *they are the professionals on public health* ? If we can understand what is blocking insights at government Departments of Health the world over, then perhaps we can find out what is blocking the *insight on preventing* pandemics in the first place. Yes, indeed "the world" is a big vague place outside of your own territory, but are you sure that you can really continue to neglect it ?

1.6.6. While there is no vaccine: options for an exit strategy

In all cases, there remains the challenge for the upcoming period till there would be a vaccine.

As an economist, I find it remarkable that the Dutch Prime Minister Mark Rutte (2020) (March 16) spoke about "herd immunity", first as one of the aims of the Dutch partial lockdown but later corrected to only a consequence and by-product of the real policy aim of flattening the curve. A crucial question is: if the government *allows* the virus to spread in a more natural way amongst the less-vulnerables, as a by-product, then why not use the virus as its own vaccine ? Apparently there is *only* the problem of *responsibility* - so to speak "*the (legal) blame*" - of *who causes an infection*, especially when it goes wrong: either some "natural cause", or the person self or a vaccinator. This issue of responsibility can better be discussed openly with an evaluation of the options, risks and costs involved. Also with an eye on future pandemics (that we rather ought to prevent though). Obviously, herd immunity is an objective in any vaccination programme, and vaccines differ from the dangerous pathogens: so the properly phrased question is whether this virus for this target subgroup of the less-vulnerables is like such vaccine and thus not the dangerous pathogen as it is for the vulnerables. One cannot avoid the impression that the Dutch government thought so, as it *allowed* the virus to spread (though asking for self-quarantine).

Another element is: a vaccine test on humans also requires deliberate infection. McKie (2020) reports in May about the Van Bunnik et al. (2020) paper: "UK scientists want to infect volunteers with Covid-19 in race to find vaccine. Trials could be speeded up by using risky strategy of deliberately introducing the virus". Critical researcher Riley at Edinburgh University: "Challenge studies are done for many diseases but only when strict criteria are followed. Firstly, the virus should be really well studied and its clinical behaviour understood in detail. It should also be incapable of causing severe illness in healthy individuals, or there should be a highly effective drug to clear the infection. None of these criteria are met for Covid-19, and I would be very concerned to hear challenge studies were being planned." The latter reasoning is within the well-defined world of medical testing and its protocols. It neglects the phenomenon that the lifting of lockdown will cause the virus to spread again *anyhow*, and that the economy collapses if we do not lift the lockdown.

There are four main scenario's on the table. All scenario's rely upon some notion of *infection status*. **Chapter 4** distinguishes actually 11 compartments with own risk profiles, and the compartments are indicated by a colouring scheme. An inspiration was my suggestion of having a "passport" for the status on sexually transmitted infections, see Colignatus (2004). At issue is not whether such compartments might exist, since any scenario uses their existence. At issue is whether it is feasible or even desirable to give *an official role* for such compartments.

(1) The current policy by the Dutch government: muddle through, wait and see, prepare ICU capacity for the Autumn, apparently especially for the vulnerable compartment that has only a 50% of survival at the ICU (see below "ad 1").

(2) Lock-on-off. The "Harvard study", or Kissler et al. (2020). RIVM commented on it favourably, but also called it risky, and it need not be the policy of the Dutch government. Kissler et al. (2020) point to the need to have (recurrent) lock-on-offs to "flatten the curve" and remain within the capacity of the health system. They suggest that over two years some "herd immunity" can arise as a by-

product as well. Thus, they allow infections, but shift the blame to "natural causes". They might also think that it is *impossible* to shield the vulnerables by quarantine so that there indeed would be a high demand for the ICU. Remarkably, RIVM & Van Dissel (2020ef) in May still discuss only the Harvard study but *not* the two following options that should already have been obvious in February too.

(3) Eradication. There is the *test, test, and test* approach, and then quarantine anybody who tests positive. Berger, Herkenhoff and Mongey (2020) provide a model-based *proof of concept*. Tabarrok (2020) supports the cost-effectiveness. Suppression fits the observation by Draulans and Piot (2020) that an infection may cause more damage than thought. It may be that the use of a SI(EY)A(CD) model is required for the proof of concept, but one might also argue that the method would be logical by itself too. The death toll could likely be calculated directly without a model run. In Holland in March, RIVM stated that there was a lack of tests. However, Alex Friedrich, head of microbiology and infection diseases of the University of Groningen medical center (UMCG), states that RIVM made a deliberate decision, and he arranged a testing strategy for Groningen, see 1V EenVandaag (2020) (April 17). RIVM misinformed the public about the options. At the end of May, the government and RIVM are making tests available for the whole country. Tweebeeke & Sietsma (2020) (June 10) report that elderly home care managers are distressed and furious that the much higher available capacity for testing has not been used at a much earlier moment.

(4) Segment and shield. A policy is to distinguish a compartment of vulnerables - say 60+ and those with comorbidity - and a compartment of less vulnerables - the younger others though perhaps still with some hidden risk. Van Bunnik et al. (2020) (May 5) call this "segmenting and shielding", see also Sample & Mason (2020). The issue is discussed by Colignatus (2020bcd) (March 31) and with a full model by epidemiologists De Vlas & Coffeng (2020) (April 12) at Erasmus MC, while Frijters (2020) (March 21) has much by implication. A non-quantitative discussion is by Eichenberger et al. (2020) (March 25), who also point to the *rising frustration by the immune compartment when they would be subjected to a string of lock-on-offs.* Criticism is: (a) Sridhar & Rafiei (2020) argue that it is *practically impossible* to shield the vulnerables. Government policies like in the UK and Holland already promised such shielding but actually only achieved this by the national lockdown. Adamik et al. (2020) alert us that families magnify the number of contacts. The proper question is: *Have we done enough to achieve the required shielding ?* (b) There is the protest that deliberate infection runs against the rules for vaccination. This is better seen in the light of the Draulans and Piot (2020) observation that it is a nasty virus that can have chronic effects. If there were no nasty effects then the principle on vaccination would actually be adhered to (which is also why the idea originated).

We arrive at the latter scenario *without* the SI(EY)A(CD) models. Thus, our praise for SI(EY)A(CD) is guarded. We could have arrived at such compartmentalisation also in February, right after the WHO declaration of the PHEIC. A modeler's question obviously is whether these two sections in the population can be put into the model, and how would they interact ? To some extent it is remarkable and even amazing that RIVM in its reports to Dutch parliament hasn't much reported on these two structural compartments, and still presents only a single $R[t]$, as if persons from the two compartments could freely mingle, and as if it should not be obvious that there is quite a distinction between the elderly and comorbids who would *always* have to be quarantined and the others for who contact reduction is *less relevant*.

Consider the following figures about the Dutch population, latest age-group data of 2019, using the level of 17.4 million of 2020. For SARS-CoV-2 there are 9000 "excess deaths" including 5748 official deaths, of which 177 below the age of 60, of which none in the ages 9-14. Let us allocate the 3252 = 9000 - 5748 unofficial deaths in proportion. The distinction of the chronically ill and multi-morbid are by Van Oostrom et al. (2011). Multi-morbidity is defined as having at least 2 chronic afflictions. Their age-group specific prevalences are applied to the age-groups of 2019. It is remarkable that the younger vulnerables have such a low SARS-CoV-2 death score, but they do not tend to live in care homes.

Category \ Age	< 60	60 +	Total
Population 2019	12.9	4.5	17.4 million
Chronically ill	3.4	3.1	6.4 million
Multi morbid	0.9	1.8	2.7 million
SARS – CoV – 2 dead	177	5571	5748 persons
Idem unofficial	100	3152	3252 persons

Let us identify the vulnerables as the 60+ and the younger chronically ill (which includes the multi-morbid), thus 4.5 + 3.4 ≈ 7.9 million (45%). Then the less-vulnerables are 9.5 million (55%).

Let us assume that the vulnerables are put into quarantine but that 1% of quarantines will be breached, with the group-specific IFF of 5.1% (observed for the 60+), meaning 7.9 million x 1% x 5.1% ≈ 4029 deaths. Potentially, all these deaths have tried the ICU with a 50% chance of survival (but still this net outcome). Regrettably there is no good indication for the IFF of the younger vulnerables.

The less-vulnerables will have the (observed) group-specific IFF of 0.06%. Avoiding the confusion of "herd immunity" we adopt 95% of the limit value Ap[∞]. With a R_0 = 4 (chosen from the literature), the limit value of all infections is Ap[∞] = 0.98. A value of 95% of this limit value is 93.1% and this is reached after some three months. This means 9.5 million x 93.1% x 0.06% ≈ 5307 deaths. All these youngsters will likely have been to the ICU, with a 90% chance at survival (and still said net outcome). The ICU will not be oversupplied (allowing for normal service too) if the infections take place in cohorts of 325,000 every 1.5 weeks, for a total of some 10 months. These cohorts are spread over hospital service areas. These assumptions give the following table, and a more detailed discussion with an Excel spreadsheet is provided by Colignatus (2020d) (update in the next week).

Category \ Compartment	Vulnerables	Less Vulnerables	Total
Population 2019	7.9	9.5	17.4 million
Infected : breaches vs deliberate	1 × %	93.1 × %	8.9 million
Infection Fatality Factor (IFF)	5.1 × %	0.06 × %	Quarantine average
SARS – CoV – 2 deaths	4029	5307	9336 persons
Maximum deaths (100 × %)	402 900	5700	408 600 persons

At issue is not the choice between *lockdown* versus "*let the pandemic run its course*". At issue is a sound (and not only preached) quarantine of the vulnerable *and possibly* deliberate (self-) infection of the less-vulnerable with sufficient medical backup if something goes wrong (because it remains risky to predict who would be less-vulnerable). This does not yet include the observation by Peter Piot on chronic effects of infection.

The scheme is not without risks, for example when people cross the quarantine barriers. This idea thus requires close discussion. See Van Bunnik et al. (2020) for their version of "segment and shield" and self-infection, see also McKie (2020). The idea to deliberately speed up the path to the limit

values provides a way to cope with the economic recession, which otherwise might develop into worse than the Depression of 1930-1941. The world has some luck that the SARS-CoV-2 pandemic hits an identifiable compartment in society. In another situation, when potential victims cannot be identified so easily, it would be much harder to think of a solution approach (other than *test, test, and test*). The sobering idea is the observation by Peter Piot that we actually may be in such a much harder situation, with a burden of disease that still must be diagnosed.

Above compartmentalisation could have been done in February with the information available at that time. Why were these policy options not formulated in this manner ? The potential answer is: RIVM might be rather a "*Bureau of Infection Statistics*" and not an "*Infection Planning Bureau*". Though there are "the three scenario's", also on the RIVM website, it still seems as if they do not have a tradition of making such scenario's, and if they do not wish to speculate about the number of deaths as if such discussion would make it seem as if they are responsible for those deaths. RIVM advises, they report and trace infections and advise on vaccinations rather than "plan" infections in the sense of advising on organising who gets what. Vaccinations are *designed to* protect the *vaccinator* from accusations of wrong-doing. By consequence, at the end of May, Holland has 9000 deaths, but precisely in the category of the vulnerables, the very group that RIVM advised to protect. Risk aversion behaviour caused the risk to actually happen (to others than RIVM itself). And while Holland in May somewhat recovers from the first wave, above scenario with another 9336 deaths is still in the waiting, at least if we want "segment and shield" as an exit strategy for the economic crisis.

(ad 1) We can compare with the Dutch planning on ICU beds, see NVIC & Gommers (2020) and NVIC (2020) (April 26), the Dutch association of intensive care doctors. Their calculation is elementary. Their population are the 20+ aged: 13,490,325. They observe that 4% has been infected (based upon testing of blood samples) and that this will grow to 60% in the next three years. (Possibly confused about herd immunity, or assuming a permanent quarantine of 40%, or assuming a R_0 = 1.5.) The 4% arose over 8 weeks, so that the other 56% would take some 112 weeks, but with a lower R_0 they now assume 156 weeks or three years. Of the infected, 0.45% arrives at the ICU. (Thus they assume that the vulnerables will not be protected, because only those have such high demand.) With 19 days per patient (indeed for the vulnerables) this translates into 700,000 bed-days. In the 8 weeks from March to April 2020 there were 2500 patients. With another 37500 persons to go, this would imply 15 other 8 week periods. While 1150 ICU beds is the Dutch standard, the new standard would be 1531 beds. A peak capacity of 2400 beds seems required. NVIC admits that regular care has suffered from the SARS-CoV-2 episode. What is remarkable in this calculation is that NVIC does not distinguish between the two compartments. It is remarkable that they do not mention the death toll: they plan beds, and if survival is only 50%, then this is not in their planning. They do not strongly advise that much more is done to *prevent* vulnerable patients to get infected in the first place. This kind of planning helps to obscure the issue. When there are sufficient beds available, then beginning-of-pipe organisations with care of the vulnerables might think that they have room to send patients downstream, while an infection may still come with quite some risk.

We will return to the issue in **Chapter 2**. In all cases, a discussion of these aspects requires an underlying understanding of epidemics. This translates as a basic understanding of models of epidemics.

1.6.7. Information that is missing now about such scenarios

Information that is missing now, and that we would need for informed decision making, is the following.

Common effect measures are lives extended ("lives saved") and (quality adjusted) life-years gained. Colignatus (2003, 2020) proposes the UnitSqrt compromise: though qaly's might no longer be needed, it still would be useful to calculate them to clarify the compromise. The issue is complicated by the Piot (2020) implication of chronic effects (an argument for eradication and an argument against deliberate infection: but there is uncertainty about the effect size), and we may include the increased awareness of the Post Intensive Care Syndrome (PICS). Such effects do not affect the death count but reduce the qaly's and increase economic costs. PM. Such effect measures require the extension of the modeling with the hospital sector, not included in the present SI(EY)A(CD) model family. Models with such enlargement are Goh (2020) and Berger et al. (2020) and Acemoglu et al. (2020). Relevant is Bodenstein, Corsetti, Guerrieri (2020), for the distinctions in economic sectors, and that rational policies, including lockdowns, are preferable above chaos. For the present modeling, I give priority on didactics above including such enlargements. (This second version includes a discussion of the life gain measures.)

While the alternative scenarios of eradication (assuming that asymptomatic infection allows this) versus "segment and shield" (with quarantine for the vulnerables and cohort-wise deliberate infection for the less vulnerables) seem rather well defined, it is still unclear what the current government policy is (except that it starts to look like the Harvard lock-on - lock-off), and it needs to be defined if we want to be able to compare outcomes.

For the base scenario, the first half of the year 2020 forms a sunk cost. We would be interested in an evaluation but there is no use in crying over spilled milk (except for restructuring the social decision making process). We better start from the present state of the economy and health system, and work from there. Since government policy can be expected to be erratic, we need a stable base scenario. The best selection is a path that avoids global warming and reduces the risks of future pandemics. It would be rather irrational to assume a path towards global warming with new pandemics, and then try to micro-manage the case of 2020 on this path. Indeed, it makes sense to assume a structural change, and then evaluate policies about SARS-CoV-2 in this context. See Colignatus (2020fg) for the Tinbergen & Hueting approach in the economics of National Accounts and ecological survival.

The required information is in the following table. We would gauge the costs not only in sizeable percentages of national income and national debt, but also in terms of unemployment, private debt and bankruptcy, and measures of inequality. We see a strange combination of both epidemiological modeling that reckons in days and macro-economic modeling that concerns the medium term (5 years) and the long term (the year 2100). A new topic is how the standard economic models would need to be adapted to the new circumstances - while it seems that we did not yet finish the discussion on including the financial crisis. PM. Mortality rates are w.r.t. the population (as opposed to the IFF that is w.r.t. the infected). PM. S&S = "segment and shield".

Category \ Compartment	Vulnerables	Less Vulnerables	Total
Population 2019	7.9	9.5	17.4 million
Mortality by SARS – CoV – 2	μ_{VC}	μ_{LC}	(2020 as sunk cost)
Mortality by reduced care	μ_{VR}	μ_{LR}	(take a longer horizon)
Mortality in eradication	μ_{VE}	μ_{LE}	(needs elaboration)
Mortality in S & S	μ_{VA}	μ_{LA}	(quarantine vs deliberate infection)
Lifeyears Lost by SARS – CoV – 2	λ_{VC}	λ_{LC}	(2020 as sunk cost)
Lifeyears lost by reduced care	λ_{VR}	λ_{LR}	(2020 as sunk cost)
Lifeyears lost in eradication	λ_{VE}	λ_{LE}	
Lifeyears lost in S & S	λ_{VA}	λ_{LA}	
UnitSqrt measure, govt policy	$\upsilon_{VC\&R}$	$\upsilon_{VL\&R}$	Colignatus (2003, 2020)
UnitSqrt in eradication	υ_{VE}	υ_{LE}	
UnitSqrt in S & S	υ_{VA}	υ_{LA}	
Economic cost, govt policy			(likely aggregate only)
Economic cost in eradication			(cost defined w.r.t. base scenario)
Economic cost in S & S			

1.7. Summary of a complex objective

In summary, this notebook has a complex objective:

(a) Redesign of didactics about SI(EY)A(CD) in order to reduce confusion.
(b) Document the latter within the computer algebra environment of *Mathematica.*
(c) Middle of the road: we may assume novice readers but do not aspire at an introductory *course* for novices, and allow for refreshment for modelers.
(d) Make this modeling and documentation very accessible for others, also for variants.
(e) Use SARS-CoV-2 parameter estimates and scenarios as examples for such SI(EY)A(CD).
(f) Place the discussion within the literature, with real data and commentary about the quality of the data.
(g) Allow for some creativity along the way, like the use of colour-coding for (quarantine) compartments.
(h) Give the context of above *exit strategies* of eradication or using the virus as its own vaccine.
(i) Identify what information is missing, which we need when we want to make informed decisions.
(j) Regard the pandemic as only an example of a much more fundamental issue, see Colignatus (2020a).

SI(EY)A(CD) is only a very basic model. It is adequate only for *outlining* the *exit strategies.* It allows us to show that the current (partial) lockdown is wise as an emergency brake - for Holland - given the otherwise deadly outbreak, with a rough indication of how many deaths would have occurred otherwise. It allows us to observe that said *exit strategies* were already feasible in February, so that the lockdown in March was an *emergency brake* indeed. It also allows us (in a later notebook) to evaluate the (quality adjusted) life-years gained, since we should not focus on the mortality count only.

1.8. Structure of this notebook

The SIA, SICD and SEYCD models are discussed sequentially in **Chapters 5-9**. Each opens with a first practical chapter on model structure and how to run the software and proceeds with a second chapter on analytical properties (supported by other routines). There is no analytical chapter on

SICD since the model merely splits A = *C* + *D.*

The SI(EY)A(CD) model family is formulated for infections but is also applied for symptomatic disease. **Chapter 10** thus discusses the distinction between the infection *generation interval* and the symptomatic *serial interval.* There is no suggestion of new didactics on this, and the discussion here is only provided for completeness, taking advantage of the best presentations that I have seen elsewhere.

Chapter 11 compares with four empirical studies, reproducing their model within the present environment. Those studies tend to use symptomatic disease, for the obvious reasons that patients turn up when they have symptoms and that (lab) testing is not easily arranged. **Chapter 11** does actual model runs, while some such comparisons are already mentioned in **Chapter 10**.

Intervention and estimation have a complexity of their own. **Chapter 12** gives a short overview. One method is the use of full separate models. Alternatively, there is also *Mathematica's* WhenEvent feature that can be used fruitfully for estimation. **Chapter 13** compares base and alternative scenarios using full SEYCD models. **Chapter 14** discusses the death toll data that might be used for estimation. **Chapter 15** uses the WhenEvent structure for R_0, with a "model insert" SEYCDT - "T" for "time of intervention". **Chapter 16** is analogous, with SEYCDB for β. The last subsection of the chapter does a "post-mortem" on γ, the result of which has been shown in the *Introduction,* section 1.4.6. In all estimation we regard $\gamma = 1/10$ as a given normalisation parameter.

This present notebook is not a developed "course". There are no *student syllabus* and *teacher's manual.* This notebook serves both such purposes now. The assumption is that new students look at all material and that experienced readers (teachers) might take bigger steps. In some sense, the *Introduction* chapter is a teacher's manual. Novice readers will understand little about this *Introduction* but can return to it after reading the subsequent chapters. PM. This notebook and package target *didactics,* and do not target *education*. Still, while writing, I have been increasingly assuming that this notebook might also be used by first year students under guidance of their teachers. Supplementary introductory comments for this have been put in **Appendices A, B** and **C**. Appendix A is advisable and B & C may be useful.

Appendix A discusses the basic *Success versus Failure* model, with exponential decay of Success, and with Failure collecting the fall-out. The model is basic for understanding the relationship between *factor* and *rate*, the relation between rate and period = 1/rate, the notion of a half-life, and for the comparison of dynamic events and the exponential probability distribution. **Appendix B** refers to my background on queueing theory and related packages in *The Economics Pack*. Epidemiological models with interarrival periods have a conceptual overlap with some queueing theory. The distinction between exposed and infectious phases concerns sequential processing but the distinction between cleared and deceased is not parallel processing but merely proportional allocation. **Appendix C** refers to other survival analysis packages in *The Economics Pac*k. **Appendix D** shows that one does not need this notebook and package. For who knows the S(E)IR(D) models and *Mathematica,* it is simple and straightforward to create such model in *Mathematica* and run a WhenEvent policy intervention with it. The present notebook and package only provide for some structure, and intended didactics. The present notebook and package also confront the user with a learning curve which might slow down needlessly. *Mathematica* is a wonderful computer algebra environment but has a learning curve of its own. There is always the issue that when some routine

doesn't (quite) do what one wishes, or doesn't work for other applications than shown in the example, then it might be easier to adapt or trace errors when you have written the routine yourself. For example, the SI(EY)A(CD) model family has extensions with Birth and background death, that allow for annual dynamics in infection processes (and a real steady state), and likely more relevant notions of herd immunity: this is not implemented here. A user faces the question whether to try to adapt what has been presented here or find alternatives.

These Chapters on SI(EY)A(CD) are preceded by the following three chapters. **Chapter 3** is on the Survival`SIA` package, its limitations, and repeats the aims in didactics and application, but now directed at the level of programming. **Chapter 4** discusses compartments and quarantine colour coding. There are more compartments in the real world than used in SI(EY)A(CD). For a policy discussion on SARS-CoV-2 it is relevant to distinguish between the more vulnerables and the less vulnerables, but these compartments are not in the models that we look at. Thinking about quarantine caused me to design a colour scheme anyway. This scheme was at least useful for the current plotting of SI(EY)A(CD) graphs.

Finally there is **Chapter 2**, following next, about the Dutch timeline and some issues of strategy and policy making. The Dutch performance is actually not so well. While the WHO issued its PHEIC on January 30, i.e. warned about the risk that a pandemic might be developing, the month of February was basically not used in Holland. The period of the (partial) lockdown since March 18 has passed much in waiting and without developing clarity about an exit strategy. Is the world going to try to exterminate SARS-CoV-2 or are we going to accept and perhaps even target some (potentially confused notion of) herd immunity or .... ? Such issues of strategy however are not in the SI(EY)A(CD) models themselves. This notebook and package were written partly to clarify that some points ought to be clear at the foundational level. The focus was on didactics and there has been no effort at economic modeling.

1.9. About the author

1.9.1. Caveat

The caveat is that I am an econometrician (Groningen 1982) and teacher of mathematics (Leiden 2008) and no medical doctor or epidemiologist. If SARS-CoV-2 is *a virus that humanity cannot live with* - like with smallpox that was eradicated in 1980, see WHO GPEI & Heymann (2020) - for example because the virus has risky properties also for future mutations, then the current low prevalence might be a good start for eradication via surveillance and quarantine for the infected. This might come at the economic price of national lockdown and the burden of disease because of the deterioration of common health care. It might be unwise to *target* for herd immunity, even though such might be seen as part of eradication. It depends upon the properties whether stable immunity can be created indeed. Corona viruses of the "common cold" tend to be forgotten by the immune system. At first I did not think that containment / suppression / eradication was really a feasible option - Taiwan would remain an island - but the Piot (2020) warning and the Berger et al. (2020) paper and Tabarrok (2020) commentary caused me to change views. For now, it is best to present and discuss scenario's, and subsequently enquire for medical views.

1.9.2. Work related to medicine

In 2002-2004 I collaborated at Erasmus Medical Center on the modeling of the Human Papilloma Virus (HPV) as the cause of cervical cancer. I gave a presentation on MISCAN modeling at a gynaecology conference in Paris and had a poster at the 4th IHEA world congress at San Francisco on public health economics. My background in modeling and also logistics was relevant because diseases may look like a Markov logistics process with stages and transition probabilities. There can be the same issues of test reliability, criteria of lives-saved or life-years-gained, and cost-effectiveness of screening and treatment. Jan Barendregt (who sadly passed away in 2017 and who will be remembered for DisMod II / EpiGear) recommended Manton & Stallard (1988) and lent me his copy (that I found excellent indeed). I also followed the discussion about the SARS epidemic of 2003. My period at Erasmus MC was too short to allow for publishing peer reviewed papers but let me mention two working papers of that period, and actually another one of 2019, all with software that eventually was included in *The Economics Pack. Applications of Mathematica,* see Colignatus (1995, 2020e). Currently I am in a phase towards retirement, and consider the purpose of the Pack as shifting towards didactics and education.

(1) Colignatus (2003, 2020): *On the value of life.* This compares the lives-saved and life-years-gained measures, and develops a compromise: a "unit-square-root" (UnitSqrt) measure, that regards each life as 100% and takes the square root of the relative gain (relative to the 100% including that gain). It compares somewhat with the "fair innings" notion in the value attached to youngsters but doesn't have a minimum age standard but a 100% standard. (Perhaps it has "fair outings".) The relation to SARS-CoV-2 is discussed by Colignatus (2020b). Discussion of the value of life meets with the problem of the "cricket and the ant". If there would be an ideal scoring measure for ants, then what to do with a cricket that has been living irresponsibly and that, facing death, wants to be treated like an ant ? The UnitSqrt method would apply for a general class of deserving persons, consisting of generally responsible behaviour, likely with a grace period up to 25 years where youngsters might need to learn from accidents, and the occasional insurable glitches for ages 25+. I have no developed ideas for chronic irresponsible behaviour of the 25+ ages who burden the public health system. However, the latter discussion should not be confused with the earlier topic of valuing normal decent lives (e.g. the lives of doctors and health workers). While the 2003 paper was written in the context of SARS, the 2020 SARS-CoV-2 pandemic caused me to update the paper with a more explicit comparison with Fair Innings and Proportional Shortfall. I now included the UnitSqrt measure in my software presentations of the life table alongside life expectancy, and the discussion below in section 7.11 shows this enhancement of clarity.

(2) Colignatus (2004): *Modifying behaviour with a passport for sexually transmitted infections* (STI). At that time there was no HPV-vaccine yet. An option was to manage human behaviour. The status of infection can be recorded in the medical dossier: free (green) or carrier (red). While children can gets warts, an assumption might be that children start out uninfected by the harmful HPV variants (status green). When adults meet and want to get into a serious relationship – in the sense of sharing their germs – then they can show each other their status of infection in their medical dossier and discuss the implications. From this working paper, we may take the idea of recording the status of infection, and using colour coding for clear communication. For SARS-CoV-2, it is

better to use "red" (alarm, or hungry in Chinese restaurants) for the barrier between zones and groups. This is discussed by Colignatus (2020c). PM. Due to the national lockdowns, we may also see a reduction in STIs, and one may wonder whether behaviour oriented policies should not have been more prominent in the past (catching HIV before it hibernates).

(3) Colignatus (2019): *Population size and climate change.* Encourage women to delay having their first child till they are of age 30+. This will give a crucial contribution in reducing climate change, especially in the crucial decade 2020-2030. Colignatus (2020a) discusses the gap between demographers and researchers on climate change.

2. Timeline and assumptions of the Dutch case

2.1. Introduction

This Chapter assumes that the reader has knowledge about the SI(EY)A(CD) models. Readers new to the subject are advised to reserve this chapter for later. This especially holds if you already had problems with the *Introduction.*

The idea of this notebook is to improve the didactics of SI(EY)A(CD) while using the SARS-CoV-2 case for Holland as an example. The ongoing pandemic is a prime motivation to look at these models, and we learn about both the models and the virus by looking at the practical issue. This Chapter considers the timeline of the Dutch case. Dutch readers may benefit from the review in NRC-Handelsblad by Derk Stokmans and Mark Lievisse Adriaanse (2020).

The timeline reported here basically stopped around June 15. The update in this 2nd version of this notebook till the end of August concerned the modeling of intervention, vaccination and herd resistance, see section 6.12, and the use of the life gain measures and life table method, see section 7.11 and **Appendix E**. In June I advised that the Summer of 2020 should not be lost, and now at the end of August it seems that it has been lost. A new development (though not unexpected for a virus like the common cold) is that there now are cases for who immunity has not been retained.

2.2. Properties of SARS-CoV-2 and some support for the Dutch policy statement of 2020-03-16

SARS-CoV-2 can be deadly for (i) *younger persons with comorbidity* and (ii) *the elderly* (60+) (often with comorbidity). This group can be called "the vulnerable".

For *younger and healthier* persons, SARS-CoV-2 may compare to the common cold - though this issue is not clear-cut, see section 1.6.3 that refers to Draulans & Piot (2020) with the warning about chronic effects. With respect to short term fatality, this group can still be called "*the less vulnerable*". There is always a risk that such person is misclassified and actually appears to be vulnerable.

In an address to the nation of March 16, Dutch prime-minister Mark Rutte (2020) formulated the policy:

(1) Protect the vulnerable
(2) Remain within the capacity of the health system, and in particular the intensive care units (ICU)

(3) Allow the less vulnerable to build up herd immunity.

At first it was presented as if herd immunity was a *policy objective*. In discussion (also) with Dutch parliament Rutte clarified that the 3rd aspect was only a *consequence* of the first two aspects.

Ever more measures were imposed, short of a *total* lockdown of Holland that would require permits for leaving one's home. Eventually, reports were focusing on the hope that the basic reproduction factor was getting below 1. However, in sum, containment / suppression hasn't been an longer term policy objective (yet).

In the press conference on April 21 2020, the policy was (re-) formulated:

(4) Wait and see, collect data and policy options, because we are still much in the dark
(5) Develop instruments for better surveillance, like an "app", for a potential long term policy objective of containment and suppression
(6) Restart the health system on normal care, which had deteriorated by the focus on the pandemic
(7) Release a few restrictions, like allowing young children to go to school and participate in sports.

Official Dutch policy is much guided by the lack of good data (see below), but it also seems that the policy options are not fully developed yet. The ship sails between the cliffs of a deadly outbreak and economic collapse (with additional collapse of the health system). This situation is not different at the beginning of June.

The following discussion basically supports points (1) - (3) above, and allows for the notion of *deliberately using the virus as its own vaccine*, namely for the less vulnerable group, while still remaining within the confines of the health system (ICU) and protecting the vulnerable. The issue is wide open however, given the nasty properties of the virus that are asking for eradication.

2.3. Relating to the RIVM data and parameters (1)

RIVM is the Dutch counterpart to the US CDC or EU ECDC. The RIVM modeling is much more complex than SI(EY)A(CD) and it is dubious to try to give translations for the relatively simple SI(EY)A(CD) model. For example, the SIA effective reproductive factor is $R[t] = R_0 \, \text{Sp}[t]$, and RIVM also publishes an effective reproduction factor $R_{\text{RIVM}}[t_{\text{Cal}}]$ indexed on the day of the calendar: which puts us at a loss whether we can take $t = 0$ for the first day of this calendar series, with $R_0 = R[0]$ for our simple case, especially when the $R_{\text{RIVM}}[t_{\text{Cal}}]$ in the graph starts below 2 while RIVM has stated that R_0 would be 2.7, then 2.3 and then 2.5. The views and estimates of RIVM are also evolving with the Dutch data and interventions over the period, so there is no "the" view. When this notebook and package were written, basically starting March, the parameters mentioned in some RIVM documents caused too many questions, and perhaps I simply did not locate the relevant documents. Thus, for the present didactic exposition, I selected $R_0 = 4$, $\beta = 0.4$, $\gamma = 1/10$ and $\alpha = 1$ based upon the international literature, in which Dutch authors were actually involved too. Since the notebook and package are oriented at didactics, the difference with (recent) views with RIVM does not matter much, though it remains relevant for the suggestion of "taking the example of Holland". The following discussion first presents the choices for this notebook and package starting in March. Section 2.9 below compares with the RIVM view of May. The following table already reviews the differences.

	RIVM (January – May)	This notebook
Objective	R[t] for intervention	Didactics
Onset in Holland	February 10 (footnote)	February 23 with Seed 100
R_0	2.7 then 2.3 then 2.5	4
Generation interval	"Generation time" first 6 then 3 – 5	$1/\alpha + 1/\gamma = 11$ days
Doubling period	5.5	2.3 days
Implied β	0.2	0.4
(Implied) γ	0.09016 (using $\alpha = 1$)	0.1
φ (IFF)	0.0138	0.015
Incubation	6	6 days
Incubation longer than generation time	yes	no
Infectious period	implied $11.1 = 1/0.09$	10 days
Asymptomatic infectiousness	1 – 3 days (related to the above ?) $1/\alpha = 6 - 3 = 3$ days ? $\alpha = 1/3$?	5 = 6 – 1 days
Period from infection to Hospital admission	14 days	

Footnote: This first date seems a result of later back-tracing. It was not widely published at the time (did not stop Carnival).

RIVM presents the following graphs as input and result from their model on $R[t]$. Their approach differs from our present simple models. The inference that $R[t]$ dropped below 1 on March 16 can only hold for the official death counts; additional CBS data show a spike of some 3000 additional "excess deaths" in April. Many patients, supported by their families and GPs, decided to no longer report to the hospital, since they were afraid of the prospect of dying alone and unconsciously in an ICU bed. It is fully unclear what this means for the dynamics of the process and the estimate. PM. See section 5.2 on the terminology of the effective $R[t]$.

gebaseerd op ziekenhuis opnames uit OSIRIS data 2020-04-17
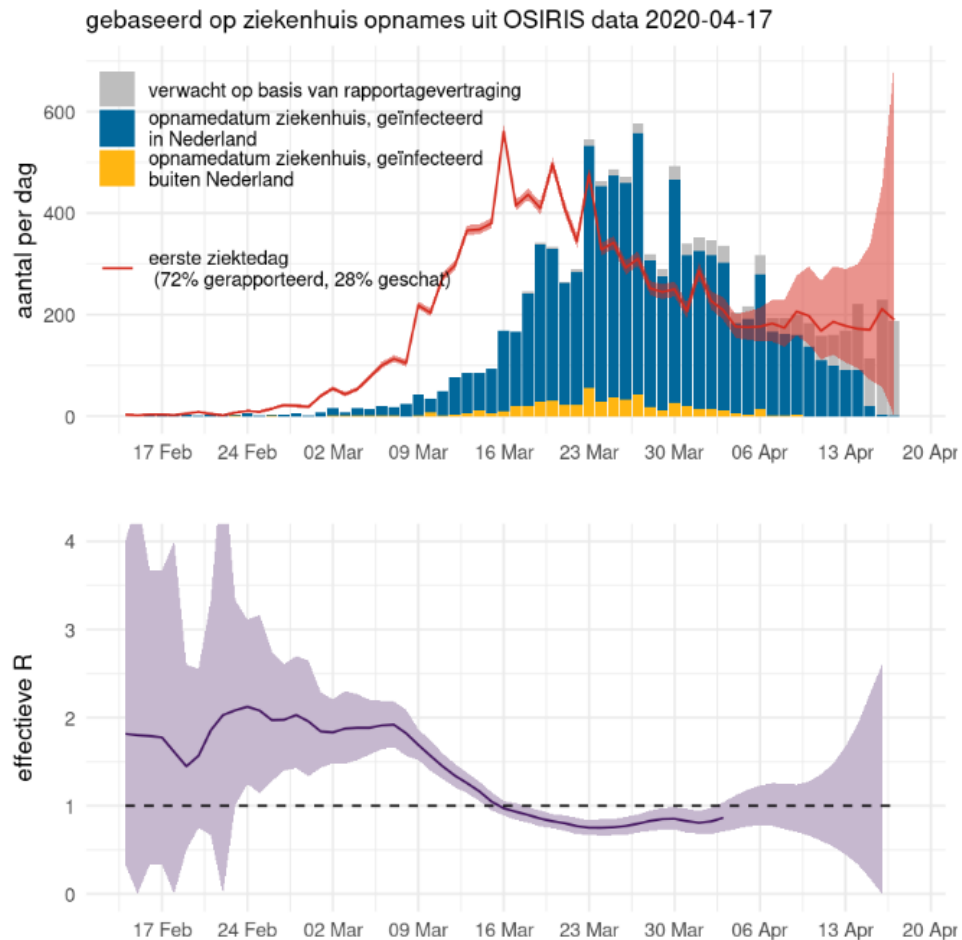
PM. Documentation of the RIVM modeling: https://www.rivm.nl/en/novel-coronavirus-covid-19/calculations/calculation-models-are-public-and-accessible

2.4. Basic parameters R_0, α, β and γ for unmitigated spreading before the lockdown

2.4.1. Choice of latency α

Anderson et al. (2020), Supplement 3, suggest a value of 1 day for the latency of the Exposed. There are not really competing alternatives. Thus α = 1 / (1 Day) = 1 too.

2.4.2. Choice of R_0

Flaxman et al. (2020), in a study on 11 European nations, find a R_0 = 3.87 for SARS-CoV-2.

Let us take 4. We write $R[0] = R_0 = 4$. We need to know the time-dependent reproduction factor $R[t]$ too. Flaxman et al. (2020:25) allow that a national lockdown takes 4 days to take full effect, and when it happens then the effect is *uncertainly* around $R0 ≈ 1$. For the smaller and closer knit country of Holland we might take 3 days for full effect.

2.4.3. Choice of γ

For this notebook we want to have a value of the infectious period $1/\gamma$ such that the natural process of clearing has been observed without confounding and intervention, so that we can subsequently insert the assumption that this clearing period would also be the (natural) infectious period.

Kretzschmar et al. (2020) (May 15) take an infectious period of $1/\gamma = 10$ days, though with variable infectiousness, which doesn't quite fit the assumptions of a constant rate in SI(EY)A(CD). But a constant rate can be seen as a special case of a variable rate.

Li et al. (2020) (January 29 & 31) gave one of the first estimates. They have 10 confirmed cases with a mean incubation period of 5.2 days. Based upon 6 pairs of cases, they find a serial interval with a mean of 7.5 days with a sd of 3.4 days and a 95% confidence interval of 5.3 to 19 days. They present an estimate of R_0 and a doubling period.

For SIA, we have the following equations for the growth rate $r = \beta - \gamma$:

```
eqs = {r0 == 2.2, r0 == beta / gamma, beta - gamma == Log[2] / 7.4}
```
$\left\{r0 = 2.2, r0 = \dfrac{beta}{gamma}, beta - gamma = 0.0936685\right\}$

```
sol = Solve[eqs, {r0, beta, gamma}] // Quiet
```
{{r0 → 2.2, beta → 0.171726, gamma → 0.0780571}}

```
1 / gamma /. sol[[1]]
```
12.8111

For SEYCD we have the following equations for the growth rate, for us with $\alpha = 1$:

```
eqs = {r0 == 2.2, r0 == beta / gamma,
  growth == Log[2] / 7.4, r0 == (1 + growth) (1 + growth / gamma) }
```
$\left\{r0 = 2.2, r0 = \dfrac{beta}{gamma}, growth = 0.0936685, r0 = (growth + 1)\left(\dfrac{growth}{gamma} + 1\right)\right\}$

```
sol = Solve[eqs, {r0, beta, gamma, growth}] // Quiet
```
{{r0 → 2.2, beta → 0.203712, gamma → 0.0925964, growth → 0.0936685}}

```
1 / gamma /. sol[[1]]
```
10.7996

Joseph Wu et al. (2020) use cases "exported from Wuhan internationally". They present an estimate of R_0 and a doubling period.

For SIA:

```
eqs = {r0 == 2.68, r0 == beta / gamma, beta - gamma == Log[2] / 6.4}
```
$\left\{r0 = 2.68, r0 = \dfrac{beta}{gamma}, beta - gamma = 0.108304\right\}$

```
sol = Solve[eqs, {r0, beta, gamma}] // Quiet
```
{{r0 → 2.68, beta → 0.172771, gamma → 0.0644668}}

```
1 / gamma /. sol[[1]]
```
15.5119

For SEYCD with $\alpha = 1$:

```
eqs = {r0 == 2.2, r0 == beta / gamma,
  growth == Log[2] / 6.4, r0 == (1 + growth) (1 + growth / gamma) }
```

$$\left\{ r0 = 2.2,\ r0 = \frac{beta}{gamma},\ growth = 0.108304,\ r0 = (growth + 1)\left(\frac{growth}{gamma} + 1\right)\right\}$$

```
sol = Solve[eqs, {r0, beta, gamma, growth}] // Quiet
```

{{r0 → 2.2, beta → 0.241894, gamma → 0.109952, growth → 0.108304}}

```
1 / gamma /. sol[[1]]
```

9.09488

Rocklöv et al. (2020) about the *Diamond Princess* state: "The cruise ship conditions clearly amplified an already highly transmissible disease." This will indeed complicate the estimate on β and R_0, but need not affect the infectious period when there is no quarantine yet. However, they do not estimate but choose $1/\gamma = 10$ days, simply stating "*In the homogeneous model, the infectious period, i, of COVID-19 was set to be 10 days based on previous findings*", where they refer to Li et al. (2020), where we see $1/\gamma = 12.8$ (SIA). They also refer to Joseph Wu et al. (2020) for their modeling (with a higher $1/\gamma$ for SIA). On the *Diamond Princess,* apparently an index case got on board around January 21-25, and by February 20 619 persons of 3700 passengers and crew (17%) tested positive. The outbreak of the infection was noted on February 3 with 10 cases, with subsequent measures on February 4. The authors write take $1/\gamma = 10$ for $t < 16$ and $1/\gamma = 4$ for $t \geq 16$, counting from (and including) January 21.

```
DayCount[{2020, 01, 21}, {2020, 02, 04}]
```

14

Rocklöv et al. (2020) also take an *incubation* period as the parameter for their "exposed" category, suggesting that their "SEIR" model does not concern *infections* but *symptomatic disease.* Their $\alpha = 1$ / (5 days) = 0.2. For infections, we took $\alpha = 1$.

We can agree that $1/\gamma = 10$ is a *conservative* assumption w.r.t. the (longer SIA) estimates by Li et al. (2020) and Joseph Wu et al. (2020), in the sense of a faster indication of the risk. This gives us the γ that we are looking for.

PM 1. The distinction between *infection* and *symptom* will be discussed in Chapter 10. Let us first develop the models with a focus on *infection* only, so that we are in a better position to understand the complication of *symptoms* (especially for parameter estimation).

PM 2. SIA, SICD and SI(EY)CD assume that the mean clearing period is also the mean infectious period. Therefor the naming and value-assignment of $1 / \gamma$ causes some difficulty. A clearing period is biologically given and cannot be easily changed. It is a property of the person rather than the process of infection, that can involve (self-) quarantine. A choice of a mean clearing period of 10 days seems acceptable for a younger population, if the virus would only circulate there. The period would be longer for elderly persons who do not cure as fast, but they would get sick and soon be less in contact with others. Overall, the actual clearing is not relevant for us, since we do not look (yet) at the further process towards (self-) isolation or hospitals. For the present modeling exercise,

it suffices that the period of infectiousness has stopped after the mean value of 10 days.

PM 3. On May 15 2020, the European CDC reports for the general public: "The infectious period may begin one to two days before symptoms appear, but people are likely most infectious during the symptomatic period, even if symptoms are mild and very non-specific. The infectious period is now estimated to last for 7-12 days in moderate cases and up to two weeks on average in severe cases." https://www.ecdc.europa.eu/en/covid-19/questions-answers. However, they do not provide a reference. It may well be that their "infectious period" doesn't quite apply to the kind of modeling that we intend. Potentially they refer to the 99% confidence value required to keep people in quarantine, while the SI(EY)A(CD) model requires a mean value.

2.4.4. Consequence for β

With $R_0 = 4$ and $\gamma = 1/10$ we find that $\beta[0] = \beta = 0.4$, with a doubling period of $\text{Log}[2] / (\beta - \gamma) = 2.3$ days. See section 5.4 for the meaning of β. In principle R_0 is a dimensionless factor, $1/\gamma$ has the dimension of *duration,* and with $R_0 = \beta / \gamma$ then β has the dimension of $1/duration$. The interpretation of β as the number of infectious contacts per (infectious) period requires that R_0 has the dimension of a number. This is immaterial for the point that contacts might be influenced by behavioural or non-pharmaceutical interventions like quarantine. In any case we get $\beta = \beta[t]$. On March 12 2020, the Dutch government decided to a (partial) national lockdown, and closed the schools on March 15, so that β has changed since then. Presently we continue with $\beta = 0.4$, and look at a change when discussing intervention and estimation.

PM. Anderson et al. (2020) look at the original situation in China, and report a doubling time in the early phase of 4-5 days, or say $r = \beta - \gamma = \text{Log}[2] / 5 \approx 0.14$. They suggest a $R_0 = 2.5$ and herd immunity of some 60%, which for a $\gamma = 0.1$ fits a $\beta = 0.25$. The value of Flaxman et al. of R_0 for Europe seems more relevant however. Liu et al. (2020) give comparisons, and there is quite some variety depending upon circumstances. It is not clear to what extent Chinese citizens already practiced (self-)quarantine. PM. Observe that α has no role here, except for delay.

NB 1. There is a difference between the internationally applied "full lockdown" (people may no longer leave their home, except for one person who gets the necessary supplies, and the personal distance is 2 meters) and the Dutch "partial lockdown" (in which people might on occasion go for a stroll but keep a distance of 1.5 meters).

NB 2. The Adamik et al. (2020) study of the MOCOS groups is important. They warn that the number of contacts should rather be based upon households, since household members infect each other. They do a micro-simulation rather than this SI(EY)A(CD) model. PM. They also nicely phrase the dichotomy between "infect nobody" and "infect towards herd immunity" though they do not mention the issue of overshoot, see our discussion of "herd resistance": "Mitigation of a novel infectious disease with the aim to reach herd immunity is a classical textbook concept in epidemiology and has been successfully applied in the past, foremost in the case of novel influenza strains (…). The idea is simple: in the absence of a vaccination for a novel infectious disease one tries to flatten the incidence curve to such an extent that the daily number of cases that require medical assistance is kept below the capacity of the health care system. The long term goals are to obtain a sufficiently large fraction of the population that has become infected and to reach herd immunity

which would lead to a less severe or even subcritical second outbreak wave. On the other hand, an extinction strategy would aim at introducing sufficient contact reductions to keep the epidemic subcritical and not lifting these restrictions until the disease becomes extinct."

2.4.5. General understanding about the choice of R_0

The following two plots give an indication how $R_0 = 4$ is located within the ranges. The *acquitted* compartment tells how large the impact of the infection will be eventually, with the limit in $A[\infty]$. Thus we concentrate on this compartment. (The distinction within the *acquitted* between the *cleared* and *deceased* is a later issue (SICD).)

The first plot below gives the contour lines of the acquitted A as a function of the basic reproduction factor. See a discussion in Section 6.10 below. Observe that the 50% level (62 days) does not have to be in the legend, and that there cannot be a number of days for the limit value because this would be infinite. (Potentially, though, we might say that the infection is exterminated when the value of the infected units drops below 1 or 0.5 (rounding error), but this does not work well for our interpolations.)

Observe: (i) After $R_0 = 4$, higher values have remarkably less impact on the outcome. Below in estimation, we will find a possible value of $R_0 = 5.7$ but this hardly affects the death toll anymore. (ii) The values of $R_0 = 4$ and $R_0 = 0.5$ (below 1) can be compared by inferring that the number of contacts must be 4 / 0.5 = 8 times reduced. This may be difficult, when people live in households, infect each other within households, and as a household have many more contacts than a single unit, see Adamik et al. (2020) of the MOCOS group.

(PM. The plot uses a the current variables. In this case, first a SI(EY)CD model with a latency of 1 day was run, which means a delay of about 14 days w.r.t. the SIA(CD) without latency. The model has $I = E + Y$ = Exposed + Infectious, so that the variable I of infected can directly be compared to the I in SIA(CD) (with $E = 0$).)

```
ApByR0Plot[] (* locked from an earlier run with SEYCD and α = 1;
see also section 6.10*)
```
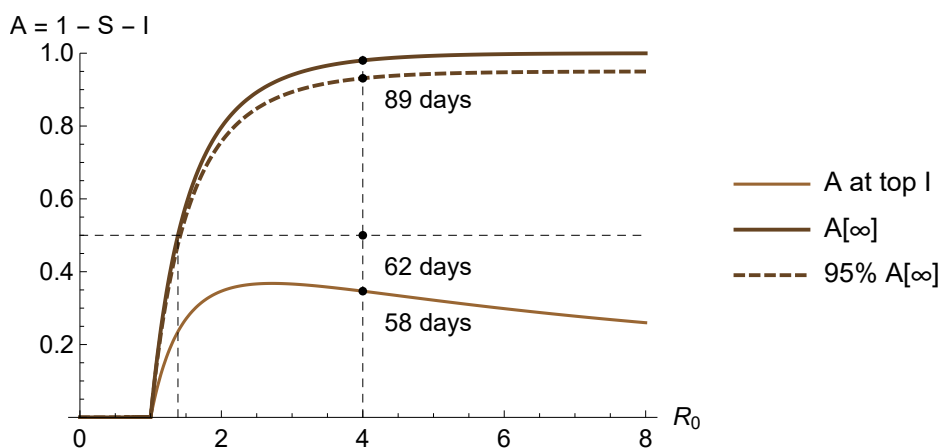

The line for the top is downward sloping after a certain point: a higher value of R_0 will cause a higher infection level for I at the top, and thus there will be a lower value for A at that time.

The second plot gives the contour lines for the number of days as a function of the basic reproduc-

tion factor. One may check the number of days at R_0 = 4. PM. The contour of 95% Ap[∞] can be crossed because it is not the limit value itself.

```
TimeByR0[ListLinePlot, SEYCDTimeAsFunctionOfR0[List]]
(* former run, read from file at startup *)
```
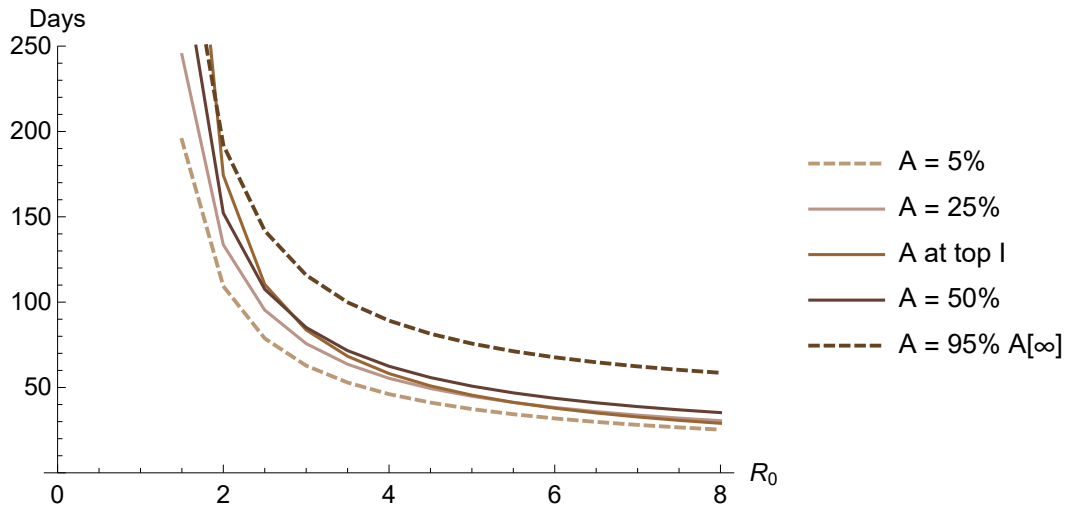

It is surprising that the dynamics of the process can be caught in a single parameter R_0. The word of caution of course is that the parameter actually changes, and cannot simply be assumed to be constant.

2.5. Choice of Infection Fatality Factor (φ = IFF)

Verity et al. (2020) find an Infection Fatality Factor (IFF) for China of 0.66%. Ferguson et al. (2020) adapt to the UK population profile and find 0.9%. Applying their age-group specific fatality factors (that they call a "ratio", their Table 1) to the Dutch age composition causes a Dutch infection fatality factor of 1.27%. Holland has on average an older society.

However, we will look at this in section 2.7 in more detail. A summary is: The Dutch data are quite unreliable, with underreporting, lack of testing, testing focused in hospital workers, changes in practices (whom to send to a hospital), and after March 15-18 the national intervention. However, the age-group of 70-79 seems most reliable, since they tend to still participate in society but do not work at hospitals. For this age-group, Ferguson et al. have an infection fatality factor of 5.1% (that they call a "ratio"). When we apply this to the Dutch deaths in this age-group (certified for SARS-CoV-2 with the current reporting practices) then the prevalence of infections for them (on some *n* days before April 2) was 0.535%. When we assume this to be the national infection prevalence, then we get an age profile of infections and deaths that is not unreasonable, see Colignatus (2020cd), with also adoption of the Ferguson IFF for the 80+. In that case, the Dutch IFF is 1.45% instead of 1.27%. This is considerably higher than the UK 0.9%, but Holland is apparently not only an older but also a more densely populated country after all. This translates into a SI(EY)CD model input parameter, rounded to φ = 0.015 = PhiSICD[] = IFF, or the Infection Fatality Rate IFR = 0.015 * γ = 0.0015.

2.6. A timeline with scores from a SEYCD model without endogenous reaction or intervention

With these parameters for the standard SEYCD model, an unhindered pandemic in Holland would have given the following path, i.e. neglecting a change in parameters due to the intervention on March 12-15 2020, with an assumed delay of 3 days, thus March 18.

The death toll could have been at 250,000, with 95% reached by May 22, and the remaining 5% spreading over the remainder of the year. Only 2% of the population would not have been infected at one stage.

Table legend: S = Susceptibles (uninfected), I = Infected (exposed or infectious), D = Deceased, C = Cleared, N = N[0] - D remaining population, A = C + D (the acquitted). Percentages (p) are w.r.t. the population at the time of onset, i.e. N[0].

```
SIABasicTable[SEYCDBasicRun, 100, {Sp, Ip, Dp, Cp, Np, Ap}]
(* former run, read from file *)
```

	Date	Day	Sp	Ip	Dp	Cp	Np	Ap
Onset	2020–02–23	0	100.0	0.0	0.0	0.0	100	0.0
1st Death	2020–03–06	12	100.0	0.0	0.0	0.0	100.0	0.0
1st Intervention	2020–03–18	24	99.9	0.1	0.0	0.0	100.0	0.0
Top if free	2020–04–21	58	25.0	40.3	0.5	34.1	99.5	34.7
Ap = 50%	2020–04–25	62	13.5	36.5	0.7	49.2	99.2	50.0
95% Ap[∞]	2020–05–22	89	2.4	4.5	1.4	91.7	98.6	93.1
Limit if free	∞	∞	2.0	0.0	1.5	96.5	98.5	98.0

The levels (q) of the relevant compartments are as follows.

```
SIABasicTable[SEYCDBasicRun, 1,
  {Sq, Iq, Cq, Dq}, TableSpacing → {1, 2}] // PopulationForm
```

	Date	Day	Sq	Iq	Cq	Dq
Onset	2020–02–23	0	17.4×10^6	100	0	0
1st Death	2020–03–06	12	17.4×10^6	1290	391	6
1st Intervention	2020–03–18	24	17.4×10^6	19438	6356	97
Top if free	2020–04–21	58	4.4×10^6	$7. \times 10^6$	5.9×10^6	90455
Ap = 50%	2020–04–25	62	2.4×10^6	6.3×10^6	8.6×10^6	130500
95% Ap[∞]	2020–05–22	89	419710	778037	$16. \times 10^6$	243034
Limit if free	∞	∞	344995	0	16.8×10^6	255825

2.7. Dutch data are unreliable, but the 70-79 age group might be stable

While epidemiologists and these modelers emphasize the uncertainty in the data, they are also the group whom we expect to provide us with estimates anyhow, whatever those uncertainties. Thus, paradoxically, we may be so focused on finally receiving some reliable information that we might perhaps not listen carefully enough to them about their warnings about these uncertainties. Some statisticians, who are (now) at some distance of involved modeling, can be more candid, see Richardson & Spiegelhalter (2020) (Royal Statistical Society) and Gill (2020) (emeritus Leiden) for example. It remains useful to identify some uncertainties.

The Dutch data are unreliable because:

(1) Tests were hardly available, and symptomatic cases were not trusted because of a partly concurrent flu episode. There was some surveillance in the beginning, but this was stopped because of the lack of tests, and the authorities decided to monitor hospital admissions and the death counts, assuming the connection between the infections, symptoms and deaths, but relying on people reporting to hospitals, and apparently underestimating those who did not report to hospitals (the

"excess deaths", which later appeared to be some 3000 of a total of 9000 deaths).

(2) Available tests were restricted to health system workers and hospital patients. The ratio "hospitalised / (confirmed infections)" thus contains confirmed infections amongst (younger) health workers.

(3) Overall (also normal) deaths, also from non-hospitalised cases, also including normal deaths, were not tested. The mixture of "normal" and "excess" components changed, as normal hospital services stagnated, and the national lockdown changed the economic process.

(4) In the beginning, GP doctors were unfamiliar with the disease, and were more inclined to send patients to the hospital, when they considered that the symptoms were more severe than a flu. Over time, when it became clear that elderly patients would suffer longer periods of isolation at the ICU, and with a lower survival probability, patients themselves, their families and GPs became more reluctant on hospitalisation. Holland has a somewhat stronger notion on the quality of life and the acceptance of death when life reaches its end. However, the change in attitude within this episode still makes for mixed data. When hospitals and ICUs reported lower arrivals, then this does not mean by itself that the infections were waning. The infected vulnerables have become more reluctant.

(5) There have been reports about outbreaks in home-care for the elderly. Personal protection equipment (PPE) and tests for care workers there were hardly available. Those patients might no longer be hospitalised because of (4).

(6) New tests are arriving and thus more infections will be reported merely because of this availability. There is also a planned uptake on surveillance now that this wave seems to be ending (May 20).

(7) Obviously, there is the distinction between pre-lockdown, the lockdown phase, and the now partial release. Modeling these phases becomes an even more complex issue.

My impression is that the data on the 70-79 age group are the most reliable. They still are relatively healthy and partake in society, they do not work in the health care system, and they are still in the phase that hospitalisation is regarded as a better option than accepting the end. Using their data as the "canary in the mine" indeed gives a profile also for the other age groups that looks reasonable and comparable with international data. See Colignatus (2020cd) for the argument and calculation.

The calculation has been as follows: The infection fatality factor (IFF) stated by the Imperial College group, Ferguson et al. (2020) (their Table 1), of 0.051 for the 70-79 age group was also applied to this Dutch age group, and the number of observed deaths thus implied a number of infected. Given the population size of this age-group, it followed that 0.0136807 had been infected. This infection rate was applied to the younger age groups, giving the numbers of infected for them. The recorded deaths and these numbers of infected gave an estimate for the IFF per age group. The outcome showed a pattern like the IFF published by Ferguson et al. (2020).

The total *number of infected* thus generated and the recorded total *number of deaths* combine into an average IFF of 1.45%. This factor will be biased, since it is based upon the group that reported to the hospital. Assuming that all Dutch people would be infected and then applying the age-specific IFFs generates a maximum number of deaths in Holland of 183,422, which translates as the population average of 1.061%. The crucial difference is by the 80+ group, that the excel sheet takes conservatively at 9.3% but which might also be double, with a large impact on numbers. Given the uncer-

tainties involved, including the possibility of overburdening the health care system when the epidemic is out of control, this (didactic) paper opts for an IFF of 1.5%, or a maximum number of 261,000 deaths in a population of 17.4 million in 2020.

Let us reproduce some aspects. Colignatus (2020c) states about Ferguson et al. (2020) (their Table 1) and these Dutch data of April 2:

"The Imperial College estimates give problems for the Dutch data. With 121 deaths in the Dutch 60-69 age group, the London age-specific IFF gives 5500 infected in the population while their "symptomatic cases per hospitalised" gives 7663 symptomatic cases in the population, which is too much since we are assuming that the flu season is over. Holland has 29 hospitalised children of age 0-9, and the London symptoms / hospital ratio for this group gives 29000 symptomatic children in the Dutch population, which would create panic if true. Looking the issue over, I cannot find a match. It must be remarked that the Dutch "reported number of cases" is rather useless, because of the lack of tests, and their preferred application to medical personnel rather than patients. Also the death count is understated since non-hospitalised deaths are not tested. See Table 1 below."

"However, the Dutch 70-79 age group may be used as canary in the mine. The number 2951 of "reported cases" will be accurate for this group, since they do not belong to medical personnel. These patients will have some symptoms (like "feeling really sick") and not be tested for nought. The reported number of 2951 means only 0.19% of the whole age group. The Imperial College IFF for this group gives an estimate that 8137 would be infected, or a share of 0.005346 or 0.5%. We arrive at the problem that we are not in the steady state. Either these elderly "infected but non-patients-yet" have a stronger immune system or they are due to arrive at the hospital at a later moment. With lack of other information, we can still presume that this is the overall prevalence of infection (haves and have-beens) in Holland. When we apply this prevalence to the whole population, then we get age-group specific ratios of hospitalisation and IFF that show the same pattern as in China and the London research group. Especially relevant is the "hospitalised per infected ratio" (H/I). See Table 2 below. NB. This uses IFR and CFR, namely as "rates" while it actually are factors IFF and sCFF (symptomatic case fatality factor)."

	CBS 2019 Holland	RIVM 2020-04-02 Reported	Hospital	deaths	Implied D/H	CFR	Imperial College "Impact" H/Smp	ICU/H	IFR	Implied by Imperial College Sympt.	Infected	Sensible
Total	17282163	14677	5778	1339	0.232	0.0912				80443	27909	No
0-9	1783504	56	29	0	0.000	0.0000	0.001	0.05	0.00002	29000	0	No
10-19	2008334	104	15	0	0.000	0.0000	0.003	0.05	0.00006	5000	0	No
20-29	2201788	901	70	0	0.000	0.0000	0.012	0.05	0.0003	5833	0	No
30-39	2108199	1170	154	1	0.006	0.0009	0.032	0.05	0.0008	4813	1250	No
40-49	2261107	1473	391	1	0.003	0.0007	0.049	0.063	0.0015	7980	667	No
50-59	2508388	2415	856	24	0.028	0.0099	0.102	0.122	0.006	8392	4000	No
60-69	2089913	2298	1272	121	0.095	0.0527	0.166	0.274	0.022	7663	5500	No
70-79	1522110	2951	1785	415	0.232	0.1406	0.243	0.432	0.051	7346	8137	Perhaps
80+	798820	3309	1206	777	0.644	0.2348	0.273	0.709	0.093	4418	8355	Okay
<60	12871320	6119	1515	26	0.017	0.0042				61018	5917	
60+	4410843	8558	4263	1313	0.308	0.1534				19426	21992	

Colignatus (2020d) performs the same analysis, now with data of April 19. The table 2 at that moment is as follows. Given the IFF of 0.051 for the age-group of 70-79, the implied infection prevalence of the 70-79 group is 20824 / 1522110 = 0.0136807, and this is applied to the groups younger than 70 years. The pattern of the age group specific IFFs is acceptable. The overall (infection weighted) IFF of 0.0145 has been rounded for this notebook and package as 0.015. (The derivation

of the population weighted IFF of 0.01061 is not shown here.)

	CBS 2019	RIVM 2020-04-19			Implied		Using Canary			China/Lancet
	Holland	Reported	Hospital	deaths	D/H	CFR	Infected	IFF	H/I	H/I
Total	17282163	31574	9593	3600	0.375	0.1140	248160	0.0145068	0.0387	
0-9	1783504	75	39	0	0.000	0.0000	24400	0.0000000	0.0016	0
10-19	2008334	274	27	1	0.037	0.0036	27475	0.0000364	0.0010	0.000408
20-29	2201788	2553	134	3	0.022	0.0012	30122	0.0000996	0.0044	0.0104
30-39	2108199	2686	264	6	0.023	0.0022	28842	0.0002080	0.0092	0.0343
40-49	2261107	3475	680	13	0.019	0.0037	30934	0.0004203	0.0220	0.0425
50-59	2508388	5717	1579	83	0.053	0.0145	34317	0.0024187	0.0460	0.0816
60-69	2089913	4388	2163	325	0.150	0.0741	28591	0.0113670	0.0757	0.118
70-79	1522110	5011	2828	1062	0.376	0.2119	20824	0.0510000	0.1358	0.166
80+	798820	7395	1879	2107	1.121	0.2849	22656	0.0930000	0.0829	0.184
<60	12871320	14780	2723	106	0.039	0.0072	176089	0.0006020	0.0155	
60+	4410843	16794	6870	3494	0.509	0.2081	72071	0.0484800	0.0953	

PM. The excel sheet is available here, and parts can be shown as follows.

```
fil = ToFileName[{$TEPDirectory[], "Applications", "Economics", "Data"},
  "2020-04-19-Dutch-Covid19-core.xls"]
```

C:\Users\Eigenaar\AppData\Roaming\Mathematica\Applications\Economics\Data\2020−04−19−Dutch−
Covid19−core.xls

```
lis =
  Import[fil][[1]] /. x_?NumberQ :> If[FractionalPart[x] == 0., IntegerPart[x], x];

rows = Take[lis, {5, 20}]; rowheaders = rows[[All, 1]]
```

{, , Total, 0−9, 10−19, 20−29, 30−39, 40−49, 50−59, 60−69, 70−79, 80+, , <60, 60+, }

These are the Dutch data with their implied ratios of hospitalised / infected from a reported class also due to symptoms, death / hospitalised, and the sCFF = death / reported.

```
TableForm[Take[rows, All, 8], TableAlignments → {"."}, TableSpacing → {1, 1}]
```

	CBS 2019	RIVM 2020−04−19			Implied		
	Holland	Reported	Hospital	Deaths	H/Rep	D/H	sCFF
Total	17 282 163	31574	9593	3600	0.303826	0.375274	0.1:
0−9	1 783 504	75	39	0	0.52	0	0
10−19	2 008 334	274	27	1	0.0985401	0.037037	0.0(
20−29	2 201 788	2553	134	3	0.0524873	0.0223881	0.0(
30−39	2 108 199	2686	264	6	0.0982874	0.0227273	0.0(
40−49	2 261 107	3475	680	13	0.195683	0.0191176	0.0(
50−59	2 508 388	5717	1579	83	0.276194	0.0525649	0.0:
60−69	2 089 913	4388	2163	325	0.492935	0.150254	0.0;
70−79	1 522 110	5011	2828	1062	0.564358	0.37553	0.2:
80+	798 820	7395	1879	2107	0.254091	1.12134	0.2:
<60	12 871 320	14 780	2723	106	0.184235	0.0389277	0.0(
60+	4 410 843	16 794	6870	3494	0.409075	0.508588	0.2(
		share 60+		0.970556			

The ratios by the Imperical College, Ferguson et al. (2020) (their Table 1), can be applied to these Dutch data, which gives implied symptomatic and infected cases. Comparison of the reported and implied data shows that the outcomes make no sense (e.g. more symptoms than infected), except for the 70+ age group.

```
TableForm [ColumnTake[rows, { {3, 5}, {12, 14}}], TableAlignments → {"."},
 TableHeadings → {rowheaders, Automatic}, TableSpacing → {1, 1} ]
```

	1	2	3		4	5	
	RIVM 2020–04–19			Implied by Imperial College			
	Reported	Hospital	Deaths		Symptom	Infected	Sensibl
Total	31 574	9593	3600		128 325.	114 919.	N
0–9	75	39	0		39 000	0	N
10–19	274	27	1		9000	16 666.7	N
20–29	2553	134	3		11 166.7	10 000	N
30–39	2686	264	6		8250	7500	N
40–49	3475	680	13		13 877.6	8666.67	N
50–59	5717	1579	83		15 480.4	13 833.3	N
60–69	4388	2163	325		13 030.1	14 772.7	N
70–79	5011	2828	1062		11 637.9	20 823.5	Perhap
80+	7395	1879	2107		6882.78	22 655.9	Oka
<60	14 780	2723	106		96 774.6	56 666.7	
60+	16 794	6870	3494		31 550.8	58 252.2	
		share 60+	0.970556				

The following assumes that 100% = 1 of the people in Holland are infected, applies the age-specific IFFs, and finds that at most 183,422 will die from the complications, which translate to the population IFF of 1.061%.

```
TableForm [ColumnTake[rows, {{2}, {16}, {19}}], TableAlignments → {"."},
 TableHeadings → {rowheaders, Automatic}, TableSpacing → {1, 1} ]
```

	1	2	3
	CBS 2019		Deaths
	Holland	IFF	1
Total	17 282 163	0.0145068	183 422.
0–9	1 783 504	0	0
10–19	2 008 334	0.0000363962	73.0957
20–29	2 201 788	0.000099595	219.287
30–39	2 108 199	0.000208033	438.574
40–49	2 261 107	0.000420256	950.244
50–59	2 508 388	0.00241866	6066.94
60–69	2 089 913	0.011367	23 756.1
70–79	1 522 110	0.051	77 627.6
80+	798 820	0.093	74 290.3
<60	12 871 320	0.000601969	7748.14
60+	4 410 843	0.04848	175 674.

2.8. Timeline of the Dutch case, to determine onset and when β changed

Reference to the time line on wikipedia (a portal and no source) may be made.

The WHO confirmed on January 12 2020 that there was a novel virus in Wuhan. The European Centre for Disease Prevention and Control (ECDC) judged the risk of introduction and spread in Europe as low. They might have been little aware that Northern Italy has workshops with cheap Chinese immigrant workers who produce luxury "Made in Italy" products (while Southern Italy has unemployment). It later appeared that Lombardy had had privatisations in the health system that caused a slow uptake of alarm. In the Dutch medical community, Bonten et al. (2020) started blogging on January 24.

Dutch RIVM (2020a) discussed the situation on January 27, and formulated a strict "case definition": (i) symptoms, and either from an infected area or having been in contact with a certified infected

person, (ii) neglect others. Cases falling under this definition however had to reported with priority A.

On January 30, WHO Director-General, Tedros Adhanom, declared the PHEIC, i.e. that there was the risk that a pandemic was developing. (The actual declaration of the pandemic was on March 11.) Holland was rather lacklustre in responding. While January was lost to the whole world, February was also lost to Holland, see Hendrickx & Modderkolk (2020). China collected medical supplies from all over the world, and the Dutch emptied their stocks and sent an airplane load to China on February 10, which effectively meant that those medical supplies and personal protection equipments (PPE) were in short supply when Holland needed them later. The February 21 conference at the Dutch royal academy of sciences KNAW (2020) had presentations that were correct on scientific content - they indicated the general lack of knowledge about the virus - but they also showed the failure by the RIVM agency for public health: namely, that while Aura Timen (RIVM) *mentions* the PHEIC, RIVM apparently neglects the cause for alarm given by the Asian data. Later the Leiden expert on Korea, professor Remco Breuker (2020), suggests that RIVM might be an expert on epidemiology but no expert on Asia. Another problem of course lies in communication. If RIVM had advised a lockdown on February 1, the Dutch population would not have understood. It is advisable to first have a national training on the situation, and broadcast reports about the situation in other countries, to generate public understanding for the WHO declaration of the pandemic. Within 10 days (before the plane leaves).

The Northern part of the country benefitted from an earlier Winter holiday and people returning from ski holidays earlier. On February 23-25 there was carnival in Southern Holland, with people participating who had returned from ski holidays and Northern Italy. For Holland, the first infection was confirmed for February 27, for a man having been in Italy. (RIVM states February 10 as the date of first recorded infection, but this seems a result of later back-tracing and was not widely published at the time.) The first official death was reported on March 6. According to the model run shown above, that day actually had six deaths in total: according to the model five were not officially recorded as SARS-CoV-2 cases. The policy of surveillance of individual cases soon broke down, since the available tests were urgently required within the health care system itself, to protect health care workers.

The government then opted for *herd immunity,* which was supported by RIVM (Jaap van Dissel), see Tweebeeke & Van Dissel (2020). Prime Minister Mark Rutte warned about the economy, and that parents would not be able to go to work when their children could not go to school. Keulemans (2020) is an interview with Jaap van Dissel (RIVM) published on March 12. Van Dissel indicates that RIVM has only limited responsibility, namely *advising* about measures and not having responsibility in managing them, and that RIVM did not protest against limited measures for the South only - apparently allowing for a spread to other parts of the country. A quote: *"If you would lock down the country then the economy stops."* This neglects partial solutions, ranging from quarantining Southern Holland to the whole range mentioned by RIVM at the end of May. Van Dissel showed little concern about the death counts in other nations and likely implied by RIVM models, and he has tended to in general to reject such modeling projections. Later that day, March 12, the government however extended the measures for the entire country. It was only on Sunday March 15 that cafes and restaurants and schools were closed, after strong advice by the Dutch federation of medical

specialists (FMS), but not so advised by RIVM.

Mark Rutte (2020) is the prime-minister's speech of March 16. He indicates three options: (1) free reign and overburdening the health care system, with many deaths as a result, (2) total lockdown, with a permit system for leaving one's home, for the coming two years, effectively closing down the economy, (3) his choice: (a) to protect the vulnerable, (b) to remain within the capacity of the health care system, (c) to allow the less vulnerable to build up herd immunity. The third option (3abc) became government policy. It is exactly the policy suggested in the original version of this present notebook, but this notebook is much stricter on quarantine barriers and deliberate use of the virus as its own vaccine. Since March, it has appeared that the virus is nastier than I thought, see Collins (2020) and De Visser (2020) and Draulans & Piot (2020). Herd immunity may be the wrong tactic, and eradication a better one.

Tweebeeke & Van Dissel (2020) is the interview that discusses the prime-minister's speech. It shows RIVM's support for the government policy towards herd immunity. Remarkably, Van Dissel suggests that children would have no role in spreading viruses, and that this would neither show from the Asian data. His statement is in contradiction with views by epidemiologists if not everyday experience that children / schools have a key role in spreading viruses. He however states that the role of children in spreading SARS-CoV-2 is being investigated now. Later in May, there is the report by German researchers, Jones et al. (2020) (including Christian Drosten).

In the discussion in the Dutch House of Parliament March 18, the Dutch health minister Bruno Bruins collapsed during his speech, and he resigned a few days later, reporting a burn-out. The prime-minister Mark Rutte restated his televised policy, though actually made a switch, by stating that herd immunity was no *policy objective* but a *consequence* of the policy, which actual objective was (3a) to protect the vulnerable and (3b) remain within health system capacity. Members of Parliament criticised him for awkward communication. The switch on herd immunity has been documented by Kustaw Bessems (2020). On March 23 mass events that require advance regulation (like Summer festivals) were forbidden. RIVM & Van Dissel (2020a) is the powerpoint presentation given by Jaap van Dissel on March 25 for the Dutch House of Parliament, with modeling done by professor Jacco Wallinga (Leiden, RIVM). Van Dissel emphasizes the process towards $R_0 \approx 1$, which is required for reducing the number of cases, if only to remain within the capacity of the health system.

Overall, the Dutch policy had a focus on remaining within the capacity boundaries of the health system, especially for the bottleneck of ICU beds. There hasn't yet been an actual policy choice on containment / suppression (forcing $R_0 \ll 1$) versus an alternative.

Over the course of April it appeared that the protection of the vulnerable wasn't well maintained in the home-care for the elderly. Personal protection equipment (PPE) supplies were directed towards hospitals and ICUs. They were not supplied to home-care institutes and workers. This is curious, *since it is the system of care that sends patients to the ICU*. There are reports about a distinction between "medicine" and "care", and a pecking order in social status, also in attention by policy makers.

RIVM stated that there was a lack of tests. However, Alex Friedrich, head of microbiology and infection diseases of the University of Groningen medical center (UMCG), states that RIVM made a

deliberate decision, and he arranged a testing strategy for Groningen, see 1V EenVandaag (2020) (April 17). At the end of May, RIVM is making tests available for the whole country. RIVM misinformed the public about the options for testing.

The number of SARS-CoV-2 confirmed deaths on April 2 was 1339 and on April 18 was 3600, and this rise predominantly was in the 60+ age group, apparently not-well-protected, even though it was the government-stated policy target. These are the official numbers only.

Also, there is a rise in *general mortality*, higher than what would be "normal" in this period like in the past, the "excess deaths". This rise in mortality includes persons who have not been tested on the virus, because they were not hospitalised. RIVM has estimated that this rise is twice as high as the officially recorded number of SARS-CoV-2 deaths. Thus the April 2-18 period may have seen a rise of some 4000 deaths, predominantly in the 60+ age with comorbidity group.

PM. With a low prevalence, and regional heterogeneity, testing the general population is costly. However, testing the deaths (the regular average of 12,000 per month) comes with a higher prevalence, and thus is relatively cheap, and can be targeted on indications by GPs.

On April 21, the prime-minister announced that the government will kick the can further down the road, with another reprise at the end of May. However, children can go to elementary school again at the end of current holidays, on May 11. RIVM suggests that they tested that children were infected via adults and not via other children. (The finding is against common sense and other epidemiological reports. One hopes that RIVM did not test only during this quarantine phase.) A personal report is that traffic intensity rose significantly in the days after April 21, which suggests that many people regard the April 21 press statement as an indication of a reduced need for quarantine. However, one might assume that people may be aware of their personal risk, and behave accordingly, though with greater likelihood with neglect of the risk of others.

Blok & Stronks (2020) (May 6) of an action group for containment, protest against the Dutch policy of mitigation, that allows gradual spreading of the virus, and the path towards release of the lockdown, with the prospect of lock-on-off cycles.

On June 3, the bloodbank Sanquin reports, based upon blood samples amongst blood donors (age 18-75), that 5-6% of the population has antigens. This outcome must be biased, and RIVM itself combines such finding with its own Pienter project.

RIVM views and the latter research are now contested by Willem Engel (https://viruswaanzin.nl), but in dubious fashion, see Van Erp (2020). Why doesn't Engel publish a paper on the issue ?

PM. Above timeline gives events that are specific for the development of the infection itself, dating from the presumed onset at the carnival of February 23, and assuming unmitigated spreading. There are some other points in time that are relevant to mention as well. In the "didactic estimation" below we will only take account of the intervention on day 24.

(a) Delay in response, from earliest indication to PHEIC to the closing of Dutch schools: this period remains uninformative since we do not have other data about the role of schools.

```
{DayCount[{2020, 01, 12 }, {2020, 3, 15}], DayCount[{2020, 01, 30 }, {2020, 3, 15}]}
{63, 45}
```

(b) Assume 100 infections for carnival February 23 and unmitigated spreading. The first official

death was recorded on March 6. There is a break in the $\beta[t]$ by the closing of schools on March 15 with an overall delay of say 3 days for societal implementation.

```
{DayCount[{2020, 02, 23}, {2020, 3, 6}],
 DayCount[{2020, 02, 23}, {2020, 3, 15 + 3}] }
```

{12, 24}

(c) The press statement of April 21 that some people interprete as a loosening of quarantine. Elementary schools start up again from May 11, though modestly, like with half-classes per day.

```
{DayCount[{2020, 3, 15 + 3}, {2020, 04, 21 + 3} ],
 DayCount[ {2020, 04, 21 + 3}, {2020, 05, 11} ]}
```

{37, 17}

2.9. Relating to the RIVM data and parameters (2)

2.9.1. Introduction

RIVM calculation models are public and accessible, see: https://www.rivm.nl/en/novel-coronavirus-covid-19/calculations/calculation-models-are-public-and-accessible

The above discussed the situation in March 2020, when we selected the parameters for this didactic version of the SI(EY)A(CD) model. Subsequently, new data have come available, and RIVM, the Dutch CDC, provided evidence to Dutch Parliament also on new modeling. Above, we already compared these findings with the parameter values in this notebook and package. In this section, we summarily discuss some reports.

There are a RIVM report from January 27 and subsequently slides from March to May with presentations at Dutch Parliament by RIVM head of the department for control of infectious diseases, professor Jaap van Dissel. We will not discuss all of these. These reports have different objectives than merely reporting on modeling but we can trace some aspects on this. The slides must be read in the context of listening to the spoken presentation. I still find it remarkable that the slides do not provide clear definitions: they could be included for reference without the need to read them aloud. Below I try to determine what some figures or graphs may mean, and mainly arrive at questions. I wonder whether the Members of Parliament would have surplus information.

RIVM & the Outbreak Management Team (OMT) (2020) (May 6 online) gives the Dutch view on asymptomatic transmission. Our own Chapter 10 below provides two common definitions. (i) The "serial interval" for symptomatic disease. (ii) The "generation interval" for infections. This was also discussed in the *Introduction,* subsection 1.4.

2.9.2. RIVM & the Outbreak Management Team (OMT) (2020) on source and contact tracing

In conventional language, "contact" means any type of contact. For epidemic "source and contact tracing" for a particular subject ("index"), the term "contact" is redefined, and there is the terminological difference between "source" (subject and unknown infector) and "contact" (subject infecting sinks). (Dutch weblink: https://lci.rivm.nl/COVID-19-bco.) There is also a difference between

asymptomatic transmission (the infector is asymptomatic) and the *incubation period* (the infectee is asymptomatic).

(1) Incubation is generally taken as 6 days. In order to trace the infector of the subject, i.e. "source tracing", contacts (in the colloquial sense though) in the six days before the onset of symptoms must be traced - or actually till the source has been found.

(2) RIVM & OMT (2020) discuss asymptomatic transmission and the consequences for "contact tracing" (i.e. that the subject can infect others). They arrive at the same conclusion of the European CDC (ECDC). The asymptomatic period can be 1-3 days. For surveillance it would suffice to back-trace contacts (recipients) to 2 days before onset of symptoms with the subject.

("Dit betekent dat voor effectieve bron- en contactopsporing (met of zonder app) contacten tot 1-3 dagen voor de eerste ziektedag moeten worden gerapporteerd en opgespoord. Het ECDC neemt hiervoor in haar rapport over contactonderzoek twee dagen voor de eerste ziektedag (ECDC 2020). Voor de Nederlandse praktijk sluiten wij aan bij het advies van de ECDC. ") https://lci.rivm.nl/-covid-19/bijlage/onderbouwing-a-pre-vroegsymptomatische-transmissie

The reasoning is unclear.

(2a) If it is acknowledged that asymptomatic transmission can occur 1-3 days earlier than why stop at 2 ? RIVM & OMT only refer to the ECDC guideline and apparently do not question it ? (If it is a matter of costs, specify those: what is the price of eradication ?)

(2b) RIVM refers to model studies, cluster studies, and cross-sectional studies. Only the first two could in theory support a 1-3 period of *asymptomatic infectiousness* on content. The cross-sectional studies indicate asymptomatic *infectedness* but this differs from *infectiousness.* RIVM & OMT (2020) notably refer to Kimball et al. (2020). Their key finding is: "Among 23 (30%) residents with positive test results, 10 (43%) had symptoms on the date of testing, and 13 (57%) were asymptomatic. Seven days after testing, 10 of these 13 previously asymptomatic residents had developed symptoms and were recategorized as presymptomatic at the time of testing." A key conclusion is that 3 / 23 = 13% remains asymptomatic. Supposedly the infection clears after a while. Nevertheless, the relevant issue is whether they were also *infectious*. This, however, was not looked at in this study. Perhaps the viral load might say something about this, but this is unclear to me. However, for the issue of contact tracing, the safest assumption is that these 13% asymptomatic *infected* persons would also be *infectious.* Thus, if contact tracing hits upon an asymptomatic infected person, then contacts *over the full infectious period* of this person must be traced, in our assumptions 10 days (though perhaps the growth and waning of the viral load might give an indication how long this person has been infectious). This notion is relevant precisely because we have seen that new outbreaks can occur via asymptomatic carriers. It is unclear why RIVM has no remark on this.

2.9.3. RIVM January 27

RIVM (2020a) (January 27) mentions a possible R_0 = 2.7 and a "generation time" of 6 days, defined as "the number of days for the next cycle of infections" (it is better to call a period a period). The reader is advised to first read Chapter 10 on the *infection* "generation interval" and the *symptomatic disease* "serial interval". We also discussed the "generation interval" in the *Introduction,* subsection 1.4. Estimates from a period with intervention generate other values for the generation interval. A

complication is that there is the generation interval for the *first* infections from patient zero, as would be measured at the individual level in empirical studies, while there is also the model that uses a generation interval (period) at the aggregate level for the generation *means:* over time one may allow changes in $\beta[t] / \gamma[t]$ so that the beginning become less relevant.

At the aggregate level, the generation interval is also the *transmission period,* which, *in the assumptions of this notebook,* consists of an exposed period of 1 day and an infectious period of 10 days, thus a total of 11 days. The value stated by RIVM also fits ranges mentioned in the literature but does not fit our choice of parameters. There is no reason to assume that the Dutch in February and the beginning of March were inclined to self-quarantine, as much as the Chinese people were at the time of some of their reports. Also, later an incubation of 6 days is mentioned: *RIVM apparently allows that the incubation period is longer than the generation interval.* Perhaps Chapter 10 should drop the assumption that *for the aggregate* the transmission interval = generation interval = incubation + infectious period = serial interval ? It may just be that we are using a SI(EY)A(CD) model while RIVM is not. RIVM later reports a "generation time" of 3-5 days, and their website later in the year refers to the Ganyani et al. (2020) estimate in which researchers at RIVM partook: this study is discussed in our section 10.6. A key comment by Ganyani et al. (2020) is: "This means that our estimates do not necessarily reflect the natural epidemiology of COVID-19, but instead reflect what is observed in the presence of these intervention measures." Instead, for the present notebook, we want to have the "natural epidemiology" first, before we consider an intervention. Ganyani et al. (2020) mention that the interventions cause a larger number of observations of negative serial intervals for individual cases, i.e. that the infectee develops symptoms before the infector does. This is okay for individual cases. However, Ganyani et al. (2020) (table 1 and 2) also have the aggregate effect that the generation interval and the serial interval have the same duration. This does not take away the large difference between 5 (RIVM) and 11 (us), and is suggestive of a distinction between "first infection" (starting a generation) versus "generation mean".

2.9.4. RIVM March 25

(1) RIVM & Van Dissel (2020a), March 25, depict the possible onset on February 4 (data by GP's) or February 10 (confidence range chart), apparently back-tracing from information by patients. They mention $R_0 \approx 2.7$, a doubling time of 5-6 days (which in the graph is extended to 4-7 days), an incubation period of 6 days, and a "generation time" of now 4-5 days. Though the RIVM may have arrived at this estimate while using a different α, it remains unclear to us which, and with our α we get this reconstruction of β and γ.

```
eqs = {r0 == 2.7, r0 == beta / gamma, r0 == (1 + r / alpha) (1 + r / gamma) } /.
  {r → Log[2] / 5.5, alpha → 1}
```

$$\left\{ r0 = 2.7, \; r0 = \frac{beta}{gamma}, \; r0 = 1.12603 \left(\frac{0.126027}{gamma} + 1 \right) \right\}$$

```
sol = Solve[eqs, {r0, beta, gamma}] // Quiet
```

$\{\{r0 \rightarrow 2.7, \; beta \rightarrow 0.243432, \; gamma \rightarrow 0.09016\}\}$

```
1 / gamma /. sol[[1]]
```

11.0914

We assume the exposed period as $1 / \alpha = 1$ day. Thus, the RIVM "generation time" with value 4-5 days differs even more from the "generation interval" in our assumptions.

```
1 / α + 1 / gamma /. (α → 1) /. sol
```
{12.0914}

With an incubation of 6 days, we have 5 = 6 - 1 days of *asymptomatic infectiousness*. RIVM would have a symptomatic period of 12.1 - 6 = 6.1 days. Since RIVM did not provide definitions or discussion, it is unclear where they stand.

(2) RIVM concentrates on "flattening the curve", to remain within the capacity of the ICU, even to the effect that they do not provide any indication of the potential number of deaths. Basically, RIVM avoids any discussion about the effect measures of "lives saved" and (quality adjusted) "life years gained". Policy makers, the "Outbreak Management Team", the medical world and RIVM have converged on the idea that medical conditions of patients imply treatment at an ICU so that this capacity must be respected. The government and RIVM had presented the "three scenario's" but in qualitative format, and there is no effort at developing more options, in also quantitative manner.

RIVM reproduces the Chinese CDC summary (Wu & McGoogan (2020) in JAMA), keeps the numbers and changes the diagnoses: 81% "mild" becomes "*little affected or without symptoms*", 14% "severe" becomes "*seriously sick*", and 5% "critical" becomes "*fatal*". While the Wu & McGoogan (2020) write "critical" and also explicitly state a case fatality of 2.3%, the RIVM interpretation is bizarre. Potentially, though, RIVM finds the CDC estimate unreliable, with its selection of mostly symptomatic patients reporting at hospitals. It is better to mention this.

RIVM oppose this Chinese "pyramid" with a finding at hospitals in Breda and Tilburg of 97% little affected and 3% seriously sick, and apparently no deaths. We can wonder why this is so relevant because in the next slide RIVM mentions 276 deceased.

Using their R_0 and our IFF = 1.5% we can find the implication of the SICD model of the limit value of deceased of the pandemic in Holland at around 240,000.

```
Explain[ SICD[Limit, BetaSIA[] → 2.7 gamma, GammaSIA[] → gamma], SICD]
```
{S → 0.0844068, I → 0, C → 0.901859, D → 0.0137339, N → 0.986266, A → 0.915593}

```
Explain[ 17.4 × 10^6 * SICD[Limit, BetaSIA[] → 2.7 gamma, GammaSIA[] → gamma], SICD]
```
$\{S → 1.46868 \times 10^6, I → 0., C → 1.56924 \times 10^7, D → 238\,970., N → 1.7161 \times 10^7, A → 1.59313 \times 10^7\}$

2.9.5. RIVM April 8

(1) RIVM & Van Dissel (2020b), April 8, now mention $R_0 \approx 2.3$, a doubling time of 5-7 days, an incubation period of 6 days, and a "generation time" of 3-5 days.

(2) Remarkably, the first slide still has the Breda and Tilburg message that there would be no deaths, and a few slides later that there are 2101 deceased, and again a few slides later there is the report of the "excess deaths". The first slide has become somewhat like a "logo".

(3) Again RIVM concentrates on "flattening the curve", to remain within the capacity of the ICU, but they do not provide any indication of the potential number of deaths.

(4) Slide 17 shows the effective *R*, in which the first day of symptomatic disease is traced back from the Osiris database, apparently February 27, which allows us to infer that the GP-data are not in this

database. The effective *R* got lower than 1 on March 23. Slide 18 shows the measures that were gradually taken in the period of February 27 to March 16.

Since R_0 changed from March 25 to April 8, let us reconstruct β and γ.

```
eqs = {r0 == 2.3, r0 == beta / gamma, r0 == (1 + r / alpha) (1 + r / gamma) } /.
  {r → Log[2] / 6., alpha → 1}
```

$\left\{r0 = 2.3,\ r0 = \dfrac{beta}{gamma},\ r0 = 1.11552 \left(\dfrac{0.115525}{gamma} + 1\right)\right\}$

```
sol = Solve[eqs, {r0, beta, gamma}] // Quiet
```

{{r0 → 2.3, beta → 0.250239, gamma → 0.1088}}

```
1 / gamma /. sol[[1]]
```

9.19121

The number of implied deceased, using IFF = 1.5%:

```
17.4 × 10^6 * SICD[Limit, BetaSIA[] → 2.3 gamma, GammaSIA[] → gamma][[4]]
```

225 093.

2.9.6. RIVM April 22

(1) RIVM & Van Dissel (2020c), April 22, still have the first "logo" slide, and correct with a later slide with 3916 deaths. Osiris mentions 34 thousand patients who tested positive on SARS-CoV-2, of which 33% are care workers, given the lack of tests and the policy that only patients and care workers are tested.

(2) Slide 9 gives the claim that without the interventions 23354 ICU admissions more would have been required, representing 90% of the total, while current admissions are 10%. A later RIVM figure is that 0.35% of all national diseases arrive at the ICU. Thus, the implied limit number of infections is Aq[∞] = Infections / Case * 23354 / .9 / 0.0035. One wishes that the slide would give full information and not only some tidbits with much to assume. Using the RIVM R_0 = 2.3 we can find the implied symptomatic Cases Per Infection (sCPI).

```
23 354 / .9 / 0.0035 ==
 CasesPerInfection 17.4 × 10^6 * SIA[Limit, BetaSIA[] → 2.3 GammaSIA[]][[3]]
```

$7.41397 \times 10^6 = 1.50062 \times 10^7$ CasesPerInfection

Keep in mind that Infection = Exposed + Asymptomatic + Symptomatic, but that we might have to make a distinction between "merely symptomatic" and "disease that might cause hospitalisation".

```
Solve[TheFormer, CasesPerInfection]
```

{{CasesPerInfection → 0.494061}}

(3) Slide 16 has the claim that the effective reproduction factor came below 1 on March 16, i.e. the first Monday when schools were closed. Apparently RIVM adapted the estimate. The estimate refers to Wallinga & Lipsitch (2006, 2007).

Contact tracing apparently allows a chart of "what age group infects what age group" (slide 20).

RIVM reports about a study done on infectious spread under and by children, but this study was done in April after schools had closed, and RIVM accurately reports that there is "selection bias". We

can be very skeptical about this study. I am much amazed that RIVM presents this biased finding as "conclusions".

(4) Tests on blood donors: On average 3.6% of the population has antigens. This is 1% for persons below 20 years and 4.2% for those older. We may presume that there has been correction for sampling selection; it leads too far now to look into the referred study (F van der Klis et al, IIV-CIb/RIVM). The issue returns on May 20, and then allows an estimate of the Infection Fatality Factor (IFF).

2.9.7. RIVM May 7

RIVM & Van Dissel (2020d), May 7, still have the first "logo" slide, corrected later with 5204 deaths.

The levels at the ICU have dropped so much that RIVM concentrates on giving scientific blessing to the government plan of phasing out lockdown measures such that a second wave ought to be avoided.

2.9.8. RIVM May 20

(1) RIVM & Van Dissel (2020e), May 20, no longer have the first "logo" slide. The death toll is 5715, not counting the "excess deaths". Osiris counted 44249 patients, and thus a Hospital Fatality Factor (HFF) of 12.9%. In the chart on slide 9, RIVM puts Holland *below* instead of *above* Spain with 12%. Below, for lack of other data, we will take the sCFF as this HFF.

HFF == 5715 / 44 249.

HFF = 0.129155

(2) Slide 9 gives the "excess deaths" as 9000, which means some 3300 cases more than the official number 5715. There is no information yet about the composition of these excess deaths. There may also be flu victims, deaths from reduced normal care, and there will be fewer traffic accidents (normally 65 per month).

Research on blood samples, now a combination of blood donors and a RIVM panel called "Pienter", shows that 3-4% of Dutch inhabitants have antigens. RIVM mentions 650,000 that "have been" infected, which means that they took 3.735%.

650 000 / (17.4 × 10^6)

0.0373563

RIVM arrives at an Infection Fatality Factor of 1.3%, and it is somewhat remarkable that they do not round to 1.4%.

IFF == 9000 / 650 000.

IFF = 0.0138462

(3) RIVM actually provides this (incomplete) table that is focused on hospitals.

	Hospital	Not	All
Have been infected	44 249 + ?	rest	650 000
Patient	44 249	?	?
Deceased	5715	rest	9000

We have: IFF = D / I = D / H * H / I, but there are different death counts, and the H / I ratio must be

corrected for this.

We can combine the Osiris data and the infections from the blood testing.

HospitalCasesPerInfection == 44 249 / 650 000.

HospitalCasesPerInfection = 0.0680754

We would like to have the symptomatic Case Fatality Factor (sCFF) too. We have IFF = D / I = D / sC * sC / I = sCFF * sCPI with the symptomatic Case Per Infection (sCPI). Above we had an indication that it might be 0.49 but this was dubious.

(4) However, let us suppose that the distribution of infections copies the distribution of deaths, or that we can use the IFF to backcast the infections. Let us suppose that the symptomatic death rate as observed for hospitals (HFF) roughly applies to the total of 9000 deaths, with sCFF = HFF, so that we can backcast the symptomatic development. Then we can fill in two blanks.

	Hospital	Not	All
Have been infected	412 750	237 250	650 000
Little affected or no symptoms	368 501	211 815	580 316
Patient or symptomatic	44 249	25 435	69 684
Cleared	38 534	22 150	60 684
Deceased	5715	3285	9000

And we can now compare with the Chinese CDC "pyramid", see Wu & McGoogan (2020). Though they also have confirmed cases, their selection has been biased by mostly symptomatic patients, though confirmed by microbiological testing.

	Chinese CDC (IFF)	Holland (IFF)
Deceased	2.3	1.38
Critical (including deceased)	5	–
Severe	14	9.34
Mild	81	89.28

Thus we now have the decomposition IFF = sCFF * sCPI with sCFF = 12.9% (HFF = 9000 / 69684) and sCPI = 69684 / 650000 = 10.7%.

(5) The claim on the number of ICU admissions that has been prevented has now risen to 44000. RIVM does not specify what percentage this is, and thus we search for a statistic on the current cumulative number of ICU admissions. For May 20 this is 2847, see https://www.stichting-nice.nl/. Thus we can do the same algorithm as above. It is actually here on May 20, slide 16, where it is stated that the ICU form 0.35% of all diseased *per day*. RIVM states that 16 /100,000 new diseases per day mean 40 hospital admissions and 10 ICU admissions per day, in the latter ratio of 1/4. Less rounded numbers are 154 / 100,000 disease new cases, 384 hospital and 96 ICU admissions, still with the latter 1/4. With a population of 17.4 million, the 0.00154 figure translates into 26796 cases per day, of which then 384 = 1.4% hospital and 96 = 0.358 ICU admissions. It still are rounded numbers. The implied figure of 1.4% "hospital case per disease" differs much from what we found above, as 9.34% "severe" or even the 6.8% "hospital cases per infection". Something happened to the definition of "disease" and it is unfortunate that the slides *do not give a definition.* (It is even unclear whether the ICU belong to the hospitals or are a separate category.) However, we can take 1/4 as the ratio and presume that 9.34 / 5 of the critical might be at ICU of which more than 50% actually dies.

```
{hosp, icu} = {384, 96} / (17.4 × 10^6 * 154 / 10^5)
```

{0.0143305, 0.00358262}

Assuming that 0.35% was rounded, and using the RIVM R_0, we find a high figure for the disease per infection, that doesn't fit even the high factor of asymptomatic cases. For us, it is a mystery what RIVM here is presenting to Dutch Parliament.

```
(44 000 + 2847) / 0.00358262 ==
 DiseasePerInfection 17.4 × 10^6 * SIA[Limit, BetaSIA[] → 2.3 GammaSIA[]][[3]]
```

$1.30762 \times 10^7 = 1.50062 \times 10^7$ DiseasePerInfection

```
Solve[TheFormer, DiseasePerInfection]
```

{{DiseasePerInfection → 0.871386}}

(8) Slide 13 has this curious diagram of the periods between the stages. Incubation has been reduced from 6 to 5 days, and instead of "symptoms" there are "complaints" ("klachten") (which I would regard as a major distinction). In the recording, Van Dissel states that the seriously ill have their own shorter path from complaints ("klachten") to ICU (9.2. days). The total period from infection to hospital admission is 5 + 7.7 + 7 = 19.7 days, suggesting a one-decimal accuracy. However, when RIVM estimates the effective reproductive factor $R[t]$, then they can back-trace hospital admissions to infections 14 days earlier (as apparently recorded in the Osiris database). I would say that 14 differs from 19.7. It may well be that various visits to the GP ("huisartsbezoek") do not result into a hospital admission, so that the averages in this scheme are much influenced by either the healthier who eventually clear by themselves or those who prefer to die at home rather than unconscious in an ICU bed. This scheme is more confusing than informative, though the period for back-tracing likely will be accurate (as accurate as the database entry system).
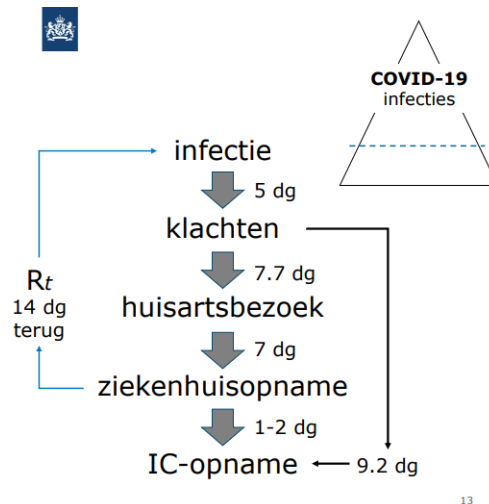
(9) The pyramid in above diagram refers to the original Chinese CDC summary that 81% mild, 14% severe, and 5% critical (that RIVM calls "fatal"). RIVM has drawn a dashed line to indicate that there are many asymptomatic infectious, i.e. "below the radar of symptoms". In the video presentation, Van Dissel very shortly remarks that *most experts think that the pandemic cannot be suppressed because of these asymptomatic infectious.* Singapore and South Korea thought that they had the infection under control but were confronted by recurrences caused by such carriers. Never-ending surveillance and contact tracing are required.

(10) RIVM then discusses Kissler et al. (2020), the Harvard study, referring to it as "Marc Lipsitch et al, Science, 2020", and subscribe to the findings. A method of *lock-on-and-off* could respect the capacity of the health care system, and would gradually result into a build-up of herd immunity of some 55% over two years. The fact that antigens have been found in Dutch blood samples is encouraging for the assumption of immunity, but it is not clear how long this will last. (Sanquin stated independently that antibodies need not say much about immunity. Corona viruses cause the common cold and the immunity system tends to forget them.) The method requires close monitoring of developments (with said delay of 14 days between onset and hospital admission), and comes with a risk of a "Fall peak" (this Fall and Winter) and overlap with the flu. Apparently "herd immunity" still is an appreciated by-product (see our deconstruction in section 1.5.6 and Chapter 6).

(11) RIVM suggests a gradual release of lockdown. There are distinctions between individual / group, inside / outside, local / regional. There are systems for "early warning" and "surveillance".

2.9.9. RIVM May 25

RIVM & Van Disssel (2020f), May 25, has the "logo" slide re-instated, though no longer mentioning Breda and Tilburg that reported no deaths, and replacing the Chinese CDC pyramid *for confirmed infections* with the RIVM *undefined "disease"*, of 98% "no or few complaints", 1.5% hospital admissions and 0.35% of ICU (which at least informs us that these are separate categories, i.e. ICU not counted as part of hospitals): (i) which emphasises the RIVM focus on the ICU, (ii) which neglects the deaths, also outside of the hospital system, and which does not mention the IFF of 1.5% as a relevant issue (their own estimate 1.38%). Remember that the sCFF was 12.9% so the RIVM undefined "disease" is something else then landing in the hospital because of symptoms. However, the slide mentions the world death toll of 350,000, so RIVM shows awareness that deaths are happening somewhere. The R_0 had been 2.7 and had been reduced to 2.3, and is now put at 2.5. What is the meaning of this change: has RIVM indeed done a re-estimation of the situation in February, or is it only the presenter of the slides who happened to pick one particular file on his computer out of a large collection of versions ? Incubation had been put at 6 days, then 5, and is back at 6 again.

2.10. A partial result of this exercise

2.10.1. A result for us

This exercise is partly useful to determine *what if we hadn't had a lockdown*. Thus, using $\beta = 4$ and $\gamma = 1/10$, what would the death toll have looked at the end of the year ? Many people would agree that a temporary lockdown in March 2020 has been sensible, if only to consider the options. Dutch society saved some time to consider the options. But did we really consider those options, and, what are the options actually ? As said, this discussion cannot be settled with a didactic exercise, but one can discuss what information would have to be improved in order to make such discussion fruitful.

2.10.2. A list of questions for others

The government and RIVM allowed the virus to spread, first formulated as an objective but quickly rephrased as a by-product of the policy of flattening the curve. Allowing a virus to spread might be

based upon medical criteria and the properties of the virus, like the common cold, but it has a distinct legal flavour of avoiding responsibility for the spread. If the virus is innocent then why not use it as its own vaccine ? The situation is a conundrum in terms of public health. Such questions should have been addressed with the WHO PHEIC, but are still waiting for an answer. The health system allowed events to evolve such that the emergency brake of a lockdown was required, so that such questions were sidetracked, but such questions still need attention.

It is sobering that the Dutch lockdown in the middle of March was an *emergency brake.* Given the statement of RIVM (2020a) of January 27, then, when the WHO declared the PHEIC on January 30, the proper course would have been to clam down on the infection, as in Taiwan, Hong Kong, Singapore and South Korea. This would have required intensive education of the Dutch population, within 10 days, but the scenes from Wuhan and Italy would have helped. The plane that left Holland on February 10 with medical supplies and personal protection equipments collected for China, would have stayed here.

The above criticises RIVM statements about herd immunity and the role of children. There was the collapse of conventional hospital care. It seems that RIVM was focused on the end-of-pipe ICU beds, with prominent doctors, and with neglect of up-stream homes of the elderly, with less status, even while RIVM stated that the policy was targeted at protecting the vulnerables. RIVM only advises, but still.

Sridhar and Rafiei (2020) indicate that it is almost impossible to shield the vulnerables, except by a national lockdown. I have my doubts whether we actually tried enough. However, RIVM in the early phase called for such protection as if it were possible without a lockdown. In the later phase RIVM provided various measures for social distancing, but rather late and even now, with RIVM supporting the release of lockdown measures, it seems that they trust that the vulnerables can be shielded. We can only assume that this is *evidence based,* so, what is this evidence ?

There are questions about source and contact tracing. Critical questions arose in the media while one would assume that RIVM should have spotted the problem itself earlier (even when it is GGD that has been assigned that task).

RIVM allowed that deaths were only counted for hospitals, and that suspected cases at home were not tested, even though RIVM knows that this distorts the information. We discussed the estimate of the dynamic $R[t]$ on hospital data only, with unreliable uncertainty because of the changing attitude w.r.t. hospitals and the peak in "excess deaths" in April.

There are questions about the balance of medical needs and economic needs. The provinces of Brabant and Limburg were not put into quarantine with the argument that a lockdown would also lock down the economy, but the lock down of two provinces does not lock down the entire nation. When RIVM had manoeuvred the process into the funnel of a national lockdown, then it seemed as if discussion with economists was superfluous, but such discussion would have been needed at all times, even in January, even now. Though it is useful to remark that this notebook looked at the didactics of SI(EY)A(CD) modeling with the example context of SARS-CoV-2 for Holland, and this notebook does not discuss the didactics of economic models for this context.

PM 1. The above basically concerned the RIVM department of infectious diseases. RIVM does more than this, and RIVM also does many more things not. In 2004, minister of Economic Affairs L.J.

Brinkhorst and public health and environment secretary P. van Geel wrote to Dutch Parliament that RIVM had been given funds to calculate each five years the figure of environmentally Sustainable National Income (eSNI) according to the definition by Hueting, see Brinkhorst & Van Geel (2004). This was done till 2005. Thereafter the calculations stopped, and it is not clear why. See Colignatus (2020f).

PM 2. The Dutch path towards lockdown is quite different from Germany and doesn't much differ from the path taken by the UK, see Wren-Lewis (2020) (March 30): " Did the government follow the science? In a way that question is irrelevant. As Ben Chu suggests, dealing with a pandemic does not involve dealing with a science where the answers are well established and known with near certainty. No scientist would want to stop a government, on seeing the first cases of COVID-19 in the UK in mid-February, distributing PPE equipment to medics, ordering new ventilators and ramping up testing capacity. That this didn't happen in the UK until mid-March is, as the Lancet suggests, a national scandal."

3. The package

3.1. Caveat: This is didactics. The package only hints at the real world

The above already indicated:

(a) We want to focus on SARS-CoV-2 but we will not get far. SIA, SICD and SI(EY)CD have few parameters, and we should not expect that these few parameters are sufficient to describe the world. It is educational that such a simple model already conveys major conceptual aspects about a pandemic, but an increased confidence in one's understanding of pandemics should not cause overconfidence.

For example, these basic models have only 1 class of susceptibles. In our own discussion there are rather two classes (SI(EY)CD2): vulnerables and less-vulnerables. Modeling with more stages and more age-groups likely requires SI(EY)CD20.

In other words: If and when elements from this notebook and package are used for real world applications then those must be documented for themselves.

(b) The notebook and package present *both* the conventional S(E)IR(D) setup *and* a *didactic redesign* to a new format SI(EY)A(CD). It you want to, you can use this package in the traditional S(E)IR(D) format. In all cases, there is no change in conceptual content or material findings. There is only the suggestion that the (underlying) notions are expressed more clearly and more accessible, extended now with user-friendly programming for an environment of *computable writing*. The principle is: *Do not program to others what you do not want to be programmed to you*.

There are obvious drawbacks to this effort at restating what is already known. An experienced reader might wonder why this redesign is needed: there is a learning curve for this notebook and package, with nothing gained on new understanding. A novice reader might benefit from a better understanding of models and programming as provided here, but then still faces the literature with its conventions. This notebook and package thus are a rather dubious endeavour. The proof is in the eating of the pudding. My expectation is that many readers will find that the reading has been

worthwhile. (For teachers: When one doesn't adopt the redesign itself then this discussion at least highlights issues in didactics.)

3.2. Design / Redesign of didactics

The following repeats the overview of the choices from didactics, as stated in the *Introduction*, but relates those choices now to some programming particulars, like the choice of options and the use of variable names.

3.2.1 Avoid using *R* for compartments, and instead use *A* and *C*

The conventional S(E)IR(D) model uses the label *R* for three applications: the *Removed,* the *Recovered* and the *Reproduction factor.* Perhaps best would be to avoid *R* altogether, so that there is not even the question what *R* might mean. However, R_0 is used in various models and it suffices to relabel the compartments. We use *A* for the *Acquitted,* and *C* for who has cleared the infectious period without dying. On occasion we may write *f* for the reproduction factor, since it is a factor indeed.

PM. The symbol *R* is also avoided in the "traditional versions" that the package recognises. However, Rq, Rp, and Rg are available, and users may set Rq = Aq, Rp = Ap and Rg = Ag. (This would have to be done after each clearing of the variables, and one might write a small routine that both clears and redefines.)

3.2.2. Have $D = \varphi A$ as an explicit proportion

Section 7.3. The conventional format of the SICD (SIRD) model has $D' = \mu I.$ This causes the question what μ means, how to calculate its value, and what happens if the overall γ would change. The best didactic format is to give the actual model $D = \varphi A.$ The deceased are a mere proportion of the acquitted, with φ = *Infection Fatality Factor* (IFF). The conventional format is oriented to Ordinary Differential Equations (ODEs) and it is an exercise in mathematics to show that this translates in the proportions. It is clearest however to state what the model entails, and regard the education about ODE as another issue.

PM. The "traditional version" in the package has $D' = \mu I.$

3.2.3. One single conceptual model SIA with two kinds of splits: *A = C + D* and *I = E + Y*

SIA is the *conceptual* model that has two kinds of nuance splits, whence the full name is SI(EY)A(CD):

(a) Section 7.3. The *acquitted A* can be split into the *Cleared* and *Deceased,* thus *A = C + D.*

(b) Section 8.3. The *infected = infectious I* can be split into the *Exposed* and *Infectious,* either *I = E + Y* (didactically clearest) or traditionally *Y = E + I* (in which *I* puts emphasis on being infectious).

Within the new didactic format, the two splits are such, that the properties of the *S, I,* and *A* compartments remain the same (taking *I = E + Y*). This warrants the use of a single set of Options[SIA] that control the functioning of *all* model formats. The routines for SIA, like finding the top of *I = E + Y* and the limit values, apply to all variants, since the same assumptions apply.

PM. The "traditional version" in the package has $Y = E + I$.

PM. For simplicity, no background births and deaths have been included, though they obviously effect conclusions about waves of infections and a steady state.

3.2.4. Separate models but joint options

Good programming requires that a new model comes with its own set of options: but this applies only when we really have different models. The SIA, SICD and SI(EY)CD models are so alike that it would be unwise to have different sets of options. Having different sets of options would require checks on e.g. the same size of the population etcetera, and this better be avoided.

It would have been possible to construct SICD by first run SIA and then split *A* into *C* and *D,* but we follow the convention (historical development) to present SICD as a separate model indeed. This has the advantage that all variables are available in both input and output. We include $A = C + D$ in the model, for this purpose*.* For SEYCD we also include $I = E + Y$.

(It may be remarked that the design of this package started out with separate options *and* the more traditional format "SEICD". This appeared to be too complex, for the reasons given. Thus the setup was simplified, and using the routines showed that this was beneficial. An example is the run on the intervention, in which the same set of Options[SIA] can be adapted to the new regime.)

PM. Still, applications of the options can differ. The user can call the routine SIA[Options, Check] for some diagnostics on the settings. For example, the parameters GammaSICD[] and MuSICD[] are used in the traditional setup of the models, and it may be useful to check that they add up to GammaSIA[] for the SIA.

```
Options[SIA]
```

$\{$Aqt0 $\to$ 0, Cqt0 $\to$ 0, DataMold $\to$ {S, I, A}, Dqt0 $\to$ 0, Iqt0 $\to$ Null, Nqt0 $\to$ 1.74 $\times 10^7$, Onset $\to$ 0, Seed $\to$ 100,

   Yqt0 $\to$ 0, AlphaSEYCD() $\to$ 1., BetaSIA() $\to$ 0.4, GammaSIA() $\to$ 0.1, PhiSICD() $\to$ 0.015, R0() $\to$ Null$\}$

```
SIA[Options, Check] // MatrixForm (* these are strings → values *)
```

$$\begin{pmatrix} \text{beta == R0} * \text{gamma (estimation)} \to 0.4 = 0.1 \text{ Null} \\ 0.1 = \text{GammaSICD() + MuSICD()} \\ \text{Aqt0 == Cqt0 + Dqt0} \to \text{True} \\ \text{Sqt0 is no parameter, Sq[t0] is in equations} \to 1.73999 \times 10^7 \\ \{\text{SIA and SICD use Seed instead of Itq0, \{Iqt0} \to \text{Null, Seed} \to 100\}\} \\ \{\text{Didactic SECYD has (Iqt0} \to \text{Null), Iqt0} \to \text{Null}\} \\ \{\text{Traditional SEYCD has (Yqt0} \to \text{Iqt0 + Seed), Yqt0} \to \text{Null + 100}\} \end{pmatrix}$$

3.2.5. Joint set of parameters and variables

The models not only have a joint set of parameters but also a joint set of variables $Sq[t]$, $Iq[t]$, $Aq[t]$, ... for the levels and $Sp[t]$, $Ip[t]$, $Ap[t]$, ... for the proportions. The routines defined for the variables of SIA can be used by SICD and SEYCD. The routines defined for *C* and *D* for the SICD model can also be used by SEYCD.

One would not expect anything less from an integrated didactic environment. Any disciplined writer would also generate such administration too. However, it is somewhat remarkable that in the many formulations of the S(E)IR(D) literature, mostly only those variable names and R_0, β and γ are the same, while there is a somewhat chaotic variety in other aspects. Let authors feel invited to

adopt more uniformity.

While variables Sq[t], Iq[t], Aq[t], ... have specifications that are also agreeable for computational writing, the package also provides for translation to shorter symbols $\mathcal{S}, \mathcal{I}, \mathcal{A}$, ... for the human eye.

3.2.6. Distinction between infections and symptomatic disease

The distinction between infection and symptomatic disease is fundamental in this discussion. Chapter 10 gives the conventional distinction between the *generation interval* and the *serial interval*. This notebook and package have no contribution to this discussion. There is only the advantage that the discussion has been rephrased within the confines of above redesign. A basic finding is that many reports in the literature apply the S(E)IR(D) → SI(EY)A(CD) model to symptomatic disease (e.g. the Goh (2020) "epidemic calculator"). For SARS-CoV-2, the crucial phenomenon is asymptomatic infectiousness.

3.2.7. Models SEYCDT and SEYCDB for estimation

Chapter 12 discusses two methods to present an intervention.

The first method has a Base SEYCD model alongside an Alternative SEYCD model, and the models differ in parameters and time of intervention.

The second method compares the Base SEYCD model with a model that uses the *Mathematica* feature of the WhenEvent statement: either SEYCDT for an intervention on R_0 or SEYCDB for an intervention on β. The WhenEvent formulation appears fruitful for estimation, so that not only the parameter before the intervention but also the parameter after the intervention can be estimated, taking into account that the observed number of deceased will be an effect of overlapping periods. Again, this estimation has educational objectives only. Practical application would require more data and also much more precise data (with e.g. regional aspects). Another advantage of this limited estimation exercise is that it clarifies the meaning of the parameters.

Having an intervention and estimation may require an administration of the *N*[0] that is used as the denominator of the proportions. This is not required for the latter SEYCDT and SEYCDB with the WhenEvent approach,

3.2.8. Modularity

A fundamental programming principle is to build with blocks that are fully functional and well-tested. The user often knows little about the building blocks and only requires that everything works as stated in the manual.

It is tempting to create flashy results (e.g. some plots below) using quick and dirty routines. In such cases, it takes only a small glitch and the whole edifice doesn't work anymore. Even worse: if programming isn't well-structured, then it is often impossible to trace its cause and repair the issue.

The present setup provides both the building blocks and the higher-up routines. The building block have been tested (by writing this notebook, also in steps) and are reliable. If there would occur glitches in the higher-up routines (e.g. entering 1/0) then it remains possible to redo the process in steps by using the underlying building blocks.

For example, in steps (available building blocks) we define the equations, assign parameter values, do the NDSolve for the path, set the level variables Sq, Iq, Aq, ... to the interpolated functions, set the proportion variables Sp, Ip, Ap, ... Normally all steps have a standard sequence, so there is also a routine to do the steps for us. The higher-up routines are less-tested than the building blocks, so if something happens, like with a parameter out of range, then it still is possible to do the steps. For example, while the package presents a main plot format with a particular choice of plotting options, and there would be a glitch, say by a newer version of *Mathematica* that changes something in plotting, it always remains possible to use the Plot routine that is standardly available within *Mathematica*.

3.2.9. Administration of scenarios

Apart from this discussion of what the models and package do, there is also the application that requires record keeping of assumptions, parameters and outcomes. This is handled in another package Survival`ApplySIA`, not shown or discussed here. My administration is organised but it has not been documented for use by others.

3.3. General setting on programming

3.3.1. Reasons for choosing *Mathematica*

Reasons for choosing *Mathematica* are the same as for writing packages in *Mathematica* in general:

- Use of a ***general environment***, so that code isn't locked-up within the confines of a particular application, but can be used in direct combination with other computable aspects, like estimation and economic implications. There is a price to pay though: this notebook and package come with the confinement of the *Mathematica* environment.

- While the ***language of mathematics*** is universal, also on the computer, Stephen Wolfram had a talent to bring it to the computer. Mathematics has been developing over millennia, and my impression is that *Mathematica* has caught this development in the best manner for application on the computer. Other applications obviously must satisfy mathematics if they are to work, and mathematics can have many dialects, so this reason need not be an universal truth itself, and likely it is only a matter of (acquired) taste.

- All this is embedded within my proposal for ***redesign of mathematics education,*** see https://zenodo.org/communities/re-engineering-math-ed/about.

I made a choice for *Mathematica* in 1993 and have much benefitted from it. Over the years, other computer algebra packages have not developed as much. New options like *R* and Python do not have the appeal of the language.

3.3.2. The Economics Pack

This notebook and its discussion are open access. At first my objective was to make the package open access too, but soon I found it more efficient to use features from *The Economics Pack,* which features had been developed in the past for good reason. The pack is available at a cost, see Colignatus (1995, 2020e). Considering that others can use notebooks in *Mathematica* only by also acquiring

Mathematica anyhow (except for the MathReader, now WolframPlayer), I dropped the notion of an open access package. The new package is basically for convenience. *The Economics Pack* already contained a section on the life sciences, and Survival`SIA` is included there. When the Pack has been installed on the computer, use the file explorer to find the notebooks. A complication is that the Help function of *Mathematica* has abolished the HelpBrowser, and switched to a paclet structure. The Pack hasn't adapted to this yet. However, the user can find the explanatory files by use of the file explorer.

```
ToFileName[{$TEPDirectory[], "Applications",
  "Economics", "Documentation", "English", "LifeSciences"}]
```

C:\Users\Eigenaar\AppData\Roaming\Mathematica\Applications\Economics\Documentation\English\
    LifeSciences\

3.4. The SIA package

```
Contents["Cool`Survival`SIA`"] (* click on symbols to find explanations *)
```

{

Cool`Survival`SIA` ⌄

Acquitted	Dp	MuSICD	QuarantineColours	SICD
Ag	Dq	Ng	*R*	Sq
AlphaSEYCD	Dqt0	Np	R0	Sqt0
Ap	Eg	Nq	R0DublinLotka	Susceptible
ApByR0Plot	Ep	Nqt0	ReadMeSIA	TforAp
Aq	Eq	Og	Recovered	TforAp50
Aqt0	Exposed	OmegaSICD	Rg	TforAp95
AttackRate	FindTime	Op	Rp	TforTopEq
BetaSIA	GammaSIA	Options$SEYCD	Rq	TforTopIq
BetaSIADivNqt0	GammaSICD	Options$SEYCDB	Seed	TforTopYq
Cg	ICU	Options$SEYCDT	SEYCD	TimeByR0
Cleared	Ig	Options$SIA	SEYCDB	TopIq
Cp	Infected	Options$SICD	SEYCDLettersOnly	Yg
Cq	Infectious	Oq	SEYCDNumericQ	Yp
Cqt0	Ip	Other	SEYCDT	Yq
Deceased	Iq	PhiSICD	Sg	Yqt0
Dg	Iqt0	PlotLegendsSIA	SIA	

}

? ReadMeSIA

Symbol

ReadMeSIA[] gives an overall explanation of the SIA\` package. See there for:

ReadMeSIA["Terminology"]

ReadMeSIA["Symbol"]

ReadMeSIA["Period"]

ReadMeSIA[model] for models SIA, SICD and SEYCD

ReadMeSIA[SEYCD, Simplify]

⌄

```
ReadMeSIA[]
```

The SIA` package recognises the levels (q) and proportions (p) with Vp = Vq / Nq[0] for variable V: N, S, E, I,
R, A, C, D, Y, O. The proportions are w.r.t. N[0] and explicitly not current Nq (thus not Vp = Vq
/ Nq). The didactic assumptions are A = C + D and I = E + Y (with Y the infectious). The package
uses the short names Vq. The long names like Exposed etcetera are available as symbols only.

The routines generally allow for user options when these are relevant.

ReadMeSIA["Terminology"] clarifies terms on mortality, fatality, factor and rate.
ReadMeSIA["Period"] gives equations on latency period and serial interval (using Strings).
ReadMeSIA["Symbol"] lists the meanings of the symbols.

The basic routines f are SIA, SICD and SEYCD. They operate via the standard Mathematica practice of
using options. We do SetOptions[SIA, ...] for the parameters, call sol = f[NDSolve, ...], call f[Set, sol]
to set Sq[t], Iq[t] etcetera, and then SetOptions[Plot, ...] and Plot[f[t], {t, 0, tmax}] and also f[Plot, ...].

We want to use Sq[t] for both numerical output at one time and as a variable
or a symbolic expression for solving at another time. To do so, the runnable equations
are put into Hold. We first clear the variables before submitting them to NDSolve.

The models can be presented in traditional or simplified form. Default is the latter.
ReadMeSIA[model] for SIA, SICD, SEYCD gives an even more simplified form, using $\beta* = \beta$ / Nqt0.
ReadMeSIA[SEYCD, Simplify] gives an even more simplified format.
For traditional form use f[Equations, TraditionalForm, ...] and f[N, TraditionalForm, ...]

PM 1. In Mathematica: N is the number function, E is the exponential number, I is the unit complex
number, C the coefficient, D the differential operatator, and O for Landau, and there are In[]
and Out[], so that we use q instead of n for the number. We avoid name conflicts, and actually
enhance clarity, by using Vq and Vp. At times, using Script font appears to work was well.
PM 2. Default parameters R0 –> 4 and γ –> 1/10 and φ –> 0.015 and α –> 1 are for Holland
and the SARS–Cov–2 pandemic, April 21 2020. PM. The SIA model uses $\beta = R0 * \gamma$

3.5. Initialisation cells for loading of The Economics Pack and Survival`SIA`

3.5.1. The Economics Pack

```
Needs["Economics`Pack`"]
```

```
ResetAll
```

{Utilities`CleanSlate`, Economics`Pack`, Cool`Tool`, Cool`Manager`, Cool`List`,
  Cool`Declare`, Cool`Context`, Cool`Common`, Cool`Inequality`, Cool`Graphics`,
  Cool`Compatibility`, System`, Global`, ComputationalGeometry`Surface`}

```
Economics[DataFile, Print → False]
```

```
Economics[LP, Print → False]
```

3.5.2. The survival packages

For documentation, see:

```
ToFileName[{$TEPDirectory[], "Applications",
  "Economics", "Documentation", "English", "LifeSciences"}]
```

C:\Users\Eigenaar\AppData\Roaming\Mathematica\Applications\Economics\Documentation\English\
    LifeSciences\

```
Economics[Survival]
```

The following packages concern the area of survival statistics and epidemiology:

> Survival`CEA`: cost−effectiveness analysis in epidemiology
>
> Survival`Common`: terms for the life table and its functions
>
> Survival`BioStatistics`: calculate m or q from observations
>
> Survival`Criteria`: life gain, time trade off, patient cost
>
> Survival`Data`: see ??DataSymbols
>
> Survival`Disease`: survival in health or disease state
>
> Survival`Epidemiology`: test statistics
>
> Survival`Graphics`: plotting routines
>
> Survival`JointTest`: test that consists of two other tests
>
> Survival`LifeTable`: life table construction and listing
>
> Survival`LifeTableByChiang`: idem, method Chiang
>
> Survival`MetaAnalysis`: pooled Relative Risk
>
> Survival`Screening`: detection phases and lead times
>
> Survival`TreatmentControl`: a.k.a. case−control studies
>
> Survival`BMI`: body mass index example
>
> Survival`SIA`: SI(EY)A(CD) version of S(E)IR(D)
>
> Survival`ApplySIA`: record keeping of such application

Symbol

SurvivalPackages[] gives an overview of the packages in survival statistics and epidemiology.

These packages aren't supported with palettes yet.

Mathematica's I is the imaginary number (complex operator), but the String "I" may be used for Incidence

˅

`$ContextPath`

3.5.3. ApplySIA` (not documented)

`Economics["Survival`ApplySIA`"]` (∗SIABasicTable∗)

˅ **Cool`Survival`ApplySIA`**

CalendarTimeLine	**SEYCDBasic**	**SIABasicRun**	**SICDBasicRun**	**UpdateSICDBasicR-un**
DateOfOnset	**SEYCDBasicRun**	**SIABasicTable**	**UpdateSEYCDBasic-Run**	
SARSCov2NLTimeli-ne	**SIABasic**	**SICDBasic**	**UpdateSIABasicRun**	

```
SetOptions[PutGetDefinition, Directory ->
   ToFileName[{$TEPDirectory[], "Applications", "Economics", "Data"}]];
PutGetDefinition[Get, SIABasicRun]; (* used in the above *)
PutGetDefinition[Get, SEYCDBasicRun]; (* used in the above *)
PutGetDefinition[Get, SIATimeAsFunctionOfR0]; (* used in the above *)
PutGetDefinition[Get, SEYCDTimeAsFunctionOfR0]; (* used in the above *)
NqSeed = 17.4 × 10^6;
```

```
SetOptions[SetDatabank, Databank -> SEYCD];
```

```
SIABasicTable[SEYCDBasicRun, 100, {Sp, Ip, Dp, Cp, Np, Ap}]
```

	Date	Day	Sp	Ip	Dp	Cp	Np	Ap
Onset	2020−02−23	0	100.0	0.0	0.0	0.0	100	0.0
1st Death	2020−03−06	12	100.0	0.0	0.0	0.0	100.0	0.0
1st Intervention	2020−03−18	24	99.9	0.1	0.0	0.0	100.0	0.0
Top if free	2020−04−21	58	25.0	40.3	0.5	34.1	99.5	34.7
Ap = 50%	2020−04−25	62	13.5	36.5	0.7	49.2	99.2	50.0
95% Ap[∞]	2020−05−22	89	2.4	4.5	1.4	91.7	98.6	93.1
Limit if free	∞	∞	2.0	0.0	1.5	96.5	98.5	98.0

4. Compartments

4.1. Principle

Principles for this chapter are:

(1) (Public health) *compartments* appear to be relevant for quarantine controls. The following is a suggestion for such compartments.

(2) The SARS-CoV-2 pandemic requires *clear communication in and over a variety of cultures in the world*. We should use as much common ground as possible.

(3) One common ground is the human sensitivity to colours. Colour coding seems rather important for effective communication. Humans are sensitive for very subtle differences in colours, but, while the rainbow shows all frequencies, or colours ordered by hues, there is a tendency to see only some 10 hues, differing by about 0.1 degrees.

4.2. The meaning of red for regulation

Like with traffic lights, we need colours to manage not only people but also *locations*. It is insufficient to paint "dangerous cars" *red* and "safe cars" *green*. Red is the general colour for alert status (likely because blood is red), and it is best to use red for the *barrier* between any quarantine group rather than for a particular group.

The basic idea is that red identifies a barrier.

The following suggested colour scheme uses identifiable patterns in the hue or RGB schemes.

Rainbow
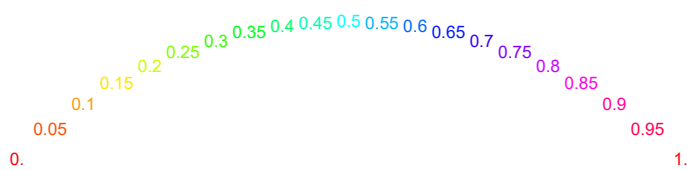
In the RGB rainbow:

- B goes from 0 (uninfected) to 1 (infected) (uphill only).

- R has the path 1-0-1 (valley).

- G has the path 0-1-0 (up and down a hill).

However, the rainbow is subsequently rearranged, namely to put red *between* the *vulnerable* and *less vulnerable* quarantine classes.

The interpretation is: *the further a quarantine class is away from the barrier, the more extreme the position will be*: either uninfected (B is 0) and more vulnerable, or infected (B is 1) and getting sicker. PM. The colour of DeeperPink for the ICU has an infringement upon the B = 0 or B = 1 dichotomy.

4.3. Suggested colour scheme

The colours might also be used for SI(EY)A(CD)-type of modeling with more compartments. The scheme allows that SI(EY)A(CD) is used for both infections and symptomatic disease. The "bluepurple" group was included for symptomatic disease. For quarantine it would likely be joined with "freshpurple".

QuarantineColours[]

Colour	Name	RGB			Colour scheme of combined Hue (no Venn diagram)
		Hue	R	G	B	
	Yellow	0.16	1.	1.	0.	Uninfected and vulnerable (elderly or diseased)
	Orange	0.09	1.	0.5	0.	Uninfected, less vulnerable (young & no disease), preferably not in
	Red	0.	1.	0.	0.	Barrrier between quarantine zones
	Green	0.3	0.	1.	0.	Immune: cured from an infection, and tested to be no longer a car
	Cyan	0.48	0.	1.	0.88	Unsymptomatic or untested, not in the vulnerable group, infectabl
	Skyblue	0.59	0.	0.46	1.	Unsymptomatic, latency or incubation (modeling)
	Blue	0.67	0.	0.	1.	Infected, (possibly) sick, and less vulnerable (modeling infectious)
	BluePurple	0.72	0.32	0.	1.	Symptomatic (modeling), or infected (tested), likely self–quarantir
	FreshPurple	0.77	0.62	0.	1.	Infected, (likely) sick, and vulnerable
	Magenta	0.83	1.	0.	1.	Hospitalised
	DeeperPink	0.91	1.	0.	0.54	Intensive care unit, complications comorbidity

For SI(EY)A(CD) modeling, mortality can be indicated by dashed Black / White lines, which fits European / Chinese conventions on these "colours".

We can put the RGB combinations in the familiar RGB-plot reminding of a Venn-diagram. (PM. Deeper pink is not shown here.)

(a) The larger disks give a Venn diagram indeed. For example, the yellow group is safe because they are protected by a quarantine barrier. For example, the blue group is infected but it is unknown whether they will become immune (green) or have to be hospitalised (magenta).

(b) The smaller disks do not form a Venn diagram. They highlight special conditions. For example, for "fresh purple", the barrier is breached, and those of the vulnerable group must be hospitalised.

(Hospitalisation itself shows that the barrier was crucial for them.)

QuarantineColours[RGBColor]
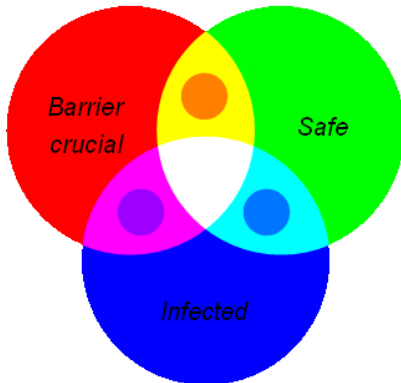
4.4. Use of these colours in the SI(EY)A(CD) plots

These colours will also be used for plotting the SI(EY)A(CD) graphs. For the Acquitted, the combination of Cleared (Green) and Deceased (Black and White) results into some brownish colour.

The SI(EY)A(CD) model, while officially named for infections, is also used by researchers for symptomatic disease. Instead of the "exposed" there is "incubation", and instead of "infectious" there is "symptomatic". In itself this is not so strange, see Chapter 10 about the distinction between the "generation interval" and the "serial interval". For the SARS-CoV-2 virus, it appears that both such applications are used indeed. These models however have different assumptions, and there is some risk of confusion. Thus it is necessary to have legends that fit the type of modeling assumptions. These are the two sets of categories, and the symptomatic case tends to be a bit darker or up in the rainbow than the infections.

The command PlotLegendsSIA[SEYCD] sets the legends and colour schemes for the standard interpretation of SIA, SICD and SEYCD for infections. PlotLegendsSIA["Symptoms"] sets those for symptomatic disease.

PlotLegendsSIA[Show, All] (∗ dashing statement deliberate ∗)

	Susceptible			Susceptible
	Exposed			Incubating
	Infectious			Symptomatic
{Dashing[{Small, Small}], ■}	Infected		{Dashing[{Small, Small}], ■}	Diseased
	Cleared			Remissed
{Dashing[{Small, Small}], ■}	Deceased		{Dashing[{Small, Small}], ■}	Deceased
	Acquitted			Acquitted
	Population			Population
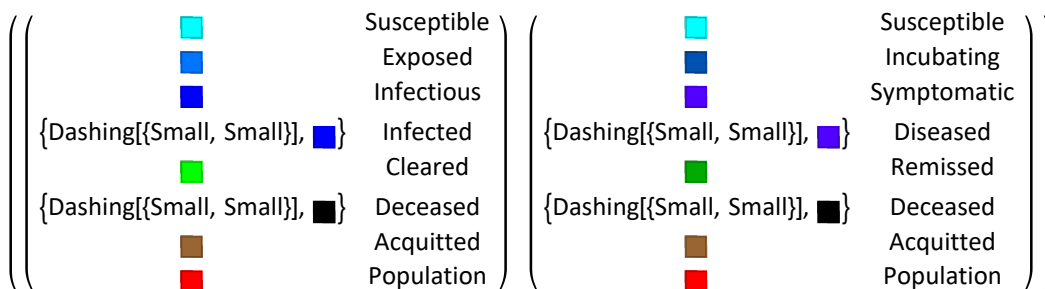
4.5. When in lack of colours

These ten or eleven "identified possible quarantine groups" and their colours will not be enough to serve the large variety of cases. However, it is always possible to combine codes. We could show colours in bands, and not blend the hues.

For example, an icon that shows both *magenta* and *freshpurple* indicates that a *hospitalised* person belongs to the *vulnerable* group, and if this is clear, then we might add a little bit of yellow to indicate that the origin is from the elderly and comorbidity group or a little bit of orange for the younger group of vulnerables. In a graphical plot we could have a line dashed with those colours (except that this hasn't been implemented yet).

```
{ColourBand[{{Magenta, 1}, { FreshPurple, 1}, {Yellow, 1/4}}, 3],
 ColourBand[{{Magenta, 1}, { FreshPurple, 1}, {Orange, 1/4}}, 3]}
```


PM 1. There is a curious effect. In everyday life, when we use colors red and green, we do not necessarily think of traffic lights. With above colour scheme for infection status, the colours in a graph on another subject may cause the question whether that other subject concerns infection status or not ... It is something to get used to.

PM 2. It would be fortunate if *Mathematica* adapts a notion of "generalised colour" that includes such band-dashing in different normal colours, so that the options for Mesh and Legend can be combined. Now a trick is to plot a function twice, in both a base colour and a dashed other, but this means that one also gets a double legend.

5. SIA

Clear

Clear all relevant variables since we may do different runs.

```
SIA[Clear] (* not Clear[SIA] ! *)
```

```
SetOptions[SetDatabank, Databank → SIA]; (* for Explain *)
```

5.1. Basic model in standard formulas

The classic SIA model has time-dependent variables S, I and A, their derivatives, their starting values $S[t_0]$, $I[t_0]$ and $A[t_0]$, and parameters. The sum of the three variables is the population N. It will be convenient for Options[SIA] to include the starting values as parameters without brackets too. The starting value of the infectious compartment $I[t_0]$ will be called the *Seed.* $A[t_0]$ will get parameter *Aqt0.* And $S[t_0]$ will not need a parameter since it can be determined as *Nqt0 - Seed - Aqt0.*

```
ReadMeSIA[SIA]
```

$\mathcal{S}' = -\beta^* \, \mathcal{S} \, \mathcal{I}$

$\mathcal{A}' = \gamma \, \mathcal{I}$

$\mathcal{I}' = -\mathcal{A}' - \mathcal{S}'$

$\mathcal{S}(t0) = -\text{Aqt0} - \text{Seed} + \mathcal{N}(t0)$

$\mathcal{I}(t0) = \text{Seed}$

$\mathcal{A}(t0) = \text{Aqt0}$

PM 1. It is logical to state *S'* and *A'* first and subsequently derive *I'*, as Smith & Moore (2001, 2004) do.

PM 2. $\beta^* = \beta / N[t_0]$. This gives a tiny value for β^* but simplifies the model. Below we immediately drop this simplification.

PM 3. Above model is in levels (q). Division by $N[t_0]$ or setting $N[t_0] = 1$ gives proportions (p) (and $\beta^* = \beta$).

5.2. Computable format

The above is pleasing for the human mind but potentially less reliable for computer algebra and computable publications.

The model is best treated in symbolic manner.

(1) Some discussions use proportions (p) in terms of $N[0]$. At base, it is more natural to use levels (q). The SIA model has population $N[0] = S + I + A$. In SICD the surviving population is $N[t] = N[0] - D$, so that over time the proportions in terms of $N[t]$ change even more, and it is important to keep the proportions w.r.t. $N[0]$. The actual model concerns *population densities*, with probabilities of contacts. Standard discussions of SICD thus properly normalize for $N = N[0]$, but working in proportions can hide changes in normalisation (e.g. changing to another year), and it is better to emphasize the notion of density from the start.

(2) Conventional presentation seems to be oriented at solving the differential equations by hand, at least as far as possible. That is, the conventional equation for *I'* actually writes out *S'* fully again. For computer algebra we can use the simplified form, like Smith & Moore (2001, 2004) do.

(3) The routines are served by symbolic expressions while readability is served by single and Greek letters. There is no fundamental difference, except that single and Greek letters are used for multiple purposes, while computer algebra expressions better be unique over a wide range. With $x_p = x_q / N_q[0]$ we can use *Script* for levels (q) (e.g. $\mathcal{S}$ = ESC scS ESC) and *Formal* (with dots below) for proportions (S = ESC .S ESC). Using different fonts will prevent conflicts with default use in *Mathematica* of *N, E, I, C* and *D.* The derivative prime has been enlarged for readability.

(4) Conventional discussions assume a start at $t = 0$ and Sp[0] ≈ 1, but for combining modeling outcomes with interventions on the parameters it is useful to have formal parameters for the values at t_0. Lin et al. (2010) show a simulation with $N = 100$ and Ip[0] = 1, and then it is important to take Sp[0] = 99.

(5) Conventional discussions have a "basic reproduction number" R_0. The use of *R* is confusing in the standard SIR formulation, while *R* is already used confusingly for both Removed and Recovered

compartments. Eric Weisstein (2020a) at *Mathworld* states: "Note that the choice of the notation R_0 is a bit unfortunate, since it has nothing to do with R." We avoid this by using $A = C + D$. There remains a lingering question on R_0 though. See Heesterbeek (2002) for a brief history, where the notion itself was started without its own symbol. Delamater et al. (2019) clarify that George MacDonald introduced the notion of an own symbol in the epidemiological literature in the 1950s, using Z_0. The term R_0 is so fundamental to the literature that it would be wrong to consider changing the symbol. (However, for programming, when we replace R in SEIR(D) by A and $C,$ there might still be confusion whether we should not also replace R_0, since confusion might still arise when readers also refer to other publications in this kind of modeling. For the computing environment below we should avoid needless confusion and computing error. On occasion within (hidden) programming we may use the neutral notion of a "factor" $F[t]$ with $F[0]$ a value at t_0 or a constant. There appears to be little need for it actually. With the use of A and C it should be much less confusing to use R_0. Perhaps the following might be a test for this.)

Subsequently, there is the issue of a time-varying factor, for example when the quarantines are installed so that the original outbreak with $R[0]$ has a different path than the new situation with $R[t]$, for time-dependent measures. The R_0 actually is time-dependent, with $R_0[t] = \beta[t] / \gamma[t]$ as the *dynamic* factor, and thus also $R_0[0]$. The literature is obviously aware that R_0 is dependent upon time and region, but for modeling it makes no sense to keep a subscript zero in all expressions. Subsequently, there is also the *effective* reproduction factor $R[t]$ that corrects $R_0[t]$ for Sp[t]. We will be using:

$R_0[t] = \beta[t] / \gamma[t] = R0[t]$ *dynamic* reproduction factor (with influence by interventions)

$R[t] = R0[t] S[t] / N[t_0]$ or in short $R = R0\, S_p$ *effective* reproduction factor (idem)

$R_{00} = R0_0 = R0[0] = R[0] = R_0$ basic reproduction factor R_0 (commonly no interventions)

In SIA the infection top is given by $R[t] = 1$.

5.3. The different formats

An overview of the combinations of *traditional / simplified* and *computable / readable* formats is as follows (dropping β^*).

```
symbolic = {SIA[Equations, t0, t],
            SIA[Equations, TraditionalForm, t0, t]} // Transpose // TableForm
```

$Sq'(t) = -\frac{BetaSIA()\, Iq(t)\, Sq(t)}{Nqt0}$ $Sq'(t) = -\frac{BetaSIA()\, Iq(t)\, Sq(t)}{Nqt0}$

$Aq'(t) = GammaSIA()\, Iq(t)$ $Iq'(t) = Iq(t) \left(\frac{BetaSIA()\, Sq(t)}{Nqt0} - GammaSIA() \right)$

$Iq'(t) = -Aq'(t) - Sq'(t)$ $Aq'(t) = GammaSIA()\, Iq(t)$

$Sq(t0) = -Aqt0 + Nqt0 - Seed$ $Sq(t0) = -Aqt0 + Nqt0 - Seed$

$Iq(t0) = Seed$ $Iq(t0) = Seed$

$Aq(t0) = Aqt0$ $Aq(t0) = Aqt0$

```
symbolic // SEYCDLettersOnly   (* without beta-star *)
```

$$S' = -\frac{\beta\,S\,I}{\mathcal{N}(\text{t0})} \qquad\qquad S' = -\frac{\beta\,S\,I}{\mathcal{N}(\text{t0})}$$

$$\mathcal{A}' = \gamma\,I \qquad\qquad I' = I\left(\frac{\beta\,S}{\mathcal{N}(\text{t0})} - \gamma\right)$$

$$I' = -\mathcal{A}' - S' \qquad\qquad \mathcal{A}' = \gamma\,I$$

$S(\text{t0}) = -\text{Aqt0} - \text{Seed} + \mathcal{N}(\text{t0}) \qquad S(\text{t0}) = -\text{Aqt0} - \text{Seed} + \mathcal{N}(\text{t0})$

$I(\text{t0}) = \text{Seed} \qquad\qquad\qquad\quad I(\text{t0}) = \text{Seed}$

$\mathcal{A}(\text{t0}) = \text{Aqt0} \qquad\qquad\qquad\quad \mathcal{A}(\text{t0}) = \text{Aqt0}$

5.4. Model formulation, choice of R_0 = 4 and acquittal period of 10 days

To be able to run the model, we must set parameters and starting values. For the model formulation, we would prefer to use constant parameters.

See the discussion above for the choice of R_0 = 4 and the mean acquittal period of $1/\gamma$ = 10 days.

In general there is the time-dependent (dynamic) transmission or reproduction factor $R_0[t] = \beta[t] / \gamma[t]$ = BetaSIA[t] / GammaSIA[t]. At the beginning, or for a constant: $R[0] = \beta / \gamma = \beta\,\delta$ with $\delta = (1 / \gamma)$ the mean acquittal period (duration).

Recall the *Introduction,* section 1.4.1, and let us now insert the dimensions: [#] for number, [#/# = 1] for factor, and [P] for period. R_0 [1 = #/#] is a factor since it gives a (new) number of infected units per (old, single) number of infected units. A linear calculation defines the (linear) growth rate as $r = (R_0 - 1) / \delta$ [1 / P], in which $R_0 - 1$ [#/#] is the increase of the next generation compared to the single originator, and δ [P] is the duration over which this increase happens. Then $R_0 = 1 + \delta\,r$ [1 = 1 + P * 1/P]. Then we may take $\gamma = 1/\delta$ [1/P] as the rate per time, and rework $R_0 = (\gamma + r) / \gamma$ and define $\beta = \gamma + r$ [1/P = 1/P + 1/P]*, and then find $R_0 = \beta / \gamma$ [1 = 1/P / (1/P)] and $r = \beta - \gamma$. (PM. Frits de Jong (1918-1976) advised dimension analysis in economics; he was the thesis supervisor of Wim Duisenberg (1959-2005), the first president of the ECB.)

In this format, the model formulation should actually not use β or $\beta[t]$ but rather $R_0[t] / \delta[t]$, since $R_0[t]$ and $\delta[t]$ are the independent parameters, while β is only a result of those two. In our case there are 4 / 10 infections per unit period. When the situation is changed, e.g. by an intervention, then $R_0[t]$ and $\delta[t]$ change, with β as a result, and it would be unwise to attach much value to β or think that it ought to remain constant in some way (unless a reason is provided). The assumption is that we can do little about the biological process of clearing and recovery to health, or the path towards death, and that $\delta[t]$ is rather the relevant observation period.

Section 1.4.4 stated: The growth of the infected = infectious in SIA can be found in $r = \text{dLog}[I] = I' / I = i = \beta\,Sp - \gamma$, see also Section 5.7. It will be useful to identify the *dynamic* reproductive factor $R_0[t] = \beta[t] / \gamma[t]$ and the *effective* reproductive factor $R[t] = R_0[t]\,Sp[t]$, where the latter corrects for the influence of the declining share of susceptibles. In SIA thus $r = (R_0\,Sp - 1) / \delta$, for $\gamma = 1/\delta$, which gives an effective version of the above. Here Sp would be a dimensionless number, as units of susceptibles per units of the base population.

SI(EY)A(CD) models are conventionally defined in terms of β and γ and not in R_0 and δ. Conventionally, the following decomposition is given to explain that $R_0[t]$ is a dimensionless number. In measurement, the number of contacts can be tallied, including the share of infectious contacts.

$R_0[t] = \tau * \kappa * \iota = (\# \frac{\text{infectious}}{\text{contact}}) * (\# \frac{\text{contact}}{\text{period}}) * (\# \frac{\text{period}}{\text{infectious}})$.

Unchanged is that $R_0[t] = \beta[t] / \gamma[t]$, so that $\beta[t] = (\# \frac{\text{infectious}}{\text{contact}}) * (\# \frac{\text{contact}}{\text{period}})$, or $\beta[t] = (\# \frac{\text{infectious}}{\text{period}})$. We can skip the reference to the infectious since it actually is a compartment, and the unit of account still is the population unit. Measuring by means of an *intermediate* step on contacts does not change the *final* result that the dimension of β is [1/P], as used in above growth accounting w.r.t. the renewal equation. The contacts are an element in the statistical measurement of β but are not part of its dimension; its dimension is [1/P] like γ.

This can also be shown by looking at the growth equations for *S* and *I* in the SIA model, that have the term *I* * *Sp.* This term can be seen as an indicator for the number of contacts. It becomes problematic when it is argued that only β is the factor that translates such contacts into a rise of *I* and a waning of *S.* Assume that β has the dimension [1/c/P] = [1/(c P)]. With a dimension analysis we get: *S'* [change in #S / P] = - β [1 / c / P] [c / (#I #S / #N)] *I Sp.* We might assume [c / (#I #S/#N)] = [1]. However, [#S / #N] = [1] since these are both population units, and [#I] is needed for *S'.* Thus a contact dimension for β hangs in the air.

It is best to think of γ as a normalisation parameter, to allow that measurements for different regions can be compared. It is rather not seen as a policy variable. This was already mentioned in the *Introduction.* See Section 12.10 on estimation. For our case of $\beta = R_0 \gamma = 4 / 10 = 0.4$, suppose that an intervention on (self-) quarantine reduces the "real infectious period" from 10 to 5 days. So we may want to use a new value of $\gamma = 1/5 = 0.2$. However, this means that the number of acquitted rises much faster, and also the number of deaths contained in them. A simulation run might cause the paradoxical result of increased death in the short run. Instead, since the argument is that the number of infections is reduced, the reduction of R_0 may also be modeled by a lower β. In fact, given that mostly the infected are put in quarantine, β might reduce by more. Thus, for modeling such a case of quarantine, it would be wiser to drop the idea that $1/\gamma$ is really associated with some "real infectious period", and we better regard it as a normalising constant.

The package includes this equation but it is only used as a reminder.

```
SIA["R0[t]"]
```

$\text{R0}(t) = \dfrac{\text{BetaSIA}(t)}{\text{GammaSIA}()}$

```
Solve[SIA["R0[t]"] //. {R0[t] → 4, GammaSIA[] → 1/10}, BetaSIA[t]]
```

$\left\{\left\{\text{BetaSIA}(t) \to \dfrac{2}{5}\right\}\right\}$

? R

> Symbol

R[t] = (♯ infectious / contact) * (♯ contact / period) * (♯ period / infectious), which is a dimensionless

number. It is the time−dependent reproduction factor. R[0] = R0. The literature has R also

for R = Recovered and/or R = Removed. We resolve the latter by taking A = Acquitted

and C = Cleared, and thus can use R and R0 unambigously for the reproduction factor.

Tau = infectious / contact = the transmissibility (probability of contact)

Kappa = contact / period = the average rate of contact between Iq and Sq

Iota = period / infectious = duration of infectiousness

SIA assumes that the acquittal period is also infectious.

BetaSIA[t] = 1 / period, or the inverse of the period between contacts

Likely the best reading is that b = R0 * g = R0 / (1/g), or that gamma

disperses an already given number of infections R0 over the 1/gamma period

⌄

? R0

> Symbol

R0[t] = BetaSIA[t] / GammaSIA[t] dynamic reproduction factor

R0[] = R0[0] = R[0]. The value at t = 0, or a constant. It is called the *basic* reproduction factor. The

number of new cases per old case (also allowing the identification of the (mean) period over which

this happens). It is the number of new infections on the brink of the first day by a single infectious

seed in an otherwise fully susceptible population. See Heesterbeek (2002), Jones (2007) and

Delamater et al. (2019). (The R stands for "reproduction" and not for "Recovered" or "Removed".)

R0[] = BetaSIA[] * 1 / GammaSIA[]

Probably the best way to understand R0 is to replace BetaSIA[] –>

R0[] * GammaSIA[] with the latter also R0[] / (1/GammaSIA[]), so that R0[] are the

*** given *** number of infectious cases that are spread of period 1/GammaSIA[]

⌄

? BetaSIA

> Symbol

BetaSIA[t] is the transmission factor per period, at time t

BetaSIA[] is a symbol in the options for a constant value

⌄

? GammaSIA

Symbol

GammaSIA[] is a constant rate of the Infectious who will be acquitted

    per day, i.e. be no longer infectious. (But might later die from pneumonia ...)

1 / GammaSIA[] is the mean acquittal period (a unit can only be acquitted once).

    NB. SIA assumes that the acquittal period is also the infectious period.

PM. Since the number of infections has been given by R0[], and R0[] determines

    all properties, 1/GammaSIA[] only spreads the given number of infections

---

5.5. Fast-track run and plot

This section directly runs the model and plots the results. The next section does the same in steps, so that the user might adapt steps and take more advantage of the environment in *Mathematica.*

The default options apply to Holland of February 23 2020 onwards. The Dutch population in 2020 is about 17.4 million people. Let us *set* NqSeed = 17.4 * 10^6, and subsequently use the *rule* Nqt0 → NqSeed, for the formal parameter Nqt0 in the equation Nq[t0] = Nqt0. NqSeed is not a parameter in the model, and only a handy variable at the user level.

```
NqSeed = 17.4 × 10^6;
SetOptions[SIA, Nqt0 → NqSeed]
```

$\{$Aqt0 → 0, Cqt0 → 0, DataMold → {S, I, A}, Dqt0 → 0, Iqt0 → Null, Nqt0 → 1.74×10^7, Onset → 0, Seed → 100,

    Yqt0 → 0, AlphaSEYCD() → 1., BetaSIA() → 0.4, GammaSIA() → 0.1, PhiSICD() → 0.015, R0() → Null$\}$

The infections are driven by the equation Iq[t0] == Seed. (Seed *is* parameter in the model.) The carnival festivities in the South of Holland were on February 23 2020. It is not unlikely that some 100 infections were seeded by people having returned from Italy and Austria from work and ski holidays. The options already have default Seed → 100. When we use a different value of Seed, then this is "baked" into this model solution, and we must take heed of routines that rely upon Options[SIA] for the value of Seed in these options. We can mention the parameter here so that the user can check more easily that model solutions change when adapting the Seed value.

When the equations are set with particular parameter values, we can use an arbitrary label, in the case "newModel" to identify the particular setting. Then newModel[t] would show the chosen model.

```
SIA[Run, Pr, newModel, t, 150, Seed → 100];
```

The variables are now available in levels and proportions. The label "store" is arbitrary. The present assignment allows us to later refer to this particular outcome.

```
store[SIA] = SIA[75] (* SIA[t] calls levels *)
```

$\{$415 556., 738 924., $1.62455 \times 10^7\}$

```
SIA[Pr, 75] (* SIA[Pr, t] calls proportions *)
```

{0.0238825, 0.0424669, 0.933651}

It is useful to have plots that show both levels and proportions. The axes assume that it is obvious what are levels and what are proportions.

```
SIA[Plot, 0, 150]
```
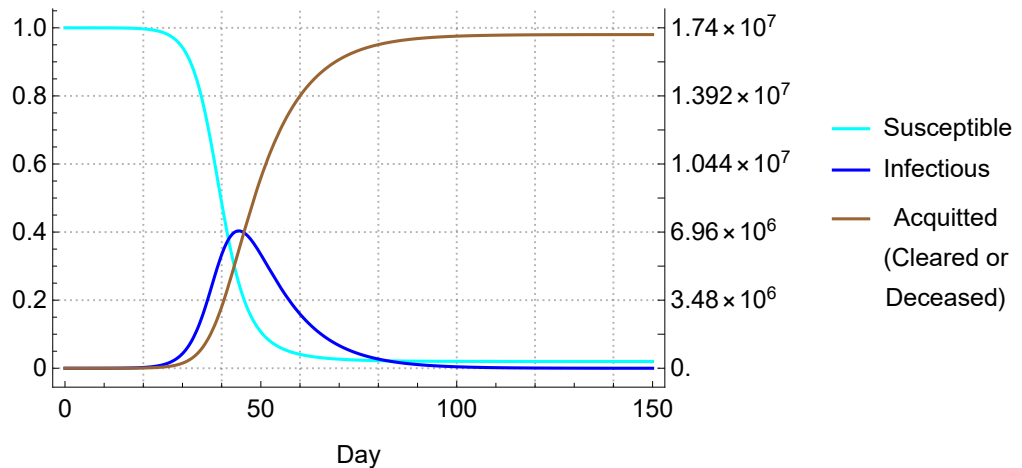

5.6. Interpretation of the result and finding crucial moments of the path

We can read values from the graph but numerical outcome is more useful.

Let us calculate the day when 95% of the Dutch population would have been acquitted from the pandemic, given the parameters and without the intervention of March 15-19. (It so happens that the R_0 is high enough so that this level can be reached.)

```
t95PercentOfN = FindRoot @@ {0.95 == Aq[t] / NqSeed, {t, 100}} // Quiet
```

$\{t \rightarrow 79.7353\}$

```
DayPlus[{2020, 02, 23}, Round[t /. t95PercentOfN]]
```

Day: Wed 13 May 2020

Above moment in time is relevant because $R_0 = 4$, with a high prevalence in the limit. For lower values, the limit value will be lower than 95%.

It is better to define t_{95} as the time when 95% of Ap[∞] is reached.

The following routines find some crucial moments on the path. When we round the days from onset then we can determine the calendar date.

NB. These dates use SIA and not SEYCD shown in the *Introduction.* (Check that $\alpha = 1$ per generation adds up over time.)

```
{TforTopIq[], TforAp50[], TforAp95[4 ]} // Quiet (* R0 = 4 *)
```

{44.3682, 48.3071, 74.4295}

```
SIABasicTable[SIABasicRun, 100] (* from another package and another run: locked *)
```

	Date	Day	Sp	Ip	Ap
Onset	2020–02–23	0	100.0	0.0	0.0
1st Death	2020–03–06	12	100.0	0.0	0.0
1st Intervention	2020–03–18	24	99.0	0.8	0.3
Top if free	2020–04–07	44	25.0	40.3	34.7
Ap = 50%	2020–04–11	48	13.5	36.5	50.0
95% Ap[∞]	2020–05–07	74	2.4	4.5	93.1
Limit if free	∞	∞	2.0	0.0	98.0

The above becomes more acute when we can also print the mortality levels, which requires SICD.

The latter table assumed no interventions. Before modeling interventions, it is better to first develop SICD and SEYCD. And before doing the latter, it is better to first understand more about SIA.

The SIA model gives particular estimates for the days of particular events. The SEYCD model will give another estimate, normally a later date because of the delay caused by latency.

```
ApByR0Plot[] (* using the days of the SIA model *)
```
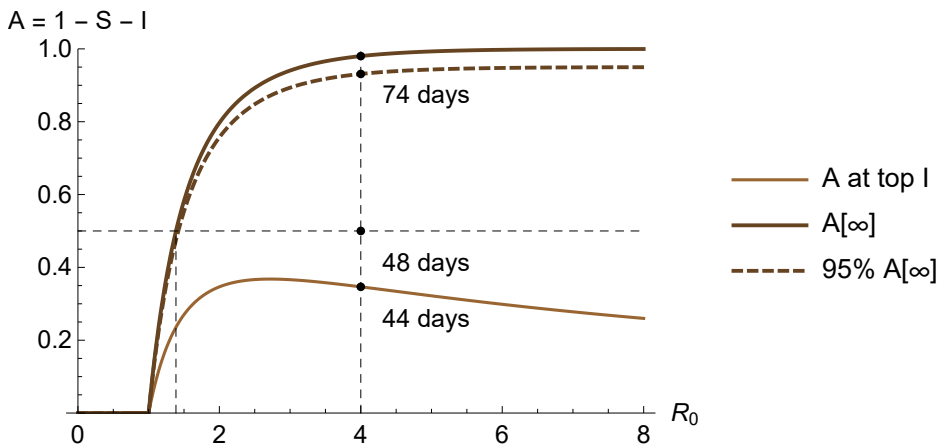

5.7. Growth rates of SIA

5.7.1. Formal

The growth rate of function f is: $g = \text{Log}[f]' = f'/f$. Fully $g[t] = d\text{Log}[f[t]] / dt$. Shortest $d\text{Log}[f]$.

A constant growth rate gives exponential growth.

The SIA equations relate their derivatives to their levels. Division by the levels gives the growth rates.

Using above numerical approximation, it is also straightforward to plot these growth rates.

The SIA equations give these growth rates.

$s = S'/S$ gives $s = -\beta I_p$

$i = I'/I$ gives $i = \beta S_p - \gamma$

$a = \mathcal{A}'/\mathcal{A}$ gives $a = \gamma I_p / \mathcal{A}_p$

An alternative but less elegant relation of *i* is:

$i = \mathcal{I}' / \mathcal{I} = (-\mathcal{S}' - \mathcal{A}') / \mathcal{I} = -s\,(\mathcal{S}_p / \mathcal{I}_p) - a\,(\mathcal{A}_p / \mathcal{I}_p)$

5.7.2. An important analytical result for the long term limit values

An important relation is $\mathcal{S} = \mathcal{S}_0\,\mathrm{Exp}[- R_0\,(\mathcal{A}_p - \mathcal{A}_p[0])]$, e.g. useful for the limit values.

Taking derivatives gives $\mathcal{S}' = \mathcal{S}_0\,\mathrm{Exp}[- R_0\,(\mathcal{A}_p - \mathcal{A}_p[0])]\,(-R_0\,\mathcal{A}_p') = -R_0\,\mathcal{S}\,\mathcal{A}_p'$. Then $s = \mathcal{S}' / \mathcal{S}$ gives $s = -R_0\,\mathcal{A}_p'$ too. Comparing the latter with the above s, we only recover already known $\mathcal{A}_p' = \gamma\,\mathcal{I}_p$. Taking the steps back however *proves* the important key relation.

5.7.3. Plotting numerical results

We use the variables *Sg, Ig* and *Ag now.*

```
SIA[GrowthRate, Equations, Hold, t]
```

$\mathrm{Hold}\Big[\Big\{\mathrm{Sg}(t) = -\dfrac{\mathrm{BetaSIA}()\,\mathrm{Iq}(t)}{\mathrm{Nqt0}},\ \mathrm{Ig}(t) = \dfrac{\mathrm{BetaSIA}()\,\mathrm{Sq}(t)}{\mathrm{Nqt0}} - \mathrm{GammaSIA}(),\ \mathrm{Ag}(t) = \dfrac{\mathrm{GammaSIA}()\,\mathrm{Iq}(t)}{\mathrm{Aq}(t)}\Big\}\Big]$

```
SIA[GrowthRate, Set];
```

```
SIA[GrowthRate, 30] (* check that there are numbers *)
```

{−0.0171973, 0.276838, 0.288342}

For plotting, avoid division by 0 at $t = 0$ when $A = 0$.

```
sirplotoptions = SIA[Plot, Options, PlotRange → All];
```

```
Plot @@ {SIA[GrowthRate, t], {t, 3, 100},
  AxesLabel → {"Day", "Growth rate"}, sirplotoptions}
```
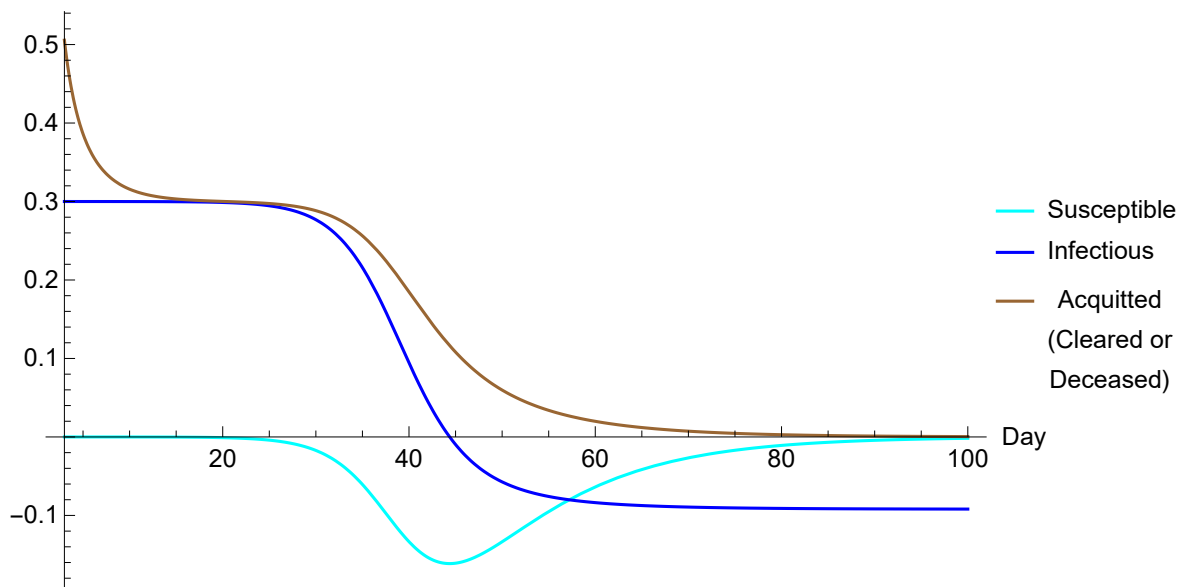

(1) When the growth rate of *I* intersects with the horizontal axis, with $i = 0$, then the infections reach their Top.

(2) The graph confirms that especially *I* has a long period (25 days) with a fairly constant exponential

growth rate of $i \approx \beta - \gamma = 0.3$. In percentages this is Exp[0.3] - 1 = 35%. Towards the end it has such growth at $i \approx -\gamma$.

(3) We may now be better aware that the decline of the susceptible has most negative growth at $s' = 0$ which gives $I' = 0$. This conforms with the top of the infectiousness. The most negative growth is at $-\beta$ Ip$_{Top}$.

(4) In terms of growth, *S* and *I* have more regular patterns, and *A* indeed follows as a remainder. This is somewhat remarkable, since the level and share plots suggest that *S* and *A* are dominant and that *I* is the remainder.

5.7.4. Difference between SIA and exponential functions

SIA has exponential phases but differs from the exponential process. There is no need for another method for numerical approximation of these equations, since we already have a good approach above. However, it helps to understand the growth rates by looking at the equations from numerical approximation too.

Let us use the notation $\hat{\mathcal{I}}_p$ for the time dependent variable that in a numerical simulation is held constant for the short moment of updating. When exponential relations do not hold exactly there can be a dynamic approximation with locally held constant shares. A stepwise simulation is: (a) start with t = 0 and calculate the shares, (b) calculate $\mathcal{S}$ and $\mathcal{I}$ from the exponential expressions, while holding the shares $\hat{\mathcal{I}}_p$ and $\hat{\mathcal{S}}_p$ constant, (c) find $\mathcal{A}$ as the remainder, (d) update the shares, (e) increase time with a small increment and return to (b). Since this method has not been tried (here), it is unclear what its value is. There is no need to try.

$s = \mathcal{S}' / \mathcal{S}$ gives $s = -\beta \mathcal{I}_p$ which gives $\mathcal{S} \approx \hat{\mathcal{S}}_0$ Exp[$-\beta \hat{\mathcal{I}}_p t$] in local approximation, e.g. when $\mathcal{I}_p \approx$ c

$i = \mathcal{I}' / \mathcal{I}$ gives $i = \beta \mathcal{S}_p - \gamma$ which gives $\mathcal{I} \approx \hat{\mathcal{I}}_0$ Exp[$(\beta \hat{\mathcal{S}}_p - \gamma) t$] in local approximation, e.g. when $\mathcal{S}_p \approx$ c

$a = \mathcal{A}' / \mathcal{A}$ gives $a = \gamma \mathcal{I}_p / \mathcal{A}_p$ while $\mathcal{A} = \mathcal{N}_0 - \mathcal{S} - \mathcal{I}$

5.8. How to use the SIA routine

? SIA

Symbol

The Susceptible–Infectious–Acquitted model. (We avoid the term SIR.)

Options[SIA] has values for Onset time, Nqt0, Seed,

    BetaSIA[], GammaSIA[] and also Aqt0 (e.g. for combinations of models).

SIA[Run, (Pr,) model_Symbol, t_Symbol, tmax, opts] makes a numerical model[t] by

    substituting the parameters, solves this with NDSolve from Onset to tmax, and makes the three

    variables available, also jointly as SIA[t]. If Pr is included, then also the proportions are set.

SIA[t] gives {Sq[t], Iq[t], Aq[t]}.

SIA[Pr, t] gives {Sp[t], Ip[t], Ap[t]} (set by above Run or by SIA[Set, Pr])

SIA[Plot, SIA[t], {t, t0, tmax}, opts] uses Frame.

The usage statement is so large that it has been cut into

    Subjects. Rather than have subroutines with own names, they are collected under:

SIA[subject, ...] for subject in SIA[Explain, List] = {Run, Plot, GrowthRate, Limit, Top, Exp}:

SIA[Explain, subject] explains subroutines of SIA[subject, ...].

SIA[Explain, All] := SIA[Explain, #]& /@ SIA[Explain, List]

SIA["R0[t]"] = (R0[t] == BetaSIA[t] * 1 / GammaSIA[]), may

    help setting parameters, while 1 / GammaSIA[] is the (mean) acquittal period

SIA[Clear] clears variables Sq, Iq, ... Sp, Ip, ...

  SIA[ListOfSymbols] is the latter list, in strings

  SIA["p–and–q"] are equations Vp = Vq / Nq[0]

⌄

Appendix: In steps (for more control of details)

Meaning of this appendix

Above routines took big steps in calculating the variables and plotting them. For applications it tends to be necessary to know how this can be done by steps.

Setting the size of the population and other options

The default options apply to Holland of February 23 2020 onwards. The Dutch population in 2020 is about 17.4 million people. Let us *set* NqSeed = 17.4 * 10^6, and subsequently use the *rule* Nqt0 → NqSeed, for the formal parameter Nqt0 in the equation Nq[t0] = Nqt0.

```
NqSeed = 17.4 × 10^6;
SetOptions[SIA, Nqt0 → NqSeed]
```

$\{$Aqt0 → 0, Cqt0 → 0, DataMold → {S, I, A}, Dqt0 → 0, Iqt0 → Null, Nqt0 → 1.74×10^7, Onset → 0, Seed → 100,

Yqt0 → 0, AlphaSEYCD() → 1., BetaSIA() → 0.4, GammaSIA() → 0.1, PhiSICD() → 0.015, R0() → Null$\}$

Choosing a model with formal t_0 or with the Onset from the options

We set the values of the parameters, so that we get equations that we can submit to NDSolve. The symbols in SI(EY)A(CD) are cleared and can be used as symbols again. We put the model equations into Hold, so that the symbols remain unevaluated when we assign values later on.

We might define the model with an arbitrary t_0.

```
model[t0_, t_] = SIA[N, t0, t, Seed → 100]
```

Hold$\big[\{$Sq′(t) = -2.29885×10^{-8} Iq(t) Sq(t), Aq′(t) = 0.1 Iq(t),

Iq′(t) = −Aq′(t) − Sq′(t), Sq(t0) = 1.73999×10^7, Iq(t0) = 100, Aq(t0) = 0$\}\big]$

```
MatrixForm @@ TheFormer
```

$$\begin{pmatrix} Sq′(t) = -2.29885 \times 10^{-8} \text{ Iq}(t) \text{ Sq}(t) \\ Aq′(t) = 0.1 \text{ Iq}(t) \\ Iq′(t) = -Aq′(t) - Sq′(t) \\ Sq(t0) = 1.73999 \times 10^7 \\ Iq(t0) = 100 \\ Aq(t0) = 0 \end{pmatrix}$$

The choice of t_0 is not arbitrary however. All parameters are determined by the Onset t_0 in the options. At a different moment, the population and the seed may have changed, and the β and γ may have been subject to interventions. The options must be consistent on this. The structure now is that the model has t_0 and that NDSolve does the actual substitution.

We can also drop the arbitrary t_0 in the model, and adopt the Onset from the Options[SIA]

```
model[t_] = SIA[N, t, Seed → 100]
```

Hold$\big[\{$Sq′(t) = -2.29885×10^{-8} Iq(t) Sq(t), Aq′(t) = 0.1 Iq(t),

Iq′(t) = −Aq′(t) − Sq′(t), Sq(0) = 1.73999×10^7, Iq(0) = 100, Aq(0) = 0$\}\big]$

```
MatrixForm @@ TheFormer
```

$$\begin{pmatrix} Sq'(t) = -2.29885 \times 10^{-8} \, Iq(t) \, Sq(t) \\ Aq'(t) = 0.1 \, Iq(t) \\ Iq'(t) = -Aq'(t) - Sq'(t) \\ Sq(0) = 1.73999 \times 10^{7} \\ Iq(0) = 100 \\ Aq(0) = 0 \end{pmatrix}$$

In either format, solving the model requires a time window, for which we need t_0 as well.

```
onset = Onset /. Options[SIA]
```

0

Solving and approximating by NDSolve for a specified time window

We solve and find a numerical approximation by SIA[NDSolve, ...] for a specified time window.

```
sol = SIA[NDSolve, model[onset, t], {t, onset, 150}]
```

$\{$Sq $\rightarrow$ InterpolatingFunction[⊞ ▢ Domain: (0. 150.) Output: scalar],

Iq $\rightarrow$ InterpolatingFunction[⊞ ▢ Domain: (0. 150.) Output: scalar],

Aq $\rightarrow$ InterpolatingFunction[⊞ ▢ Domain: (0. 150.) Output: scalar]$\}$
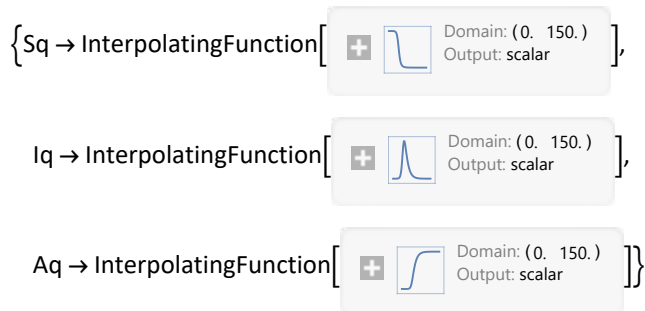
Making the variables directly available

We want to be able to call Sq[t], Iq[t] and Aq[t] separately, and SIA[t] = {Sq[t], Iq[t], Aq[t]} jointly and directly. We use above "sol" to set these functions.

```
SIA[Set, sol];
```

After this setting, we can call the functions directly for different moments in time. We arbitrarily select day 75.

```
SIA[75]
```

$\{415\,556., 738\,924., 1.62455 \times 10^{7}\}$

Plotting in levels and proportions

Plotting of outcomes of SIA[t] is served by colour-coding and legend. The following uses Plot and not SIA[Plot ....].

```
sirplotoptions = SIA[Plot, Options, PlotRange → All];
```

`Plot @@ {SIA[t], {t, 0, 150}, sirplotoptions}`
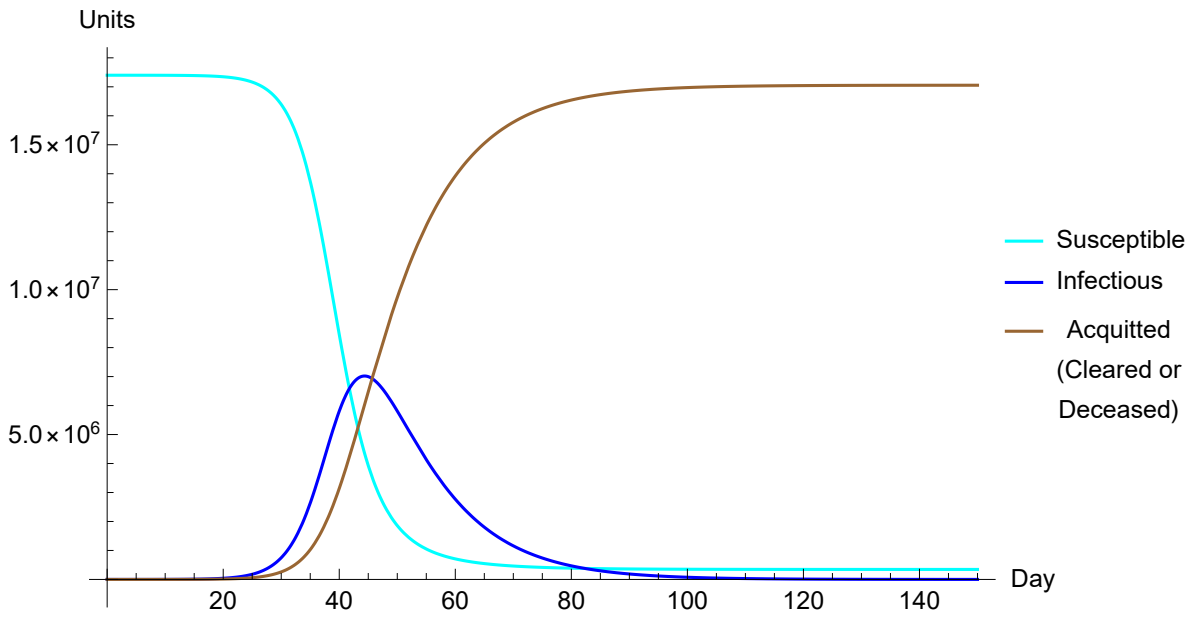
PM. When using SetOptions[Plot] then other analyses have to unset those options again. Alternatively, new options can be put into a variable. The standard Plot routine can be used, but the input must be evaluated before Plot can access it. (Default ImageSize → Full to make plots fit the PDF, even though they might be overly large on screen.)

Given the earlier setting of NqSeed, it is easy to plot the proportions.

`Plot @@ {SIA[t] / NqSeed, {t, 0, 150}, sirplotoptions}`
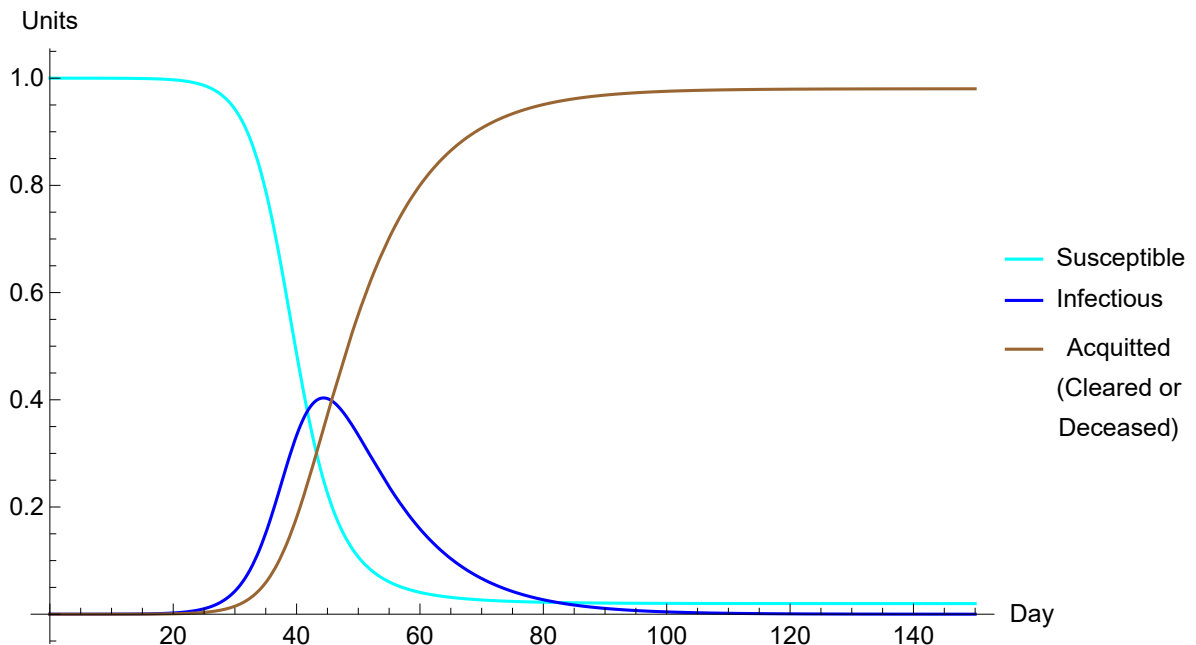
PM. Separate levels and proportions

The following selects the equations, and solves only for the levels.

```
SIA[Run, newModel, t, 150]
```

$\Big\{$InterpolatingFunction$\Big[$ ▦ Domain: (0. 150.) Output: scalar $\Big]$(t\$), InterpolatingFunction$\Big[$ ▦ Domain: (0. 150.) Output: scalar $\Big]$(t\$),

InterpolatingFunction$\Big[$ ▦ Domain: (0. 150.) Output: scalar $\Big]$(t\$)$\Big\}$

Not surprisingly, the model and the outcomes are the same as above.

```
newModel[t]
```

Hold$\Big[\{$Sq$'(t) = -2.29885 \times 10^{-8}$ Iq(t) Sq(t), Aq$'(t) = 0.1$ Iq(t),

Iq$'(t) = -$Aq$'(t) -$ Sq$'(t)$, Sq$(0) = 1.73999 \times 10^7$, Iq$(0) = 100$, Aq$(0) = 0\}\Big]$

```
SIA[75]
```

$\{415\,556., 738\,924., 1.62455 \times 10^7\}$

This sets the proportions.

```
SIA[Set, Pr, NqSeed] (* or leaving NqSeed out, and take the options *)
```

$\Big\{5.74713 \times 10^{-8}$ InterpolatingFunction$\Big[$ ▦ Domain: (0. 150.) Output: scalar $\Big]$(t\$),

5.74713×10^{-8} InterpolatingFunction$\Big[$ ▦ Domain: (0. 150.) Output: scalar $\Big]$(t\$),

5.74713×10^{-8} InterpolatingFunction$\Big[$ ▦ Domain: (0. 150.) Output: scalar $\Big]$(t\$)$\Big\}$

```
SIA[Pr, 50]
```

{0.1068, 0.334003, 0.559197}

Appendix: Comparison with the traditional form for the equations

We might do:

```
tradmodel[t0_, t_] = SIA[N, TraditionalForm, t0, t, Seed → 100]
```

Hold$\Big[\{$Sq$'(t) = -2.29885 \times 10^{-8}$ Iq(t) Sq(t), Iq$'(t) =$ Iq(t) $(2.29885 \times 10^{-8}$ Sq$(t) - 0.1)$,

Aq$'(t) = 0.1$ Iq(t), Sq$(t0) = 1.73999 \times 10^7$, Iq$(t0) = 100$, Aq$(t0) = 0\}\Big]$

```
trad = SIA[NDSolve, tradmodel[0, t], {t, 0, 150}];
```

```
SIA[Set, trad];
```

After this setting, we can call the functions directly for different moments in time.

```
store[SIA, tradmodel] = SIA[75]
```

$\{415\,556., 738\,924., 1.62455 \times 10^7\}$

There is no difference in outcome with the simplified form.

```
store[SIA] (* we had stored this outcome of the standard run *)
```

$\{415\,556., 738\,924., 1.62455 \times 10^7\}$

Appendix: Technical notes

PM 1. We might do a normal call of NDSolve. For this, we must be sure that the variables are cleared.

```
Clear[Sq, Iq, Aq]; NDSolve[ReleaseHold[model[0, t]], {Sq, Iq, Aq}, {t, 0, 150}]
```

$\Big\{\Big\{$ Sq $\rightarrow$ InterpolatingFunction$\Big[$ Domain: (0. 150.) Output: scalar $\Big]$,

Iq $\rightarrow$ InterpolatingFunction$\Big[$ Domain: (0. 150.) Output: scalar $\Big]$,

Aq $\rightarrow$ InterpolatingFunction$\Big[$ Domain: (0. 150.) Output: scalar $\Big]\Big\}\Big\}$

The SIA[NDSolve, ...] format has been made simpler: (i) without the need for Clear, ReleaseHold and the mentioning of the variables, (ii) while the routine also stores the outcome in Results[SIA] so that we can later retrieve earlier runs if needed.

PM 2. We might solve for the proportions, but division by NqSeed can be simpler (as has been done above).

```
SIA[Clear]
```

```
SIA["p-and-q"]
```

$\Big\{$ Np$(t) = \dfrac{\text{Nq}(t)}{\text{Nq}(0)}$, Sp$(t) = \dfrac{\text{Sq}(t)}{\text{Nq}(0)}$, Ep$(t) = \dfrac{\text{Eq}(t)}{\text{Nq}(0)}$, Ip$(t) = \dfrac{\text{Iq}(t)}{\text{Nq}(0)}$, Rp$(t) = \dfrac{\text{Rq}(t)}{\text{Nq}(0)}$,

Ap$(t) = \dfrac{\text{Aq}(t)}{\text{Nq}(0)}$, Cp$(t) = \dfrac{\text{Cq}(t)}{\text{Nq}(0)}$, Dp$(t) = \dfrac{\text{Dq}(t)}{\text{Nq}(0)}$, Yp$(t) = \dfrac{\text{Yq}(t)}{\text{Nq}(0)}$, Op$(t) = \dfrac{\text{Oq}(t)}{\text{Nq}(0)}\Big\}$

6. Analytical properties of SIA

The SIA model has analytical solutions for some key aspects.

PM. $x_p = x_q / N[t_0]$ is the proportion of x_q with respect to the original population.

6.1. Proportionality between numbers and proportions

With $x_p = x_q / N_q[0]$ we have $\frac{\partial}{\partial t} x_p = (\frac{\partial}{\partial t} x_q) / N_q[0]$. The equations are proportional so that the parameters are the same for the model in levels and the model in proportions.

Crucial conditions are derived from the model of proportions, while it still seems more natural to work with the levels. There is no need to choose since we actually do both anyway. It remains important to check translations from one format to the other.

6.2. Exponential phases, doubling period and half-life

When the susceptible would be a constant fraction $\hat{S}_p$ then $\mathcal{I} = \hat{\mathcal{I}}_0 \, \text{Exp}[\,(\,\beta\,\hat{S}_p - \gamma\,)\,t]$. Especially at the beginning of the process $\hat{S}_p \approx 1$, and $\mathcal{I}$ grows at rate $i \approx \beta - \gamma$. For our parameters this is 0.3 (see the plot). We can write $\mathcal{I} \approx$ Seed * $\text{Exp}[(\beta - \gamma)\,t]$ in the exponential first phase. For the proportion:

```
SIA[Ip, Exp, t] (*exponential function*)
```

$5.74713 \times 10^{-6} \, e^{0.3\,t}$

In that phase $\beta > \gamma$ so it has a *doubling period* of Log[2] / (β - γ).

```
Log[2] / (BetaSIA[] - GammaSIA[]) /. Options[SIA]
```

2.31049

At the end, when most infections are over, then *I* has a *half-life* of Log[2] / γ, i.e. almost a week.

```
Log[2] / GammaSIA[] /. Options[SIA]
```

6.93147

When the lockdown in Holland was imposed, it had a relatively fast effect, but there was growing impatience when it took so long to actually wane.

6.3. *Effective reproduction factor $R[t]$, R_0, R_{00}*

6.3.1. Dynamic $\beta[t]$ / $\gamma[t]$ and effective reproduction factor $R[t]$

Section 5.2 already stated these definitions.

$R_0[t] = \beta[t] / \gamma[t]$ is the *dynamic* reproduction factor.

The *effective* reproduction factor also reckons with the reduction of the compartment of the susceptibles.

It is $R = R[t] = R_0[t]\,S[t] / N[t_0]$, which for constant parameters reduces to $(\beta / \gamma)\,S[t] / N[t_0] = (\beta / \gamma)\,S_p$.

6.3.2. Basic reproduction factor $R_0 = R[t_0] = \beta[t_0] / \gamma[t_0] = \beta / \gamma$ (assuming $Sp[t_0] \to 1$)

For $t \to t_0$ (i.e. apart from seeding) $S_p = 1$. Define $R_0 = R[t_0] = \beta / \gamma$ as the *basic reproduction factor.* It is in the literature known as the basic reproduction *number* R_0.

For the SI(EY)A(CD) default of constant parameters, we can also write $R = R_0\,S_p$.

The basic reproduction factor is relevant (i) at $t = 0$ (whether the infection will take off or not) and (ii) possibly when the infections have a top other than at $t = 0$, (iii) for the values in the limit.

6.4. Growth rate and steady state

6.4.1. Dynamics of infection

Reworking the equation for *I'* into an expression with *R,* we find $I' = \gamma\,I\,(R - 1)$.

Observe that $\frac{\partial}{\partial t}\ln[x] = (1/x \; \frac{\partial}{\partial t}x) = x' / x$ is the growth rate of *x*.

We have found the *growth rate of the infections* $(\beta \; S_p - \gamma) = I' / I = \gamma \, (R - 1)$.

(a) If $R > 1$ then the number of infections rises. If $R < 1$ then it wanes. At $R = 1$ there is a top and growth is 0. (Instead of inventing calculus, Newton might also have invented growth accounting.)

(b) *R* eventually gets below 1 by itself.

(c) If *prevention is better than cure*, then interventions choose $\beta[t+1] < \gamma / S_p[t+1]$ as their target.

While we have $I \to 0$ we also have $I' = I \, (\beta \, S_p - \gamma) \to I[\infty] \, (\beta \, S_p[\infty] - \gamma) = I[\infty] = 0$. Thus *I* has the "*asymptotic steady state*" towards 0, which actually means that it is not a steady state in proper definition. The growth rate may still be nonzero, at a limiting value $i[\infty] = \beta \, S_p[\infty] - \gamma \neq 0$. Reasons are: (i) we always divide by a nonzero *I*, (ii) $Sp[\infty] = 1 / R_0$ only holds when $Sp[\infty] = 1$ which cannot be the case (see below).

6.4.2. Comparison to a disease with a steady state of inflow and outflow

There is a *steady state* where *inflow = outflow,* because otherwise there would be no steadiness. The SIA model differs from a model with a steady state. It is useful to shortly review a steady state in another disease than an infection. We might take the incidence quotient as βI and the recovery quotient as γ but let us use different symbols because this is another type of situation.

DiseaseBasics[Set]

$\Big\{$Equations $\to \Big\{$Population = Diseased + Healthy,

    Diseased = Population Prevalence, CumulatedIncidence = Healthy IncidenceQuotient,

    CumulatedRecovery = Diseased RecoveryQuotient, CumulatedIncidence – CumulatedRecovery = 0,

    DiseaseDuration = $\dfrac{1}{\text{RecoveryQuotient}}$, HealthDuration = $\dfrac{1}{\text{IncidenceQuotient}}\Big\}\Big\}$

DiseaseBasics[{DiseaseDuration → "Pinf", IncidenceQuotient → q, Population → N}] //
   Last // Last // MatrixForm

$\begin{pmatrix}
\text{CumulatedIncidence} \to \frac{N\,q}{\text{Pinf}\,q+1} \\
\text{CumulatedRecovery} \to \frac{N\,q}{\text{Pinf}\,q+1} \\
\text{Diseased} \to \frac{N\,\text{Pinf}\,q}{\text{Pinf}\,q+1} \\
\text{HealthDuration} \to \frac{1}{q} \\
\text{Healthy} \to \frac{N}{\text{Pinf}\,q+1} \\
\text{Prevalence} \to \frac{\text{Pinf}\,q}{\text{Pinf}\,q+1} \\
\text{RecoveryQuotient} \to \frac{1}{\text{Pinf}}
\end{pmatrix}$

? DiseaseBasics

> Symbol
>
> DiseaseBasics[{x___Rule}, elims___List, opts] is a SolveFrom application of the
>
> > basic Disease equations, for the steady state when there is no disease–specific death.
>
> By definition Prevalence == Diseased / (Healthy + Diseased).
>
> In the steady state, the outflow from the healthy
>
> > must also be equal to the recovery outflow from the diseased. Then also:
>
> > Prevalence == IncidenceQuotient / (IncidenceQuotient + RecoveryQuotient);
>
> where IncidenceQuotient and RecoveryQuotient
>
> > are measured as (perunage) fractions of the relevant subpopulations.
>
> Implied are also durations. With d = DiseaseDuration
>
> > = 1 / RecoveryQuotient, and λ = IncidenceQuotient, we find:
>
> > Prevalence = λ d / (λ d + 1)
>
> The latter notation may be more acceptable when there are
>
> > deaths involved and the 'cure' is rather the replacement with newborns. The
> >
> > screen detected duration concerns a prevalence screen with 100% sensitivity.
>
> Options[DiseaseBasics] contains the equations.
>
> DiseaseBasics[Set] sets Equations –> DiseaseBasics[Equations] (default).
>
> DiseaseBasics[Set, n] sets Equations –> DiseaseBasics[Equations, n] such that aggregate Healthy and
>
> > Diseased are decomposed in n subgroups that have the same (mutual) steady state properties.
>
> See also ToDiseaseSymbols and SteadyStateFlows
>
> ⌄

> Symbol

6.5. Maximal infections at the origin or later in the process

6.5.1. A maximum at the seed when $S_p[0] \leq 1 / R_0$ and thereafter decay

At the first seed, $S_p[0] \approx 1$. If $R < 1$ or $R_0 < 1 / S_p[0]$ (conventionally $R_0 < 1 < 1 / S_p[0]$) then the infection does not develop into an epidemic.

PM. When $S_p[0] = 1 / R_0$: when $R_0 < 1$ then $S_p > 1$, which is impossible for a share. For $R_0 < 1$ we already have at the start $S_p[0] < 1 / R_0$ so that the infection dwindles down.

6.5.2. A maximum after the seed when $S_p[0] > 1 / R_0$

The traditional model format directly shows that the infections find their top (other than at the origin) when:

$(I' = 0 \ \& \ I \neq 0) \Longleftrightarrow (\beta \, S / N[t_0] - \gamma = 0) \Longleftrightarrow (R_0 \, S_p - 1 = 0) \Longleftrightarrow (R = 1) \Longleftrightarrow (S_p = 1 / R_0)$

6.6. Later in the process, consequences of $R_0 > 1 / S_p[0]$, "the" top of infections

The following assumes that at the origin $R_0 > 1 / S_p[0] \approx 1$.

6.6.1. The values of the shares at the top

For SARS-CoV-2, the top of I_p for $R_0 = 4$ lies at $S_p = 1/4$.

```
SpTop == GammaSIA[] / BetaSIA[] /. Options[SIA]
```

SpTop = 0.25

What about the other two variables ?

Above we mentioned that $S = S_0 \, \text{Exp}[- R_0 \, (\mathcal{A}_p - \mathcal{A}_p[0])]$, e.g. useful for the limit values. At the top of infections:

$$S_p = \gamma / \beta = \text{Exp}[- \beta/\gamma \, (\mathcal{A}_{p,\,\text{Top}} - \mathcal{A}_p[0])]$$

$$S_p = 1/R_0 = \text{Exp}[- R_0 \, (\mathcal{A}_{p,\,\text{Top}} - \mathcal{A}_p[0])]$$

$$-\text{Log}[R_0] = - R_0 \, (\mathcal{A}_{p,\,\text{Top}} - \mathcal{A}_p[0])$$

$$\mathcal{A}_{p,\,\text{Top}} = \text{Log}[R_0] / R_0 + \mathcal{A}_p[0].$$

Values of the proportions thus are, and we cannot neglect Aqt0:

$$\{S_p, \mathcal{I}_p, \ \mathcal{A}_p\}_{\text{Top}} = \tfrac{1}{R_0} \, \{1, \ R_0 - 1 - \text{Log}[R_0], \ \text{Log}[R_0]\} \ + \ \{0, \ -\mathcal{A}_p[0], \ \mathcal{A}_p[0]\}$$

The normal case has $\mathcal{A}_p[0] = 0$, and check that $R_0 = 1$ causes the point $\{1, 0, 0\}$:

$$\{S_p, \mathcal{I}_p, \ \mathcal{A}_p\}_{\text{Top}} = \tfrac{1}{R_0} \, \{1, \ R_0 - 1 - \text{Log}[R_0], \ \text{Log}[R_0]\} \ + \ \{0, \ 0, \ 0\}$$

$$\{S_p, \mathcal{I}_p, \ \mathcal{A}_p\}_{\text{Top}} = \tfrac{1}{R_0} \, \{1, \ R_0 - 1 - \text{Log}[R_0], \ \text{Log}[R_0]\}$$

```
SIA[Ip, Top, R₀] == SIA[Ip, Top, 4.]
```

$$\left\{\frac{1}{R_0}, \frac{R_0 - \log(R_0) - 1}{R_0}, \frac{\log(R_0)}{R_0}\right\} = \{0.25, 0.403426, 0.346574\}$$

```
SIA[Explain, Top]
```

SIA[Ip, Top, f, atq0:0] gives the top of Ip for f = R0 formally: {1, f − 1 − Log[f], Log[f]} / f + {0, −atq0, atq0}

A related routine is:

TopIq[SIA, ...] for the t when Iq has its top when R0 > 1 (inverse problem)

    with comparison of theoretical values and those generated by the interpolation run

TforTopIq[...] applies FindRoot to the equation Sp[t] == 1/R0

For SARS-CoV-2 with $R_0 = 4$ and $\gamma = 1/10$, remarkably $\mathcal{I}_{p,\text{Top}} \approx \beta$ but the above shows that this is not exactly so. We can solve $\gamma = \beta / R_0 = \mathcal{I}_{p,\text{Top}} / R_0$ for those values of γ that cause such outcome, and it just so happens that we selected values close to that contour.

```
Plot[(f - Log[f] - 1) / f^2, {f, 1, 8}, AxesOrigin → {0, 0},
 AxesLabel → {R0, "γ that causes Ip,Top ≈ β"}, BaseStyle → {FontSize → 13}]
```

γ that causes $\mathcal{I}_{p,\text{Top}} \approx \beta$
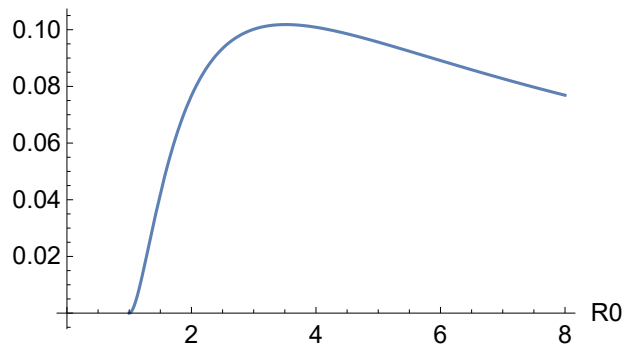
6.6.2. The value of t_{Top}

(1) Let us calculate t_{Top} using above solution of SpTop for the equations SIA[t]. Given that the value of Sp at the top of Iq follows from R_0 and is not affected by Aq[0], the fast calculation is - and this only works when the model has been run and Sq[t] has values:

```
SIA[30] / NqSeed (* arbitrary point to check that this gives numbers *)
```

{0.942096, 0.0429933, 0.0149105}

```
FindRoot[1 / 4 == Sq[t] / NqSeed, {t, 50, 100}] // Quiet
```

$\{t \rightarrow 44.3682\}$

It will be useful to have this available under its own function call.

```
TforTopIq[]   (* not Quiet *)
```

⸱⸱⸱   InterpolatingFunction: Input value {−114.655} lies outside the range of data in the interpolating function. Extrapolation will be used.

44.3682

(2) When we have found t_{Top} then there are two estimates for the proportions of the variables at the

top.

(2a) The theoretical values directly from β and γ above.

(2b) Those calculated from the numerically approximated equations, that allowed us to find t_{Top}.

The numerical approximation might cause a difference. The following routine prints the difference and checks whether the sum of the absolute differences is larger than (default) 0.0001.

Check that the top for $R_0 = 4$ lies at $S_p = 1/4$ indeed.

```
res = TopIq[SIA, t]
```

⋯ InterpolatingFunction: Input value {−84.1178} lies outside the range of data in the interpolating function. Extrapolation will be used.

$$\left\{t \rightarrow 44.3682, \text{Proportion} \rightarrow \begin{array}{c|ccc} & \text{Sp} & \text{Ip} & \text{Ap} \\ \hline \text{From } \beta \text{ and } \gamma & 0.25 & 0.403426 & 0.346574 \\ \text{Equations} & 0.25 & 0.403428 & 0.346572 \\ (1) - (2) & 0 & -1.39436 \times 10^{-6} & 1.39436 \times 10^{-6} \end{array} \right.,$$

$$\text{Nqt0} \rightarrow 1.74 \times 10^7, \text{Seed} \rightarrow 100, \text{Aqt0} \rightarrow 0,$$

$$\text{List} \rightarrow \begin{pmatrix} 0.25 & 0.403426 & 0.346574 \\ 0.25 & 0.403428 & 0.346572 \\ 0 & -1.39436 \times 10^{-6} & 1.39436 \times 10^{-6} \end{pmatrix}, \text{AcceptableErrorQ} \rightarrow \text{True}\Bigg\}$$

(3) Relating the top to real world events.

```
tTop = t /. res ;
```

Above, we mentioned the carnival festivities in the South of Holland of February 23 2020. It is not impossible that some 100 infections were seeded. Taking the daycount, we can determine the expected day of the calendar when the top of infections would have been, had there been no intervention in the mean time.

```
DayPlus[{2020, 02, 23}, tTop // Round]
```

Day: Tue 7 Apr 2020

(4) Use of this function.

? TopIq

> Symbol

TopIq[SIA || SICD, t] is a more involved routine than TforTopIq, and not only finds the

time but also compares numerical approximations with the theoretical outcomes at the

top of Iq. If you have SICD[t] or SEYCD[t] available, then you might choose SIA or SICD.

There are no parametric outcomes for E and Y in SEYCD, and thus there is no comparison

TopIq[SIA, t] assumes that SIA[t] gives {Sq, Iq, Aq} as functions of t (which will also happen when SICD[t]

and SEYCD[t] have been set), and uses FindRoot to find t from Sq[t] == Nqt0 / R0, i.e. when the

infections reach their top (assuming that R0 > 1). Options for Nqt0, BetaSIA[], GammaSIA[] and Aqt0

can be supplied or are taken from Options[SIA]. Default search startvalues for t are SIA[Ip, Exp, Top]

and 3 Log[beta/gamma] / beta for Ip = 1/3. Output gives both t and the {Sp, Ip, Ap} at the top, both

theory = SIA[Ip, Top, beta/gamma (, aqt0)] and calculated (from the Equations / interpolations).

Options[TopIq] have Max –> 10^–4 to give a message when Total[Abs[theory – calculated]] > max

TopIq[SIA, {t, tmin, tmax}] uses FindRoot for that window.

TopIq[SICD, t] similar, allocates {c, d} = {labda, mu} / (labda + mu) Ap.

TopIq[SICD, {t, tmin, tmax}] similar

Legacy:

TopIq[SIA, NSolve, t] uses NSolve

TopIq[SIA, NSolve, Sq[t], t] only solves for t

TopIq[SIA, NSolve, Sq[t], Nqt0, R0, t] solves directly without options

⌄

(5) PM. TopIq uses FindRoot. NSolve does not always generate outcomes.

```
nsol = NSolve[Sq[t] == NqSeed / 4 , t] // Quiet
```

$$\left\{\left\{t \to \text{InverseFunction}\left[\text{InterpolatingFunction}\left[\boxed{\begin{array}{c} \text{Domain: (0. 150.)} \\ \text{Output: scalar} \end{array}}\right], 1, 1\right]\left[4.35 \times 10^6\right]\right\}\right\}$$

```
SIA[t] / NqSeed /. nsol[[1]]; (* show if there is a numerical outcome *)
```

⋯ InverseFunction: Inverse functions are being used. Values may be lost for multivalued inverses.

6.7. The exponential growth phase revisited

Above we discussed exponential growth. Now that we have discussed the top, we can include the following comment.

The infections have a first phase of exponential growth. The growth slows down because the lowering number of susceptibles reduces the impact of β. The phase is interesting for us for two reasons. A comparison may give us information when the true SIA path starts deviating, perhaps because of intervention. Secondly, rather technically, the exponential outcome provides a starting point for the FindRoot routine above.

Since we know the proportion Ip at the top, we can calculate how long it would take to get there if the exponential phase had continued. This routine does this calculation for us.

tExp = SIA[Ip, Exp, Top]

37.1968

Thus, the SIA-properties cause a delay of the top of infections with some 7 days. With above doubling time of about 2.3 days, the difference is close to 8 times the size.

tTop − tExp

7.17132

SIA[Explain, Exp] (* also relevant for finding starting values for FindRoot *)

SIA[Ip, Exp, t] gives Ip for the first exponential phase
    seed / Nqt0 Exp[(beta − gamma) t], with the values taken from the options

SIA[Ip, Exp, t, beta, gamma, ntq0, seed] takes values

SIA[Ip, Exp, Top] calculates the time for Ip to reach the top if it were using
    exponential growth only (but falling Sp causes delay), with values taken from options

SIA[Ip, Exp, Top, ntq0, seed, beta, gamma, a0:0] takes values

6.8. Exponential growth and exponentially distributed acquittal periods

6.8.1. One function for two purposes: probability and dynamics

Appendix A contains the fundamental "success and failure" model of *exponential decay of Success*, with *Failure collecting the decay*. While this is a dynamic *growth process*, the shapes of the functions are such that we can interpret the proportion of failure as the *cumulative probability* (CDF) of the random variable "whether the event of failing will take a particular duration". The population is homogeneous and we do not know which event of decay will take what duration. The model assumes a constant rate of decay λ which happens to conform with the exponential distribution. The PDF of Failure then is the exponential density, with $\text{PDF}_{\text{Failure}} = \lambda$ Success. See *Appendix* A for a short review of the Success and Failure model, and the distinction between *calendar time* (with

prevalence) and *duration* (age). (Events at different calendar times might be presented in an age distribution.)

6.8.2. Plotting with mean acquittal period 1 / γ

For SARS-CoV-2, the Success / Failure plot with $\gamma = 0.1$ is as follows. The plot gives the proportions of Success and Failure, with Success having exponential decay and Failure = 1 - Success. The compartment of the *infected* is taken as the *Success* category, and acquittal is counted as "failure": acquittal means a decay of Success. Thus if $S_p = 0$ (e.g. they are quarantined or flee the country) and there is no new inflow of infectiousness, then the remaining pool shows exponential decay as $\mathcal{I}$ = $\mathcal{I}_0$ Exp[- γ t]. The mean acquittal period is $\mu = 1 / \gamma$ and it is no surprise that 10% of the patients may take more than 25 days to exit from the scheme.

The half-life can be found at the intersection of Success = Failure = 1/2. Clearing has a half-life of h = Log[2] / γ = μ Log[2] $\approx$ 0.7 μ. Thus $\mu \approx 1.4$ h. In this case $\mu = 10$ is easy to spot. PM. The plot uses S for Success and not for Susceptible, and uses $\lambda = \gamma$.

SuccessFailurePlot[0.1, E → False]
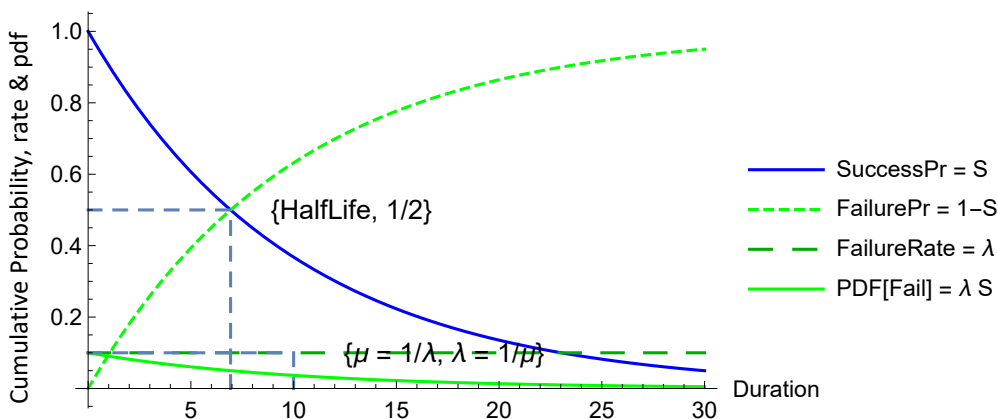
6.8.3. Interpretation of the mean acquittal period 1 / γ in SIA

We can interprete γ as a *rate* (always a mean) and 1 / γ as a *mean duration*.

For SIA, 1 / γ is best interpreted as the "mean acquittal *and* infectious period", to express its dual effect.

(1) SIA makes the additional assumption that the acquittal period is also the "mean infectious period". This does not hold by identity. (The birth certificate of Paul does not state that he and Mary later married, though he still is Paul who did so.)

(2) Apart from this assumption on the acquittal period, there are processes of *clearing* and *returning to health*, or the *path towards death.* Those concern the medical circumstances of the persons involved, and those are not directly relevant for the process of infecting other people. A part will be relevant for hospital costs. Such elements thus can be used for additional modeling on those aspects.

The model formulation that $\mathcal{A}$' = γ $\mathcal{I}$ still is correct for the *rate of acquittal* (removal) from its

source, i.e. $\gamma = \mathcal{A}' / \mathcal{I}$.

The SIA assumptions obviously remain relevant for the first steps in modeling. It is remains important to be aware of more complex models and such discussions in the literature. The SEYCD model already modifies the infectious period by including the Exposed.

There is no assumption of a *constant* acquittal period, e.g. two weeks for everyone.

For SARS-CoV-2, it now has been reported in Holland that some 200 patients have been in the ICU for longer than 30 days. Note that a patient who is released from the ICU may take a year to recover from ICU (with the deterioration of muscles etc.).

6.9. The limit with the *Lambert W* or *ProductLog* function

6.9.1. Finding the limit values

SIA has the property that $I \to 0$ but never becomes zero. We write the limit of I as Iq[∞] = 0, a value never reached. Thus we have Sp[∞] + Ap[∞] = 1. For SIA this can be solved by the Lambert W or the *ProductLog* function.

As said, the relation $\mathcal{S} = \mathcal{S}_0$ Exp[- R_0 ($\mathcal{A}_p$ - $\mathcal{A}_p$[0])] is useful for the limit values.

The outcome depends upon R_0 and the starting values Sp[t0] and Ap[t0]. The latter have defaults 1 and 0, but might be different after an intervention.

The following generates the three proportions {Sp[∞], Iq[∞], Ap[∞]}, using the default option settings.

```
SIA[Limit]
```

{0.0198273, 0, 0.980173}

We may also insert unevaluated parameters, and the discussion below provides an explanation.

```
SIA[Limit, BetaSIA[] → R0 GammaSIA[]]
```

$$\left\{-\frac{1.\ W\left(-0.999994\ e^{-1.\ R0}\ R0\right)}{R0},\ 0,\ \frac{1.\ W\left(-0.999994\ e^{-1.\ R0}\ R0\right)}{R0}+1\right\}$$

6.9.2. Difference between top and limit outcomes

The following gives the difference in shares between the limiting values and those at the top of infections. Notice the value for Ap, meaning that many units still have to end their infectious period, including those who are infected after the top.

```
TopSIA = SIA[Ip, Top, (BetaSIA[] / GammaSIA[] /. Options[SIA])];
```

```
Thread[ {Sp, Ip, Ap} → (SIA[Limit] - TopSIA)]
```

{Sp → −0.230173, Ip → −0.403426, Ap → 0.633599}

6.9.3. Deduction

As said, the relation $\mathcal{S} = \mathcal{S}_0$ Exp[- R_0 ($\mathcal{A}_p$ - $\mathcal{A}_p$[0])] is useful for the limit values. At the limit we have {Sp[∞], 0, Ap[∞] = 1 - Sp[∞]}.

Solving vor Sp[∞], and substituting the symbols for the limiting values, and taking Aqt0 zero, we get:

SpLimit == SIA[ProductLog, R0]

$$\text{SpLimit} = -\frac{W(-e^{-R0}\,R0)}{R0}$$

SpLimit == SIA$\big[$ProductLog, $\big($BetaSIA[] / GammaSIA[]$\big)\big]$ /. Options[SIA] // N

SpLimit = 0.0198274

The number of units (supposedly living till eternity, in an eternal population of 17.4 million):

Last[TheFormer] * NqSeed

344 997.

In practice, we cannot simply assume that Aqt0 is zero. Hence, the routine allows for the appropriately full input.

SIA[Explain, Limit]

SIA[Limit (, opts)] uses opts or default Options[SIA] to determine the infinite
    values. The default options have the Iq seed and n = Nqt0 and Aqt0, so that it follows
    that Sqt0 = n − seed − Aqt0; then it calls SIA[ProductLog, b/g, Sqt0 / n, Aqt0 /n];

SIA[ProductLog, R0, Spt0:1, Apt0:0] = Sp[∞] is the ProductLog finding for Sp[inf] on R0 and those
    proportions. With the limit of Iq zero, the counter part of the limit of Sp is the limit of Ap
SIA[Sp, Limit, x__] uses the productlog with such values of x, finds s = Sp[inf], and returns {s, 0, 1−s}
See TforAp95

6.10. The importance of t50 and t95

The following provides additional explanation for the figure already shown above.

For SIA, we have two phenomena that directly depend upon R_0, namely the top and the limit. Let us look now at Ap, since this determines the eventual burden of disease. The two equations are (with some blanks between them):

$$\text{PointAtTop} = \left\{\frac{1}{f},\ \frac{f - \log(f) - 1}{f},\ \frac{\log(f)}{f}\right\} \qquad \text{PointAtLimit} = \left\{-\frac{W(-e^{-f}f)}{f},\ 0,\ \frac{W(-e^{-f}f)}{f} + 1\right\}$$

Both the *top of infections* and the *limit* values are determined by R_0. The overall burden of disease is determined by the limit value. The top of infections says rather little about the overall burden of disease, which only appears in the limit. The limit is a long way off. Thus it seems useful to compare also some other values than the top only:

(1) t_{50} when A_p reaches 50% of the population

(2) t_{95} when A_p reaches 95% of A_∞.

These durations give us a sense how the burden of disease is spread over time. Observe that the indicators relate to different bases: (i) for t_{50} we do a numerical approximation based upon the parameters, (ii) for t_{95} we directly use the parameters to find the R_0 value. Obviously, $t \to \infty$ cannot be calculated anyway.

Thus we find these special moments in time. They all require R_0 but the first ones include it in the numerical approximations and the latter ones allow a parametric solution, whence we insert the numerical value of our choice for R_0.

```
pnts = {TforTopIq[],
   TforAp50[],
   TforAp95[4],
   1 – SIA[ProductLog, 4.]} // Quiet
```

{44.3682, 48.3071, 74.4295, 0.980173}

We already showed the following plot twice, and there is adequate reason to plot it again, here. The horizontal axis shows the different values of R_0 with a vertical dashed line at the value 4. For t_{50} there is no contour in the legend but we still get a dot.

```
ApByR0Plot[]
```
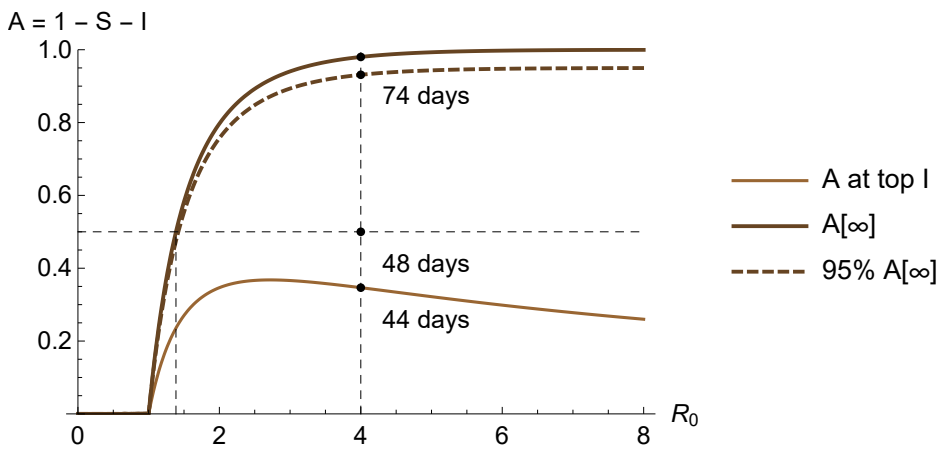

For the contour of Ap at the top of Ip, a higher value of R_0 will cause a higher infection level for Ip at the top, and then there will be a lower value for Ap.

It is surprising that the dynamics of the process can be caught in one parameter. The word of caution of course is that we may be assuming a constant value of this parameter, while it actually changes.

The explanation of the plot is straightforward:

? ApByR0Plot

> Symbol
>
> ApByR0Plot[R0chosen:4, R0max:8] plots Ap[R0]
>
> > for {R0, 0, R0max}, with R0 the basic reproduction factor. Plotted are:
>
> (1) The limit value of Ap[Infinity]
>
> (2) Ap95 = 95% of Ap[Infinity], and the days t95 to get there
>
> (3) Ap50 = 50% of Population, and the days t50 to get there
>
> (4) ApTop, i.e. at the top of Iq, and the days tTop to get there
>
> PM. Options Text and Line (defaults True) control plotting of text and lines
>
> PM. The plot requires numerical proportions, e.g. SIA[Set, Pr, (, Nq[0])]
>
> NB. The graphs are fully determined by R0
>
> > on the horizontal axis. See TimeByR0 for time as a function of R0
>
> ApByR0Plot["HerdImmunity", fmax:8] gives the plot with the Ap[infinity] and
>
> > the claimed ApHI = 1 − 1/R0: the difference between the lines times population times
> >
> > the Infection Fatality Factor is the overshoot in deaths not accounted for by ApHI

6.11. Attack rate vs force of infection

Breda et al. (2012:105): "the force of infection is, by definition, the probability per unit of time that a susceptible becomes infected." For the SIA model, this appears to be the (absolute value of the) growth rate $-s = \beta\, \mathcal{I}_p$. Authors are free to define terms of course, but the association with the term "force" does not seem to enhance clarity. It suffices to identify this expression as the growth rate or the *rate of decline of the remaining susceptible compartment*.

The growth rate $s = -\beta\, \mathcal{I}_p$ can also be related to the value of $\mathcal{S}_p$ itself via $s = -\beta\, \mathcal{S}_p\, \psi$ with now the (cumulated) "attack rate" $\psi = \mathcal{I}_p / \mathcal{S}_p$, i.e. the number of infectious units encircling the susceptible. The term "attack rate" seems also to be used for Ip[t] itself, but we already have a name for Ip.

```
AttackRate[Plot, t, 0, 150, BaseStyle → {FontSize → 14}]
```

Iq / Sq
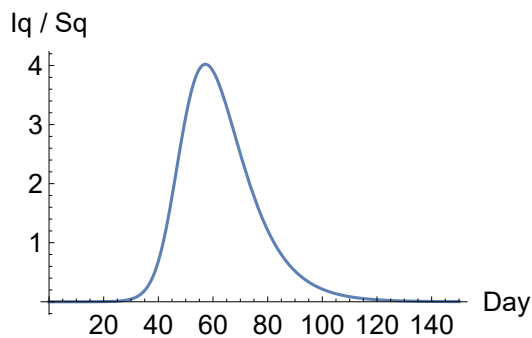
6.12. Herd protective effect, resistance and intervention

A core point of this section is that the term "herd immunity" is better replaced by the term "herd resistance". The term "herd immunity" has technical meanings that are clear to specialists in epidemiology but the term "herd immunity" itself causes confusion for policy makers and the general public. The reference to immunity suggests protection but "herd immunity" does not mean protection for all members of the herd. It only indicates a protective effect by a decline in the epidemic so that the herd itself survives, but potentially in a (much) smaller size and at the cost of still many more members.

6.12.1. Individual and population, immunity and protection

Some elementary definitions are in order. *Protection* is more general than *immunity*.

- Immunity for an unit protects the unit against an infection (at a normal dose): it means that an infection does not apply.

- When an infection might apply then protection may be provided by an endogenous response or an intervention like quarantine that causes that the effective reproduction factor $R = 0$ or when the virus has been eradicated (e.g. by draining swamps and burning clothes).

Let us also distinguish *protection* (dichotomy: yes or no) and a *protective effect* (continuous probability):

- Protection is a protective effect of 100%. When protection fails then it actually wasn't given, and it was wrong to call it "protection" in the first place.

- When a protective effect was provided and fails then it still was given though it did not result into protection.

A protective effect for a person differs from a protective effect at the population level, since a population can survive even when some units perish, and since what some units do has effects on other units. A unit might seek protection from infection by quarantine. A surviving unit that recovers from an infection has gained protection by immunity. At issue in this section is what a population can do to get the best protective effect.

The definition for "*population immunity*" is that $I = 0$, i.e. that everyone is protected because there are no infections. This end state is not reached in the SIA model, because it has only the limiting process $I \to 0$, which we may also denote as Iq[∞] = 0. SIA has no steady state (in its strict defini-

tion); there is only an *asymptotic* state. However, SIA uses a continuous approximation to a discrete problem, and perhaps it is better to say that the infection becomes extinct when *Iq* drops below 0.5 (rounding down).

6.12.2. Avenues for a protective effect

The SIA limiting values give $1 = Sp[\infty] + Iq[\infty] + Ap[\infty]$ and $1 = Sp[\infty] + Ap[\infty]$ because $Iq[\infty] = 0$.

Of interest are the following avenues for a protective effect:

(1) When the effective reproduction factor $R = Sp\, R_0 < 1$, then the epidemic wanes, and eventually a susceptible fraction $Sp[\infty] = 1 - Ap[\infty]$ would no longer be at risk of that particular epidemic. This can be called a "***herd protective effect***" or "***herd resistance***". We better use the term "protective effect" rather than "protection", since the latter might cause a false sense of security. For example, when an infection has run its course till 99% $Ap[\infty]$ then we might assume a high protective effect, but when new infectious seeds are introduced (e.g. by foreign travellers) then those may still cause a new (small) epidemic, especially when $Ap[\infty] << 1$, so that the protective effect is not 100% (which is required for the definition of "protection").

(2) The latter herd protective effect can be enhanced by deliberate intervention at the personal level, in particular by ***vaccination*** or ***social distancing*** (reduction of contacts and potentially also the infectiousness of contacts).

(3) There are other measures like eradication of the virus (e.g. by draining swamps or burning clothes and the like). In practice it might be difficult to prove that a virus has been eradicated though. (PM. A virus needs a host for reproduction. When hosts clear and the virus cannot find new units to infect, then basically the virus becomes extinct. Potentially, a virus might persist in a "repository" (a location or carrier animal) but this would not pose a threat if it is at adequate distance of the susceptible units.)

Interventions under (2) apply to the same mechanisms as are involved in the natural process of an infection under (1), they thus involve the same mathematical structure and formulas, and thus it becomes near to impossible to discuss (1) and (2) separately. We will not say much on (3).

6.12.3. Better use the term "herd resistance" than "herd immunity"

The above three avenues can be compared to the notion of "***herd immunity***" as this is used in the literature in epidemiology.

This notion of "*herd immunity*" seems to allow for a deviation of *population immunity,* presumably to make up for the notion that population immunity is unattainable in reality. However, some key papers and a deconstruction of their definitions show that "herd immunity" appears to be a creative but confusing linguistic expression for a situation (i) when there is some *herd protective effect*, see above under (1), but (ii) without the *protection* as suggested by the word "immunity". A herd can survive as a herd by sacrificing many members, and humanity with 7 billion people may survive as humanity when 6 billion perish, and perhaps history shows that humanity survived in the past in similar ways, but there is no need to create confusion about this by using a confusing term. We can assume that experts in epidemiology know that "herd immunity" does not imply such protection, but the term will create confusion when it is used in communication with the general public, like in

the speech by the Dutch Prime Minister Rutte (2020), see section 1.5.6.

It is inadvisable that the literature in epidemiology uses a creative linguistic expression that creates a false sense of security, which term is also used in communication with the general public, and which term does not convey a proper message to an audience that is not familiar with the true meaning of the term. The term "herd (infection) resistance" is better, also because the general public might also mistake the "herd protective effect" for "herd protection".

Define $\mathbf{Ap_{Herd,\infty}} \overset{\text{def}}{=} Ap[\infty]$. We might consider the idea that this herd "protects" the still susceptible $\mathbf{Sp_{Herd,\infty}} \overset{\text{def}}{=} Sp[\infty]$. However, this is a wrong use of the word "protects". $Ap[\infty]$ is only a limiting value. In the process $Sp[t] \rightarrow Sp[\infty]$ there are still units infected. For some units there is not the promised protection. It might be someone dear to you. The problem with the use of the limiting values is that one must specify *when* the "herd immunity" is reached and for limits to infinity this is impossible.

```
ApLimit == 1 - SIA[ProductLog, BetaSIA[] / GammaSIA[]] /. Options[SIA] // N
```

ApLimit = 0.980173

Alternatively, we might use a criterion like 95% of the limit value, and take $\mathbf{Ap_{Herd,95}} \overset{\text{def}}{=} 95\% \, Ap[\infty]$. We discussed in section 5.6 at what moment in time this level is reached for our combination of parameter values.

```
t95 = TforAp95[4]
```

74.4295

Thus we can describe this situation (of "*an increased protective effect to 95% of the limit value*") without using, or trying to save, the term "herd immunity" with its implied false sense of security. Below in subsections 6.12.7-9 we will consider some alternative efforts at definition (notably based upon a notion of a steady state) that turn out to have drawbacks too. Before discussing this, it is better to first look into interventions like vaccination and social distancing, since the mathematics is so close to this notion of herd resistance.

Different authors may introduce different particulars when they use the term "herd immunity". Some definitions are:

- "the resistance of a group to attack by a disease to which a large proportion of the members are immune, thus lessening the likelihood of a patient with a disease coming into contact with a susceptible individual" (Dorland's Illustrated Medical Dictionary 1965 quoted by Fox et al. (1971)). (PM. The "resistance" apparently is a degree. Note that the term "herd infection resistance" could also have been adopted in 1971.)

- Fox et al. (1971) criticise the assumption of homogeneity and argue for heterogeneity and clusters: that "open populations are made up of innumerable definable but often interlocking subgroups which differ in respect to proportions of immunes and intimacy of contacts". They do not question that there can be a protective effect. The authors develop a view on the level of vaccination - but this is something else than providing a sound definition for "herd immunity" (which is our focus of interest in this subsection).

- Fine, Eames & Heymann (2011) refer to a "variety of meaning" (with 7 references). Their abstract: "Some authors use it to describe the proportion immune among individuals in a population. Others

use it with reference to a particular threshold proportion of immune individuals that should lead to a decline in incidence of infection. Still others use it to refer to a pattern of immunity that should protect a population from invasion of a new infection. A common implication of the term is that the risk of infection among susceptible individuals in a population is reduced by the presence and proximity of immune individuals (this is sometimes referred to as ''indirect protection'' or a ''herd effect'')."

Heesterbeek et al. (2015) define "herd immunity" as "*state of the population where the fraction protected is just sufficient to prevent outbreaks* (R_e < 1)". (We write R instead of R_e.) This takes an "outbreak" as $R \geq 1$. However, *protection* (in general) is more than "*preventing an outbreak*", while we also would want to see reference to the particular effect of immunity of units within the herd. It is also possible that this formulation may confuse "protection" with "protective effect" (i.e. as in the definitions above). The core notion rather is not that a fraction is protected (has become immune) but that their immunity has a protective effect, like preventing an outbreak, which however should not be confused with "immunity of the herd" so that there would be no more infections (because of an overshoot).

In section 1.5.6 we already stated that the effective reproduction factor $R = Sp\, R_0 = (1 - Ip - Ap)\, R_0 < 1$, iff $Ip + Ap > 1 - 1 / R_0$. When $Ip > 0$, this still means that *infections continue*. The literature calls this "overshoot". When an epidemic is raging and $Ip \gg 0$ then the epidemic still has quite an impact. The definition by Heesterbeek et al. (2015) thus cannot convince.

This deconstruction of key papers establishes (i) that there is no useful definition of "herd immunity" and (ii) that the use of this term only contributes to confusion. Let us define "***pseudo herd immunity***" when $R > 1$ and $Ap > 1 - 1 / R$. Since $R = Sp\, R_0$ and $Sp < 1$ (with nonzero seed), it may suffice that $Ap > 1 - 1 / R_0$. This particular formula, that neglects the role of *Ip,* will be discussed below but can best be mentioned here at the outset. Authors writing about "herd immunity" might intend the herd protective effect, and when they refer to this formula then they would be advised to speak about "pseudo herd immunity" since there is no immunity of the herd and there is no reason to suggest that there is such (other than that the herd may survive as a herd and avoid extinction perhaps even when 99% succombs).

In the following, we will first discuss *vaccination* and *social distancing* before we return to the issue concerning the confusion as has been generated by the term "herd immunity". When there is confusion, it may remain vague what the confusion actually is or what causes it. Speculation about what a confusion is might be even more confusing. Thus it is better to first present a proper analysis and then secondly perhaps indicate what possible confusions might be.

6.12.4. Burden of disease, sacrifice ratio and intervention

Let the disease have a (monetised) burden b per infected unit. The total burden of disease is b Aq[∞]. An example is the loss of income over the disease period. The risk of death, with the infection fatality factor (IFF), may cause different considerations and decisions than about money. Here for the SIA model we first assume IFF = 0, then we might generalise to $b = bm + bd$ IFF for a pure monetary component bm and the monetary aspects bd for IFF ≠ 0, and later for the SICD model we focus on fatalities.

The share of the infected itself is already an indicator of the burden of disease. By setting norms, we may formulate an acceptable burden of disease (ABD). We can reduce the burden of disease *b* Aq[∞] by an intervention. The advantage of monetisation is that we can now relate the benefits of the intervention to the costs of the intervention. The total cost in level (q) is:

$$TCq = b \, Aq[\infty \mid \textit{Intervention}] + \text{Cost}[\textit{Intervention}] \leq \text{ABD}$$

With ABD the acceptable level of the burden of disease for the whole population then *a* = ABD / *N*[0] is the per unit value. The goal can be reformulated in proportions (p) as:

$$TCp = b \, Ap[\infty \mid \textit{Intervention}] + \text{Cost}[\textit{Intervention}] / N[0] \leq a$$

When the acceptable level is determined per infected unit, then this causes us to consider policy intervention that are directed at the units, e.g. on *treatment* or on the *duration of the disease*. When the acceptable level is determined for the whole population, then the burden of the infected is dispersed over the wider population, which may be regarded as an element of insurance. A *reduction of the number of infected* doesn't change the burden per infected but does change the burden on the population.

In the following we will look only at the burden for the population and at the reduction of the number of infected. A further simplification is possible by dividing by the burden per unit.

$$\text{UnitCost}[\textit{Intervention}] = \text{Cost}[\textit{Intervention}] / (b \, N[0])$$

$$TCpb = Ap[\infty \mid \textit{Intervention}] + \text{UnitCost}[\textit{Intervention}] \leq a / b$$

The latter *a* / *b* will be called the *sacrifice ratio*. For example, a disease might cost one month of illness and recovery from illness, valued at an average salary of EUR 2000 per infected person. When society deems *a* = EUR 100 per inhabitant acceptable, then the sacrifice ratio is *a* / *b* = 100 / 2000 = 5%. Conceivably 19 persons might offer EUR 100 each so that the diseased also has EUR 1900. It might also be that the diseased units remain dependent upon both own savings and charity so that the "cost for society as a whole" is calculated only in abstract manner.

The effect of the intervention can be measured as Ap[∞] - Ap[∞ | *Intervention*] in non-monetary units, but also as Ap[∞] - *TCpb* = Ap[∞] - (Ap[∞ | *Intervention*] + UnitCost[*Intervention*]) in monetary format that also accounts for the cost of the intervention. A *maximal effect* means a *minimal TCpb,* given that Ap[∞] is constant.

These relations provide an inspiration to derive a rule for *a.* Namely, when *TCpb* is minimal then costs cannot be reduced further, and society would certainly regard *unavoidable minimal costs* as minimally acceptable, i.e. *a/b* = Minimum[*TCpb* | *Intervention*]. Society may of course prefer a higher value for ABD, which value might be attained by *more* but also *less* intervention. For equal costs it may be possible to save more or less units from an infection.

There are at least two ways to reduce the number of infected units: (i) vaccination, i.e. a pharmaceutical intervention, (ii) reducing contacts, i.e. a behavioural or non-pharmaceutical intervention. Above condition on *TCpb* must be refined in order to account for the particulars of these two different methods. Potentially these methods might be mathematically equivalent e.g. when $R_{0,\text{vac}} \approx (1 - v) R_0 \approx R_{0,\text{int}}$, for *v* the degree of vaccination, but the methods could still differ because of the cost structure.

6.12.5. Vaccination

6.12.5.1. It is possible to avoid a separate compartment

We assume that vaccination is fully effective and results into immunity directly. A way to model vaccination is to introduce a new compartment V so that $N = V + S + I + A$. We can avoid the introduction of a new compartment by accounting the vaccinated as "having been infected" (and thus immune). Then we take $Aq[0] = v\, Nq[0]$ and $Ap[0] = v$. By consequence $Sp[0] = (1 - v) - \text{Seed} / Nq[0]$, in which $Iq[0] = \text{Seed}$. Vaccination then results into the limit value $Ap[\infty \mid Ap[0] = v]$, see our discussion of the limit values.

The first infectious unit (the seed) has an offspring of R_0 units, via direct infection of its contacts. When those contacts have been vaccinated with degree v in the range $[0, 1]$, then this means that $v\, R_0$ would not be infected and $(1 - v)\, R_0$ would still be infected. The infection will remain constant when $(1 - v)\, R_0 = 1$; it will grow then $(1 - v)\, R_0 > 1$; and it will reduce when $(1 - v)\, R_0 < 1$. The latter can be rewritten into a condition for the degree of vaccination to warrant that the infection will be reduced: $v > 1 - 1 / R_0$.

6.12.5.2. Acceptable and minimum overall costs

Above cost-benefit calculation must be adapted for the particulars of vaccination. The vaccinated units do not suffer the burden of disease. Thus this burden has been reduced to $b\, (Ap[\infty \mid v] - v)$.

Vaccination comes with a cost c per unit. The total cost (TC) is, in level (q) and proportion (p):

$$TCq = c\, v\, Nq[0] + b\, (Aq[\infty \mid v] - v\, Nq[0]) \le ABD$$

$$TCp = c\, v + b\, (Ap[\infty \mid v] - v) \le a$$

$$TCpb = c/b\, v + (Ap[\infty \mid v] - v) = (p - 1)\, v + Ap[\infty \mid v] \le a/b \quad \text{(i.e. normalised with } b, \text{ using } p =$$

c/b)

6.12.5.3. Cost-benefit table

Before discussing details, the main result can be presented upfront.

(i) The proportion of units saved from an infection is $v + (Ap[\infty] - Ap[\infty \mid v])$.

(ii) The proportion of units not saved is $Ap[\infty \mid v] - v$.

(iii) The sacrifice ratio is $TCpb = p\, v + (Ap[\infty \mid v] - v)$. The latter can be multiplied by $b\, N[0]$, with b the burden of disease per unit, to find the aggregate "acceptable burden of disease" (ABD).

(iv) The gain is the reduction of per unit costs, given as $Ap[\infty] - TCpb$, which consists of the reduction in Ap corrected for the normalised costs p per vaccination.

For example, if we vaccinate 50% of the units at a cost of $p = 1/10$, then the costs and benefits can be tabulated as follows.

```
SIA["Vaccination", Table, .5, 1/10]
```

	Population	Proportion
Ap[∞] = Ap[∞	R0 = 4.]	17.1×10^6	0.980173
Vaccination (v)	8.7×10^6	0.5
Ap[∞	v]	15.6×10^6	0.898408
Ap[∞	v] − v (not saved)	6.93×10^6	0.398408
Ap[∞] − Ap[∞	v]	1.42×10^6	0.0817647
Ap[∞] − Ap[∞	v] + v (saved)	10.1×10^6	0.581765
TCpb = p v + Ap[∞	v] − v	7.8×10^6	0.448408
Ap[∞] − TCpb	9.25×10^6	0.531765

Minimising *TCpb* gives the following level of vaccination:

```
SIA["Vaccination", NMinimize, 1/10][[3]]
```

Vaccination (v) (min TCpb) → 0.753497

```
SIA["Vaccination", Table, 1/10]
```

	Population	Proportion
Ap[∞] = Ap[∞	R0 = 4.]	17.1×10^6	0.980173
Vaccination (v) (min TCpb)	13.1×10^6	0.753497
Ap[∞	v]	13.1×10^6	0.753886
Ap[∞	v] − v (not saved)	6767	0.000388901
Ap[∞] − Ap[∞	v]	3.94×10^6	0.226287
Ap[∞] − Ap[∞	v] + v (saved)	$17. \times 10^6$	0.979784
TCpb = p v + Ap[∞	v] − v	1.32×10^6	0.0757386
Ap[∞] − TCpb	15.7×10^6	0.904434

```
SIA[Explain, "Intervention"]
```

SIA["Vaccination", v, p] for numerical values gives TCpb = p v + (Ap[inf | v] − v) /.

(Sp0 –> 1 − v − seed). This value of the sacrifice ratio can be multiplied by b Nq[0],

for b the burden of disease per unit. Output is stored in Results[SIA, "Vaccination"]

SIA["Vaccination", NMinimize, p] finds v by minimizing over TCpb

SIA["Vaccination", Table, (v,) p] presents the outcome both in levels and proportions. The proportion of

units saved from an infection is v + (Ap[inf] − Ap[inf | v]). The proportion of units not saved is Ap[inf |

v] − v. The costs and benefits can be evaluated in both units and money (units minus cost of change)

SIA["R0intervention", R0int, p] for numerical values gives TCpb = p (R0

− R0int)^2 + Ap[inf | R0int]. The sacrifice ratio can be multiplied by b Nq[0], for b

the burden of disease per unit. Output is stored in Results[SIA, "R0intervention"]

SIA["R0intervention", NMinimize, p] finds R0int by minimizing TCpb

SIA["R0intervention", Table, (v,) p] presents the outcome in levels and proportions. The proportion

saved from an infection is Ap[inf | R0] − Ap[inf | R0int]. The proportion not saved is Ap[inf | R0int].

The costs and benefits can be evaluated in both units and money (units minus cost of change)

6.12.5.4. Minimisation

The following discusses how to find *a*/*b* = Minimise[*TCpb* | *v*] subject to *v* in [0, 1]. We look at the

shape of the cost function and check whether we can use algebraic or numerical methods.

```
apinf = 1 - SIA[ProductLog, R0, Sp0 , v]
```

$$\frac{W\left(R0\, Sp0\left(-e^{R0\,(v-1)}\right)\right)}{R0} + 1$$

```
TCpb =  p v + (apinf - v) /. (Sp0 -> 1 - v - seed)
```

$$p\, v + \frac{W\left(R0\left(-e^{R0\,(v-1)}\right)(-seed - v + 1)\right)}{R0} - v + 1$$

An example is that the cost of a vaccination is $p = 1/10$ of the cost of the disease, or that vaccinating 10 units is as costly as letting one unit suffer the disease (and asking 19 other units for help with compensation). For a single unit there is the choice of either vaccination at the cost of EUR 200 or taking the risk of the infection and disease at the cost of EUR 2000. The risk is determined by how many would be vaccinated.

```
TCpbnum[v_, p_] = TCpb /.  {R0 → 4, seed → 100 / NqSeed}
```

$$\frac{1}{4} W\left(-4\, e^{4\,(v-1)}\,(0.999994 - v)\right) + p\, v - v + 1$$

This cost function is nonlinear due to the term $Ap[\infty \mid v]$. When the function would be smooth and convex to the origin then we can find a minimum from the first order conditions. In other cases we may use a numerical approach.

Mathematica cannot solve the first order condition.

```
Solve[D[TCpb, v] == 0 , v]
```

⋯ Solve: This system cannot be solved with the methods available to Solve.

$$\text{Solve}\left[p - \frac{e^{-R0\,(v-1)}\left(R0\, e^{R0\,(v-1)} - R0^2\, e^{R0\,(v-1)}(-seed - v + 1)\right) W\left(R0\left(-e^{R0\,(v-1)}\right)(-seed - v + 1)\right)}{R0^2\,(-seed - v + 1)\left(W\left(R0\left(-e^{R0\,(v-1)}\right)(-seed - v + 1)\right) + 1\right)} - 1 = 0,\, v\right]$$

A plot for numerical values is:

```
Plot @@ {TCpbnum[v, #] & /@ {0, 1 / 5, 1 / 10, 1 / 20} , {v, 0, 1}, AxesOrigin → {0, 0},
   AxesLabel → {"v", "p v + Ap[∞ | v] - v"}, BaseStyle → {FontSize → 13}}
```
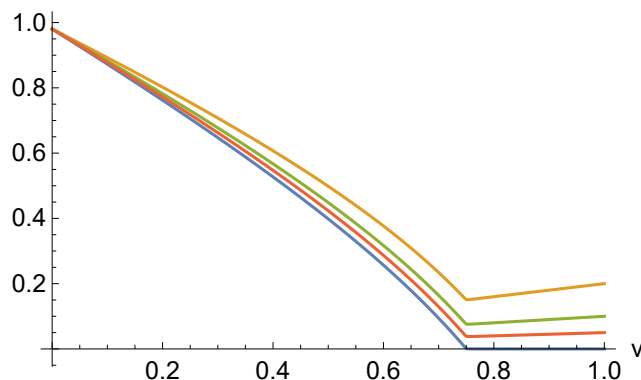

Remarkably, the normalised total cost reaches a minimum around $v = 1 - 1/R_0$.

Substitution of $v \to 1 - 1/R0$ gives:

```
nc = TCpb /. v → 1 - 1 / R0
```

$$p\left(1 - \frac{1}{R0}\right) + \frac{W\left(-\frac{R0\left(\frac{1}{R0}-\text{seed}\right)}{e}\right)}{R0} + \frac{1}{R0}$$

Observe that when seed = 0, then we get:

```
ProductLog[-1 / e]
```

−1

When seed = 0 (no need for vaccination) then the normalised cost at $v = 1 - 1 / R_0$ reduces to:

```
nc /. seed → 0
```

$$p\left(1 - \frac{1}{R0}\right)$$

Numerical minimisation generates proper values.

```
fmsol = NMinimize @@ {TCpb /. {R0 → 4 , seed → 100 / NqSeed, p → 1 / 10}, v}
```

{0.0757386, {$v \to 0.753497$}}

The extreme value of seed = 0 would generate a situation without infections, but the minimum of this particular function still gives the expected outcome.

```
fmsol = NMinimize @@ {TCpb  /. {R0 → 4 , seed → 0, p → 1 / 10}, v}
```

{0.075, {$v \to 0.75$}}

Conclusions are:

(i) the infection will be reduced when $v > 1 - 1 / R_0$
(ii) for some cost functions it would also be possible to attain minimum costs
(iii) in this particular example (with $R_0 = 4$, IFF = 0, etcetera) a sacrifice ratio of 0.075386 would be minimally acceptable (since it forms an overall minimum) and it is attained for the value $v = 0.753497 > 1 - 1 / R_0$
(iv) vaccination is also possible at levels $v \le 1 - 1 / R_0$ but then there will be a higher burden of disease and costs would not be minimal
(v) let us not forget that it was assumed that all units are equal so that it does not matter which unit is vaccinated and which is not.

6.12.5.5. Implied R0 (1 - *v*)

A way to imagine the impact of vaccination is that the original R_0 is reduced to $R_{0,\,\text{vac}} = R_0\,(1 - v)$. When we check what the implied $\text{Ap}[\infty \mid R_{0,\,\text{vac}}]$ is, then we find a slight difference with the proportion as calculated above for those who were not saved by the vaccination. In levels, the below gives 7046.79 units instead of 6767 units above that are not saved from infection, starting with the Seed = 100. It is tempting to attribute the difference to numerical precision, but given the general precision of *Mathematica* it may be better to regard the use of this $R_{0,\,\text{vac}} = R_0\,(1 - v)$ as a (very decent) approximation only.

```
vac = "Vaccination (v) (min TCpb)" /. Results[SIA, "Vaccination"]
```

0.753497

```
impliedR0v = BetaSIA[] / GammaSIA[] (1 - vac) /. Options[SIA]
```

0.986012

```
impliedApinf = 1 - SIA[ProductLog, impliedR0v, 1 - Seed / Nqt0] /. Options[SIA]
```

0.000404988

```
TheFormer Nqt0 /. Options[SIA]
```

7046.79

6.12.5.6. PM. 1. Vaccination and Options[SIA] for running the model

For running and plotting of the model, we can handle vaccination and Ap[∞ | v] by setting Aqt0 $\rightarrow$ v NqSeed in the Options[SIA]. In the SIA model, Sq[0] = N[0] - Aq[0] - Seed. The SIA[Limit, ...] statement uses the same relations on the ProductLog as used above. Perhaps it is useful to show this.

```
ResetOptions[SIA]; SetOptions[SIA, Aqt0 → v NqSeed]
```

{Aqt0 $\rightarrow$ 1.74×10^7 v, Cqt0 $\rightarrow$ 0, DataMold $\rightarrow$ {S, I, A}, Dqt0 $\rightarrow$ 0, Iqt0 $\rightarrow$ Null,

  Nqt0 $\rightarrow$ 1.74×10^7, Onset $\rightarrow$ 0, Seed $\rightarrow$ 100, Yqt0 $\rightarrow$ 0, AlphaSEYCD() $\rightarrow$ 1.,

  BetaSIA() $\rightarrow$ 0.4, GammaSIA() $\rightarrow$ 0.1, PhiSICD() $\rightarrow$ 0.015, R0() $\rightarrow$ Null}

```
explv =
  (Explain[SIA[Limit, BetaSIA[] → R0 GammaSIA[]], SIA] /. {1. → 1, -1. → -1} //
    Simplify) /. 1.` → 1
```

$$\left\{ S \rightarrow -\frac{W\big(R0\,(v - 0.999994)\,e^{R0\,(v-1)}\big)}{R0},\ I \rightarrow 0,\ A \rightarrow \frac{W\big(R0\,(v - 0.999994)\,e^{R0\,(v-1)}\big) + R0}{R0} \right\}$$

Above expression generates Ap[∞ | v] - v as follows:

```
("A" - v) /. explv /. R0 → 4 // Simplify
```

$$\frac{1}{4}\big(W(0.0732626\,e^{4\,v}\,(v - 0.999994)) - 4\,v + 4\big)$$

While above we used:

```
(apinf - v) /. (Sp0 -> 1 - v - seed) /. {R0 → 4, seed → 100 / NqSeed} // Simplify
```

$$\frac{1}{4}\big(W(0.0732626\,e^{4\,v}\,(v - 0.999994)) - 4\,v + 4\big)$$

Thus, setting the options in above manner: (i) generates the situation for the vaccination as we have discussed in the above, and (ii) allows to run the model and plot the course of the infection.

```
SetOptions[SIA, Aqt0 → vac NqSeed] (* the value of vac was set above *)
```

{Aqt0 $\rightarrow$ 1.31108×10^7, Cqt0 $\rightarrow$ 0, DataMold $\rightarrow$ {S, I, A}, Dqt0 $\rightarrow$ 0,

  Iqt0 $\rightarrow$ Null, Nqt0 $\rightarrow$ 1.74×10^7, Onset $\rightarrow$ 0, Seed $\rightarrow$ 100, Yqt0 $\rightarrow$ 0, AlphaSEYCD() $\rightarrow$ 1.,

  BetaSIA() $\rightarrow$ 0.4, GammaSIA() $\rightarrow$ 0.1, PhiSICD() $\rightarrow$ 0.015, R0() $\rightarrow$ Null}

```
SIA[Run, Pr, newModel, t, 150, Seed → 100];
```

```
SIA[Plot, 0, 150]
```
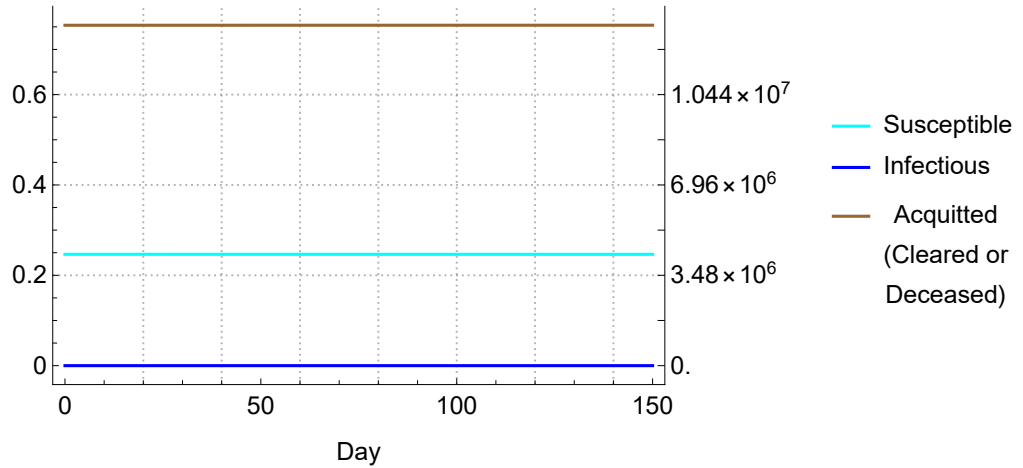


```
ResetOptions[SIA];
```

6.12.5.7. PM. 2. When the sacrifice ratio is determined exogenously

Society can determine the sacrifice ratio directly. In that case, more than minimal costs can be made by vaccinating more or less units.

```
eqval[p_] =
 sacrificeratio == p v + (apinf - v) /. (Sp0 -> 1 - v - seed) /. {seed → 100 / NqSeed}
```

$$\text{sacrificeratio} = \frac{W\big(R0\,(0.999994 - v)\,\big(-e^{R0\,(v-1)}\big)\big)}{R0} + p\,v - v + 1$$

For example, when $p = 1/10$ and $R_0 = 4$ and society accepts a sacrifice ratio of 0.9 which is much higher than above 0.0757386, then (see above plot) there are two solutions for the degree of vaccination, one to the left of 1 - 1 / R_0 and one to the right. The two values can be found by using different starting values. It depends upon society whether the low v or the high v solution is chosen.

```
FindRoot @@ {eqval[1 / 10] /. {R0 → 4, sacrificeratio → .9}, {v, .5}}
```

$\{v \to 0.0823181\}$

```
FindRoot @@ {eqval[1 / 10] /. {R0 → 4, sacrificeratio → .9}, {v, 1}}
```

$\{v \to 9.\}$

PM. In some cases we can solve the equation to find an explicit expression for *v*. Apparently, *Mathematica* does not solve for nonzero values of $p = c/b$, but a solution appears to be possible when $p = 0$, i.e. when the cost of vaccination is neglected.

```
sol = Solve[eqval[1 / 10], v]
```

⋯ Solve: Solve was unable to solve the system with inexact coefficients or the system obtained by direct rationalization of inexact numbers present in the system. Since many of the methods used by Solve require exact input, providing Solve with an exact version of the system may help.

$$\text{Solve}\left[\text{sacrificeratio} = \frac{W\left(\text{R0} \left(0.999994 - v\right) \left(-e^{\text{R0}\,(v-1)}\right)\right)}{\text{R0}} - \frac{9\,v}{10} + 1, v\right]$$

```
sol = Solve[eqval[0] , v]
```

$$\left\{\left\{v \to \frac{0.999994\ \text{R0}\ e^{-1.\ \text{R0}\ \text{sacrificeratio}} + 1.\ \text{R0}\ \text{sacrificeratio} - 1.\ \text{R0}}{\text{R0}\ e^{-1.\ \text{R0}\ \text{sacrificeratio}} - 1.\ \text{R0}}\right\}\right\}$$

A maximum-condition prevents negative outcomes.

```
vaccination[R0_, sacrificeratio_] = Max[0, v /. sol[[1, 1]]]
```

$$\max\left(0, \frac{0.999994\ \text{R0}\ e^{-1.\ \text{R0}\ \text{sacrificeratio}} + 1.\ \text{R0}\ \text{sacrificeratio} - 1.\ \text{R0}}{\text{R0}\ e^{-1.\ \text{R0}\ \text{sacrificeratio}} - 1.\ \text{R0}}\right)$$

For example, with $R_0 = 4$ and an acceptable sacrifice ratio of 1/4 = 0.25 (much higher than above minimum 0.0758), and neglecting the cost of vaccination, then vaccination of 60% means that 15% less units are vaccinated. The higher accepted cost arises because of the (individual) burden of disease b (and it is not said that society compensates the infected).

```
vaccination[4, 1 / 4]
```

0.604509

The following contour plot indicates the general shape of the curves. The plot relates to what has been plotted for "herd immunity".

```
ContourPlot[Evaluate[vaccination[r0, val]],
 {r0, .5, 8}, {val, 0, 1}, FrameLabel → { "R0", "a / b"},
 BaseStyle → {FontSize → 13}, AspectRatio → 1 / GoldenRatio]
```


PM. Compare.

```
ApByR0Plot["HerdImmunity"]
```

A_{HI} called Herd Immunity
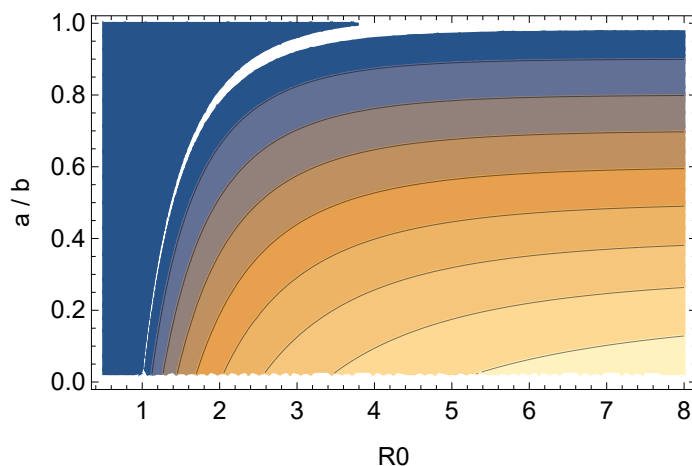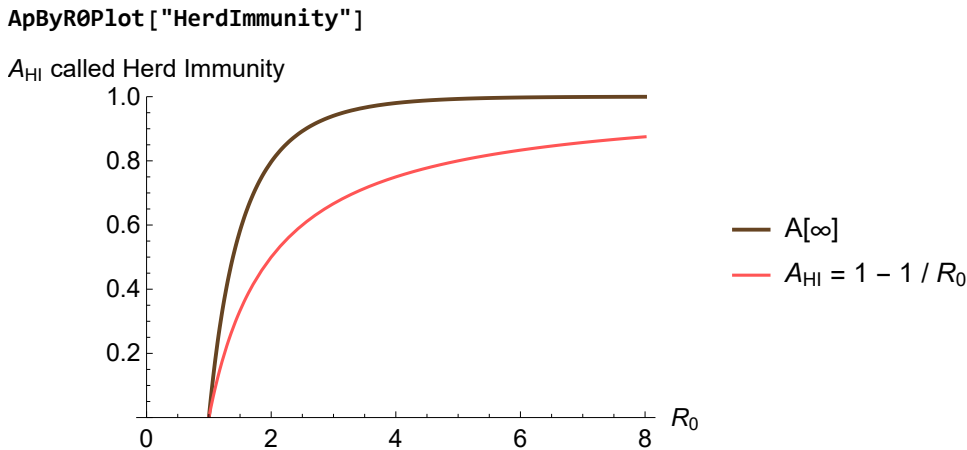
6.12.6. Social distancing, reducing contacts and their infectiousness

A behavioural intervention reduces both *contacts* and the *infectiousness of contacts*, so that R_0 is adapted to $R_{0,int}$. The impact is measured by comparing the two final states, in this case Ap[∞] = Ap[∞ | R_0] and Ap[∞ | $R_{0,int}$]. This comparison can be done directly in terms of units but also in monetary form when we also subtract the cost of the intervention.

6.12.6.1. Acceptable and minimum overall costs

Above cost-benefit calculation must be adapted for the particulars of the non-pharmaceutical intervention. The crucial question is what costs might arise when R_0 is adapted. A much used format in cost minimisation is quadratic costs. Let us assume that the reduction of contacts has quadratic costs $c\,(R_0 - R_{0,int})$^2 per unit.

The total cost (TC) is, in level (q) and proportion (p):

$$TCq = c\,(R_0 - R_{0,int})\text{^2 } Nq[0] + b\,Aq[\infty \mid R_{0,int}] \;\leq ABD$$

$$TCp = c\,(R_0 - R_{0,int})\text{^2} + b\,Ap[\infty \mid R_{0,int}] \;\leq a$$

$$TCpb = c/b\,(R_0 - R_{0,int})\text{^2} + Ap[\infty \mid R_{0,int}] \;\leq a/b \quad \text{(i.e. normalised with } b, \text{ using } p = c/b)$$

6.12.6.2. Cost-benefit table

Before discussing details, the main result can be presented upfront.

(i) The proportion of units saved from an infection is Ap[∞] - Ap[∞ | $R_{0,int}$].

(ii) The proportion of units not saved is Ap[∞ | $R_{0,int}$].

(iii) The sacrifice ratio is *TCpb* (see above). The latter can be multiplied by $b\,N[0]$, with b the burden of disease per unit, to find the aggregate "acceptable burden of disease" (ABD).

(iv) The gain is the reduction of per unit costs, given as Ap[∞] - *TCpb*, which consists of the reduction in Ap corrected for the quadratic costs of the behavioural adjustment.

For example, if we target a $R_{0,int}$ = 0.80 while the quadratic cost parameter is p = 1/10, then the costs and benefits can be tabulated as follows, indicating that monetary costs will be higher than the burden of disease itself.

```
SIA["R0intervention", Table, .8, 1/10]
```

	Population	Proportion
Ap[∞] = Ap[∞ | R0 = 4.]	17.1×10^6	0.980173
Ap[∞ | R0int = 0.8] (not saved)	500	0.0000287337
Ap[∞] − Ap[∞ | R0int] (saved)	17.1×10^6	0.980144
TCpb = p (R0−R0int)^2 + Ap[∞ | R0int]	17.8×10^6	1.02403
Ap[∞] − TCpb	−763 095	−0.043856

Minimising *TCpb* gives the following $R_{0,\,int}$.

```
SIA["R0intervention", NMinimize, 1/10][[3]]
```

Ap[∞ | R0int = 0.998526] (not saved) (min TCpb) → 0.00222168

```
Results[SIA, "R0intervention", NMinimize][[2, 1, 2]]
```

0.998526

```
SIA["R0intervention", Table, 1/10]
```

	Population	Proportion
Ap[∞] = Ap[∞ | R0 = 4.]	17.1×10^6	0.980173
Ap[∞ | R0int = 0.998526] (not saved) (min TCpb)	38 657	0.00222168
Ap[∞] − Ap[∞ | R0int] (saved)	$17. \times 10^6$	0.977951
TCpb = p (R0−R0int)^2 + Ap[∞ | R0int]	15.7×10^6	0.903107
Ap[∞] − TCpb	1.34×10^6	0.0770662

```
SIA[Explain, "Intervention"]
```

SIA["Vaccination", v, p] for numerical values gives TCpb = p v + (Ap[inf | v] − v) /.

    (Sp0 −> 1 − v − seed). This value of the sacrifice ratio can be multiplied by b Nq[0],

    for b the burden of disease per unit. Output is stored in Results[SIA, "Vaccination"]

SIA["Vaccination", NMinimize, p] finds v by minimizing over TCpb

SIA["Vaccination", Table, (v,) p] presents the outcome both in levels and proportions. The proportion of

    units saved from an infection is v + (Ap[inf] − Ap[inf | v]). The proportion of units not saved is Ap[inf |

    v] − v. The costs and benefits can be evaluated in both units and money (units minus cost of change)

SIA["R0intervention", R0int, p] for numerical values gives TCpb = p (R0

    − R0int)^2 + Ap[inf | R0int]. The sacrifice ratio can be multiplied by b Nq[0], for b

    the burden of disease per unit. Output is stored in Results[SIA, "R0intervention"]

SIA["R0intervention", NMinimize, p] finds R0int by minimizing TCpb

SIA["R0intervention", Table, (v,) p] presents the outcome in levels and proportions. The proportion

    saved from an infection is Ap[inf | R0] − Ap[inf | R0int]. The proportion not saved is Ap[inf | R0int].

    The costs and benefits can be evaluated in both units and money (units minus cost of change)

6.12.6.3. Minimisation

We now focus on finding a/b = Minimize[$TCpb\,|R_{0,\,int}$] subject to $R_{0,\,int}$ in $[0, R_0]$.

```
apinf = 1 − SIA[ProductLog, R0int, 1 − seed, 0]
```

$$\frac{W\!\left(-e^{-R0int}\ R0int\ (1 - seed)\right)}{R0int} + 1$$

```
TCpb = p (R0 - R0int) ^2 + apinf
```

$$p\,(R0 - R0int)^2 + \frac{W\left(-e^{-R0int}\,R0int\,(1 - seed)\right)}{R0int} + 1$$

Below we will consider values for *p* of 0, 1/5, 1/10 and 1/20.

```
TCpbnum[R0int_, p_] = TCpb /. {R0 → 4, seed → 100 / NqSeed}
```

$$p\,(4 - R0int)^2 + \frac{W\left(-0.999994\,e^{-R0int}\,R0int\right)}{R0int} + 1$$

Special cases are (a) $R_{0,\,int} = 1$ with no adaptation costs ($p = 0$), and (b) $R_{0,\,int} = R_0$ for any *p.* In the first case, the Seed = 100 number of infections is continued till eternity, and we tend to expect that eventually still all susceptible units will be infected, but eventually there are two effects: (i) the reduction of the *effective* reproductive number, (ii) the seed appears to be a smaller fraction of the acquitted with exponential decay that may also take a very long duration.

```
{TCpbnum[1, 0], TCpbnum[4, p]}
```

{0.00338649, 0.980173}

```
TheFormer * NqSeed
```

$\{58\,925.,\ 1.7055 \times 10^7\}$

This cost function is nonlinear. When the function would be smooth ("well-behaved") and convex to the origin then we can find a minimum from the first order conditions. In other cases we may use a numerical approach.

Mathematica cannot solve the first order condition.

```
Solve[D[TCpb, R0int] == 0 , R0int]
```

⋯ Solve: This system cannot be solved with the methods available to Solve.

$$\text{Solve}\left[-2\,p\,(R0 - R0int) - \frac{W\left(-e^{-R0int}\,R0int\,(1 - seed)\right)}{R0int^2} - \right.$$
$$\left. \frac{e^{R0int}\left(e^{-R0int}\,R0int\,(1 - seed) - e^{-R0int}\,(1 - seed)\right)\,W\left(-e^{-R0int}\,R0int\,(1 - seed)\right)}{R0int^2\,(1 - seed)\,\left(W\left(-e^{-R0int}\,R0int\,(1 - seed)\right) + 1\right)} = 0,\ R0int\right]$$

Recall that the function for Ap rises sharply from $R_0 = 1$ (see below plot for *p* = 0). When this format is combined with quadratic costs then we get a special form. A plot for numerical values *p* in {0, 1/5, 1/10, 1/20} is:

```
Plot @@ {TCpbnum[R0int, #] & /@ {0, 1/5, 1/10, 1/20},
  {R0int, 0, 6}, AxesOrigin → {0, 0}, PlotRange → {0, 2}, AxesLabel →
  {"R0int", "p (R0-R0int)^2 + Ap[∞ | R0int]"}, BaseStyle → {FontSize → 13}}
```
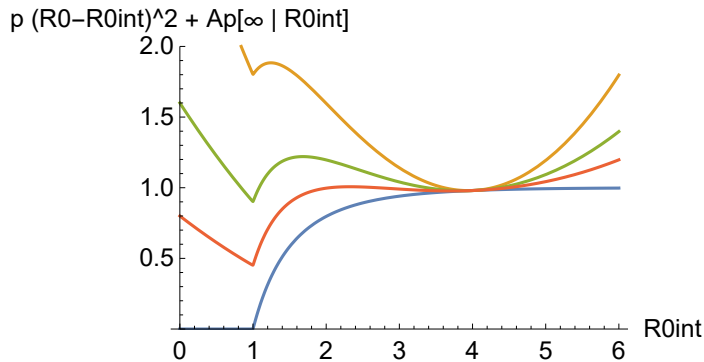

Some conclusions are:

(i) Given the notion of quadratic costs of adjustment and the special format of other costs b Aq[∞ | $R_{0,\,int}$] we may accept that a *local minimum* may be found around $R_{0,\,int}$ = 1. Thus, we can accept that values around $R_{0,\,int}$ = 1/2 cannot be generated by this form of the cost function.

(ii) The value of p apparently has a switching effect on the *global minimum*. For a particular $p = p_{crit}$, that local minimum will have the same value as for the original reproduction number. Alternatively put: TCpb[$R_{0,\,int}, p_{crit}$] = Ap[∞ | R_0]. For $p > p_{crit}$ it is better to leave R_0 like it is, and for $p < p_{crit}$ it is better to switch to the value around 1. Given the interventions for SARS-CoV-2 apparently the latter is the case.

```
TCpbnum[R0int, pcrit] == TCpbnum[4, p]
```

$$\text{pcrit}\,(4 - \text{R0int})^2 + \frac{W\!\left(-0.999994\,e^{-\text{R0int}}\,\text{R0int}\right)}{\text{R0int}} + 1 = 0.980173$$

Numerical minimisation of *TCpb* generates proper values.

```
fmsol = NMinimize[TCpbnum[R0int, 1/10], R0int]
```

{0.903107, {R0int → 0.998526}}

The cost reduction would be:

```
TCpbnum[4, p] - fmsol[[1]]
```

0.0770662

Scaled up to the aggregate level, this amounts to a sizeable sum, of 7.7% of the population with each a burden of disease of b.

```
TheFormer * b * NqSeed
```

$1.34095 \times 10^6\,b$

6.12.6.4. PM. When the sacrifice ratio is determined exogenously

Society can determine the sacrifice ratio directly. In that case, more than minimal costs can be made by saving *more* or *less* units.

```
eqval[p_] = sacrificeratio == TCpb /. {seed → 100 / NqSeed}
```

$$\text{sacrificeratio} = p\,(\text{R0} - \text{R0int})^2 + \frac{W(-0.999994\,e^{-\text{R0int}}\,\text{R0int})}{\text{R0int}} + 1$$

On occasion it appears that more than one solutions is possible - see the plot. It depends upon society what point is selected. (When the infected are compensated by society then it would make sense to select the minimum number of infected units.) An example is:

```
FindRoot @@ {eqval[1 / 10] /. {R0 → 4, sacrificeratio → 1}, {R0int, .5}}
```

{R0int → 0.837778}

```
FindRoot @@ {eqval[1 / 10] /. {R0 → 4, sacrificeratio → 1}, {R0int, 2}}
```

{R0int → 3.37532}

PM. Apparently, *Mathematica* does not solve for nonzero values of *p*. A solution appears to be possible when $p = 0$, i.e. when the cost of change is neglected. This appears to be none other than the relation for Ap[$\infty \mid R_0$] itself.

```
sol = Solve[eqval[1 / 10], R0int]
```

⚫ Solve: Solve was unable to solve the system with inexact coefficients or the system obtained by direct rationalization of inexact numbers present in the system. Since many of the methods used by Solve require exact input, providing Solve with an exact version of the system may help.

$$\text{Solve}\Big[\text{sacrificeratio} = \frac{1}{10}\,(\text{R0} - \text{R0int})^2 + \frac{W(-0.999994\,e^{-\text{R0int}}\,\text{R0int})}{\text{R0int}} + 1,\ \text{R0int}\Big]$$

```
sol = Solve[eqval[0] , R0int]
```

$$\Big\{\Big\{\text{R0int} \to \frac{-1.\,\log(1.0000000000 - 1.0000000000\,\text{sacrificeratio}) - 5.74714 \times 10^{-6}}{\text{sacrificeratio}}\Big\}\Big\}$$

```
TheFormer /. x_?NumberQ :⧴ (Chop[x + 1] - 1)
```

$$\Big\{\Big\{\text{R0int} \to \frac{-\log(1.0000000000 - \text{sacrificeratio}) - 5.74714 \times 10^{-6}}{\text{sacrificeratio}}\Big\}\Big\}$$

```
sol = TheFormer /.  y_?NumberQ :⧴ (Chop[y - 1] + 1)
```

$$\Big\{\Big\{\text{R0int} \to \frac{-\log(1 - \text{sacrificeratio}) - 5.74714 \times 10^{-6}}{\text{sacrificeratio}}\Big\}\Big\}$$

```
R0intervention[sacrificeratio_] = R0int /. sol[[1, 1]]
```

$$\frac{-\log(1 - \text{sacrificeratio}) - 5.74714 \times 10^{-6}}{\text{sacrificeratio}}$$

This appears to be none other than the inverse of the plot for Ap[$\infty \mid R_0$].

```
Plot[R0intervention[val], {val, 0, 1},
 AxesOrigin → {0, 0}, AxesLabel → {"a / b", "Required R0"},
 PlotRange → {0, 8}, BaseStyle → {FontSize → 14}]
```

Required R0

PM. Compare.
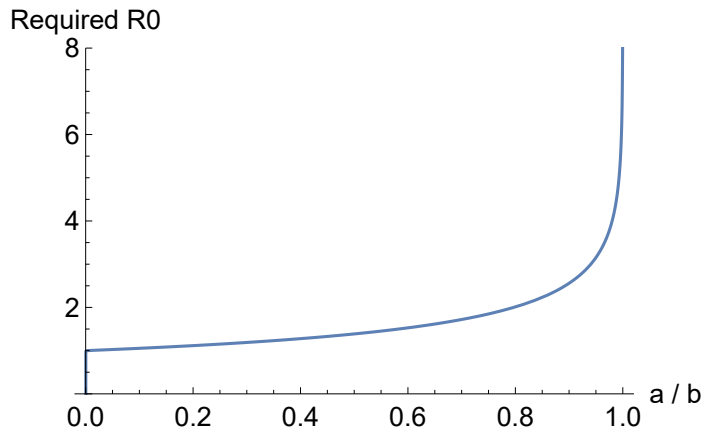
ApByR0Plot["HerdImmunity"]

A_{HI} called Herd Immunity
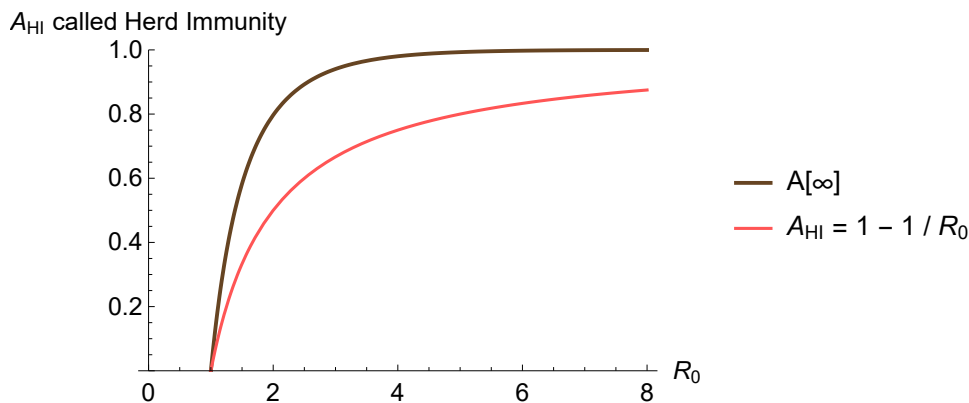
6.12.7. The problematic term "herd immunity" - continued

In subsection 6.12.1 we mentioned **(1)** "population immunity", and in subsection 6.12.3 we mentioned **(2) Ap$_{Herd, \infty}$** $\overset{\text{def}}{=}$ Ap[∞] and **(3) Ap$_{Herd, 95}$** $\overset{\text{def}}{=}$ 95% Ap[∞]. To prevent confusion, it appeared useful to first discuss intervention on vaccination and social distancing. We now continue in listing other options for what authors might mean by the term "herd immunity".

(4) *Other models* may have a *steady state,* meaning that $I = c$ so that $I' = 0$.

A "steady state" for infections is defined as follows. R_0 are the victims of an infectious unit in a situation without immunity. When only 1 contact results into a new infectious unit when there is immunity, then R_0 - 1 end up nowhere because of immunity. The ratio $(R_0 - 1) / R_0 = 1 - 1 / R_0$ gives the same notion as stated above. However, this also means that *infections continue*.

We can relate to the formulas for a disease with a steady state in section 6.2.4. When we take the IncidenceQuotient q there as βI for $I = c$ and the DiseaseDuration as $1/\gamma$, then we get the prevalence = $R_0 c / (R_0 c + 1)$. Using factor $f = R_0 c + 1$, the prevalence translates as $(f - 1) / f = 1 - 1/f$, which has the same structure but $f = R_0 c + 1$. This however reduces to $f = R_0$ if $c = 1 - 1/R_0$, which is the same as the above.

(a) If $I = c \neq 0$ then it is dubious to call this "herd immunity" since there will be infections at rate c / Nq[0].

(b) Such model is not SIA, and we should not mix models.

(5) For the SI(EY)A(CD) family, there is an *asymptotic steady state* because $I' \to 0$. For SIA we still have growth $i[\infty] = \beta\, S_p[\infty]$ - γ. This would be zero if $S_p[\infty] = 1 / R_0$. The only basic reproduction factor R_0 that allows that Sp[∞] = 1 / R_0 is $R_0 = 1$. The outcome that Sp = 1 however is incompatible with the idea that Iq[0] = Seed $\neq$ 0.

$$\text{Sp}[\infty] == 1 / R_0 \quad \text{or} \quad -\frac{W(-e^{-f}f)}{f} = \frac{1}{f} \quad \text{solves into} \quad \{\{f \to 1\}\}$$

For the SI(EY)A(CD) family, the literature still shows discussions that combine these ideas in confusing manner. It may be thought that Sp[∞] = 1 / R_0 and then we see the use of the formula $\text{Ap}_{\text{Herd}} \overset{???}{=} 1 - 1 / R_0$.

For SI(EY)A(CD), the latter combines *both* **Ip$_{\text{Herd}, \infty}$** = Ip[∞] = 0 *and* **Ip$_{\text{Herd}, \text{Top}}$** = Ip$_{\text{Top}}$ $\neq$ 0 (see point (6)). This combination is illogical.

(6) Define **Sp$_{\text{Herd}, \text{Top}}$** $\overset{\text{def}}{=} 1 / R_0$, with the advantage of an identifiable moment in time when this is reached. This notion of "*pseudo herd immunity*" is to focus on the idea that the disease gradually dies out when the *effective reproduction factor* drops below 1, or $R < 1$, either at the very start or after the top of infections.

For SARS-CoV-2 we find the proportions at the top as follows.

```
SIA[Ip, Top, BetaSIA[] / GammaSIA[]] /. Options[SIA]
```

{0.25, 0.403426, 0.346574}

However in the downward phase there will still be new infections for Sp$_\infty$ - Sp$_{\text{Top}}$. We already calculated this above. It is a curious form of "herd immunity" when one implies that when this "immunity" is reached, thereafter still 23% of the susceptibles will be infected. (This percentage depends upon the R_0, and observe that we cannot take the value from the lockdown period).

```
TopSIA = SIA[Ip, Top, BetaSIA[] / GammaSIA[] /. Options[SIA]];
Thread[ {Sp, Ip, Ap} → (SIA[Limit] - TopSIA)]
```

{Sp → −0.230173, Ip → −0.403426, Ap → 0.633599}

(7) The following only gives details about the above basic observation.

6.12.8. Solving for an assumed herd resistance

For SIA, we find Iq[∞] = 0 in above Sp[∞] + Ap[∞] = 1. Thus Iq[t] has a limit value 0 but never becomes 0.

Let us consider what value of R_0 would be relevant as a public health goal. Let us target a value of Sp[∞] = 40%, so that that the proportion of the acquitted (i.e. having had an infection but no longer infectious) is Ap[∞] = 60%.

```
eqSpLimit = 0.4 == SIA[ProductLog, R₀]
```

$$0.4 == -\frac{W\left(-e^{-R_0}\,R_0\right)}{R_0}$$

```
res = NSolve[eqSpLimit, R₀] // Quiet
```

$\{\{R_0 \to 1.52715\}\}$

```
SIA[Limit, BetaSIA[] → (GammaSIA[] R₀ /. res[[1]])]
```

{0.399994, 0, 0.600006}

Thus the authorities might intervene to achieve such lower value for R_0.

For SARS-CoV-2, Dutch authorities (RIVM) have indicated a herd resistance (so-called "herd immunity") of 60% of the population. If they used the SIA model and above reasoning then they took Ap[∞] = 40% as in above example, and apparently they used a $R_0 \approx 1.5$. However, RIVM published a $R_0 \approx 2.5$, and it is more likely that they used the steady state formula 1 - 1/R_0, which is another model.

6.12.9. A confusing notion of so-called "herd immunity" from a consideration of the steady state

In section 1.5.6 we already stated: With Ap the proportion of immune, the effective reproduction (1 - Ap) $R_0 \leq 1$ if $Ap \geq 1 - 1 / R_0$. There is the following type of reasoning in the literature that might pertain to a model with a steady state, another model than SIA but that has overlapping properties with SIA, so that the reasoning becomes confusing on different assumptions. The confusion is about using different models, and making assumptions about a steady state that would not exist in the SIA model.There is no need to give references.

Consider *another model* with a steady state, in which $I' = 0$ for $t > T$ means that I_p has gotten a constant value $I_p = c$. Let the proportional variables be $1 = S_p + I_p + A_p$. Assume that the other model has a similar relation as SIA to the effect that with $I_p = c \neq 0$, we have $R = 1$ or $S_p = 1 / R_0$. Define $h = I_p + A_p$ as the herd proportion, of those already acquitted or soon-to-be-acquitted (from infectiousness). Conveniently $R < 1$ for $t > T$.

Then $1 = 1 / R_0 + h$ which gives $h = 1 - 1 / R_0$.

Somehow, there creeps in the assumption that $c \to 0$ while the assumption for the steady state was $c \neq 0$.

This type of reasoning combines the long run Sp[∞] + Ap[∞] = 1 relation with the assumption that $S_p = 1 / R_0$ to generate Sp[∞] = 1 / R_0.

```
eqSpLimit = (GammaSIA[] / BetaSIA[] == SIA[ProductLog, R₀] /. Options[SIA])
```

$$0.25 == -\frac{W\left(-e^{-R_0}\,R_0\right)}{R_0}$$

The latter however generates another R_0. Thus curiously, the information about the current reproduction factor (0.25 = 1 / R_0 but holding for the top) is used to determine a result under another hypothetical reproduction factor.

```
res = NSolve[eqSpLimit, R₀] // Quiet
```

$\{\{R_0 \to 1.84839\}\}$

```
SIA[Limit, BetaSIA[] → (GammaSIA[] R₀ /. res[[1]])]
```

$\{0.249997, 0, 0.750003\}$

The problem with this "reasoning about the steady state" is:

(i) Unless the constant value of *I* is 0, there still will be (dwindling) infections, and possible deaths, so that this herd proportion does not provide the protection which one associates with the notion of *immunity.*

(ii) SIA has no steady state, only an approximation in the long run, so that the above assumes other models.

In the following plot, the area between the curves gives the "overshoot", which can be multiplied by 0.015 * the population size to find the death toll of confusion.

```
ApByR0Plot["HerdImmunity"]
```


A_{HI} called Herd Immunity
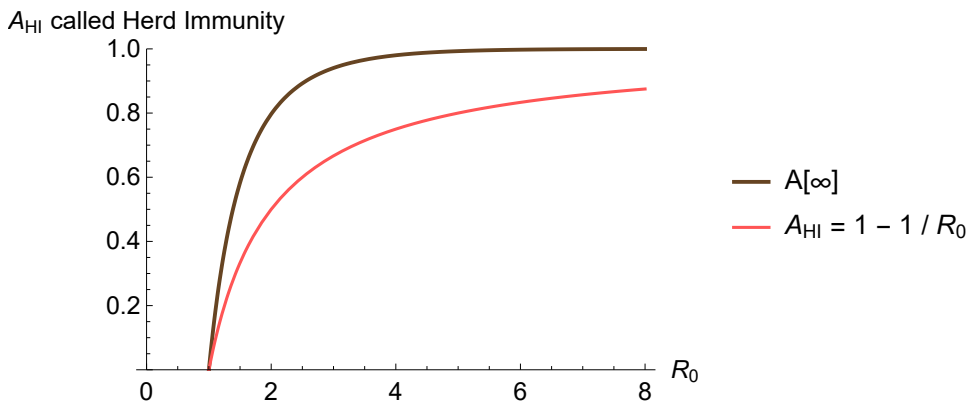
6.12.10. In sum

(1) The condition of zero growth of the infections, $i = 0$, does not only hold for the transient top but also for a steady state. A steady state in SIA is only possible when $\beta = 0$ or when $\mathcal{S}_p$ or $\mathcal{I}_p$ are quarantined. For SIA we basically have only limit values when time goes to infinity.

(2) In the limit $i = \beta\, Sp[\infty] - \gamma \neq 0$. $Sp[\infty] = 1 / R_0$ only holds when $Sp[\infty] = 1$, which is illogical since we seeded the population with an infection.

(3) There may be confusion in the discussion, but also the use of another model.

(4) It seems likely that the notion of "herd immunity" has been used in the epidemiological literature in well-defined cases, like vaccination for measles, in which discussions the experts know what they are doing, so that in practice no relevant confusion arose in those practices. However, in the SARS-CoV-2 pandemic, the term was applied by less-expert people in a situation new to them.

PM. The following is something of a mnemonic summary, that allows the substitution of parameter values.

```
eqs = { PeriodBetweenInfections[t] == 1 / BetaSIA[t],
    RecoveryPeriod[] == 1 / GammaSIA[],
    R0[t] == BetaSIA[t] / GammaSIA[],
    HerdImmunity["ConfusionOrUnspecifiedOtherModel"] == 1 - 1/ R₀,
    HerdImmunity["Rather misleading: reached R[t] < 1"] == "Sp[t] < 1/ R₀[t]",
    HerdImmunity[SIA, ApLimit] == 1 - SIA[ProductLog, R₀], (* perhaps include 95% *)
    "R₀ ≤ 1" → (R₀ ≤ 1), (* the left does not change when replacing the right *)
    TopIq → SIA[Ip, Top, R₀]};
```

```
eqs  /. Options[SIA] // MatrixForm
```

$$
\begin{pmatrix}
\text{PeriodBetweenInfections}(t) = \frac{1}{\text{BetaSIA}(t)} \\
\text{RecoveryPeriod}() = 10. \\
\text{R0}(t) = 10.\,\text{BetaSIA}(t) \\
\text{HerdImmunity}(\text{ConfusionOrUnspecifiedOtherModel}) = 1 - \frac{1}{R_0} \\
\text{HerdImmunity}(\text{Rather misleading: reached R[t] < 1}) = \text{Sp}[t] < 1/\,R_0[t] \\
\text{HerdImmunity}(\text{SIA, ApLimit}) = \frac{W\left(-e^{-R_0\,R_0}\right)}{R_0} + 1 \\
R_0 \le 1 \to R_0 \le 1 \\
\text{TopIq} \to \left\{ \frac{1}{R_0},\ \frac{R_0 - \log(R_0) - 1}{R_0},\ \frac{\log(R_0)}{R_0} \right\}
\end{pmatrix}
$$

7. SICD

Clear

Let us first clear all relevant variables, since we may do different runs. When we call a "run" then the routine does so actually itself too, but it seems to be good discipline not to wait till such routine guidance.

```
SIA[Clear]
```

```
SetOptions[SetDatabank, Databank → SICD]; (* for Explain *)
```

7.1. Fast-track run and plot

This section directly runs the model and plots the results. An appendix to this chapter does the same in steps, so that the user might take more advantage of the environment in *Mathematica.*

The Dutch population in 2020 is about 17.4 million people. It will be convenient to use formal parameter Nqt0 → the size of the population, while Npt0 → 1. Subsequently, for quick calculation of proportions it will be useful to set NqSeed = Nq[t0] at a numerical value. NqSeed is not a parameter in the model, and only a handy variable name at the user level.

```
NqSeed = 17.4 × 10^6;
SetOptions[SIA, Nqt0 → NqSeed]
```

$\{$Aqt0 → 0, Cqt0 → 0, DataMold → {S, I, A}, Dqt0 → 0, Iqt0 → Null, Nqt0 → 1.74 × 10^7, Onset → 0, Seed → 100, Yqt0 → 0, AlphaSEYCD() → 1., BetaSIA() → 0.4, GammaSIA() → 0.1, PhiSICD() → 0.015, R0() → Null$\}$

The infections are driven by the equation Iq[t0] == Seed. (Seed is parameter in the model.) The carnival festivities in the South of Holland were on February 23 2020. It is not unlikely that some 100

infections were seeded by people having returned from Italy and Austria from work and ski holidays. The options have default Seed → 100. When we use a different value of Seed, then this is "baked" into this model solution, and we must take heed of routines that rely upon Options[SIA] for the value of Seed in these options. We can mention the parameter here so that the user can check more easily that model solutions change when adapting the Seed value.

```
SICD[Run, Pr, newModel, t, 150, Seed → 100];
```

The variables are now available in levels and proportions.

```
res = SICD[75]
```

$\{415\,556., 738\,924., 1.60018 \times 10^7, 243\,683., 1.71563 \times 10^7, 1.62455 \times 10^7\}$

```
Explain[res, SICD]
```

$\{S \to 415\,556., I \to 738\,924., C \to 1.60018 \times 10^7, D \to 243\,683., N \to 1.71563 \times 10^7, A \to 1.62455 \times 10^7\}$

```
SICD[Pr, 75]
```

{0.0238825, 0.0424669, 0.919646, 0.0140048, 0.985995, 0.933651}

```
Explain[TheFormer, SICD]
```

$\{S \to 0.0238825, I \to 0.0424669, C \to 0.919646, D \to 0.0140048, N \to 0.985995, A \to 0.933651\}$

We now have the additional information that by day 75 some 240,000 people would have died, had the government not intervened.

It is useful to have plots that show both levels and proportions. The plot axes assume that it is obvious what are levels and what are proportions.

In the following plot, the population drops by about 1.5% because of the deceased. The difference between the population (red) and the cleared compartment (green) are the remaining 2% susceptible (cyan).

```
SICD[Plot, 0, 150]
```

7.2. Relation to capacity of ICU beds

The national challenge has been for the intensive care units (ICU beds). Assume that there are 1500 beds on the ICU, and that each death, after an infectious period of $1/\gamma$, still requires $1/h$ = 10 days in the hospital. (Admission to a hospital acquits from the infectious stage for the public at large.) The rate per bed is 3 patients per month or h = 3/30 per day. A good way to look at this is to take a cohort of 10 beds, with one new patient arriving every day. In a steady state with input = output, the ICUs can accept only 1500 h = 150 new patients per day, as each new patient takes the open spot in a cohort of 10 beds. Capacity remains at 1500, and we can plot above graph for a lower range at that level to show how the ICU system cannot take the new arrivals who are clearly at risk of dying.

```
sicdplotoptionsICU = SICD[Plot, Options, ICU, PlotRange → {0, 10000}];
Plot @@ {SICD[ICU, 1500, t], {t, 0, 150}, sicdplotoptionsICU}
```


People & ICU capacity
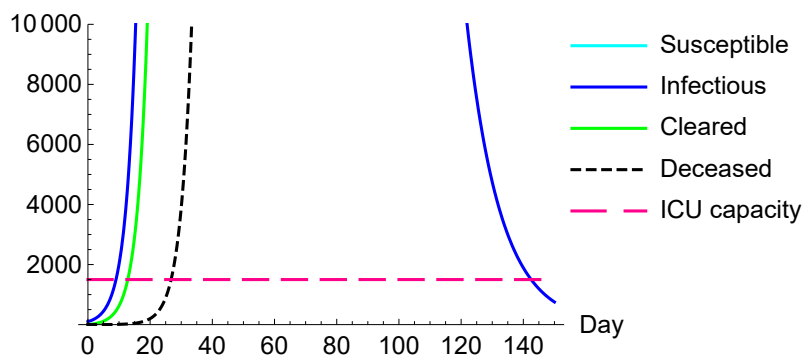
```
SICD[Explain, ICU]
```

SICD plots can be enlightened by including a line for the ICU capacity. This is simple by itself
    but becomes a bit more complicated because of the colouring and legend. In this plotting,
    the level of population is far out of range, and not included. Two main subroutines are:
SICD[Plot, Options, ICU, opts] combines plotting options and allows user options
SICD[ICU, level, t] does Append[Take[SICD[t], 4], level], thus removes N & A, inserts level

Subroutines used by those two main subroutines are:
SICD["Colours", ICU] gives DeeperPink for the ICU capacity
SICD[PlotLegends, ICU] uses the label "ICU capacity"
SICD[Plot, PlotLegends, ICU] puts the labels into a Placed object
SICD[Plot, Options, ICU,"Other"] chooses ImageSize and AxesLabel

SICD[Plot, SetOptions, ICU, opts] may set Plot options (inadvisable for other Plots)

7.3. Inclusion of *C* and *D* in SIA

7.3.1. Introduction

The Acquitted compartment of SIA can be divided into the compartments of *C* Cleared and *D*

Deceased. The sum of $A = C + D$ in SICD has the same value as A in SIA.

The SICD model has an endogenous population $N[t] = N[0] - D[t]$. This endogenous population is only used for plotting and not for the ratios. The (infectious) contacts between the units depend upon the *population density*, normalised for $N[t_0]$ and not $N[t]$. For S_p we have $N[t_0]$ in the denominator and not $N[t]$. For, if we would use $N[t]$ in the denominator then a decreasing population due to deaths would increase the intensity of contacts and thus infections, while actually the population density decreases. However, including endogenous N clarifies what the outcomes imply.

7.3.2. The traditional format for *C* and *D* causes conceptual questions

In the traditional presentation of SICD, the development of the two new compartments is given by:

$C' = \lambda I$, for those cleared of the infection

$D' = \mu I$, for the deceased

Then the acquitted are $A' = (\lambda + \mu) I = \gamma I$ as in SIA, with $\gamma = \lambda + \mu$. Conventionally the sum A is not mentioned.

This causes conceptual questions about how to calculate the values of λ and μ. When we have a variant $\gamma[t]$ e.g. depending upon an intervention, how would we adapt the parameters ? It would be tempting to keep μ constant, and let λ take all changes, since the intervention cannot change the biological death rate. This however can cause inexplicable outcomes.

7.3.3. A clearer format for *C* and *D* but still not perfect

A clearer format is:

$D' = \gamma \, \varphi \, I$, for the deceased

$C' = \gamma \, (1 - \varphi) \, I$, for those cleared of the infection

 $A' = \gamma \, I$ as in SIA

for φ the *Infection Fatality Factor* (IFF).

The model has an outcome $\text{IFF}[\infty] = D[\infty] / (C[\infty] + D[\infty]) = D[\infty] / A[\infty] = \text{Dp}[\infty] / (1 - \text{Sp}[\infty])$. We can make the model fit better when we choose $\varphi = \text{IFF}[\infty] = \text{IFF}[\text{observed}]$. This presentation of the model immediately clarifies the meaning of the parameters and the properties at the limit and for variants. A change of γ due to an intervention cannot change the biological IFF and the shares within the Acquitted department. (Check what keeping $\mu = \gamma[0] \, \varphi$ constant at the cost of $\lambda[t] = \gamma[t] - \mu$ would generate.)

In SIA, the assumption was that the duration before acquittal $\delta_{\text{SIA}} = 1 / \gamma_{\text{SIA}}$ was also the infectious period. This assumption is not changed here. The population is homogeneous: whether one belongs to the survivors or deceased cannot be said in advance. The final weights in the population can only be determined after the whole infection episode is over. The event of death only happens at the end of the acquittal period. The dying are partaking in spreading the disease thus just as much as the eventual survivors, at least as long as they are alive (and not quarantined, like being hospitalised). When we regard C and D as mere proportions of A then they can have the same infectious period.

However, in this (still rather traditional) formulation with $D' = \gamma \varphi I$, a student might consider that the mean infectious period for the deceased compartment is $1 / (\gamma \varphi)$, with a conventional value of $1 / 0.0015 = 667$ days. This idea arises purely by analogy with $A' = \gamma I$, and has not been well thought about, but can arise nevertheless merely because of such thinking by analogy. The proper reasoning is that a proportion φI is taken, with a mean period of $\varphi / (\gamma \varphi)$. Nevertheless, given the model formulation, we cannot immediately reject the possible interpretation that there is a distinction between the *lethal period* and the (survivor) *clearing period.* The traditional formulation invites the interpretation of exponential decay with different half-lives, at least until we can spot an inconsistency. The $\delta_{SIA} = 1 / \gamma_{SIA}$ acquittal duration could be an average of these two other durations.

This line of thinking opens up a can of worms, with a need for deeper discussion of the assumptions of the underlying distributions and such. All of this still results in the same finding that we actually have only proportions $D = \varphi I$ and $C = (1 - \varphi) I$, but for a moment we might also have to consider alternatives.

7.3.4. Didactically clearest format for *C* and *D*

The conclusion is that the above formulations have been inspired by the mathematical elegance of the use of ordinary differential equations. They have not been targeted at empirical and didactic clarity.

The didactically best formulation is:

$A' = \gamma I$

$D = \varphi I$

$C = (1 - \varphi) I = A - D$

This formulation makes immediately clear that the dynamics of SIA are unchanged and that C and D are only proportions, with a clear interpretation of φ = IFF. It is also immediately clear that there is no special new application of differential equations, and that we only apply a little bit of administration within the compartment of the acquitted.

Analytical clarity requires us to include the expression $A = C + D$ in the differential equations anyway, because we must clarify the meaning of the variables from the start. This will also reduce clutter in later routines on programming and runtime calculation of the simple addition. Conventional presentations of the set of equations leave out these administrative equations, apparently with the objective to enhance clarity by focusing on the core mechanism. This core however is quite clear, precisely when the administrative equations have been included. The core may be less clear when there is uncertainty about what the variables actually mean.

When SICD is presented in this didactic form, some readers might think that this would not be the proper SICD model from the literature. If those readers come from a background of already having used SI(E)R(D) models, then it would appear that they haven't quite understood the underlying mechanism yet, apparently. Nevertheless, the discussion below also provides for the traditional formulation, and we can easily check that the model outcomes are the same. Other users might want to present the original form with differential equations merely to show that it *results* into such proportions, but having a course on differential equations is a different objective than presenting SICD. When the model is presented, then it must be presented in its clearest format, and not with

some additional hidden objectives on education in mathematics.

7.4. Basic model

```
SIA[Clear]
```

```
ReadMeSIA[SICD]
```

$\mathcal{S}' = -\beta^* \, \mathcal{S} \, \mathcal{I}$

$\mathcal{A}' = \gamma \, \mathcal{I}$

$\mathcal{I}' = -\mathcal{A}' - \mathcal{S}'$

$\mathcal{D} = \mathcal{A} \, \varphi$

$C = \mathcal{A} \, (1 - \varphi)$

$\mathcal{N} = \mathcal{N}(t0) - \mathcal{D}$

$\mathcal{N}(t0) = \mathcal{N}(t0)$

$\mathcal{S}(t0) = -Cqt0 - Dqt0 - Seed + \mathcal{N}(t0)$

$\mathcal{I}(t0) = Seed$

$C(t0) = Cqt0$

$\mathcal{D}(t0) = Dqt0$

$\mathcal{A}(t0) = Cqt0 + Dqt0$

PM 1. It is logical to first state *S'* and *A'* and subsequently derive *I'*, as Smith & Moore (2001, 2004) do.

PM 2. $\beta^* = \beta \, / \, N[t_0]$. This gives a tiny value for β^* but simplifies the model.

PM 3. Above model is in levels. Division by $N[t_0]$ or setting $N[t_0] = 1$ gives proportions (and $\beta^* = \beta$).

PM 4. The model might be run for symptomatic disease. Then $\mu = \gamma$ sCFF and $\lambda = \gamma$ (1 - sCFF), for the *symptomatic case fatality factor* (sCFF). Not all infected will develop symptoms, so that sCFF = D / sC = (D / I) *(I / sC) = IFF / (sC / I).

PM 5. It is analytically and didactically relevant, namely for the comparison with SIA, to include *A*. This also reduces clutter in subsequent programming and execution (namely to always include what wasn't stated at the start).

7.5. Computable format (*Copy from SIA*)

See the corresponding section in the discussion of SIA.

7.6. The didactic and traditional formats

```
SIA[Clear]
```

An overview of the traditional / simplified and computable / readable formats is as follows.

```
symbolic = {SICD[Equations, t0, t],
            SICD[Equations, TraditionalForm, t0, t]} // Transpose // TableForm
```

$Sq'(t) = -\frac{\text{BetaSIA() Iq}(t)\, Sq(t)}{\text{Nqt0}}$ $\qquad$ $Sq'(t) = -\frac{\text{BetaSIA() Iq}(t)\, Sq(t)}{\text{Nqt0}}$

$Aq'(t) = \text{GammaSIA() Iq}(t)$ $\qquad$ $Iq'(t) = Iq(t)\left(\frac{\text{BetaSIA() Sq}(t)}{\text{Nqt0}} - \text{GammaSICD()} - \text{MuSICD()}\right)$

$Iq'(t) = -Aq'(t) - Sq'(t)$ $\qquad$ $Cq'(t) = \text{GammaSICD() Iq}(t)$

$Dq(t) = \text{PhiSICD() Aq}(t)$ $\qquad$ $Dq'(t) = \text{MuSICD() Iq}(t)$

$Cq(t) = (1 - \text{PhiSICD()}) \, Aq(t)$ $\qquad$ $Nq(t) = \text{Nqt0} - Dq(t)$

$Nq(t) = \text{Nqt0} - Dq(t)$ $\qquad$ $Aq(t) = Cq(t) + Dq(t)$

$Nq(t0) = \text{Nqt0}$ $\qquad$ $Nq(t0) = \text{Nqt0}$

$Sq(t0) = -\text{Cqt0} - \text{Dqt0} + \text{Nqt0} - \text{Seed}$ $\qquad$ $Sq(t0) = -\text{Cqt0} - \text{Dqt0} + \text{Nqt0} - \text{Seed}$

$Iq(t0) = \text{Seed}$ $\qquad$ $Iq(t0) = \text{Seed}$

$Cq(t0) = \text{Cqt0}$ $\qquad$ $Cq(t0) = \text{Cqt0}$

$Dq(t0) = \text{Dqt0}$ $\qquad$ $Dq(t0) = \text{Dqt0}$

$Aq(t0) = \text{Cqt0} + \text{Dqt0}$ $\qquad$ $Aq(t0) = \text{Cqt0} + \text{Dqt0}$

```
symbolic // SEYCDLettersOnly
```

$\mathcal{S}' = -\frac{\beta\, \mathcal{S}\, \mathcal{I}}{\mathcal{N}(t0)}$ $\qquad$ $\mathcal{S}' = -\frac{\beta\, \mathcal{S}\, \mathcal{I}}{\mathcal{N}(t0)}$

$\mathcal{A}' = \gamma\, \mathcal{I}$ $\qquad$ $\mathcal{I}' = \mathcal{I}\left(-\lambda - \mu + \frac{\beta\, \mathcal{S}}{\mathcal{N}(t0)}\right)$

$\mathcal{I}' = -\mathcal{A}' - \mathcal{S}'$ $\qquad$ $C' = \lambda\, \mathcal{I}$

$\mathcal{D} = \mathcal{A}\, \varphi$ $\qquad$ $\mathcal{D}' = \mu\, \mathcal{I}$

$C = \mathcal{A}\, (1 - \varphi)$ $\qquad$ $\mathcal{N} = \mathcal{N}(t0) - \mathcal{D}$

$\mathcal{N} = \mathcal{N}(t0) - \mathcal{D}$ $\qquad$ $\mathcal{A} = C + \mathcal{D}$

$\mathcal{N}(t0) = \mathcal{N}(t0)$ $\qquad$ $\mathcal{N}(t0) = \mathcal{N}(t0)$

$\mathcal{S}(t0) = -\text{Cqt0} - \text{Dqt0} - \text{Seed} + \mathcal{N}(t0)$ $\qquad$ $\mathcal{S}(t0) = -\text{Cqt0} - \text{Dqt0} - \text{Seed} + \mathcal{N}(t0)$

$\mathcal{I}(t0) = \text{Seed}$ $\qquad$ $\mathcal{I}(t0) = \text{Seed}$

$C(t0) = \text{Cqt0}$ $\qquad$ $C(t0) = \text{Cqt0}$

$\mathcal{D}(t0) = \text{Dqt0}$ $\qquad$ $\mathcal{D}(t0) = \text{Dqt0}$

$\mathcal{A}(t0) = \text{Cqt0} + \text{Dqt0}$ $\qquad$ $\mathcal{A}(t0) = \text{Cqt0} + \text{Dqt0}$

```
Options[SIA]
```

$\{\text{Aqt0} \to 0,\ \text{Cqt0} \to 0,\ \text{DataMold} \to \{S, I, A\},\ \text{Dqt0} \to 0,\ \text{Iqt0} \to \text{Null},\ \text{Nqt0} \to 1.74 \times 10^7,\ \text{Onset} \to 0,\ \text{Seed} \to 100,$
$\quad \text{Yqt0} \to 0,\ \text{AlphaSEYCD()} \to 1.,\ \text{BetaSIA()} \to 0.4,\ \text{GammaSIA()} \to 0.1,\ \text{PhiSICD()} \to 0.015,\ \text{R0()} \to \text{Null}\}$

```
TheFormer // SEYCDLettersOnly
```

$\{\text{Aqt0} \to 0,\ \text{Cqt0} \to 0,\ \text{DataMold} \to \{S, I, A\},\ \text{Dqt0} \to 0,\ \text{Iqt0} \to \text{Null},\ \mathcal{N}(t0) \to 1.74 \times 10^7,$
$\quad \text{Onset} \to 0,\ \text{Seed} \to 100,\ \text{Yqt0} \to 0,\ \alpha \to 1.,\ \beta \to 0.4,\ \gamma \to 0.1,\ \varphi \to 0.015,\ \text{R0} \to \text{Null}\}$

```
Options[SICD] (* used by Explain on standard output of SICD[t] *)
```

$\{\text{DataMold} \to \{S, I, C, D, N, A\}\}$

NB. When running the traditional model forma, its parameters can take values from Options[SIA] by first doing a substitution using the following rule.

```
SICD[GammaSIA, Rule]
```

$\{\text{GammaSICD()} \to \text{GammaSIA()} \, (1 - \text{PhiSICD()}),\ \text{MuSICD()} \to \text{GammaSIA()} \, \text{PhiSICD()}\}$

```
TheFormer // SEYCDLettersOnly
```

$\{\lambda \to \gamma\, (1 - \varphi),\ \mu \to \gamma\, \varphi\}$

7.7. Choice of R_0 = 4 and acquittal period of 10 days (*Copy from SIA*)

See the corresponding section in SIA.

7.8. Infection Fatality Factor (IFF) and symptomatic Case Fatality Factor (sCFF)

7.8.1. Infections versus symptomatic cases

The *Infection Fatality Factor* (IFF) is the *proportion of who dies after having contacted the infection.*

For Holland, a rounded estimate of the *Infection Fatality Factor* (IFF) of SARS-CoV-2 for the Dutch population, in its composition of 2020, is 1.5% (at least up to the beginning of April, see Colignatus (2020d)).

Commonly, it would also be the *Case Fatality Rate,* but we better distinguish:

(i) (level) *factor* and (instantaneous) *rate* (see Appendix A),

(ii) "Cases" are often defined by (flu-like) symptoms, which is the *symptomatic rate.* Not all infected develop symptoms.

(iii) To avoid confusion about "cases" and infections, we follow the suggestion of the London Imperial College to use infections in the denominator. However, there still is the distinction between rate IFR and factor IFF. (It does not help that the Imperial College calls the factor a rate.)

7.8.2. IFF[t] and the eventual limit value IFF[∞]

There is also the distinction between the factor at a point in time and the value in the limit.

iIFF[t] = Dp[t] / (1 - Sp[t]) observed at time t, *inclusively* counting the *haves* and *hads,* influenced by changing Ip[t]

IFF[t] = Dp[t] / (Cp[t] + Dp[t]) observed at time t, counting the *hads* only

IFF[∞] = Dp[∞] / (1 - Sp[∞]) = Dp[∞] / (Cp[∞] + Dp[∞]) the eventual limit value (the key parameter) (Ip[∞] = 0).

7.8.3. Short conversion from IFF (observation) to IFR (model parameter μ)

In traditional formulation, the SICD model requires the input of μ. The model generates outcome of a model-IFF that allows to check that the model fits the observed-IFF. The conversion is rather simple, once we understand the properties of SICD.

When Ap arises because of decay by Ip with rate $\gamma = \mu + \lambda$ then it may be assumed now *without proof* that Cp and Dp are fractions $\mu / (\mu + \lambda)$ Ap and $\lambda / (\mu + \lambda)$ Ap over the whole range, and thus also in the limit.

Consider the Acquitted Ap[∞] = Cp[∞] + Dp[∞] = 1 - Sp[∞] (since Ip[∞] = 0). These four relations apply:

```
λ + μ == γ_SIA == 0.1;
```

```
IFF[∞] == Dp[∞] / Ap[∞] == 0.015;

Dp[∞] / Ap[∞] == µ / (λ + µ)  == µ / γ_SIA  == IFF[∞] ;

µ == IFF[∞] * γ_SIA == 0.0015;
```

Importantly: it is pointless to say that IFF[∞] was found by the model, since it was input, either directly or via a hidden manner by putting in $µ$ and $λ$.

Maugeri et al. (2020) have the elegant formulation that $µ = γ$ IFF and $λ = γ$ (1 - IFF) in the model from the start. This is relevant for model variants, when we adapt $γ$ by means of interventions. Since we can do little about the IFF, the change in $γ$ affects both $µ$ and $λ$. For Options[SIA] it suffices to use $γ$ and $φ$ = IFF, and we can always calculate $µ = γ φ$ if we would need it for comparison, e.g. with publications.

It is possible to call $µ = φ γ$ the "infection fatality rate" (IFR) as long as it is clear that this rate has been calculated with a SICD model, because other models might require other rates. It is a bit dubious indeed to use the term IFR for such $µ$, since actually $µ = φ / T_{inf}$ so that the value is spread over the infectious period. The discussion about $1/γ$ would be multiplied by a discussion on $µ$. It seems best to avoid the traditional format overall.

7.8.4. An expected death toll of 250,000, if there had been no endogenous reaction or intervention

We thus also know the expected death toll when the pandemic is over. The estimate of the final prevalence is Ap[∞] = 98%. The IFF of 1.5% gives, both as a percentage and in numbers for Holland:

```
.98 * 0.015 * {1, NqSeed}
```

{0.0147, 255 780.}

Thus:

(i) Modeling with SIA(CD) helps to identify the limit prevalence given R_0 (while the IFF was already known). It remains important to keep the endogenous reactions with $R_{0,end}$ and the government intervention with $R_{0,int}$ in mind.

(ii) In this type of research it remains important to consider the toll in terms of (quality adjusted) life-years lost too.

7.8.5. Traditional limit, IFF (observation) → (model parameter $µ$), starting values

The discussion about SIA used $S = -β S_p I$ to get $S = S_0 \text{Exp}[- R_0 (\mathcal{A}_p - \mathcal{A}_p[0])]$ for an analytical result on the limit outcomes Sp[∞] and Ap[∞].

We can repeat this analysis for the two new compartments C and D, using the conventional formulation with:

$C' = λ I$, for those cleared of the infection

$D' = µ I$, for the deceased.

We have $I = (C_p' / λ) = (D_p' / µ)$ and thus two ways to substitute in the first SICD equation for S,

giving:

$$S = -\beta\, S_p (C_p{}' / \lambda) = -\beta\, S_p (D_p{}' / \mu).$$

Instead of using $R_0 = \beta / \gamma$ and A_p we now express:

$$\mathcal{S} = f[S_0, \beta / \lambda,\ C_p,\ C_p[0]]$$

$$S = g[\mathcal{S}_0, \beta / \mu, D_p, D_p[0]].$$

Thus the value of $S[\infty]$ given by SIA and new $S[\infty] = f[C_p[\infty]]$ and $S[\infty] = g[D_p[\infty]]$ allows us to find the values of $C_p[\infty]$ and $D_p[\infty]$.

It is a bit of algebra to show that Cp[∞] and Dp[∞] divide Ap[∞] up into the proportions $\varphi = \mu / \gamma$ and $1 - \varphi = \lambda / \gamma$ for $\gamma = \lambda + \mu$. There will be clutter when Cp[0] and Dp[0] are nonzero. While this derivation is rather useless when we define C and D as proportions from the start, the derivation remains useful for an understanding of the influence of the clutter.

We conclude that we have a well-defined and calculable result for the limit value of the IFF according to the model. Not writing the clutter:

$$\text{IFF}(\infty) = \frac{Dp(\infty)}{Cp(\infty) + Dp(\infty)} = \frac{Dp(\infty)}{1 - Sp(\infty)} = \frac{\mu}{\lambda + \mu} = \varphi$$

It follows that we can set the fatality rate in SICD as $\mu = \gamma_{\text{SIA}} * \text{IFF}[\infty]$, using above observation for IFF[March] = 0.015, and skipping the idea of waiting till eternity. Instead, though, it appears better to formulate C and D from the outset as proportions. This is the mathematical portent of the model, and it is better expressed so clearly then.

PM. In SICD we might use *death* statistics to trace the *development* of *infections,* but when we do not know the *level* of infections then we cannot determine the IFF, and thus neither μ. We might use the property that infections and thus deaths have exponential growth at the start with rate $\beta - \gamma$. Then we are still lacking a fundamental piece of information. Using the share of hospital patients who clear the infection might be dubious when this is no representative sample.

? PhiSICD

> Symbol
>
> PhiSICD[] stands for the Infection Fatality Factor (IFF), such that
>
> > GammaSIA[] * PhiSICD[] is the rate MuSICD[] of the infectious who decease
>
> ⌄

? MuSICD

> Symbol
>
> MuSICD[] is a symbol, no longer in the options,
>
> > for a constant rate of the Infectious who will decease (not in SIA)]
>
> PM. Numerically GammaSIA[] == GammaSICD[] + MuSICD[] means different clearing periods per subgroup
>
> ⌄

? GammaSICD

> Symbol
>
> GammaSICD[] is a symbol, no longer in the options, for a constant
>
>   rate of the Infectious per day who will no longer be infectious and not dead
>
> PM. GammaSIA[] == GammaSICD[] + MuSICD[], or MuSICD[] = GammaSIA[] PhiSICD[], with the latter
>
>   the Infection Fatality Factor (IFF). GammaSICD[] may be represented by symbol Lambda (life)
>
> ⌄

7.8.6. A small check on consistency of the starting values

The β and γ in the options for SIA apply for SICD as well. We also know the limit properties, since these are determined by the parameter values (without running the model). The IFF has been put directly into the parameters as PhiSICD[]. However, the starting values of SICD may still affect the eventual proportion of the assigned compartments. The basic warning is that model runs should have consistent starting values.

{sp1, ip1, ap1} = SIA[Limit]

{0.0198273, 0, 0.980173}

{sp2, ip2, cp2, dp2, np2, ap2} = SICD[Limit, Cqt0 → 10 000.]
 (* assume a somewhat larger nonzero starting value *)

{0.0198644, 0, 0.965442, 0.0146934, 0.985307, 0.980136}

dp2 / ap2

0.0149912

dp2 / ap1

0.0149906

PM. SICD[Limit, ...] tests whether the starting values for Cqt0 and Dqt0 are zero, and if so, allocates the shares {Cp[∞], Dp[∞]} via {1 - φ, φ} Ap[∞] , since there is no need to calculate via the Sp what already is known.

Thus both models have the same Ap[∞] while IFF[∞] = Dp[∞] / Ap[∞] fits the input parameters.

The following might be different from the input PhiSICD[] when the starting values in the Options[SIA] are nonzero.

SICD[Limit, "IFF"]

0.015

7.8.7. Summary of terms

Our terminology has been:

```
ReadMeSIA["Terminology"]
```

This terminology applies to the SICD model:

SICD improves on SIA by distinguishing A = C + D, for D deceased and C cleared.

> (−Sp') is the *incidence* of infections at moment t.

> Ip is the *prevalence* of infections at moment t.

> 1 − Sp = I + C + D is the *cumulated prevalence* at moment t (i.e. the haves and hads).

> 1 − Sp[∞] = Ap[∞] in both SIA and SICD is the *limit prevalence*, since Ip[∞] = 0.

> Cp is the cumulated survival factor at moment t (but the Sp also count as living).

> Dp is the *cumulated mortality factor* at moment t (mortality ~ population).

> Cp[∞] is the limit cumulated survival, and part of the burden of disease (infection).

> Dp[∞] is the *limit (specific) Infection Mortality Factor* (IMF) (prevalence of death).

> Dp[∞] / Ap[∞] = Dp[∞] / (Cp[∞] + Dp[∞]) is the *limit Infection Fatality Factor* (IFF).

> MuSICD[] is the *(instantaneous) Infection Fatality Rate* (IFR) (fatality ~ infection).

> MuSICD[] = GammaSIA[] * PhiSICD[] in the model, with PhiSICD[] the IFF.

> In addition for SEYCD:

> Ip = Ep + Yp subdivides into Exposed (but not infectious) and Infectious.

7.9. Interpretation of the result and finding crucial moments of the path

The major finding here is the death toll. More sobering is that we put the 1.5% Infection Fatality Factor into the model and only reproduce it. The key contribution of the model is (i) the limit prevalence Ap[∞] and (ii) that it provides a time-path, how the death toll will be reached over the course of time. If we have an intervention at a particular moment, there will still be consequences for who are in the pipeline. (See section 1.4.6 and the intervention and estimation chapters that this remains complex.)

We already presented the table with the crucial moments along the path. For some persons, a table with figures might convey the message better than indicating the moments in a graph. For other persons a graph works best. For a graph, it might be advisable not to use the full vertical axis, since 1.5% might not seem like a large difference. But it is better shown too, precisely for this reason, and thus SICD is a real improvement over SIA.

Further conclusions are best discussed when we have also looked at the SEYCD model.

7.10. How to use the SICD routine

```
? SICD
```

> **Symbol**
>
> SICD is SIA with A = C + D = Cleared + Deceased.
>
> R0 = BetaSIA / GammaSIA remains the same.
>
> New is that D = PhiSICD[] A, and D = A −
>
> > C, with phi the Infection Fatality Factor (IFF) = D[Infinity] / A[Infinity]
>
> SICD[(Pr,) t] gives {S, I, C, D, N, A}, with N = N[0] − D and A = C + D. Observe that it is best to include N
>
> > and A in the equations from the start, so that their values are directly available via the interpolated
> >
> > functions, and we do not have to burden later routines and calculations with performing the same
> >
> > arithmetic. Analytically, the meaning of N and A should also be clear from the start. PM. A DataMold
> >
> > has been set, so that one can call Explain[lis, SICD] for such output lis (see Databank package)
>
> SICD[Plot, ...] is like SIA{Plot, ...], see there. The routine takes the
>
> > first 5 elements of the list. Input of SICD[t] is okay, while A is not in the legend.
>
> SICD["Colours"] gives the colours for plotting. The deceased are given
>
> > in dashed black and white, for European / Chinese conventions. The remaining
> >
> > population (after subtracting the deceased) is plotted in red (quarantine barrier).
>
> For traditional form use SICD[Equations, TraditionalForm, ...] and SICD[N, TraditionalForm, ...].
>
> MuSICD[] = GammaSIA[] * PhiSICD[], with the latter the Infection Fatality Factor (IFF).
>
> GammaSICD[] = GammaSIA[] * (1 − PhiSICD[]).
>
> In Greek letters: mu = gamma * phi, labda = gamma (1 − phi).
>
> SICD[MuSICD, gammaSIA, IFF] := gammaSIA * IFF for the
>
> > Infection Fatality Factor IFF (limit value), is a reminder, using the limit property
> >
> > that IFF[inf] = Dp[inf] / (1 − Sp[inf]) = mu / gammaSIA = mu / (labda + mu).
>
> SICD[Limit] has Ip[inf] = 0, finds s = SIA[ProductLog, b/g, ..], and applies the same reasoning to C and D as
>
> > SIA does for A = C+D, so that c = f[s] and d = g[s], and then we easily find {s, 0, c, d}. If Cqt0 = Dqt0
> >
> > = 0, then {c, d} = (1−s) {phi, 1−phi}, otherwise there is clutter. SICD[Limit, Differences] shows the
> >
> > difference between the actual values and the outcome if those starting values were zero. SICD[Limit,
> >
> > "IFF"] returns d / (c + d) at the limit, expected to be phi but possibly deviating if not Cqt0 = Dqt0 = 0.
>
> See SICD[Explain, ICU] for plotting with a capacity level of ICU.

7.11. Cost effectiveness and reconsideration of herd resistance

Here we return to the issue of Section 6.12 on policy, intervention like vaccination, and the properties of herd resistance. In Section 6.12 the objective function was minimal costs by using *counts,* with e.g. vaccination cost per item and a burden of disease per infected item. For the SICD model it is fitting to have a more general objective function that uses *amounts.* We may now move from the *lives saved* (lives extended) criterion to the (quality adjusted) *life-years gained* criterion.

A more general objective function is especially relevant for SARS-CoV-2. The death toll consists of mostly the elderly so that the life-years gained are relatively small, so that the economic burden of an economic lockdown can be disproportional. Other diseases may have a better *cost-effectiveness per life-year gained*.

The point to keep in mind for an infectious disease is that it can explode. A lockdown creates a benefit that may be difficult to prove since an explosion is a counterfactual. In a cost-benefit calculation, e.g. for Holland, we should not reckon with the 9000 actual deaths till June but with perhaps 25,000 deaths, if there had been an endogenous behavioural response, or with perhaps 250,000 deaths if the infection had fully run its course unrestrictedly (given the values of the parameters at onset).

For the purposes of this present discussion it leads too far to develop a specific more general objective function and framework for policy optimisation. This more general framework would include the different age-groups (namely to be able to find the life-years) and the contribution of those age-groups to economic production. This is too complex for our present purposes. The present purpose is to mention some points that rather should not be forgotten for such a framework. Awareness of these points will be helpful for the interpretation of reports on cost-effectiveness and on the use of the notion of "herd resistance".

While the SICD model focuses on the lives extended, the alternative life gain measure of (quality adjusted) life-years gained, UnitSqrt and proportional shortfall highlight the relevance of other perspectives.

Mangen et al. (2014) started the ECDC-project "Estimating the current and future burden of communicable diseases in the European Union and EEA/EFTA", that also looked at this issue, so that "the effects of disease and death on the burden of disease together are expressed in a single number rather than separately". Cassini et al. (2018) give their result. The choice for the DALY framework fits the classic WHO study by Murray and Lopez. Cassini et al. (2018:15) state: "We did not apply time discounting, which is generally applied in economic studies, because we did not consider there to be reasons justifying the decline of healthy life years over time. Similarly, age-weighting was also not applied because it was considered that a healthy life year should be valued equally, irrespective of the age at which it is lived or lost. Both choices are consistent with current methodologies used by the World Health Organization's (WHO) Foodborne Disease Burden Epidemiology Reference Group (FERG) and the Global Burden of Disease studies [ref]." The next subsection clarifies that neutrality w.r.t. persons can imply non-neutrality w.r.t. life-years (e.g. as women tend to live longer than men).

Appendix E contains details about the life table and SARS-CoV-2, and how to run some routines.

The present section gives a selection of results.

7.11.1. Lives extended, life-years gained, UnitSqrt, proportional shortfall

The following is taken from Colignatus (2003, 2020). The notions of *lives saved* (lives extended) and (quality adjusted) *life-years gained* are commonly known. Perhaps lesser known are the *fair innings* (FI) by Williams (1997) and the *proportional shortfall* (PS) by the Dutch national health insurance, see Van de Wetering et al. (2013) and in Dutch ZIN (2018ab). Least known will be my own suggestion of the *UnitSqrt* measure that regards each life as 100% and that is neutral to age and sex.

The LifeGain function in *The Economics Pack* collects the various measures in a uniform format. This has the advantage that outcomes for particular cases can be compared without uncertainty about the input and the calculation. The LifeGain function uses a standard set of variables, and particular aspects, like the rate of interest for discounting or the *fair innings* age target, are in options.

The LifeGain function has been formulated in terms of an intervention leading to recovery. With age *a* and remaining life expectancy *d* because of an accident or disease, then the intervention would recover life expectancy with an additional *x* to the value *e* = *d* + *x*. The variable *x* can be seen as *added to d* or as *taken away from e* because of the accident or disease. We can distinguish the average age-specific life expectation *e*[*a*] versus some potentially other value *e* due to treatment.

The LifeGain format thus can also represent a *burden of disease calculation*. Namely, the burden of a particular disease is the loss in life expectancy, *burden* = *e*[*a*] - *d* = *x,* where *d* is the remaining expectancy after incurring the disease. Prevention of the accident or disease can be seen as an intervention that "restores" the original life expectancy that otherwise would have been lost. The LifeGain format is more general than for only *e*[*a*], since it also allows for some differing *e.*

It appears that most common and relevant criteria use these variables. Some are:

```
Absolute    = x = e - d              (life-years gained)
Discounted = PV[x, r] / (1+r)^d
Relative    = x / (a + d)
Unit        = x / (a + d + x)        = (e - d) / (a + e)
Valued      =  u[x] / u[a + d + x]   (becomes UnitSqrt when u = Sqrt)
Marginal    = (u[e] - u[d]) / u[e]   (becomes proportional shortfall when u = Identity)
```

The absolute measure gives the level of life-years gained *x* = *e* - *d*. For the example agents of age 16 and 60, with their addition of 6 years, there is an equal score, namely 6. The absolute measure does not explicitly refer to the age of the person. In practice it relies on it, since often *e* = *e*[*a*].

The discounted value of *x* is the present value of the constant stream of *x* years, PV[*x, r*], subsequently discounted for the period *d* that precedes it.

The relative measure is *x* / (*a* + *d*). The relative measure has the interpretation that a person's life expectancy after the accident or disease would be *a* + *d* if nature had its course, and the relative increase of life with *x* towards *a* + *e* is taken with respect to that origin. For the relative measure with *d* = 0, an increase of 6 years involves a 37.5% increase for a 16-years old but only a 10% increase for a 60-years old. One 16-years old then is about as much as four 60-years olds. The relative measure also implies a dependence upon sex, since women will tend to have higher percent-

age increases.

An approach is to take each life as a whole, and to allocate 100% to each life. The relative addition to each life, or the share of the gain in total life, then is the unit measure $x / (a + e)$, where $a + e$ gives the individual life expectancy. For example with $d = 0$, for the 16-years old the score is 6 / 22 = 27% and for the 60-years old the score is 6 / 66 = 9%.

The different denominators $(a + d)$ and $(a + e)$ depend upon age and are not fixed at birth. This is especially relevant since life expectancy tends to differ with age. One might want to impose a norm that the relative impact of an intervention is related to the life expectation at birth e^*, giving x / e^*. This however has the consequence that older people receive preferential treatment, since $e^* < (a + e[a])$. Making the denominator dependent upon the individual life, has the benefit of imposing more equality amongst the different ages.

Comparing these measures, one cannot escape the impression that the unit measure still is too sensitive to the young. Adding 6 years to a 60-year old or to a 16-year old is neutral when considering the absolute measure, but is highly favourable to the 16-year old when considered relatively, even when unitized. The absolute and relative measures do not yet account for the utility function $u[.]$.

The valued measure $u[x] / u[a+d+x]$ expresses the idea that one refocuses on x. The numerator values the addition to life in a way that is independent of age, while it takes account of diminishing marginal utility. The denominator values the total duration of life. This approach seems theoretically most satisfying. The value is always in the range [0, 1]. With the square root, the measure becomes $\sqrt{x / (a + d + x)}$ which is the square root of the unit measure. This expression will be called the UnitSqrt measure. The UnitSqrt measure orders events in the same manner as the Unit measure, but the relative differences are smaller. For $x = 6$ and $d = 0$, for the 60-years old the score is 30% and for the 16-years old the score is 52%. The age groups are not equal, as in the absolute measure, but they are not so unequal as in the other measures.

Since the UnitSqrt measure takes a life as 100%, it can be weighted by the numbers of people in the age groups, rather than by more complex weights such as the number of (quality adjusted) life-years in the age groups.

The marginal addition to utility that is age-sensitive is $(u[a+d+x] - u[a+d])$, and this measure is biased against old age due to diminishing marginal utility. An application is in "*fair innings*". The unitised measure (i.e. the addition as a share of the total) is $(1 - u[a+d] / u[a+d+x])$, and this enhances the bias since the old age has a higher denominator. In case of the square root, the unitised measure is $(1 - \sqrt{(a + d) / (a + d + x)})$ and it seems much like the percentage increase.

A final measure of success to mention here considers only the remaining life-years. Any memory of age is forgotten here, except for the lingering impact when $e = e[a]$. Rather than taking the increase e / d, it is useful to be more sex-neutral and normalise to 100%, which gives $1 - d / e = x / (d + x)$, which measure is known as "*proportional shortfall*". It is more general to incorporate the notion of marginal utility and time preference, which gives $(1 - u[d] / u[e])$. The measure gives the utility share of x in the total remaining utility (assuming lack of memory of age).

Life expectancy is a key element in above criteria. Thus we arrive at life table methods.

7.11.2. Life table and SARS-CoV-2

7.11.2.1. Life table

It was already mentioned in section 1.1.2 that actuarial methods pertain to survival curves that are closely related to the mathematics for the SI(EY)A(CD) models, see Weisstein (2020ab).

The first table below reproduces a calculation on the WHO 2000 standard population, revised to a cohort of 100,000. The life expectancy of the 85+ group is managed by the options. Discounting of the life expectancy is done with $r = 3\%$ default. The UnitSqrtLE[] = UnitSqrt[a, 0, $e[a]$] = $\sqrt{e[a] / (a + e[a])}$, i.e. saving a person from immediate death $d = 0$ and restoring $e[a]$. PM 1. There is no clear way how to include the measure of proportional shortfall $1 - d / e = x / (d + x)$ in this life table; with $d = 0$ the values are all 1. PM 2. The table headings apparently require dots too, when elements are aligned at the decimal dot.

When we compare the scores for life expectancy (LE), discounted LE, and the UnitSqrt with LE, then the ratio of the first / last elements for LE is 5.48, discounting reduces this to 3.13, and the latter happens to be similar for UnitSqrt at 3.08. According to the latter two criteria for this dataset, saving one baby / toddler may cost as much as saving three persons of 85+.

```
lab = "WHO 2000"; (* the following assumes the evaluation in Appendix E *)

First[#] / Last[#] & /@ (#[lab] & /@ {LifeExpectancy, DiscLifeExp, UnitSqrtLE}) // N
```

{5.48296, 3.1349, 3.08221}

	Living	q	Death	LifeYears	.MortalityRate	Life.Exp	Disc.LE	Unit.SqrtLE
0–4	100 000.	0.01136	1136.	497 160.	0.00228498	54.8	27.1	1.
5–9	98 864.	0.0115006	1137.	491 478.	0.00231343	50.4	26.2	0.953833
10–14	97 727.	0.0116242	1136.	485 795.	0.00233843	46.	25.1	0.906307
15–19	96 591.	0.0352931	3409.	474 433.	0.00718543	41.5	23.9	0.857037
20–24	93 182.	0.0365843	3409.	457 388.	0.0074532	37.9	22.8	0.809158
25–29	89 773.	0.0379736	3409.	440 343.	0.0077417	34.3	21.5	0.760405
30–34	86 364.	0.0526377	4546.	420 455.	0.0108121	30.5	20.1	0.710174
35–39	81 818.	0.0833313	6818.	392 045.	0.0173909	27.1	18.6	0.660486
40–44	75 000.	0.0909067	6818.	357 955.	0.0190471	24.3	17.3	0.614891
45–49	68 182.	0.0999971	6818.	323 865.	0.021052	21.5	15.9	0.568601
50–54	61 364.	0.148149	9091.	284 093.	0.0320001	18.6	14.3	0.520819
55–59	52 273.	0.195665	10 228.	235 795.	0.0433767	16.4	13.	0.479406
60–64	42 045.	0.189178	7954.	190 340.	0.0417884	14.8	12.	0.444783
65–69	34 091.	0.266669	9091.	147 728.	0.061539	12.7	10.6	0.403842
70–74	25 000.	0.3182	7955.	105 113.	0.0756808	11.4	9.7	0.373714
75–79	17 045.	0.4	6818.	68 180.	0.1	10.5	9.	0.350438
80–84	10 227.	0.333333	3409.	42 612.5	0.08	10.8	9.3	0.345349
85+	6818.	1.	6818.	68 180.	0.1	10.	8.7	0.324443

The following table focuses on adjustment for disability and quality of life. The disability scores have been chosen arbitrarily. Conventionally, DALY = YLL + YLD, with the distinction between the Years of Life Lost (YLL) and Years Lost due to Disability (YLD). For a life table, the deaths relevant for the YLL are already included in the deaths, and thus not specified here. (For the European burden of infectious diseases the YLL are 71%, see Cassini et al. (2018:3&7).) The vector DALY[] that is used here, with arbitrarily chosen disability factors, only concerns YLD. The values of (1 - disability) LifeYears[] can be cumulated and generate the DALE. (An alternative for YLL is LE[YLL = 0] - DALE,

using a life table LE in which each YLL death is recorded at the otherwise expected age.) It has also been used here that qol = Sqrt[1 - disability], to allow for some accommodation to the disability. The UnitSqrtQALE[] = UnitSqrt[QAAge[*age*], 0, QALE[*age*]], with calendar *age* as an index, saving a person from immediate death and restoring QALE[*age*]. The disability and lower quality of life of the later ages causes a reduction of their value. Again the UnitSqrt measure treats lives as most equal with the lowest first / last ratio.

First[#] / Last[#] & /@ (#[lab] & /@ {DALE, QALE, DiscQALE, UnitSqrtQALE})

{16.7564, 10.2439, 5.58865, 4.07843}

Rather than these first / last weights, we should use the averages of ages <60 and 60+.

	Living	LifeExp	Disability	DALE	QoL	QALE	DiscQALE	QAAge	UnitSqrtQALE
0–4	100 000.	54.8	0.046	51.1	0.977	52.8	26.7	0.	1.
5–9	98 864.	50.4	0.036	46.9	0.982	48.5	25.8	4.9	0.953187
10–14	97 727.	46.	0.034	42.6	0.983	44.2	24.7	9.8	0.904696
15–19	96 591.	41.5	0.04	38.2	0.98	39.7	23.4	14.7	0.854305
20–24	93 182.	37.9	0.04	34.7	0.98	36.2	22.2	19.6	0.805387
25–29	89 773.	34.3	0.044	31.2	0.978	32.6	20.9	24.5	0.755456
30–34	86 364.	30.5	0.05	27.5	0.975	28.9	19.4	29.4	0.703975
35–39	81 818.	27.1	0.05	24.2	0.975	25.5	17.9	34.3	0.653002
40–44	75 000.	24.3	0.054	21.4	0.973	22.7	16.5	39.1	0.605827
45–49	68 182.	21.5	0.06	18.6	0.97	19.9	15.	44.	0.557639
50–54	61 364.	18.6	0.068	15.7	0.965	16.9	13.3	48.9	0.507516
55–59	52 273.	16.4	0.084	13.4	0.957	14.6	11.9	53.7	0.463027
60–64	42 045.	14.8	0.102	11.5	0.948	12.8	10.7	58.5	0.424421
65–69	34 091.	12.7	0.132	9.1	0.932	10.6	9.1	63.2	0.378235
70–74	25 000.	11.4	0.19	7.3	0.9	8.9	7.8	67.9	0.340218
75–79	17 045.	10.5	0.27	5.7	0.854	7.5	6.7	72.4	0.306072
80–84	10 227.	10.8	0.36	4.7	0.8	6.8	6.1	76.6	0.284906
85+	6818.	10.	0.695	3.	0.516	5.2	4.8	80.6	0.245193

LifeTableEquations[]

$$\Big\{ \text{Living}(t) = \text{Birth}(0)\,\text{Survival}(t),\ \text{Death}(t, n) = \text{Living}(t) - \text{Living}(n + t),$$

$$\text{MortalityQuotient}(t, n) = \frac{\text{Death}(t, n)}{\text{Living}(t)},\ \text{PeriodHazard}(t, n) = -\log(1 - \text{MortalityQuotient}(t, n)),$$

$$\text{LifeYears}(t, n) = \int_t^{n+t} \text{Living}(\tau)\,d\tau,\ \text{LifeYears(Line)}(t, n) = n\Big(\text{Living}(t) - \frac{1}{2}\text{Death}(t, n)\Big),$$

$$\text{MortalityRate}(t, n) = \frac{\text{Death}(t, n)}{\text{LifeYears}(t, n)},\ \text{LifeExpectancy}(t) = \frac{\text{LifeYears}(t, \infty)}{\text{Living}(t)}\Big\}$$

TheFormer /. ToLifeTableSymbols // MatrixForm

$$\begin{pmatrix} l(t) = b(0)\,S(t) \\ d(t, n) = l(t) - l(n + t) \\ q(t, n) = \frac{d(t,n)}{l(t)} \\ h(t, n) = -\log(1 - q(t, n)) \\ L(t, n) = \int_t^{n+t} l(\tau)\,d\tau \\ L(\text{Line})(t, n) = n\Big(l(t) - \frac{1}{2} d(t, n)\Big) \\ m(t, n) = \frac{d(t,n)}{L(t,n)} \\ e(t) = \frac{L(t,\infty)}{l(t)} \end{pmatrix}$$

7.11.2.2. Life table Holland 2018

This is the abridged table for Holland 2018, copied from Appendix E.

LifeTable["Holland 2018 abridged"]

	Living	q	Death	LifeYears	MortalityRate	LifeExp	DiscLE	UnitSqrtL
0–4	100 000.	0.00394	394.	448 574.	0.000878338	81.8	30.8	1.
5–9	99 606.	0.000381503	38.	497 802.	0.0000763356	77.6	30.4	0.969261
10–14	99 568.	0.000311345	31.	497 946.	0.0000622558	72.6	29.9	0.937539
15–19	99 537.	0.000924279	92.	497 482.	0.000184931	67.6	29.3	0.904716
20–24	99 445.	0.00133742	133.	496 956.	0.00026763	62.7	28.5	0.870742
25–29	99 312.	0.0016715	166.	496 400.	0.000334408	57.8	27.7	0.835483
30–34	99 146.	0.00212817	211.	494 950.	0.000426306	52.9	26.7	0.798768
35–39	98 935.	0.00289079	286.	494 146.	0.000578777	48.	25.6	0.760436
40–44	98 649.	0.00443998	438.	492 264.	0.000889767	43.1	24.4	0.720256
45–49	98 211.	0.00733115	720.	489 683.	0.00147034	38.3	22.9	0.678074
50–54	97 491.	0.0121037	1180.	484 558.	0.00243521	33.6	21.3	0.633741
55–59	96 311.	0.0210672	2029.	476 389.	0.00425913	28.9	19.4	0.587171
60–64	94 282.	0.0347999	3281.	464 442.	0.0070644	24.5	17.4	0.53854
65–69	91 001.	0.0537247	4889.	442 785.	0.0110415	20.3	15.3	0.487744
70–74	86 112.	0.0867823	7473.	413 561.	0.0180699	16.3	12.9	0.434599
75–79	78 639.	0.141151	11 100.	367 355.	0.030216	12.6	10.5	0.379128
80–84	67 539.	0.245399	16 574.	298 572.	0.0555109	9.2	8.1	0.321465
85–89	50 965.	0.430433	21 937.	201 059.	0.109107	6.4	5.8	0.263846
90–94	29 028.	0.65268	18 946.	94 546.7	0.200388	4.2	4.	0.212112
95–99	10 082.	0.747173	7533.	22 924.3	0.328604	2.8	2.7	0.170082
100+	2549.	1.	2549.	5607.8	0.454545	2.2	2.1	0.147442

For this table we do not have an adjustment for disability and quality of life yet. Given SARS-CoV-2 and the distinction between <60 and 60+ age groups, it still seems relevant to determine the weighted averages of (discounted) life expectancy and UnitSqrtLE for these two groups. Life expectancy concerns years, and the weights will be the LifeYearsSum[] = Living[] * LifeExpectancy[]. The UnitSqrt[] measure takes each life as 100%, and the weights will be the Living[]. Currently we will use cohort weights but there is a good argument that the actual population is used. Appendix E will actually using the non-abridged table of length 100 for the weighted averages. The outcomes are:

	LifeExpectancy	DiscLifeExp	UnitSqrtLE
age < 60	58.98	26.58	0.79
age ≥ 60	16.97	12.79	0.39
All	54.98	24.56	0.68
Ratio	3.48	2.08	2.02

When we compare the <60 and the 60+ groups, then, in terms of life expectancy, saving a person from the younger group may cost as much as saving 3.5 persons from the elderly group. Discounting by 3% reduces that ratio to 2.08. The UnitSqrtLE measure that takes each life as 100% reduces the ratio a bit more to 2.02.

7.11.2.3. Life table and SARS-CoV-2

Within the SI(EY)A(CD) context, the level of current infections (*I* or *Y*) is the prevalence. The sum of those over time is the cumulated prevalence (*A*). The death toll after the epidemic can be found with the cumulated prevalence. For the life table the cumulated prevalence is important. In the following discussion the reference will be to the (cumulated) prevalence, in which discussion the

adjective may often be dropped.

Marois et al. (2020) at IIASA assess the potential impact of Covid-19 on life expectancy. For Northern Europe (Holland) a prevalence around 2% would hardly have an impact, of 10% might cause a loss in life expectancy of above 1 year, a prevalence of 50% a loss in the range of 3-9 years. This is roughly linear with 1 year of life expectancy per 10% points of prevalence. Some aspects of their model seem needlessly complex. They have a probability distribution over the year for getting infected, and the probability of death only applies to the infectious period of two weeks. For now it seems that the prevalence with the infection fatality factor (IFF) would be sufficient. However, it may well be that SARS-CoV-2 is like the common cold, without immunity, and with the possibility of getting it 4-6 times per year, see Worrall (2011). Subsequently, they allow for high prevalences of even 70%, so that the impact on life expectancy can be large.

A necessary distinction exists between a (dynamic) population model and a life table cohort. In a (dynamic) population model, the epidemic might run a few years and then be resolved by vaccination and the evolutionary pressure upon the virus (so that it becomes less deadly). A cohort model represents a stationary population instead. The disease-specific effect per age-group in the current life table setup means that the infection returns each year, without immunity, with a high likelihood of getting the disease at least once over one's life. The assumed IFF of 1.5% in the aggregate and a cumulated prevalence of 100% mean that the cohort dwindles on average by 1.5% per 1-year age group. With this rate, eventually 78% would succumb to the virus. In the actual calculation it appears that the effect is larger since the IFF is higher for the higher age groups.

```
1 - (1 - .015) ^100
```

0.779391

In a life table, the number of deaths remains the same, equal to the number of births, here 100,000. Deaths only shift up or down the ages. All reports indicate that SARS-CoV-2 works via comorbidity. The virus has an accelerating effect on mortality (rather than adding to the 100,000). Its victims thus can no longer die from the normal effect of comorbidity later on. What was "all-cause mortality" in 2018 becomes "other-cause mortality" in 2020, and this will be reduced for the higher age groups.

Assume this scenario:

- Immunity is lost after a year
- A cumulated prevalence p is infected *each year* and they suffer the disease & age specific fatality, *year after year*. The specific mortality quotient is initially $q = p$ IFF for the first year (with different values per age group). The worst scenario is that $p = 1$.
- SARS-CoV-2 also is a nasty virus. For survivors it leaves scars and can negatively effect future survival prospects. Let us assume a 10% fraction that enhances the IFF in the subsequent year and 10% in the second year (as a proxy for effects much later in life). Thus we get $q = p$ (IFF + 0.1 IFF[t-1] + 0.1 IFF[t - 2]).
- Given that these deaths are attributed to the virus rather than the underlying comorbidity, and since the disease returns every year, eventually a large part of all deaths will be attributed to the virus.
- Other-cause deaths become $d_{OC,tot}$ = 100,000 - $d_{SARS-CoV-2}$. The life table column of all-cause deaths can be proportionally reduced to this number, at least from the age when SARS-CoV-2 starts

making victims. The assumption is that the virus accelerates deaths that would otherwise arise later because of comorbidity anyway. (An alternative assumption, not looked into here, is that the other-cause deaths would continue independently, so that SARS-CoV-2 adds per age group, with the effect that the population dwindles down even faster, and that the virus wipes out the last age groups.)

In the first half of 2020, Holland managed to keep the infection with SARS-CoV-2 below 10%. By comparison, the common cold corona viruses may have a prevalence of some 25% for the preceding two months, see Worrall (2011). Let us now consider a scenario in which discipline becomes lax and the prevalence rises to 15% per annum. Subsequently, we will allow *p* be a parameter, and use the symbolic manipulation of *Mathematica* to plot outcomes as functions of *p.*

Our result can be found in section 7.11.2.5. For life expectancy, it is remarkable that the outcome is quite linear. In the life table computation, a 10% rise in (annual cumulated) prevalence means about a 0.5 year drop in life expectancy. This can be explained by the phenomenon that the fatalities are mostly in the higher age groups. This outcome is much lower than the 1% average drop size stated by Marois et al. (2020), though it may be their lowest outcome in their range. It is not clear to me now how the difference arises. The method followed here is the straightforward life table method, using the disease specific mortality quotient $q = p$ IFF and turning the 2018 all-cause mortality into the 2020 other-cause mortality in proportional manner. Our assumption that the disease returns every year without immunity also is quite strong, i.e. for the numerical outcome though perhaps not strong in terms of realism.

7.11.2.4. Survival curves 2018, 2020 with p = 0.15, and other causes

In the following survival plot, the blue line gives the situation in 2018. The red line gives the all-cause reduced survival due to SARS-CoV-2 with a prevalence of 15%. The orange line is the other-cause in 2020, which is the all-cause of 2018 but now shifted because of the earlier deaths by the virus. The incremental formula is $S_{OC} = S_{All} / S_{Specific}$.
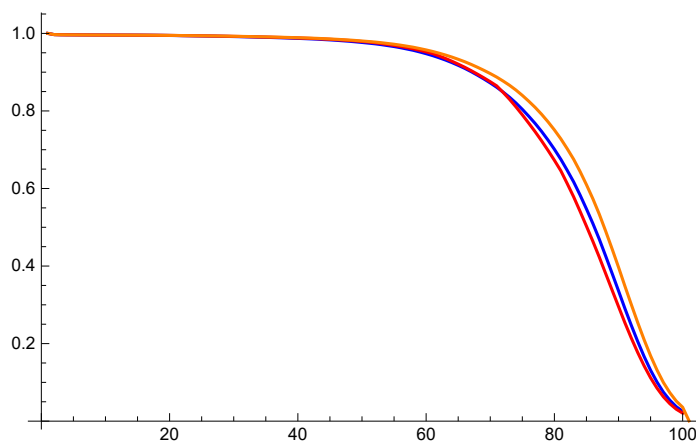
— All causes before SARS–CoV–2   — All causes incl. SARS–CoV–2
— Other causes excl. SARS–CoV–2

The disease-specific death toll was 20,271 on a cohort of 100,000, within a stationary population of 8.27 million. Even for a prevalence of 15% the effect on life expectancy is relatively small because of

the concentration in the higher age groups. At birth the LE is reduced from 81.8 to 81.1, a difference of 0.66. At age 50, LE is reduced from 33.6 to 32.7, a difference of 0.80.

For the UnitSqrtLE = UnitSqrt[a, 0, e] the effect is even smaller given that the effect on life expectancy was small to start with. The curves for the two situations (almost) overlap. Decisions that use this measure will hardly be in need of revision because of the new situation of this epidemic.

UnitSqrtLE
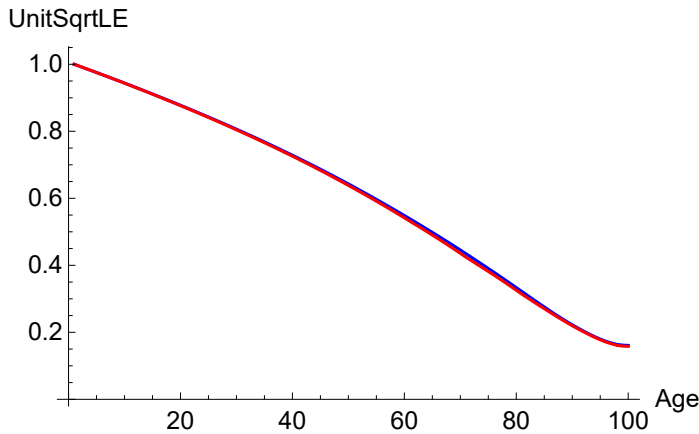
7.11.2.5. Routine with symbolic parameter

The advantage of *Mathematica* is that above calculations can also be put into a routine that uses p as a symbol.

When prevalence p rises to 100% then at some point the number of deaths cannot be larger than the birth cohort itself (and the life table becomes shorter because of empty age groups).
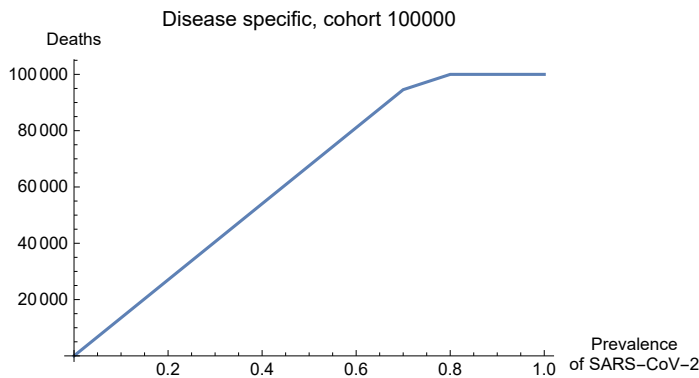
For life expectancy, it is remarkable that the outcome is quite linear. In this computation, a 10% rise in prevalence means about a 0.5 year drop in life expectancy, which is much lower than the 1% drop size stated by Marois et al. (2020). See above for a discussion of the difference.

Life expectancy at birth drops from 81.8 to 77.4, a difference of 4.4.

At birth
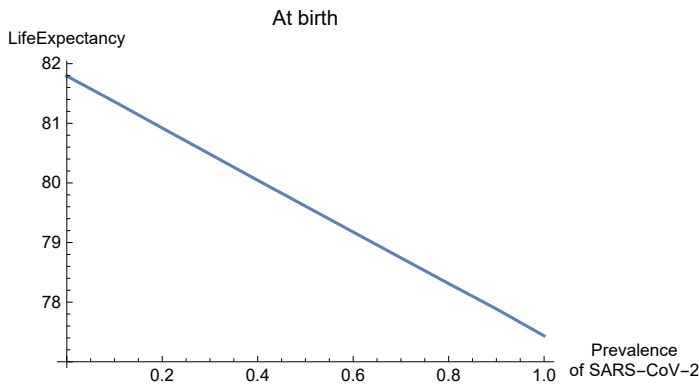
LifeExpectancy

Life expectancy at 50 drops from 33.6 to 28.3, a difference of 5.3.

At age 50

LifeExpectancy
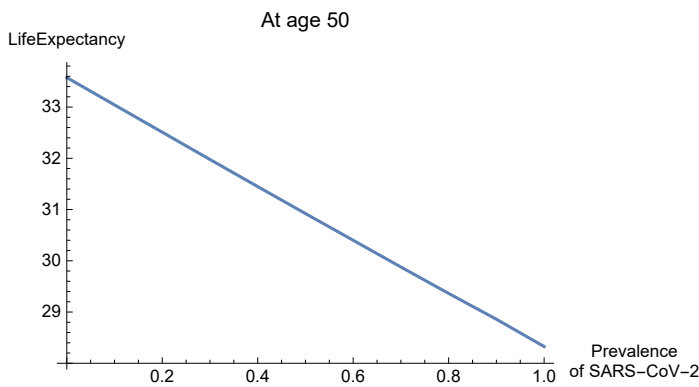
For each value of *p* we can plot the UnitSqrt[*a*, 0, *e*[*a*]] = $\sqrt{e[a] / (a + e[a])}$ curve. For the higher ages, the difference is of the order of 0.05.

Prevalence in steps of 0.1

UnitSqrtLE
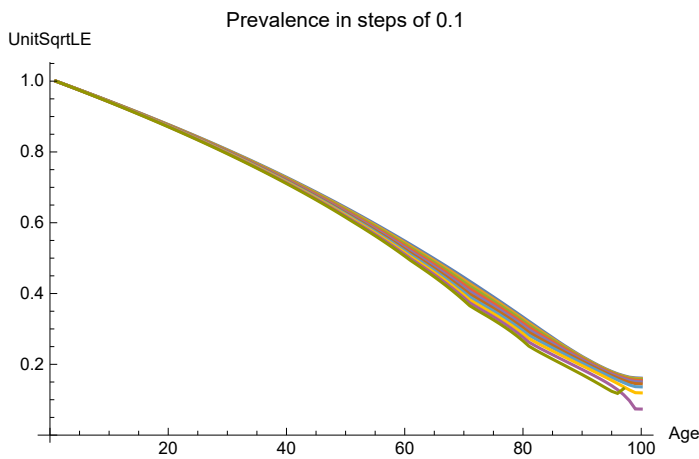
7.11.3. Optimal allocation, micro and macro

The present discussion concerns the Public Health discussion of optimal allocation of resources at the macro or aggregate (national) level from the viewpoint of social welfare. The discussion by Colignatus (2003, 2020) *On the value of life* and the notion of the UnitSqrt life gain measure pertain to that perspective. However, there can be interaction between the macro and micro level. Presumably macro rules might cause micro implementation, and general developments at the micro level show up as general developments at the macro level.

Because of the focus on intensive care units (ICU) for SARS-CoV-2, Colignatus (2020b) makes some comments about the bedside micro-level. The Dutch medical specialists and the general medical

association issued the report FMS & KNMG (2020) on "stage 3 C" when "non-medical" considerations might be considered for the ICU triage. The "medical" stages of "3 A & B" already pertain to a survival probability that can be empirically related to age, sex or social economic status (SES). It may be suggested that the UnitSqrt measure might also be used at the micro level as one of the indicators to assess the value of treatment in a way that is more neutral to age and sex and other factors that affect life expectancy. This type of decision making relies much upon digital information handling but such patient dossiers tend to be available nowadays. Some general rules obviously cannot replace the acumen that is required in the hectic of decisions on actual patients.

One remarkable comment in the FMS & KNMG (2020) report is that a patient with a shorter (expected) use of an ICU bed would be preferred to another patient with a longer (expected) use, since this "would allow to save more lives". It so happens that SARS-CoV-2 patients of age <60 may require 10 days of ICU bed time and that such patients of age 60+ may require 20 days, so that this rule favours the young. One would be hard put though if some young would require 20 days to survive and some elderly would require 10 days to die even when treatment is started in the hope that they would survive.

The presently discussed SI(EY)A(CD) models do not include the phase of hospitalisation and ICUs, which phase is important for the actual recovery from the infection and disease. A first step in such modeling is the distinction by age groups. The following may highlight this aspect from above discussion.

Consider an Intensive Care Unit (ICU) with 10 beds. A young person Y requires 10 days to recover and an elderly person E requires 20 days to recover. Let us consider a month of 30 days, and divide this month in 3 periods of 10 days. For the first 10 days, we can allocate Y_1 and E_1 persons from each group. For the second week, 10 - E_1 beds remain, and we can allocate Y_2 and E_2. For the final week, 10 - E_2 beds remain, and we can allocate those to Y_3.

In the above, it was a suggested that one person of age <60 is equivalent to 2 persons of 60+, in terms of discounted QALEs or UnitSqrtLE. The value of an allocation thus is 2 (Y_1 + Y_2 + Y_3) + E_1 + E_2. The subsequent objective is to maximize this value. This gives the *Linear Programming* problem Max $w.x$ subject to $A.x \leq c$ and $x \geq 0$.

Max $2 Y_1 + E_1 + 2 Y_2 + E_2 + 2 Y_3$

S.t. $Y_1 + E_1 \leq 10$

$\quad\quad E_1 + Y_2 + E_2 \leq 10$

$\quad\quad\quad\quad E_2 + Y_3 \leq 10$

$\quad Y_i, E_i \geq 0$

```
w = {2, 1, 2, 1, 2}; (* weights in the objective function *)
c = 10 {1, 1, 1};   (* constraints *)
A = {{1, 1, 0, 0, 0},
     {0, 1, 1, 1, 0},
     {0, 0, 0, 1, 1}};
```

The linear programming solution is found as follows. Given the high appreciation of the youngsters, i.e. double as valuable as the elderly, they would be given priority.

```
NumLP[w, A, c]
```

{Solve → 60, Variables → {10, 0, 10, 0, 10}, Slack → {0, 0, 0}, Basis → {1, 3, 5}, ShadowPrices → {2, 2, 2}}

The elderly are disadvantaged because of their greater use of resources. The weight in the objective function provides little compensation. The "ranges of optimality" are the ranges for which each coefficient in the objective function can change without affecting the solution, given the values of the other coefficients. The value of the elderly must rise to 3 before they enter into the solution. Given that the elderly are already so much lower valued than the youngsters we can hardly expect that they rise above them.

```
RangeOfOptimality[]
```

$$\begin{pmatrix} 0 & \infty \\ 0 & 3 \\ 0 & \infty \\ 0 & 3 \\ 0 & \infty \end{pmatrix}$$

This problem definition is deficient, of course, since the health care system already takes account of the conventional appreciation of youngsters. Health care capacity has been built with the purpose to look after the elderly too. Thus in practice there is an additional restriction $Y_1 + Y_2 + Y_3 \leq Y^*$, and resources above this are intended for the elderly. The SARS-CoV-2 epidemic causes an additional inflow of Y and E, which causes the allocation question to pop up again. The answer could have been that the capacity above Y^* was built on purpose for the elderly. It remains a moot point that such capacity might still be used to save more youngsters because of their higher LifeGain score. In addition, there is the competition between the usual elderly and the new elderly patients because of SARS-CoV-2 (who are the same persons who would have shown up a few year later anyway). A key point is that one might wish to help infected patients, since, if they are not helped, they might remain outside of the health care system and infect more people. If the virus remains with us then capacity must be enlarged or made flexible towards enlargement, while the other options are containment and eradication.

The discussion *On the value of life,* Colignatus (2003, 2020), distinguishes between private and national health insurance. When one is young and healthy then such insurance seems fine just in case, even when it isn't used, since it better not to have a disease than be in need of using the insurance. When one grows older and more sickly then using the insurance feels like getting what one has been paying for. The sad moment is when one hasn't really used the insurance and wants to use it for serious illness, but has become too old and sickly for it. In such cases one has paid one's whole life but receives no benefit (other than being insured all the time). In private insurance, the balancing of net contributors and net users is left to the market. In national insurance, this requires explicit discussion and agreement in Parliament. For general acceptance of national insurance it seems unavoidable that a bonus / malus system is used that also considers health behaviour and actual use. Infectious diseases however remain the key reason why also the rich have a good motive to also protect the poor.

7.11.4. Conclusions

This section looked at a wider perspective for the *lives saved* (lives extended) measure as used in SI(EY)A(CD) modeling. A wider perspective has been given in the paper *On the value of life,* Coligna-

tus (2003, 2020). The present discussion cannot replace the latter paper, and one is advised to consider it. We discussed some aspects relating to SARS-CoV-2. **Appendix E** applied the age specific infection fatality factors (IFF) in a life table.

Conclusions are:

(1) *Life expectancy* ($e[a]$) appears to decrease linearly at the average rate of 0.5 year per 10% increase in the (cumulated) prevalence of the virus, assuming the annual loss of immunity and the recurrence of the virus every year. The linear change occurs in all age groups, and is steeper in the higher ages. If the infection is evenly spread over the population then mortality is proportional to the life-years lost and change in life expectancy. For rule-of-thumb reasoning, the SI(EY)A(CD) model doesn't have to be extended with different age groups (e.g. <60 and 60+), but it obviously must be for scenarios that pertain to such age groups (such as above consideration of stopping hospital treatment because of a bad prognosis related to age).

For herd resistance: if cumulated prevalence would be 60% as RIVM suggested, then life expectancy would reduce with 3 years. If it would be Ap[∞ | R_0 = 4] = 98% then life expectancy would reduce by nearly 5 years. The annual death toll then would be seen as an acceleration of the mortality by comorbidity.

(2) The *UnitSqrt* measure ($\sqrt{(e[a] - d) / (a + e[a])}$ here using $e = e[a]$ and $d = 0$) is hardly affected by the impact of the virus. The measure originated from the philosophy that each life is 100%, i.e. similar to the 1 of the *life saved* but with the possibility of some grading within. The measure appears to have a relevant position in the life table along (discounted) (quality adjusted) life expectancy. The measure is neutral to age and sex, and apparently also neutral to an epidemic. When persons die because of an epidemic then their lives are still counted as 100% even when the duration is shorter. Policy decisions about Public Health that are based upon the measure apparently would not need to be revised because of the epidemic. The measure helps to distinguish such issues from those that are required for the evaluation of an epidemic (for which the loss of life expectancy would be a consideration). A result of the UnitSqrt measure is that saving the life of a <60 youngster is twice as valuable as saving the live of an elderly 60+, since the proportion of restored life expectancy is so much larger: but this conclusion is not affected by the epidemic (except that SARS-CoV-2 causes us to choose these age groups).

(3) The *Proportional Shortfall* measure (($e[a]$ - d) / $e[a]$), used by the Dutch national health insurance ZIN (2018ab), appears to have no added value for this discussion, since one either would have died d = 0 or doesn't die $d = e[a]$, so that the measure reduces to the *lives saved* criterion.

? LifeGain

Symbol

LifeGain[a, d, x (, u) (, r) (, f)] for

a = age

d = the number of years to death due to a disease

x = the added life expectancy due to treatment

e = d + x = the life expectancy after treatment

u = utility (with default option Utility –> "Sqrt")

r = rate of interest (default taken from Options[LifeGain])

f = an age–dependent Life Expectancy function, such that e = f[a]

gives the treatment score according to the utility

"Absolute" = e – d = x

"Discounted" = PV[x, r] / (1+r)^d

"Relative"   = (e–d) / (a+d) = x / (a+d)

"Unit"   = (e–d) / (a+e) = x / (a+d+x)

"Sqrt"   = Sqrt[e–d] / Sqrt[a+e] = Sqrt[x / (a+d+x)]

"UnitLog"   = Log[1+e–d] / Log[1+a+e] = Log[1+x] / Log[1+a+d+x]

"Marginal" = (u[e] – u[d]) / u[e] for default Utility

other "u"   = u[x] / u[a+d+x] for u = ToExpression["u"]

When f is specified, then x is recalculated as Max[f[a]

    – d, x'] where x' is the original input (when left out, it will get default 0)

Extended with PrShortfall and FairInnings, see there.

"Rescue"   = d   (the lower the more urgent)

"PrShortfall" = x / (d + x)   (Marginal with Identity)

"FI Unit"   = 1 – Min[a + d, norm] / Min[a + d + x, norm]

⌄

Appendix. In steps (for management of details)

Setting the size of the population and other options

The Dutch population in 2020 is about 17.4 million people. It will be convenient to use formal parameter Nqt0 → the size of the population, while Npt0 → 1. Subsequently, for quick calculation of proportions it will be useful to set NqSeed = Nq[t0] at a numerical value.

```
NqSeed = 17.4 × 10^6;
SetOptions[SIA, Nqt0 → NqSeed]
```

$\{$Aqt0 → 0, Cqt0 → 0, DataMold → {S, I, A}, Dqt0 → 0, Iqt0 → Null, Nqt0 → 1.74×10^7, Onset → 0, Seed → 100,

   Yqt0 → 0, AlphaSEYCD() → 1., BetaSIA() → 0.4, GammaSIA() → 0.1, PhiSICD() → 0.015, R0() → Null$\}$

Choosing a model with formal t_0 or with the Onset from the options

We put the model equations into Hold, so that the symbols remain unevaluated when we assign values later on.

We seed the model with Seed → 100 as in the supplied options, but may choose another value later on.

```
model[t0_, t_] = SICD[N, t0, t, Seed → 100]
```

Hold$\big[\{$Sq′$(t) = -2.29885 \times 10^{-8}$ Iq(t) Sq(t), Aq′$(t) = 0.1$ Iq(t), Iq′$(t) = -$Aq′$(t) -$ Sq′(t),

   Dq$(t) = 0.015$ Aq(t), Cq$(t) = 0.985$ Aq(t), Nq$(t) = 1.74 \times 10^7 -$ Dq(t), Nq$(t0) = 1.74 \times 10^7$,

   Sq$(t0) = 1.73999 \times 10^7$, Iq$(t0) = 100$, Cq$(t0) = 0$, Dq$(t0) = 0$, Aq$(t0) = 0\}\big]$

```
MatrixForm @@ TheFormer
```

$$\begin{pmatrix} \text{Sq}′(t) = -2.29885 \times 10^{-8}\ \text{Iq}(t)\ \text{Sq}(t) \\ \text{Aq}′(t) = 0.1\ \text{Iq}(t) \\ \text{Iq}′(t) = -\text{Aq}′(t) - \text{Sq}′(t) \\ \text{Dq}(t) = 0.015\ \text{Aq}(t) \\ \text{Cq}(t) = 0.985\ \text{Aq}(t) \\ \text{Nq}(t) = 1.74 \times 10^7 - \text{Dq}(t) \\ \text{Nq}(t0) = 1.74 \times 10^7 \\ \text{Sq}(t0) = 1.73999 \times 10^7 \\ \text{Iq}(t0) = 100 \\ \text{Cq}(t0) = 0 \\ \text{Dq}(t0) = 0 \\ \text{Aq}(t0) = 0 \end{pmatrix}$$

Solving and approximating by NDSolve for a specified time window

We can run the model with SICD[NDSolve, ...], which gives a numerical approximation.

```
sol = SICD[NDSolve, model[0, t], {t, 0, 150}]
```

$\{$Sq → InterpolatingFunction[⊞ ⎍ Domain: (0. 150.) Output: scalar],

   Iq → InterpolatingFunction[⊞ ⋀ Domain: (0. 150.) Output: scalar],

   Cq → InterpolatingFunction[⊞ ⌐ Domain: (0. 150.) Output: scalar],

   Dq → InterpolatingFunction[⊞ ⌐ Domain: (0. 150.) Output: scalar],

   Nq → InterpolatingFunction[⊞ ⌐ Domain: (0. 150.) Output: scalar],

   Aq → InterpolatingFunction[⊞ ⌐ Domain: (0. 150.) Output: scalar]$\}$

Making the variables directly available

We want to be able to call *Sq*[*t*], *Iq*[*t*], *Cq*[*t*], *Dq*[*t*], *N*[*t*] and SICD[*t*] directly. We use above "sol" to set these functions.

```
SICD[Set, sol];
```

```
SICD[75]
```

$\{415\,556., 738\,924., 1.60018 \times 10^7, 243\,683., 1.71563 \times 10^7, 1.62455 \times 10^7\}$

The values for S & I are the same with SIA, and thus also A = C + D.

Plotting in levels and proportions

Options[Plot] better have some default settings for the layout of the SICD plots.

```
sicdplotoptions = SICD[Plot, Options];
```

```
p0 = Plot @@ {Drop[SICD[t], -1], {t, 0, 150}, sicdplotoptions}
  (*Drop the last element A*)
```
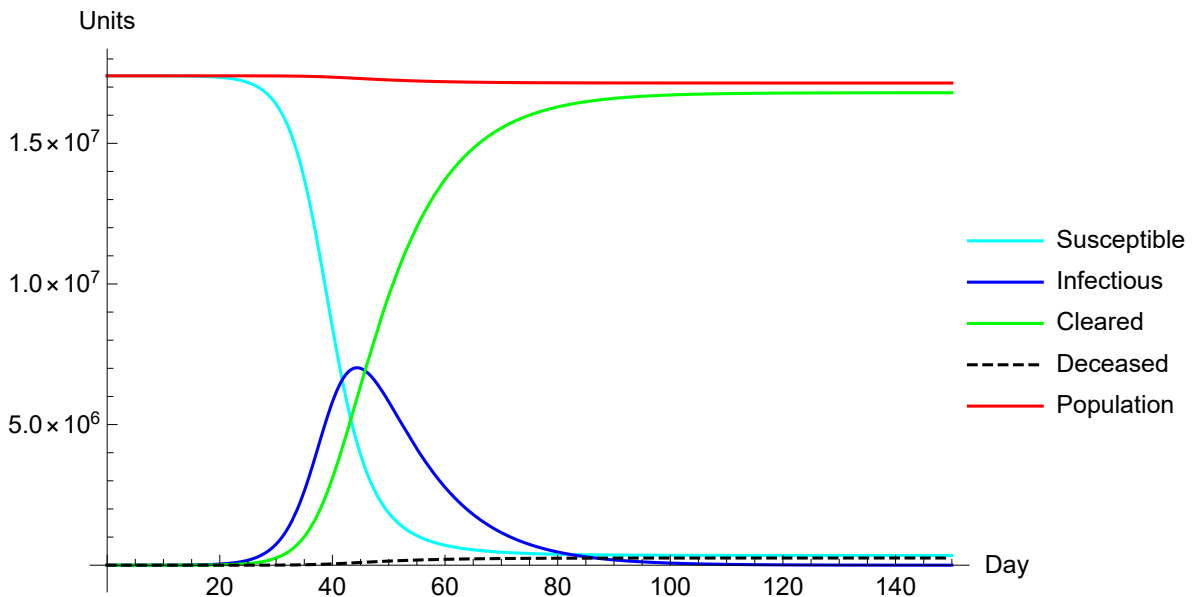

Given our earlier setting of NqSeed, it is easy to plot for the proportions. An axis with proportions indeed helps to focus on the proportions.

```
Plot @@ {Drop[SICD[t], -1] / NqSeed, {t, 0, 150}, sicdplotoptions}
```
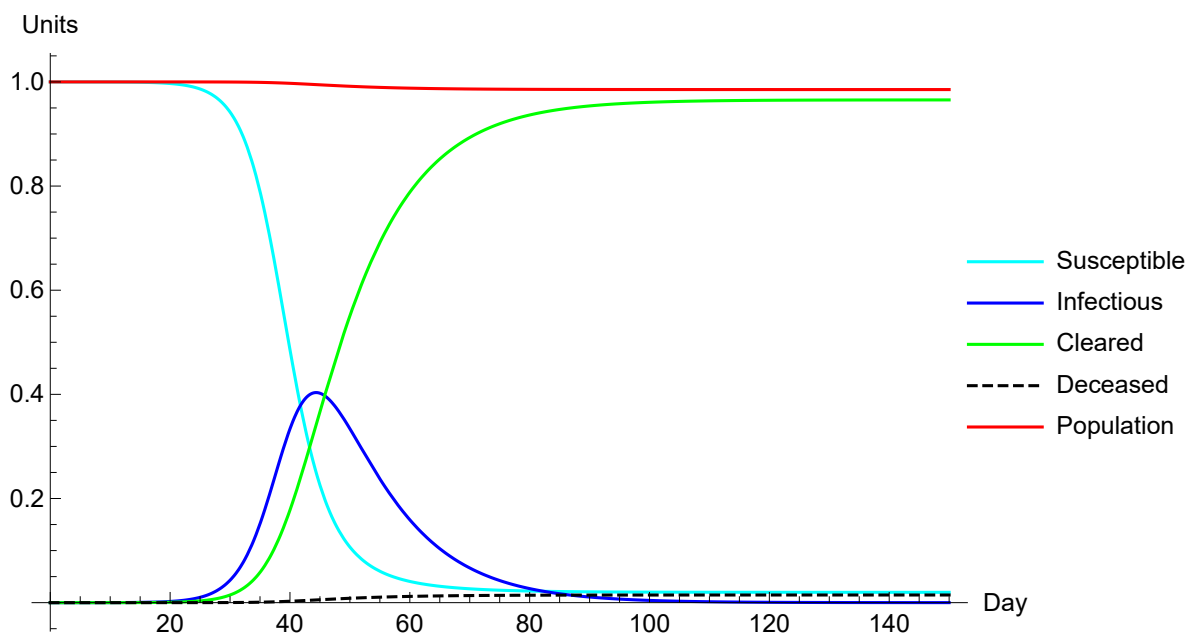

Recapitulation

Above steps to run the model are combined in this routine, for the levels only. Let us use another label for the model.

```
SICD[Run, newModel, t, 150]
```

$\Big\{$InterpolatingFunction$\Big[$ ⊞ ⌐ Domain: (0. 150.) Output: scalar $\Big]$(t$), InterpolatingFunction$\Big[$ ⊞ ⋀ Domain: (0. 150.) Output: scalar $\Big]$(t$),

InterpolatingFunction$\Big[$ ⊞ ⌐ Domain: (0. 150.) Output: scalar $\Big]$(t$), InterpolatingFunction$\Big[$ ⊞ ⌐ Domain: (0. 150.) Output: scalar $\Big]$(t$),

InterpolatingFunction$\Big[$ ⊞ ⌐ Domain: (0. 150.) Output: scalar $\Big]$(t$), InterpolatingFunction$\Big[$ ⊞ ⌐ Domain: (0. 150.) Output: scalar $\Big]$(t$)$\Big\}$

Not surprisingly, the model and the outcomes are the same as above.

```
newModel[t]
```

Hold$\Big[\{$Sq$'(t) = -2.29885 \times 10^{-8}$ Iq(t) Sq(t), Aq$'(t) = 0.1$ Iq(t),

Iq$'(t) = -$Aq$'(t) - Sq'(t)$, Dq$(t) = 0.015$ Aq(t), Cq$(t) = 0.985$ Aq(t), Nq$(t) = 1.74 \times 10^{7} - Dq(t)$,

Nq$(0) = 1.74 \times 10^{7}$, Sq$(0) = 1.73999 \times 10^{7}$, Iq$(0) = 100$, Cq$(0) = 0$, Dq$(0) = 0$, Aq$(0) = 0\}\Big]$

Proportions

It appears useful to have the proportions available. We set them in the following way:

```
SICD[Set, Pr, NqSeed] (* or leaving NqSeed out, and take the options *)
```

$\Big\{5.74713 \times 10^{-8}$ InterpolatingFunction$\Big[$ ⊞ ⌐ Domain: (0. 150.) Output: scalar $\Big]$(t$),

5.74713×10^{-8} InterpolatingFunction$\Big[$ ⊞ ⋀ Domain: (0. 150.) Output: scalar $\Big]$(t$),

5.74713×10^{-8} InterpolatingFunction$\Big[$ ⊞ ⌐ Domain: (0. 150.) Output: scalar $\Big]$(t$),

5.74713×10^{-8} InterpolatingFunction$\Big[$ ⊞ ⌐ Domain: (0. 150.) Output: scalar $\Big]$(t$),

5.74713×10^{-8} InterpolatingFunction$\Big[$ ⊞ ⌐ Domain: (0. 150.) Output: scalar $\Big]$(t$),

5.74713×10^{-8} InterpolatingFunction$\Big[$ ⊞ ⌐ Domain: (0. 150.) Output: scalar $\Big]$(t$)$\Big\}$

```
SICD[Pr, 50]
```

{0.1068, 0.334003, 0.550809, 0.00838795, 0.991612, 0.559197}

Appendix. Comparison with the traditional form for the equations

We would clear the variables first.

```
SIA[Clear]
```

We might do:

```
tradmodel[t0_, t_] = SICD[N, TraditionalForm, t0, t, Seed → 100]
```

$\text{Hold}\big[\{Sq'(t) = -2.29885 \times 10^{-8}\,Iq(t)\,Sq(t),\ Iq'(t) = Iq(t)\,(2.29885 \times 10^{-8}\,Sq(t) - 0.1),$

$\quad Cq'(t) = 0.0985\,Iq(t),\ Dq'(t) = 0.0015\,Iq(t),\ Nq(t) = 1.74 \times 10^{7} - Dq(t),\ Aq(t) = Cq(t) + Dq(t),$

$\quad Nq(t0) = 1.74 \times 10^{7},\ Sq(t0) = 1.73999 \times 10^{7},\ Iq(t0) = 100,\ Cq(t0) = 0,\ Dq(t0) = 0,\ Aq(t0) = 0\}\big]$

MatrixForm @@ TheFormer

$$\begin{pmatrix}
Sq'(t) = -2.29885 \times 10^{-8}\,Iq(t)\,Sq(t) \\
Iq'(t) = Iq(t)\,(2.29885 \times 10^{-8}\,Sq(t) - 0.1) \\
Cq'(t) = 0.0985\,Iq(t) \\
Dq'(t) = 0.0015\,Iq(t) \\
Nq(t) = 1.74 \times 10^{7} - Dq(t) \\
Aq(t) = Cq(t) + Dq(t) \\
Nq(t0) = 1.74 \times 10^{7} \\
Sq(t0) = 1.73999 \times 10^{7} \\
Iq(t0) = 100 \\
Cq(t0) = 0 \\
Dq(t0) = 0 \\
Aq(t0) = 0
\end{pmatrix}$$

```
trad = SICD[NDSolve, tradmodel[0, t], {t, 0, 150}];
```

```
SICD[Set, trad];
```

After this setting, we can call the functions directly for different moments in time.

```
store[SICD, tradmodel] = SICD[75]
```

$\{415\,556.,\ 738\,924.,\ 1.60018 \times 10^{7},\ 243\,683.,\ 1.71563 \times 10^{7},\ 1.62455 \times 10^{7}\}$

TheFormer // Explain

$\{S \to 415\,556.,\ I \to 738\,924.,\ C \to 1.60018 \times 10^{7},\ D \to 243\,683.,\ N \to 1.71563 \times 10^{7},\ A \to 1.62455 \times 10^{7}\}$

Perhaps it is enlightening to show the constant share of the deceased in the acquitted. We used the traditional format here, and still find that the share is constant at the implied IFF.

```
Plot[Dq[t] / Aq[t], {t, 0, 150}, PlotRange → {.014, .016},
 AxesLabel → {"Day", "Share D / (C + D)"}]
```
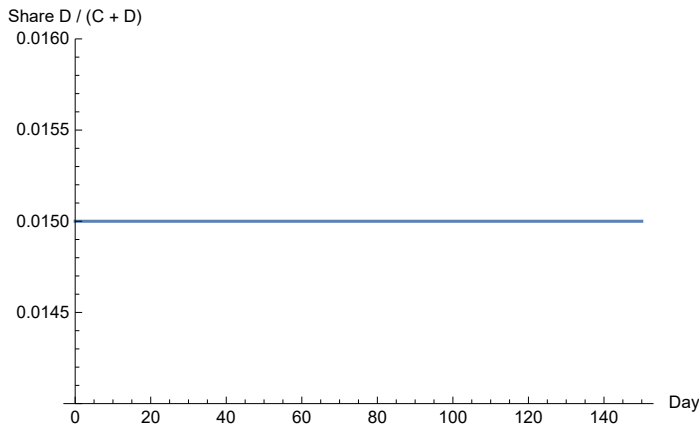

Appendix. The mistake to regard $1/\mu$ and $1/\lambda$ as implied infectious periods

The formula $\gamma = \gamma_{SIA} = (1 - \varphi)\,\gamma + \varphi\,\gamma = \gamma_{SICD} + \mu = \lambda + \mu$ is unambiguous. What is its meaning ?

We already said that the best model formulation has $A' = \gamma\, I$ and $D = \varphi\, A$ and $C = (1 - \varphi)\, A$. This immediately clarifies that C and D are mere proportions of A. In the formulation with $D' = \mu\, I$ or even $D' = \gamma\, \varphi\, I$, we might consider that the mean infectious period for the deceased compartment is $1 / \mu = 1 / (\gamma\, \varphi)$, with a conventional value of $1 / 0.0015 = 667$ days. This is only based upon an analogy with $A' = \gamma\, I$. The conceptual error is that we cannot take I as the base, but must take $\varphi\, I$ which thus gives period $\varphi / (\gamma\, \varphi)$.

Still, we cannot simply reject the idea of different infectious periods for the cleared and deceased. When we take this idea serious then this gives rise to implied parameter ω, namely the weight of the lethal infectious period in the average infectious acquittal period. Let us see how far this goes.

If we disregard distributions and only consider rates and means:

```
{eqgamma = {γ == μ + λ, 1 / γ == (1 - ω) / λ + ω / μ},
  Solve[eqgamma, {ω, γ}]} // TableForm
```

$\gamma = \lambda + \mu \qquad \frac{1}{\gamma} = \frac{1 - \omega}{\lambda} + \frac{\omega}{\mu}$

$\omega \to \frac{\mu^2}{\mu^2 - \lambda^2}$
$\gamma \to \lambda + \mu$

```
(SICD[OmegaSICD] /. SEYCD["GreekLetters"]) // MatrixForm
```

$$
\begin{pmatrix}
\text{Period(Acquittal, SIA)} = \frac{1}{\gamma} \\
\text{Period(Survivor acquittal = clearing, SICD)} = \frac{1}{\lambda} \\
\text{Period(Lethal acquittal, SICD)} = \frac{1}{\mu} \\
\gamma = \lambda + \mu \\
\frac{1}{\gamma} = \frac{1}{\lambda + \mu} \\
\text{Period(Acquittal, SICD)} = \frac{1}{\lambda + \mu} \\
\text{Period(Acquittal, SICD)} = \frac{1 - \omega}{\lambda} + \frac{\omega}{\mu} \\
\omega = \frac{\mu^2}{\lambda^2 - \mu^2}
\end{pmatrix}
$$

This scheme would only work if the weight ω would fit the shares of C and D in A, so that we have a proper mean indeed. This requires $\omega = \varphi$.

We also have $\mu = \gamma\, \varphi$. Substitution gives:

```
eq = (SICD[OmegaSICD] // Last) /.
    {OmegaSICD[] → PhiSICD[], MuSICD[] → PhiSICD[] × GammaSIA[],
     GammaSICD[] → GammaSIA[] (1 - PhiSICD[])} /.
   SEYCD["GreekLetters"] (*this uses strings*)
```

$\varphi = \dfrac{\gamma^2\, \varphi^2}{\gamma^2\, (1 - \varphi)^2 - \gamma^2\, \varphi^2}$

```
seq = Simplify[eq] /. "φ" → x
```

$x = \dfrac{x^2}{1 - 2x}$

```
Solve[seq, x]
```

$\left\{ \{x \to 0\}, \left\{ x \to \frac{1}{3} \right\} \right\}$

Thus, this line of interpretation might gain validity if the infection is lethal for a third of the infected. (The zero point is not have SICD in the first place.) This still leaves the question about the underlying distributions. With the common assumption on the distributions (*C* and *D* are proportions of *A*), the two lines of interpretations just happen to agree when the infection is lethal for a third of the population - but the logic of the interpretation still cannot be regarded as sound.

8. SI(EY)A(CD)

Clear

Clear all relevant variables since we may do different runs.

```
SIA[Clear]

ResetOptions[SIA];

SetOptions[SetDatabank, Databank → SEYCD]; (* for Explain *)

PlotLegendsSIA[SEYCD] (* base versus intervention scenario *)
```

8.1. Fast-track run and plot

This section directly runs the model and plots the results. An appendix to this chapter does the same in steps, so that the user might take more advantage of the environment in *Mathematica.*

The Dutch population in 2020 is about 17.4 million people. It will be convenient to use formal parameter Nqt0 → the size of the population. Subsequently, for quick calculation of proportions it will be useful to set NqSeed = Nq[t0] at a numerical value. NqSeed is not a parameter in the model, and only a handy variable at the user level.

```
NqSeed = 17.4 × 10^6;
SetOptions[SIA, Nqt0 → NqSeed]
```

$\{$Aqt0 → 0, Cqt0 → 0, DataMold → {S, I, A}, Dqt0 → 0, Iqt0 → Null, Nqt0 → 1.74×10^7, Onset → 0, Seed → 100, Yqt0 → 0, AlphaSEYCD() → 1., BetaSIA() → 0.4, GammaSIA() → 0.1, PhiSICD() → 0.015, R0() → Null$\}$

With $1/\gamma = 10$, the parameter $\alpha = 1$ adds another $1/\alpha$ day of non-infectious exposure to the total transmission period, then 11. This means that every new "generation of infections" adds another 1 day to the whole process, and the limit value may shift by some 20 days.

There is no other important difference between SIA(CD) and SEYA(CD). There are no analytical results on the E and Y components. Values can be found by numerical interpolation.

The infections are driven by the equation Eq[t0] == Seed. (Seed is parameter in the model.) The carnival festivities in the South of Holland were on February 23 2020. It is not unlikely that some 100 infections were seeded by people having returned from Italy and Austria from work and ski holidays. The options have default Seed → 100. When we use a different value of Seed, then this is "baked" into this model solution, and we must take heed of routines that rely upon Options[SIA] for the value of Seed in these options. We can mention the parameter here so that the user can check more easily that model solutions change when adapting the Seed value. (For larger values of α a longer time window could be needed.)

```
SEYCD[Run, Pr, newModel, t, 150, Seed → 100];
```

The variables are now available in levels and proportions.

```
store[SEYCD] = SEYCD[75]
  (*store this for later comparison with the traditional format *)
```

$\{688\,760., 48\,491.2, 2.61518 \times 10^6, 1.38369 \times 10^7, 210\,714., 1.71893 \times 10^7, 2.66367 \times 10^6, 1.40476 \times 10^7\}$

```
TheFormer // Explain
```

$\{S \rightarrow 688\,760., E \rightarrow 48\,491.2, Y \rightarrow 2.61518 \times 10^6, C \rightarrow 1.38369 \times 10^7,$

$\quad D \rightarrow 210\,714., N \rightarrow 1.71893 \times 10^7, I \rightarrow 2.66367 \times 10^6, A \rightarrow 1.40476 \times 10^7\}$

We now have the additional information that by day 75 some 210,000 people would have died, had the government not intervened.

```
SEYCD[Pr, 75]
```

{0.0395839, 0.00278685, 0.150297, 0.795222, 0.01211, 0.98789, 0.153084, 0.807332}

```
TheFormer // Explain
```

{S → 0.0395839, E → 0.00278685, Y → 0.150297,

$\quad$ C → 0.795222, D → 0.01211, N → 0.98789, I → 0.153084, A → 0.807332}

It is useful to have a plot that show both levels and proportions. The plot axes assume that it is obvious what are levels and what are proportions.

In the following plot, the population drops by about 1.5% because of the deceased. The difference between the population (red) and the cleared compartment (green) are the remaining 2% suscepti-ble (cyan).

Some points to observe are: (i) The exposure delay of 1 day has little effect, but if it would be 5 days then we would be advised to plot for 200 instead of 150 days. Such delays cumulate. (ii) Also, observe the top of *Y* is after the top of *E.* At first this makes little sense, because a high availability of infectious units would cause more infections and thus also more exposed units. However, at the top of *Y* the susceptibles have decreased so much that the effective contribution to new infections has been much reduced. (iii) Since the tops of *E* and *Y* are at different locations, there is no proportional relation with *I*. Potentially relevant is the time between the tops. (iv) Conventional plots may not show *I* = *E* + *Y* but it is important in relation to SIA and the analytical properties (e.g. the top).

```
SEYCD[Plot, 0, 150]
```


8.2. Relation to capacity of ICU beds

The national challenge has been for the intensive care units (ICU beds). Assume that there are 1500 beds on the ICU, and that each death, after an infectious period of $1/\gamma$, still requires $1/h$ = 10 days in the hospital. (Admission to a hospital acquits from the infectious stage for the public at large.) The rate per bed is 3 patients per month or h = 3/30 per day. A good way to look at this is to take a cohort of 10 beds, with one new patient arriving every day. In a steady state with input = output, the ICUs can accept only 1500 h = 150 new patients per day, as each new patient takes the open spot in a cohort of 10 beds. Capacity remains at 1500, and we can plot above graph for a lower range at that level to show how the ICU system cannot take the new arrivals who are clearly at risk of dying.

The difference with the plot in section 7.2 is that there now are the exposed *E* in the middle, while the infectious *Y* have shifted to the right. The delay by the exposed has not much effect on the dynamics for the ICU.

```
seicdplotoptionsICU = SEYCD[Plot, Options, ICU, PlotRange → {0, 10000}];
Plot @@ {SEYCD[ICU, 1500, t], {t, 0, 150}, seicdplotoptionsICU}
```
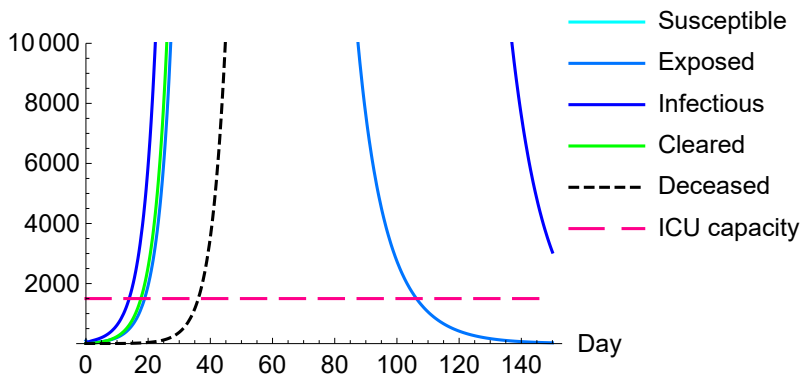
```
SEYCD[Explain, ICU]
```

Replace SICD by SEYCD:

SICD plots can be enlightened by including a line for the ICU capacity. This is simple by itself
    but becomes a bit more complicated because of the colouring and legend. In this plotting,
    the level of population is far out of range, and not included. Two main subroutines are:

SICD[Plot, Options, ICU, opts] combines plotting options and allows user options

SICD[ICU, level, t] does Append[Take[SICD[t], 4], level], thus removes N & A, inserts level

Subroutines used by those two main subroutines are:

SICD["Colours", ICU] gives DeeperPink for the ICU capacity

SICD[PlotLegends, ICU] uses the label "ICU capacity"

SICD[Plot, PlotLegends, ICU] puts the labels into a Placed object

SICD[Plot, Options, ICU,"Other"] chooses ImageSize and AxesLabel

SICD[Plot, SetOptions, ICU, opts] may set Plot options (inadvisable for other Plots)

8.3. Keeping the same compartments or recompartmentalise ?

The SIA model assumes that *infected = infectious,* but now for SEYCD we distinguish between (i) a period of *latency,* of being infected but without being infectious, and (ii) the subsequent infectious period thereafter.

8.3.1. The traditional choice of keeping the same word "infectious"

Traditionally the SEYCD model creates a new compartment, so that $N[0] = S + E + I + A$, with Exposed E and Infectious I, and also $A = C + D$. We may introduce variable $Y = E + I$ to get the SYA compartments for comparison with earlier results of SIA (in which I then means $Y,$ or $E = 0$).

Indeed, we can prove that the S and A compartments in SE(Y)I(A)CD and SIA, are the same, so that that the $Y = E + I$ in SEYCD would be the same as the I in SIA. To check this, also in programming, causes a lot of translating, with sources of confusion. There is no *need* for this complexity, except when one wishes to adhere to the dogma that $I = infectious.$

8.3.2. The didactic approach to maintain the same compartments

Another perspective is to argue that we already had the SIA compartments. If nothing changes in S and $A,$ then nothing changes in I too, since the sum to $N[0]$ remains the same. We can sub-compartmentalise $I = E + Y,$ with Exposed E and Infectious $Y.$ This is comparable to how we split $A = C + D.$ This has the advantage that we do not have to keep in memory each time that we changed compartments. We do not have to adapt routines. For example TopIq still gives the top of $I = E + Y.$ (Humans may be more flexible than computers, but in the interaction with inflexible computers humans may become even more inflexible.) We can agree that I in SIA also means $I = Y$ because $E = 0$, but we do not stick to a dogma of *vocabulary.*

8.3.3. The choice (leave it up to the user)

The two views cause a curious conflict in perspectives *Y = E + I* or *I = E + Y*. It concerns only "different names for the same notions". A choice comes with discussions like (one side) "we called *I* "infectious" and thus it will remain so" versus (other side) "there is already this *I = N*[0] - *S* - *A* compartment and you only gave it a wrong name".

In practice, compartments are not set in stone.

(i) Researchers must create as many compartments as is necessary to describe the actual process, and they are advised to use names that convey the key properties of those compartments. For these users, we allow a "traditional formulation of the SEYCD model".

(ii) For education and didactics, it is useful to build upon earlier foundations. The foundation in SIA is not the name "*I* = Infected = Infectious" but the foundation are also the properties of two classes *S* and *A* that remain the same, while we now consider a change within the third compartment *I*. Thus, this follows the *I = E + Y* route. This notebook and package follow the didactics track.

(iii) In SICD we included *A = C + D* in the set of equations. This clarifies what the variables mean, in comparison with SIA, and has the added advantage that subsequent programs and runtime applications do not have to do the addition every time over and over again. For SEYCD we better include the administrative equation *Y = E + I* (traditional) or *I = E + Y* (didactic).

8.4. Basic model in standard formulas (didactic format)

SIA[Clear] (* important since we have run the model above *)

The traditional SEYCD model, in its didactic SEYCD format, is as follows, with *Infected = Exposed + Infectious* or *I = E + Y.*

ReadMeSIA[SEYCD, Simplify]

$\mathcal{S}' = -\beta^* \, \mathcal{S} \, \mathcal{Y}$

$\mathcal{E}' = \alpha \, (-\mathcal{E}) - \mathcal{S}'$

$\mathcal{Y}' = \alpha \, \mathcal{E} - \mathcal{A}'$

$\mathcal{A}' = \gamma \, \mathcal{Y}$

$\mathcal{D} = \mathcal{A} \, \varphi$

$C = \mathcal{A} - \mathcal{D}$

$\mathcal{N} = \mathcal{N}(t0) - \mathcal{D}$

$\mathcal{I} = \mathcal{Y} + \mathcal{E}$

$\mathcal{N}(t0) = \mathcal{N}(t0) - \text{Dqt0}$

$\mathcal{S}(t0) = -\text{Cqt0} - \text{Dqt0} - \text{Seed} + \mathcal{N}(t0) - \text{Yqt0}$

$\mathcal{E}(t0) = \text{Seed}$

$\mathcal{Y}(t0) = \text{Yqt0}$

$C(t0) = \text{Cqt0}$

$\mathcal{D}(t0) = \text{Dqt0}$

$\mathcal{A}(t0) = \text{Cqt0} + \text{Dqt0}$

$\mathcal{I}(t0) = \text{Seed} + \text{Yqt0}$

PM 1. An alternative is to first state *S',E',* and *A'* and subsequently derive *Y',* but now *E* invites *Y* to be higher up in the ranking, and it has an advantage to keep *A* with *C* and *D.*

PM 2. $\beta^* = \beta / N[t_0]$. This gives a tiny value for β^* but simplifies the model. Below we immediately drop this simplification.

PM 3. Above model is in levels. Division by $N[t_0]$ or setting $N[t_0] = 1$ gives proportions (and $\beta^* = \beta$).

PM 4. See the discussion about φ = IFF for the SICD model. We now adopt subtraction $C = A - D$.

PM 5. The model might be run for symptomatic disease. Then φ = sCFF for the *symptomatic case fatality factor* (sCFF). Not all infected will develop symptoms, so that sCFF = D / sC = (D / I) *(I / sC) = IFF / (sC / I). SARS-CoV-2 studies who apply the SEIRD or SEYCD model for symptoms actually are Kucharski et al. (2020) and the Goh (2020) "epidemic calculator", and Maugeri et al. (2020): see the chapter on comparison with other studies.

PM 6. SEYCD reduces to SIA if we take $\alpha \to \infty$ or $1/\alpha \to 0$, though $1/\alpha = 1/24$ for one hour might suffice. If we take $\alpha = 0$ then all infections are stopped immediately. The effect of α is only to delay events. The same properties of SIA(CD) apply, except for this time delay.

PM 7. Yan & Chowell (2019:167) nicely show how the rescaling of time into $\tau = \gamma t$ changes the SEYCD model into one that is dominated by only two parameters: β / γ and α / γ. Thus we basically still need three parameters, the latter two, and the scaling of time. Their deduction however helps us to focus on R_0 and these ratios, and not on β / α.

NB. There is a small computational trick for Nq[0]. Below we will have an intervention, and switch from $t_0 = 0$ to $t_0 = 24$ for the day of intervention. This means that Nq[t] = Nq[0] - Dq[t] will have nonzero deaths. By the present specification, the proportions are still relative to the option parameter *Nqt0* but we can link the values of Nq[t] in the base and intervention scenarios.

8.5. Computable format (*Copy from SIA*)

See the corresponding section in the discussion of SIA.

8.6. The different formats

An overview of the traditional / simplified and computable / readable formats is as follows.

```
symbolic = {SEYCD[Equations, t0, t],
    SEYCD[Equations, TraditionalForm, t0, t]} // Transpose // TableForm
```

$Sq'(t) = -\dfrac{\text{BetaSIA}()\, Sq(t)\, Yq(t)}{\text{Nqt0}}$ $Sq'(t) = -\dfrac{\text{BetaSIA}()\, Iq(t)\, Sq(t)}{\text{Nqt0}}$

$Eq'(t) = -\text{AlphaSEYCD}()\, Eq(t) - Sq'(t)$ $Eq'(t) = \dfrac{\text{BetaSIA}()\, Iq(t)\, Sq(t)}{\text{Nqt0}} - \text{AlphaSEYCD}()\, Eq(t)$

$Yq'(t) = \text{AlphaSEYCD}()\, Eq(t) - Aq'(t)$	$Iq'(t) = \text{AlphaSEYCD}()\, Eq(t) - (\text{GammaSICD}() + \text{MuSICD}())\, Iq(t)$
$Aq'(t) = \text{GammaSIA}()\, Yq(t)$	$Cq'(t) = \text{GammaSICD}()\, Iq(t)$
$Dq(t) = \text{PhiSICD}()\, Aq(t)$	$Dq'(t) = \text{MuSICD}()\, Iq(t)$
$Cq(t) = Aq(t) - Dq(t)$	$Nq(t) = \text{Nqt0} - Dq(t)$
$Nq(t) = \text{Nqt0} - Dq(t)$	$Yq(t) = Eq(t) + Iq(t)$
$Iq(t) = Eq(t) + Yq(t)$	$Aq(t) = Cq(t) + Dq(t)$
$Nq(t0) = \text{Nqt0} - \text{Dqt0}$	$Nq(t0) = \text{Nqt0}$
$Sq(t0) = -\text{Cqt0} - \text{Dqt0} + \text{Nqt0} - \text{Seed} - \text{Yqt0}$	$Sq(t0) = -\text{Cqt0} - \text{Dqt0} - \text{Iqt0} + \text{Nqt0} - \text{Seed}$
$Eq(t0) = \text{Seed}$	$Eq(t0) = \text{Seed}$
$Yq(t0) = \text{Yqt0}$	$Iq(t0) = \text{Iqt0}$
$Cq(t0) = \text{Cqt0}$	$Cq(t0) = \text{Cqt0}$
$Dq(t0) = \text{Dqt0}$	$Dq(t0) = \text{Dqt0}$
$Aq(t0) = \text{Cqt0} + \text{Dqt0}$	$Aq(t0) = \text{Cqt0} + \text{Dqt0}$
$Iq(t0) = \text{Seed} + \text{Yqt0}$	$Yq(t0) = \text{Iqt0} + \text{Seed}$

symbolic // SEYCDLettersOnly

$$S' = -\frac{\beta S \mathcal{Y}}{\mathcal{N}(t0)}$$

$$\mathcal{E}' = \alpha(-\mathcal{E}) - S'$$

$$\mathcal{Y}' = \alpha \mathcal{E} - \mathcal{A}'$$

$$\mathcal{A}' = \gamma \mathcal{Y}$$

$$\mathcal{D} = \mathcal{A} \varphi$$

$$C = \mathcal{A} - \mathcal{D}$$

$$\mathcal{N} = \mathcal{N}(t0) - \mathcal{D}$$

$$\mathcal{I} = \mathcal{Y} + \mathcal{E}$$

$$\mathcal{N}(t0) = \mathcal{N}(t0) - Dqt0$$

$$S(t0) = -Cqt0 - Dqt0 - Seed + \mathcal{N}(t0) - Yqt0$$

$$\mathcal{E}(t0) = Seed$$

$$\mathcal{Y}(t0) = Yqt0$$

$$C(t0) = Cqt0$$

$$\mathcal{D}(t0) = Dqt0$$

$$\mathcal{A}(t0) = Cqt0 + Dqt0$$

$$\mathcal{I}(t0) = Seed + Yqt0$$

$$S' = -\frac{\beta S \mathcal{I}}{\mathcal{N}(t0)}$$

$$\mathcal{E}' = \frac{\beta S \mathcal{I}}{\mathcal{N}(t0)} - \alpha \mathcal{E}$$

$$\mathcal{I}' = \alpha \mathcal{E} - \mathcal{I}(\lambda + \mu)$$

$$C' = \lambda \mathcal{I}$$

$$\mathcal{D}' = \mu \mathcal{I}$$

$$\mathcal{N} = \mathcal{N}(t0) - \mathcal{D}$$

$$\mathcal{Y} = \mathcal{E} + \mathcal{I}$$

$$\mathcal{A} = C + \mathcal{D}$$

$$\mathcal{N}(t0) = \mathcal{N}(t0)$$

$$S(t0) = -Cqt0 - Dqt0 - Iqt0 - Seed + \mathcal{N}(t0)$$

$$\mathcal{E}(t0) = Seed$$

$$\mathcal{I}(t0) = Iqt0$$

$$C(t0) = Cqt0$$

$$\mathcal{D}(t0) = Dqt0$$

$$\mathcal{A}(t0) = Cqt0 + Dqt0$$

$$\mathcal{Y}(t0) = Iqt0 + Seed$$

NB. Repeat: There is a small computational trick here. Below we will have an intervention, and switch from $t_0 = 0$ to $t_0 = 24$ for the day of intervention. This means that Nq[t] = Nq[0] - Dq[t] will have nonzero deaths. By the present specification, the proportions are still calculated w.r.t. the option parameter *Nqt0* but we can link the values of Nq[t] in the base and intervention scenarios. (This trick is not applied to the traditional format, since we will not run an intervention on this.)

NB. (i) The Seed is allocated to Eq[0]. We have to, otherwise the model will not run. (ii) To allow for the traditional approach, Options[SIA] contain key Iqt0, which would not be used in the simpler didactic version. (iii) Depending upon the value of α, the plotting period may need to be extended from 150 to 200 days to see the same phasing out.

Options[SIA]

$\{Aqt0 \rightarrow 0, Cqt0 \rightarrow 0, DataMold \rightarrow \{S, I, A\}, Dqt0 \rightarrow 0, Iqt0 \rightarrow Null, Nqt0 \rightarrow 1.74 \times 10^7, Onset \rightarrow 0, Seed \rightarrow 100,$
$Yqt0 \rightarrow 0, AlphaSEYCD() \rightarrow 1., BetaSIA() \rightarrow 0.4, GammaSIA() \rightarrow 0.1, PhiSICD() \rightarrow 0.015, R0() \rightarrow Null\}$

TheFormer // SEYCDLettersOnly

$\{Aqt0 \rightarrow 0, Cqt0 \rightarrow 0, DataMold \rightarrow \{S, I, A\}, Dqt0 \rightarrow 0, Iqt0 \rightarrow Null, \mathcal{N}(t0) \rightarrow 1.74 \times 10^7,$
$Onset \rightarrow 0, Seed \rightarrow 100, Yqt0 \rightarrow 0, \alpha \rightarrow 1., \beta \rightarrow 0.4, \gamma \rightarrow 0.1, \varphi \rightarrow 0.015, R0 \rightarrow Null\}$

When running the traditional model, its parameters can take values from Options[SIA] by first doing a substitution using the following rule.

SICD[GammaSIA, Rule]

$\{GammaSICD() \rightarrow GammaSIA()(1 - PhiSICD()), MuSICD() \rightarrow GammaSIA() PhiSICD()\}$

TheFormer // SEYCDLettersOnly

$\{\lambda \rightarrow \gamma(1 - \varphi), \mu \rightarrow \gamma \varphi\}$

Options[SEYCD] (* for Explain of SEYCD[t] *)

$\{DataMold \rightarrow \{S, E, Y, C, D, N, I, A\}\}$

? AlphaSEYCD

> Symbol
>
> AlphaSEYCD[] is the symbol in the options for a
>
> > constant rate of the Exposed who will become infectious (only in SEYCD)
>
> ⌄

8.7. Choice of parameters such that SEYCD is similar to SIA(CD)

When we choose the parameters of SI(EY)A(CD) as the same as SIA(CD) then the model also behaves like SIA(CD), except for the time delay caused by α.

For SEYCD, it appears that $R_0 = \beta / \gamma$ as before, see the deduction below. For comparison with SIA, we thus can maintain the γ as in SIA, and φ or $\gamma = \lambda + \mu$ as in SICD.

The new element in the discussion is that the process now consists of *latency* and *infectious* periods. Latency is when one is infected but not infectious yet. There is no parameter for the sum of these two, relevant for the "transmission period" = $1/\alpha + 1/\gamma$, since it has no further implication or use in the model. The parameter α is useful to understand the delay in events, and should help in accuracy of estimation of parameters (though we don't do so below). Above we already explained the choice of $1/\alpha = 1$ day.

8.8. The infection fatality rate (*Copy from SICD*)

See the corresponding section in SICD.

8.9. Choice of $R_0 = 4$ and acquittal period of 10 days (*Copy from SIA*)

See the corresponding section in SIA.

8.10. How to use the SEYCD routine

? SEYCD

Symbol

SEYCD is SICD with Exposed. See their for formats. In addition:

Implemented is SEYCD, with I = E+Y and A = C+D, thus with same SIA compartments.

This allows identification of which parameters have changed in what fashion.

SEYCD[t] puts out {S, E, Y, C, D, N, I, A} (8 variables). PM. A DataMold has been
set, so that one can call Explain[lis, SEYCD] for such output lis (see Databank package)

SEYCD[Limit] only rearranges SICD[Limit]

AlphaSEYCD[] is the basic new parameter for the step from Exposed to Infectious. This also causes the
options Yqt0 (didactic) and Iqt0 (traditional format). The Exposed concern the latent period 1/alpha,
which appears to be about 1 day (not to be confused with the serial interval of about 5 days)

SEYCD["R0[t]"] gives equations to help the choice of parameter values. R0 = beta / gamma
can be kept the same, and there is only a longer duration of the overall process because
of the insertion of the latency period 1/ alpha before the infectious period 1/gamma

SEYCD[Variables] gives the list of first letters of the variables, as strings

SEYCD["GreekLetters"] is a rule to change symbols into Greek letters

SEYCD[Plot, Options] or SEYCD[Plot, Setoptions]

SEYCD[Plot, Options, All] are defaults for plotting of SI(EY)A(CD).

SEYCD["Colours"]

SEYCD[Equations, ...] has the small trick that Nq[t0] == Nqt0 − Dqt0, which has no effect while Dqt0 −> 0.
However, when an intervention is run, with Onset −> nonzero, then there is a smooth transition
between the paths. Observe that Sq' still depends upon Nqt0 and not the value reduced by the Dqt0

For traditional form use SEYCD[Equations, TraditionalForm, ...] and SEYCD[N, TraditionalForm, ...]

⌄

Appendix. Run in steps (for management of details)

The Dutch population in 2020 is about 17.4 million people. These SI(EY)A(CD) models use a constant population. It will be convenient to use formal parameters Nq → the size of the population, and Np → 1. Subsequently, for quick calculation of proportions it will be useful to have a NqSeed available as a variable a the user level.

```
NqSeed = 17.4 × 10^6;
SetOptions[SIA, Nqt0 → NqSeed]
```

$\{$Aqt0 → 0, Cqt0 → 0, DataMold → {S, I, A}, Dqt0 → 0, Iqt0 → Null, Nqt0 → 1.74 × 10^7, Onset → 0, Seed → 100, Yqt0 → 0, AlphaSEYCD() → 1., BetaSIA() → 0.4, GammaSIA() → 0.1, PhiSICD() → 0.015, R0() → Null$\}$

We put the model equations into Hold, so that the symbols remain unevaluated when we assign values later on.

We seed the model with Seed → 100 as in the supplied options, but may choose another value later on.

```
model[t0_, t_] = SEYCD[N, t0, t, Seed → 100]
```

Hold$\big[\{$Sq′(t) = −2.29885 × 10^{-8} Sq(t) Yq(t), Eq′(t) = −1. Eq(t) − Sq′(t),

Yq′(t) = 1. Eq(t) − Aq′(t), Aq′(t) = 0.1 Yq(t), Dq(t) = 0.015 Aq(t), Cq(t) = Aq(t) − Dq(t),

Nq(t) = 1.74 × 10^7 − Dq(t), Iq(t) = Eq(t) + Yq(t), Nq(t0) = 1.74 × 10^7, Sq(t0) = 1.73999 × 10^7,

Eq(t0) = 100, Yq(t0) = 0, Cq(t0) = 0, Dq(t0) = 0, Aq(t0) = 0, Iq(t0) = 100$\}\big]$

```
MatrixForm @@ TheFormer
```

$$\begin{pmatrix}
\text{Sq}'(t) = -2.29885 \times 10^{-8}\,\text{Sq}(t)\,\text{Yq}(t) \\
\text{Eq}'(t) = -1.\,\text{Eq}(t) - \text{Sq}'(t) \\
\text{Yq}'(t) = 1.\,\text{Eq}(t) - \text{Aq}'(t) \\
\text{Aq}'(t) = 0.1\,\text{Yq}(t) \\
\text{Dq}(t) = 0.015\,\text{Aq}(t) \\
\text{Cq}(t) = \text{Aq}(t) - \text{Dq}(t) \\
\text{Nq}(t) = 1.74 \times 10^7 - \text{Dq}(t) \\
\text{Iq}(t) = \text{Eq}(t) + \text{Yq}(t) \\
\text{Nq}(t0) = 1.74 \times 10^7 \\
\text{Sq}(t0) = 1.73999 \times 10^7 \\
\text{Eq}(t0) = 100 \\
\text{Yq}(t0) = 0 \\
\text{Cq}(t0) = 0 \\
\text{Dq}(t0) = 0 \\
\text{Aq}(t0) = 0 \\
\text{Iq}(t0) = 100
\end{pmatrix}$$

We can run the model with SICD[NDSolve, ...], which gives a numerical approximation.

```
sol = SEYCD[NDSolve, model[0, t], {t, 0, 150}]
```

$\{$Sq $\rightarrow$ InterpolatingFunction[⊞ ⎍ Domain: (0. 150.) Output: scalar],

Eq $\rightarrow$ InterpolatingFunction[⊞ ⋀ Domain: (0. 150.) Output: scalar],

Yq $\rightarrow$ InterpolatingFunction[⊞ ⋀ Domain: (0. 150.) Output: scalar],

Cq $\rightarrow$ InterpolatingFunction[⊞ ⌠ Domain: (0. 150.) Output: scalar],

Dq $\rightarrow$ InterpolatingFunction[⊞ ⌠ Domain: (0. 150.) Output: scalar],

Nq $\rightarrow$ InterpolatingFunction[⊞ ⎍ Domain: (0. 150.) Output: scalar],

Aq $\rightarrow$ InterpolatingFunction[⊞ ⌠ Domain: (0. 150.) Output: scalar],

Iq $\rightarrow$ InterpolatingFunction[⊞ ⋀ Domain: (0. 150.) Output: scalar]$\}$

We want to be able to call $Sq[t]$, $Eq[t]$, $Iq[t]$, $Cq[t]$, $Dq[t]$, $Nq\{t\}$ and SEYCD[t] directly. We use above "sol" to set these functions.

```
SEYCD[Set, sol];
```

After this setting, we can call the functions directly for different moments in time. Let us also set the proportions.

```
SEYCD[Set, Pr];
```

It is somewhat useless to compare with earlier outcomes, because we know that the functions have shifted.

```
SEYCD[75] // Explain
```

$\{$S $\rightarrow$ 688 760., E $\rightarrow$ 48 491.2, Y $\rightarrow$ 2.61518 $\times 10^6$, C $\rightarrow$ 1.38369 $\times 10^7$,

D $\rightarrow$ 210 714., N $\rightarrow$ 1.71893 $\times 10^7$, I $\rightarrow$ 2.66367 $\times 10^6$, A $\rightarrow$ 1.40476 $\times 10^7\}$

There are some default settings for the layout of the SEYCD plots. NB. These options assume that the last element in SEYCD[t] is dropped, which are the Acquitted. They are relevant for top and limit but not for the present plotting.

```
seycdplotoptions = SEYCD[Plot, Options];
```

```
Plot @@ {Drop[SEYCD[t], -1], {t, 0, 150}, seycdplotoptions}
 (*DROP the last A, not legended*)
```
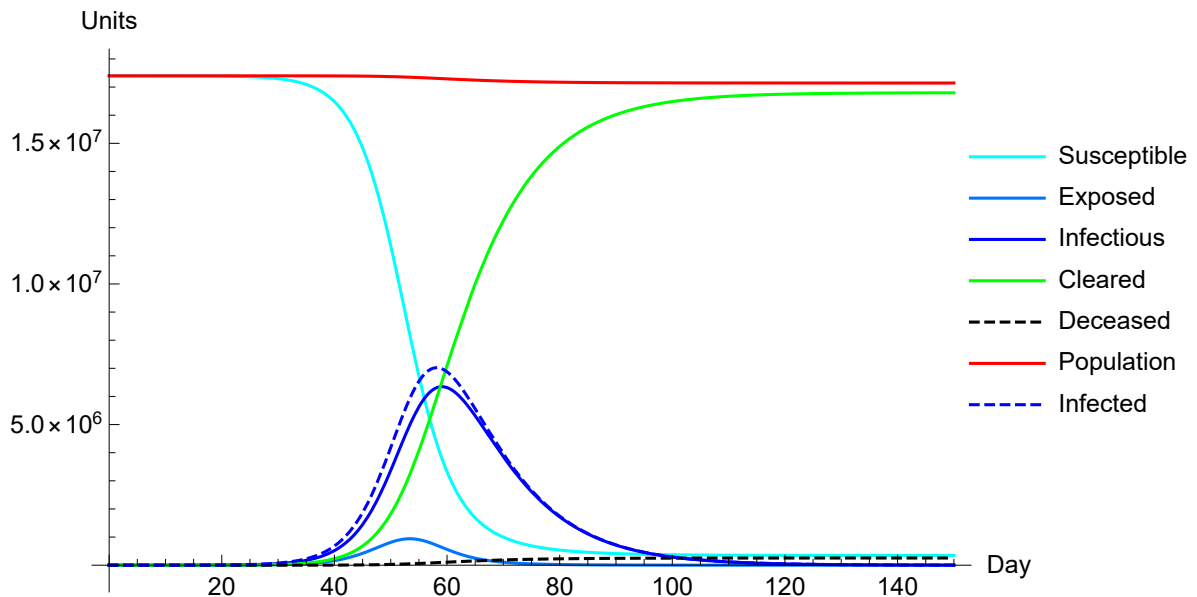

Given our earlier setting of NqSeed, it is easy to plot for the proportions (standard Plot doesn't have an option for two vertical axes).

```
Plot @@ {Drop[SEYCD[t], -1] / NqSeed, {t, 0, 150}, seycdplotoptions} (*DROP A*)
```
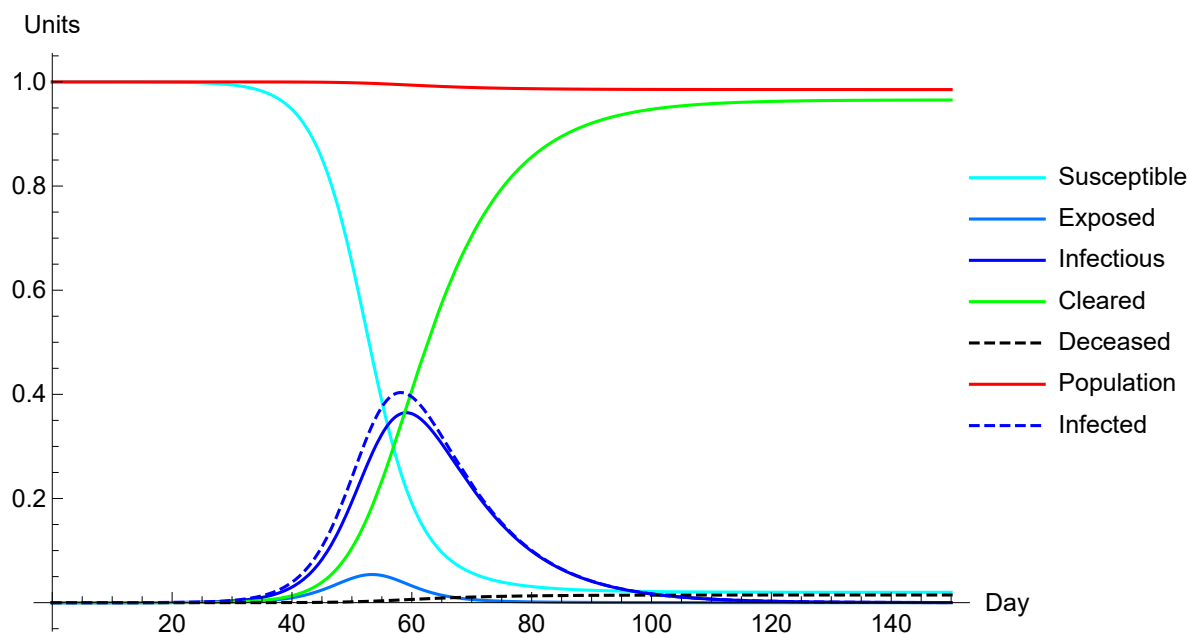

Appendix. Comparison with the traditional form for the equations ($Y = E + I$)

The traditional format of SEYCD has $Y = E + I$ instead of $I = E + Y$.

```
SIA[Clear]
```

```
SetOptions[SetDatabank, Databank → SEYCD];
```

```
Options[SIA]
```

$\{Aqt0 \rightarrow 0, Cqt0 \rightarrow 0, DataMold \rightarrow \{S, I, A\}, Dqt0 \rightarrow 0, Iqt0 \rightarrow Null, Nqt0 \rightarrow 1.74 \times 10^7, Onset \rightarrow 0, Seed \rightarrow 100,$
$Yqt0 \rightarrow 0, AlphaSEYCD() \rightarrow 1., BetaSIA() \rightarrow 0.4, GammaSIA() \rightarrow 0.1, PhiSICD() \rightarrow 0.015, R0() \rightarrow Null\}$

```
? Iqt0
```

> Symbol
>
> Iqt0 = Iq[t0], an option in SEYCD traditional, not in SIA or SICD, default Null
>
> ⌄

It is important to choose Iqt0 → 0 now.

```
tradmodel[t0_, t_] = SEYCD[N, TraditionalForm, t0, t, Seed → 100, Iqt0 → 0]
```

$\text{Hold}\big[\{Sq'(t) = -2.29885 \times 10^{-8} \, Iq(t) \, Sq(t), Eq'(t) = 2.29885 \times 10^{-8} \, Iq(t) \, Sq(t) - 1. \, Eq(t),$
$Iq'(t) = 1. \, Eq(t) - 0.1 \, Iq(t), Cq'(t) = 0.0985 \, Iq(t), Dq'(t) = 0.0015 \, Iq(t), Nq(t) = 1.74 \times 10^7 - Dq(t),$
$Yq(t) = Eq(t) + Iq(t), Aq(t) = Cq(t) + Dq(t), Nq(t0) = 1.74 \times 10^7, Sq(t0) = 1.73999 \times 10^7,$
$Eq(t0) = 100, Iq(t0) = 0, Cq(t0) = 0, Dq(t0) = 0, Aq(t0) = 0, Yq(t0) = 100\}\big]$

```
MatrixForm @@ TheFormer
```

$$\begin{pmatrix} Sq'(t) = -2.29885 \times 10^{-8} \, Iq(t) \, Sq(t) \\ Eq'(t) = 2.29885 \times 10^{-8} \, Iq(t) \, Sq(t) - 1. \, Eq(t) \\ Iq'(t) = 1. \, Eq(t) - 0.1 \, Iq(t) \\ Cq'(t) = 0.0985 \, Iq(t) \\ Dq'(t) = 0.0015 \, Iq(t) \\ Nq(t) = 1.74 \times 10^7 - Dq(t) \\ Yq(t) = Eq(t) + Iq(t) \\ Aq(t) = Cq(t) + Dq(t) \\ Nq(t0) = 1.74 \times 10^7 \\ Sq(t0) = 1.73999 \times 10^7 \\ Eq(t0) = 100 \\ Iq(t0) = 0 \\ Cq(t0) = 0 \\ Dq(t0) = 0 \\ Aq(t0) = 0 \\ Yq(t0) = 100 \end{pmatrix}$$

```
trad = SEYCD[NDSolve, tradmodel[0, t], {t, 0, 150}];
```

```
SEYCD[Set, trad];
```

After this setting, we can call the functions directly for different moments in time.

```
(store[SEYCD, tradmodel] = SEYCD[75]) // Explain
```

$\{S \rightarrow 688\,760., E \rightarrow 48\,491.2, Y \rightarrow 2.66367 \times 10^6, C \rightarrow 1.38369 \times 10^7,$
$D \rightarrow 210\,714., N \rightarrow 1.71893 \times 10^7, I \rightarrow 2.61518 \times 10^6, A \rightarrow 1.40476 \times 10^7\}$

Let us create a small databank using the same SEYCD[DataMold] to compare the findings of the traditional model with those of the didactic SEYCD setup.

```
anylabel[Data] = {store[SEYCD] (*former run*), store[SEYCD, tradmodel] };
```

```
SetDatabank[SEYCD[DataMold], anylabel[Data], anylabel]
```

$\{Databank \rightarrow anylabel\}$

Check that *I* and *Y* have changed values.

```
ShowData[] (* cell locked / non-evaluatable, in case a former run wasn't stored *)
```

	1	2
S	688 760.	688 760.
E	48 491.2	48 491.2
Y	2.61518×10^6	2.66367×10^6
C	1.38369×10^7	1.38369×10^7
D	210 714.	210 714.
N	1.71893×10^7	1.71893×10^7
I	2.66367×10^6	2.61518×10^6
A	1.40476×10^7	1.40476×10^7

To prevent possible confusion, run the educational SEYCD again.

```
ResetOptions[SIA]
```

$\{$Aqt0 $\to$ 0, Cqt0 $\to$ 0, DataMold $\to$ {S, I, A}, Dqt0 $\to$ 0, Iqt0 $\to$ Null, Nqt0 $\to$ 1.74×10^7, Onset $\to$ 0, Seed $\to$ 100, Yqt0 $\to$ 0, AlphaSEYCD() $\to$ 1., BetaSIA() $\to$ 0.4, GammaSIA() $\to$ 0.1, PhiSICD() $\to$ 0.015, R0() $\to$ Null$\}$

```
SEYCD[Run, Pr, newModel, t, 150, Seed → 100];
```

9. Analytical properties of SEYCD

9.1. Using didactic SEYCD with I = E + Y rather than traditional SEYCD with Y = E + I

The following establishes:

(a) SEI(Y)CD is just SIA(CD) with a delay between exposure and infectiousness. The properties for the limit and top values are maintained. This is shown most clearly by using *I* = *E* + *Y* and SEYCD,

(b) There are no analytical solutions of *E* and *Y,* but there are equations that can be solved numerically to find the tops of their curves: TforTopEq and TforTopYq.

(c) There are some other comments that show some properties of the model, but the main result is that we cannot find other analytical results, whence one may be advised to skip the discussion that there are no results.

(d) When we have SEYCD[t] available, then we can also call SIA[t] and SICD[t] for their selections of variables. If we would want to make a pure run of SIA, then store those in an own new routine PureSIA[t], because SEYCD[...] will use all standard variables here.

The following shows that it is didactically simpler to use SEYCD with *I* = *E* + *Y*, so that SEYCD = SIA with some delay.

PM. In section 1.4.4 we had a relation for the early phase of exponential growth. Now we extend for the whole domain.

PM. When some reset is needed, the key commands for the standards SEYCD setup are:

```
ResetOptions[SIA];
```

```
SEYCD[Run, Pr, newModel, t, 150, Seed → 100];
```

```
SetOptions[SetDatabank, Databank → SEYCD];

PlotLegendsSIA[SEYCD] (* base versus intervention scenario *)
```

9.2. The growth rate of *I* depends upon *Y / I*

One track is to look at the growth equations. When we add the equations for *E'* and *Y'* then we get *I'*:

$I' = E' + Y' = \beta \, Sp \, Y - \gamma \, Y = (\beta \, Sp - \gamma) \, Y$

$dLog[I] = I' / I = (\beta \, Sp - \gamma) \, (Y / I)$ which in SIA was $i = \beta \, S_p - \gamma$

With $0 \le Y / I \le 1$ we can expect lower growth (over a longer period), while an early estimate now is more involved: $r \approx (\beta - \gamma) \, (Y / I)$.

For the growth of *S,* there is the next section.

9.3. The key equation on Sp, with consequences for top and limit

In SIA we have the key relation $\mathcal{S} = \mathcal{S}_0 \, Exp[- R_0 \, (\mathcal{A}_p - \mathcal{A}_p[0])]$. We back-traced this via $A' = \gamma \, I,$ and substituting $I = A' / \gamma$ in $S' = -\beta \, Sp \, I = R_0 \, Sp \, A'$.

For SEYCD, we have $A' = C' + D' = \gamma \, Y,$ and $S' = -\beta \, Sp \, Y.$ Thus we find: $S' = R_0 \, Sp \, A',$ which is the same expression as in SIA.

Thus we have a relation for 2 of the 3 SIA compartments that is unchanged, as in SIA.

There are two consequences, one in combination with the top of Iq.

9.3.1. For the limit

For the limit, we have {Sp[∞], 0, Ap[∞]} like in SIA, with the same R_0 and dependency upon the starting values. In the limit both *E* and *Y* are zero, and Cp[∞] and Dp[∞] are like SICD.

Thus we have all limit values {Sp[∞], 0, 0, Cp[∞], Dp[∞]}.

With the same parameters, there is no *need* for a separate SEYCD[Limit] routine, but it is obviously convenient, and provided.

```
SEYCD[Limit] // Explain
```

{S → 0.0198273, E → 0, Y → 0, C → 0.96547, D → 0.0147026, N → 0.985297, I → 0, A → 0.980173}

9.3.2. For the top

We best discuss the top of *I* separately.

9.4. The top of *I* at *I'* = 0 is unchanged

9.4.1. Basics unchanged

We find a top of *I* when its growth is zero. At the start, *I = E* and *Y* = 0. Depending upon the value of $\beta \, Sp - \gamma$ it will grow or decline. This analysis is no different than in SIA. If $R_0 > 1$ and given that $Y \ne 0$, we find the same $Sp_{Top \, Iq} = 1 / R_0$ as in SIA.

In steps: $I' = E' + Y' = 0$ or $I' = (- S' - \alpha E) + (\alpha E - \gamma Y) = - S' - \gamma Y = (\beta \, Sp - \gamma) \, Y = 0$.

9.4.2. $\{S, \ I, \ A\}_{Top}$ are the same

For the top of I, $Sp = 1 / R_0$ again implies $Ap = Log[R_0] / R_0$ (with the same nuance on starting values).

With the same parameters, it is doubtful whether there is need for a separate SEYCD routine.

```
SIA[Ip, Top, R₀] == SIA[Ip, Top, 4.] (* not inserting the starting values *)
```
$$\left\{ \frac{1}{R_0}, \ \frac{R_0 - \log(R_0) - 1}{R_0}, \ \frac{\log(R_0)}{R_0} \right\} = \{0.25, \ 0.403426, \ 0.346574\}$$

```
tTopI = TforTopIq[]
```
58.1472

```
pnt = SEYCD[tTopI] // Explain
```
$\{S \to 4.35 \times 10^6, \ E \to 710\,236., \ Y \to 6.30941 \times 10^6, \ C \to 5.9399 \times 10^6,$

$\quad D \to 90\,455.3, \ N \to 1.73095 \times 10^7, \ I \to 7.01964 \times 10^6, \ A \to 6.03036 \times 10^6\}$

The values $Eq[t]$ and $Yp[t]$ can be found by applying the t that applies to the TopIq. Let us try to find an analytical expression though.

9.4.3. To plot I or not to plot I

Thus, I still is an important variable for the process. Perhaps it is no longer plotted, traditionally, when only E and Y are plotted. In such an approach, researchers might doubt whether I really is a relevant variable, and perhaps only a legacy of SIA. However, it appears that I still is an important indicator for the overall behaviour of the infection.

9.4.4. Main point: delay in the top

In SEYCD, every person who gets infected still has the same infectious period. The difference is only the time delay. While the Iq[t] of SIA and SEYCD differ at a particular point in time, the integral values are still the same, whence the overall properties remain the same: which sameness is caught by the parameter values.

(At first, there is a bit of a paradox. In SIA, the I stands for *infected = infectious,* and in SEYCD this is no longer the case, with $I = E + Y$. But the same value of the top still applies, how can this be ? The cause is that we maintain the same $1/\gamma$ for the infectious period. What is important is that the horizontal axis shifts, not necessarily the vertical one.)

9.4.5. Trying to find analytical values for E and Y at the top

The following will not work, but it gives some information anyhow.

Take $Y' = \alpha E - \gamma Y$, and use $I = E + Y$ to also get $I' = E' + Y' = 0$ at the top, to that $E' = - Y'$. Thus we have a system of two differential equations:

$\{ Y' = \alpha E - \gamma Y, E' = - Y' \}$

Writing x for the exposed:

```
res =
 DSolve[{y'[t] == α x[t] - γ y[t], x'[t] == -y'[t]}, {x[t], y[t]}, t] // FullSimplify
```

$$\left\{\left\{x(t) \to \frac{e^{-t(\alpha+\gamma)}\left(\alpha c_1 - \gamma c_2 + \gamma (c_1 + c_2) e^{t(\alpha+\gamma)}\right)}{\alpha + \gamma}, \ y(t) \to \frac{e^{-t(\alpha+\gamma)}\left(-\alpha c_1 + \gamma c_2 + \alpha (c_1 + c_2) e^{t(\alpha+\gamma)}\right)}{\alpha + \gamma}\right\}\right\}$$

Since we are evaluating at the point at the top (we assumed $I' = 0$), let us take $\tau = 0$. The following holds for any parameters α and γ.

```
res /. t → 0 // Simplify
```

$$\{\{x(0) \to c_1, \ y(0) \to c_2\}\}$$

This merely rearranges around the point of approximation. We haven't found a simplifying analytical result, like for example:

```
{Eq[t], Yq[t]} == {γ, α} / (α + γ) Iq[t] /. t → tTopI /.
  {γ → GammaSIA[], α → AlphaSEYCD[]} /. Options[SIA]
```

False

```
plottable[t_] =
 ({x[t], y[t]} /. res[[1]] /. C[1] → Eq[tTopI] /. C[2] → Yq[tTopI] /.
    {γ → GammaSIA[], α → AlphaSEYCD[]} /. Options[SIA] // Simplify)
```

$$\left\{638\,149. + 72\,086.\,e^{-1.1\,t}, \ 6.38149 \times 10^6 - 72\,086.\,e^{-1.1\,t}\right\}$$

```
Plot @@ {Flatten[{
    {Legended[Eq[t], "Exposed"], Legended[Yq[t], "Infectious"]},
    plottable[t - tTopI]
   }], {t, 40, 80}, PlotRange → {0, 10^7}, PlotStyle →
   {SkyBlue, Blue, {Dashed, SkyBlue}, {Dashed, Blue}}, AxesLabel → {Day, "Units"}}
```
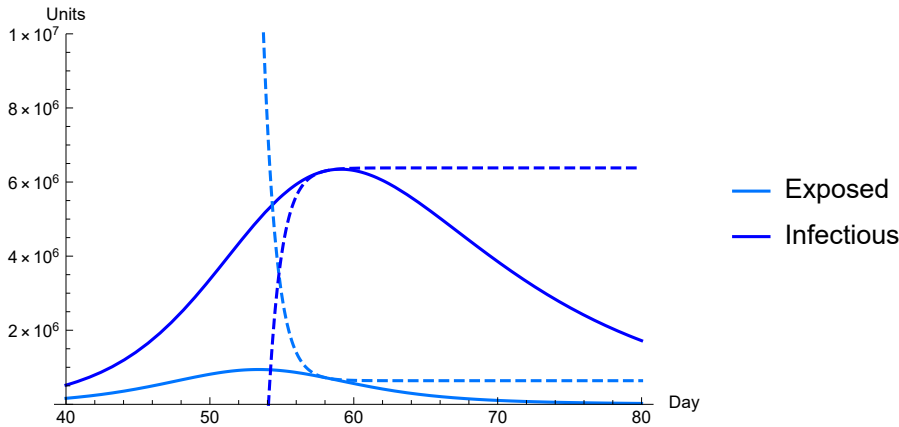

Local conclusion:

(i) These are exponential relations, and thus will not work outside of the particular point at the top.

(ii) In itself it may be a useful observation that e.g. *E* at the top of *I* differs from the value $\gamma / (\alpha + \gamma)\, I$. But this value has no analytical expression, as far as now, and thus we can do little with this. The idea was that we might perhaps spot an analytical relation, but we end up with an approximation at a point of an expression that already is an approximation …

The issue can be filed. The Appendix to this section contains the same story but then in terms of growth rates.

9.5. Summary on the top of *I* at *I'* = 0

The comments on the top of Iq are somewhat spaghetti because there are separate contributions from Sp and the derivative itself. Let us collect the consequences for the routines.

Once SEYCD[*t*] has been set, and all variables are available, we also have:

(a) The time when Iq reaches its top. For SIA(CD) we had 44.4 days. This is now 14 days later, due to only 1 day of latency per generation.

tTopI = TforTopIq[]

58.1472

(b) All values at the top of Iq are available via the routine. For SICD we had 90 thousand deaths on day 44, if there had been no intervention. The very same value now is delayed to day 58.

topI = SEYCD[tTopI] // Explain

$\{S \to 4.35 \times 10^6, E \to 710\,236., Y \to 6.30941 \times 10^6, C \to 5.9399 \times 10^6,$
$\quad D \to 90\,455.3, N \to 1.73095 \times 10^7, I \to 7.01964 \times 10^6, A \to 6.03036 \times 10^6\}$

SEYCD[Pr, tTopI] // Explain

$\{S \to 0.25, E \to 0.0408181, Y \to 0.36261, C \to 0.341374,$
$\quad D \to 0.00519858, N \to 0.994801, I \to 0.403428, A \to 0.346572\}$

(c) There is the more involved comparison of the theoretical values and the actual approximations. There is no need to further develop this for SEYCD since we have no relation (yet) for a theoretical expectation other than for SIA(CD).

NB. We can call the routine for SICD since it uses the same variable names. Observe however that the input is not a pure SICD model, since the variables derive from a SEYCD calculation. This does not make a difference for the algorithm.

TopIq[SICD, t] // Quiet

$\{t \to 58.1472, \text{Proportion} \to$

	Sp	Ip	Cp	Dp
From β and γ Equations	0.25	0.403426	0.341375	0.0051986
	0.25	0.403428	0.341374	0.00519858
(1) − (2)	0	-1.42684×10^{-6}	1.40543×10^{-6}	2.14025×10

$\text{Dqt0} \to 0, \text{List} \to \begin{pmatrix} 0.25 & 0.403426 & 0.341375 & 0.0051986 & 0.346574 \\ 0.25 & 0.403428 & 0.341374 & 0.00519858 & 0.346572 \\ 0 & -1.42684 \times 10^{-6} & 1.40543 \times 10^{-6} & 2.14025 \times 10^{-8} & 1.42684 \times 10^{-6} \end{pmatrix},$

$\text{AcceptableErrorQ} \to \text{True}\}$

SIA[Explain, Top]

SIA[Ip, Top, f, atq0:0] gives the top of Ip for f = R0 formally: {1, f − 1 − Log[f], Log[f]} / f + {0, −atq0, atq0}
A related routine is:
TopIq[SIA, ...] for the t when Iq has its top when R0 > 1 (inverse problem)
    with comparison of theoretical values and those generated by the interpolation run
TforTopIq[...] applies FindRoot to the equation Sp[t] == 1/R0

(d) There is no analytical expression *E* and *Y* yet.

9.6. The subsequent issues concern *E* and *Y,* and no longer *S, I* and *A*

PM. Observe that it is *easier* to make these analytical points by the choice of variables *I* = *E* + *Y*. We have been able to directly refer to earlier results in SIA without having to translate variable names. The mathematical result stands, whatever we call the variables.

9.7. The top of *E,* or when *E' =* 0, giving routine TforTopEq

At the top of *E,* the derivative is *E'* = - *S'* - α *E* = 0. To find *t,* we rather have variables in levels than in derivatives.

With *S'* = -β *Sp Y* we get α *E* = β *Sp Y,* or *E* / *Y* = β/α *Sp,* or α / β = *Sp Y* / *E,* where the latter might make a bit more sense since *E* starts with a positive Seed. There does not seem to be a further analytical solution for this. The equation can be solved by numerical approximation. There appear to be actually two solutions, see the intersections in the plot below.

```
{alpha, beta} = {AlphaSEYCD[], BetaSIA[]} /. Options[SIA];
tTopE = FindTime[Eq[t] == beta / alpha Sp[t] × Yq[t], {t, 50, 150}] // Quiet
```

53.3892

This has been turned into this routine, for ease.

```
TforTopEq[]
```

53.3892

```
? TforTopEq
```

Symbol
TforTopEq[t0:50, tmax:150] is FindTime for alpha
Eq[t] == beta Sp[t] Yq[t], with the parameters taken from Options[SIA]
⌄

9.8. The top of *Y,* or when *Y' =* 0, giving routine TforTopYq

9.8.1. Analytically

At the top of *Y,* the derivative is *Y'* = α *E* - γ *Y* = 0 and α *E* = γ *Y.*

Using *I* = *E* + *Y,* we also find *I* = (1 + α/γ) *E* or *E* = γ / (α + γ) *I.* Similarly, at this top: *Y* = α / (α + γ) *I.* This is comparable to the result in the appendix to this section. Thus we have part of an algebraic solution for the values at the top of *Y.* We can check the values in the plot.

```
{Eq, Yq} == {γ, α} / (α + γ) Iq    (* at the top of Y *)
```

$$\{Eq, Yq\} = \left\{ \frac{\gamma \, Iq}{\alpha + \gamma}, \frac{\alpha \, Iq}{\alpha + \gamma} \right\}$$

It seems like a lot but it hangs in the air since we do not know *I.*

9.8.2. PM. On taking second derivatives

PM. It is tempting to reason that when $\alpha E = \gamma Y$ then also $\alpha E' = \gamma Y'$, which in this case gets down to $E' = 0$ too. However, these are differential equations. In the graphs we can see that the tops are not at the same location. See the example where $y' = x + y$ also seems to suggest that $y' = 0$ implies (differentiating at a derivative) that $x' + y' = x' = 0$, while the solution shows that this only holds for a special constant $C = 0$. Instead $x' = y'' - y'$. (See https://math.stackexchange.com/question-s/537629/is-the-differential-equation-y-xy-separable.)

9.8.3. Numerically, giving routine TforTopYp

The equation $E / Y = \gamma / \alpha$ can be solved numerically for the top.

```
{alpha, beta, gamma} = {AlphaSEYCD[], BetaSIA[], GammaSIA[]} /. Options[SIA];
tTopY = FindTime[Eq[t] == gamma / alpha Yq[t] , {t, 50, 150}] // Quiet
```

59.086

For convenience:

```
TforTopYq[]
```

59.086

```
? TforTopYq
```

Symbol
TforTopYq[t0:50, tmax:150] is FindTime for alpha   Eq[t] == gamma Yq[t], with the parameters taken from Options[SIA]
⌄

9.8.4. Time between the tops of *E, I* and *Y*

The following is straightforward. The distances between the tops are not symmetric around *I.*

```
{tTopY - tTopE, " == ", tTopI - tTopE, " + ", tTopY - tTopI} // Row
```

5.69685 == 4.75807 + 0.938778

9.8.5. Substituting *Y* in *S'*

Substituting $Y = \alpha / (\alpha + \gamma) I$ into $S' = -\beta Sp Y = - (\alpha \beta / (\alpha + \gamma)) Sp I$.

At the top of *Y* the model behaves as a SIA model for *I,* with $\beta^* = \alpha \beta / (\alpha + \gamma)$ and $R_0 = \alpha \beta / (\alpha + \gamma) / \gamma$.

```
AlphaSEYCD[] × BetaSIA[] / (AlphaSEYCD[] + GammaSIA[]) / GammaSIA[] /. Options[SIA]
```

3.63636

However, this is local. Thus, substitution of this local R_0 into the formula will be a category mistake.

```
test = SIA[Ip, Top, R₀ ] == SIA[Ip, Top, TheFormer]
```

$$\left\{ \frac{1}{R_0}, \frac{R_0 - \log(R_0) - 1}{R_0}, \frac{\log(R_0)}{R_0} \right\} = \{0.275, 0.369979, 0.355021\}$$

The actual values at the top of Y are, in fact:

```
SIA[Pr, tTopY]
```

{0.218059, 0.401195, 0.380746}

```
SEYCD[Pr, tTopY] // Explain
```

{S → 0.218059, E → 0.0364723, Y → 0.364723, C → 0.375035,

   D → 0.00571119, N → 0.994289, I → 0.401195, A → 0.380746}

9.8.6. Plotting $\{\gamma, \alpha\} / (\alpha + \gamma)$ Iq

While $Y = \alpha / (\alpha + \gamma) I$ holds at the top of Y, the following plot shows that this is not generally the case. Though Y looks rather like a delay of I (which includes E), it is not *just* a delay.

```
{alpha, beta, gamma} = {AlphaSEYCD[], BetaSIA[], GammaSIA[]} /. Options[SIA];
Plot @@ {{
    Legended[ Yp[t], "Yp[t]"],
    Legended[alpha / (alpha + gamma) Ip[t], "α / (α + γ) Ip[t]"]}, {t, 0, 150}}
```
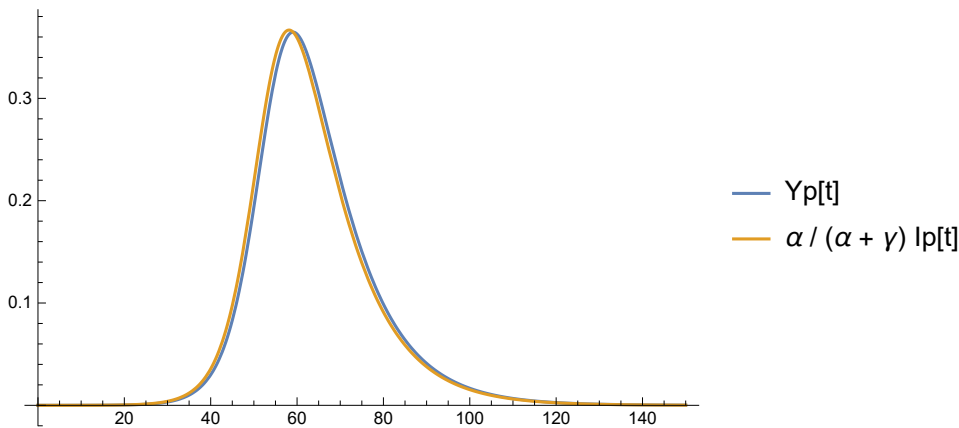

Just to be sure: the vertical axis for *E* is a factor 10 smaller than the vertical axis for *Y*.

```
{alpha, beta, gamma} = {AlphaSEYCD[], BetaSIA[], GammaSIA[]} /. Options[SIA];
Plot @@ {{
    Legended[ Ep[t], "Ep[t]"],
    Legended[gamma / (alpha + gamma) Ip[t], "γ / (α + γ) Ip[t]"]},
  {t, 0, 150}, PlotRange → All}
```
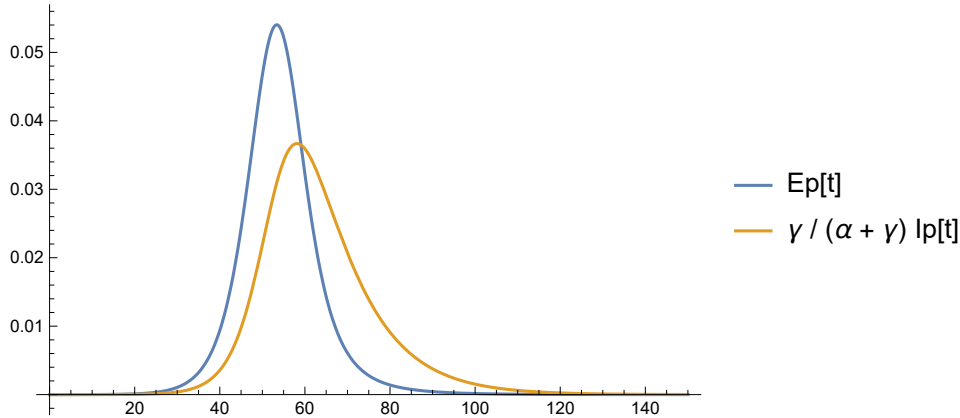

9.9. Plots for *E'* = 0 and *Y'* = 0 (but generally not both at the same time)

The following plots *Y* / *E* and the intersecting line and curve that determine the tops. (NB. *E* starts nonzero with the Seed, while *Y* starts at 0.)

```
{alpha, beta, gamma} = {AlphaSEYCD[], BetaSIA[], GammaSIA[]} /. Options[SIA];
Plot @@ {{
    Legended[Yq[t] / Eq[t] , "Y / E"],
    Legended[ alpha / gamma, "α / γ"],
    Legended[ alpha / beta, "α / β"],
    Legended[Yq[t] / Eq[t] * Sp[t], "Sp Y / E"]}, {t, 0, 150},
  PlotRange → {0, 1.2 alpha / gamma}
}
```
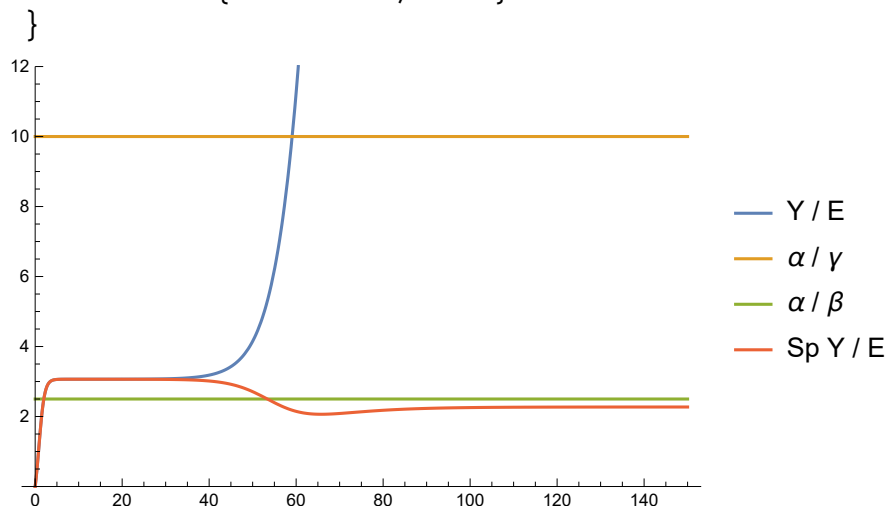

The following plots *E* / *Y* and the intersecting line and curve that determine the tops.

```
{alpha, beta, gamma} = {AlphaSEYCD[], BetaSIA[], GammaSIA[]} /. Options[SIA];
Plot @@ {{
    Legended[Eq[t] / Yq[t], "E / Y"],
    Legended[gamma / alpha, "γ / α"],
    Legended[beta / alpha * Sp[t], "ß / α Sp"]},
   {t, 0, 150}, PlotRange → {0, 1.5 beta / alpha}}
```
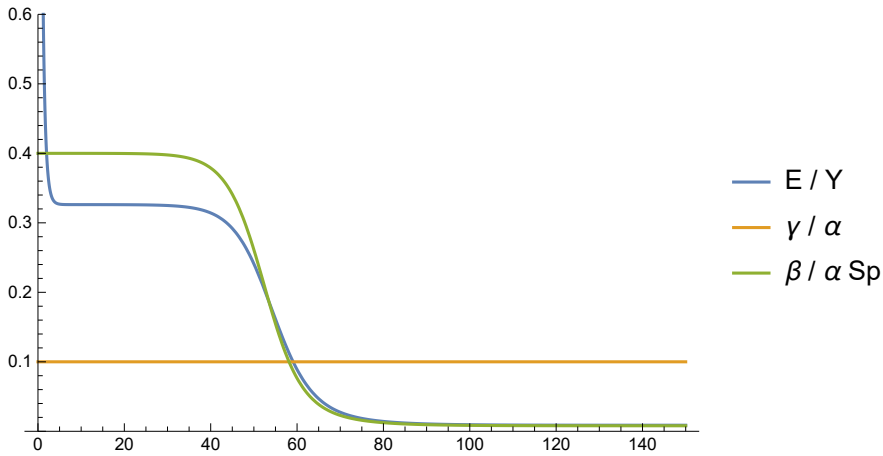

9.10. Conclusion of this chapter

The main relations on the limit values and the top of Iq do not differ from SIA, see the beginning of this chapter

Subsequently there are two numerical routines to find the time for the tops of Eq and Yq in SEYCD in particular.

There are no analytical results here.

Appendix. An exercise on growth (for the record only)

We have the rates of growth:

dLog[E] = E' / E = β Sp Y/E - α
dLog[Y] = α E/Y - γ

Thus we can eliminate E/Y:

β Sp = (dLog[E] + α) E/Y = (dLog[E] + α) (dLog[Y] + γ) / α. Thus:

Sp = (dLog[E] + α) (dLog[Y] + γ) / (α β)

And recall that also $Sp = f[R_0, Ap]$ above.

(1) Perhaps this might be helpful when we observe the rates of growth of E and Y, and know the parameters, so that we can estimate Sp. The growth rates are monotone but observation without error would be difficult.

(2) For the top of I, with $I' = 0$, we had all {S, I, A}p values, like $Sp = 1 / R_0$.

```
SIA[Ip, Top, R0]
```
$$\left\{ \frac{1}{R0}, \frac{R0 - \log(R0) - 1}{R0}, \frac{\log(R0)}{R0} \right\}$$

Then, using $(\alpha\,\beta)\,/\,R_0 = \alpha\,\gamma$:

$\alpha\,\gamma$ = (dLog[E] + α) (dLog[I - E] + γ) evaluated with I at *IqTop*.

This is a variant to the same equation that we already subjected to DSolve, above.

It is tempting to simply substitute the evaluated value as a constant c = *IqTop*. Then dLog[c - E] = -E' / (c - E).

However, I is a variable, and we have dLog[I - E] = (I' - E') / (I - E). However, we are evaluating at the point I' = 0, and we can substitute these values, which gives -E' / (c - E) anyhow. (Recall the equation for dLog[I] and that we took the value $\beta\,Sp = \gamma$.) Thus we arrive at a proper expression for an approximation at the top of I.

Mathematica finds a solution for this equation (a local relation at the top of I). With x for the exposed, and assuming y'/y = -x' / (constant - x), and using "I" for the value at the top:

```
dsol =
 DSolve[α γ == (x'[t] / x[t] + α) (-x'[t] / ("I" - x[t]) + γ), x[t], t] // FullSimplify
```

$$\left\{\{x(t) \to c_1\},\ \left\{x(t) \to \frac{\gamma\,I}{\alpha + \gamma} + c_1\,e^{-t\,(\alpha+\gamma)}\right\}\right\}$$

The equation without parameters is:

```
eq = "E" == (x[t] /. dsol[[2]]) /. topI /.
    {α → AlphaSEYCD[], γ → GammaSIA[]} /. Options[SIA]
```

$710\,236. = 638\,149. + c_1\,e^{-1.1\,t}$

The following generates a huge number.

```
nsol = Solve[eq /. t → tTopI, C[1]]
```

$\{\{c_1 \to 4.32687 \times 10^{32}\}\}$

Let us put in t = 0 for a local approximation.

```
nsol = Solve[eq /. t → 0, C[1]]
```

$\{\{c_1 \to 72\,086.\}\}$

Thus an approximate relation for E at the top is, provided that the t = 0 corresponds to t = tTopI ≈ 58.

```
Eprox[t_] = eq[[2]] /. nsol[[1]]
```

$638\,149. + 72\,086.\,e^{-1.1\,t}$

```
{Eq[t], Eprox[t - tTopI]} /. t → tTopI
```

$\{710\,236.,\ 710\,236.\}$
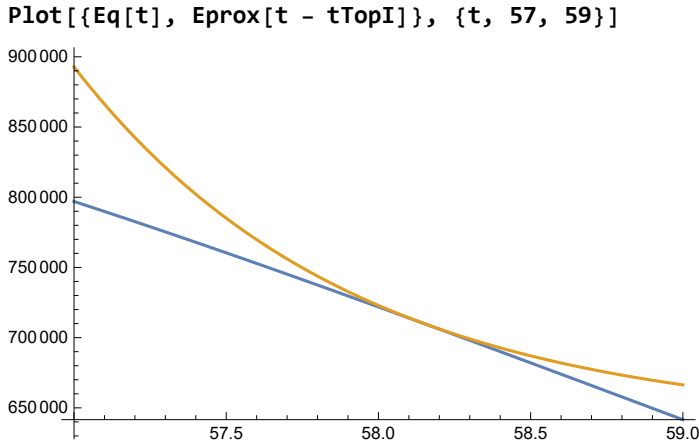
```
Plot[{Eq[t], Eprox[t - tTopI]}, {t, 57, 59}]
```


The idea was that we might perhaps spot an analytical relation, but we end up with an approxima-tion at a point of an expression that already is an approximation …

PM. Mathematica has no result for these joint equations, while the above shows the solution, using a small substitution (elimination of a variable).

```
dsol = DSolve[{α γ == (x'[t] / x[t] + α) (y'[t] / y[t] + γ), C[1] == x[t] + y[t]},
   {x[t], y[t]}, t] // FullSimplify
```

$$\text{DSolve}\left[\left\{\alpha\, \gamma = \left(\alpha + \frac{x'(t)}{x(t)}\right)\left(\gamma + \frac{y'(t)}{y(t)}\right), c_1 = x(t) + y(t)\right\}, \{x(t), y(t)\}, t\right]$$

10. Infections versus symptoms

10.1. Introduction

It is basic to distinguish between *infection* and *symptoms*. SI(EY)A(CD) modeling focuses on the biological mechanism of infection. There are authors who apply the SI(EY)A(CD) model to symp-tomatic disease however. The key consideration remains that symptoms have an impact on behaviour and hospital treatment and reporting, and thus are "easier to measure" than infection status.

Leonhard Euler in 1767 already formulated a "*renewal equation*" for population dynamics, and the SI(EY)A(CD) model appears to be a reformulation in terms of (ordinary) differential equations (ODE). Relevant papers are Fine (2003), Svensson (2005, 2007), Breda et al. (2012) and Champredon, Dushoff & Earn (2018). The latter paper is quite recommendable, though they regrettably do not refer to Svensson while a comparison of results would have been informative.

It appears that the discussion can be structured best by first discussing the distinction between infection and symptoms, and in particular the distinction between the *generation interval* and the *serial interval.* When this has been clarified, we can focus on formulas for infections only. The following discusses what aspects of this analysis have been included in the package, and compares parameter estimates by different authors: now with greater attention for the distinction between infections and symptoms.

10.2. Comparing IFF and sCFF

With the Chinese *infection fatality factor* (IFF) of 0.66 (Verity et al. (2020)), and the adjusted UK IFF of 0.9 (Ferguson et al. (2020)), the Goh (2020) parameter of 2% concerns symptoms and is the *symptomatic Case Fatality Factor* (sCFF). (They call it a rate but it is a factor.) When one misses many cases without symptoms then the sCFF indeed is higher than the IFF.

We can use: IFF = D / I = (D / sC) * (sC / I) = sCFF * #symptomatic cases per infected.

Thus we have 0.66% = 2% * sC / I, so that sC / I = 0.33 symptomatic cases per infected, or three infected units per 1 symptomatic unit.

However, when we change to the Dutch IFF of 1.5%, we cannot simply assume that this ratio stays the same.

There remains the question whether symptoms might also have been caused by common flu. There is not much discussion in this literature about confounding with flu symptoms, even though January had the flu season in China. However, authors refer to "confirmed cases", which implies microbiological testing, which would exclude such confounding (at the reporting stage at least).

10.3. Infection versus symptomatic disease

10.3.1. Arrival periods for a single unit

The basic distinction is:

- An infection starts with a *latency* period, during which one isn't infectious, and continues with the infectious stage.

- A symptomatic disease starts with an *incubation*, during which one doesn't have symptoms yet, and continues with the symptomatic stage.

We may assume that the total lengths of both paths are the same. This is a big assumption. A practical explanation for modeling is that when a person is put into quarantine, then for the purpose of modeling both periods end. However, this is more complex when we are interested in the period before intervention. For SARS-CoV-2 the incubation period will tend to be longer than the latency period: whence there are *asymptomatic infectious* cases. For the following diagram we rather not introduce new colours but can use *colour bands* that fit the previous colour scheme.

QuarantineColours["Period"]

While each unit will have its own values for these periods, the diagram can also display the population averages.

With lack of microbiological testing, symptoms may be all that we have. There are key issues in measurement. (a) Persons with symptoms might also suffer from another (flu) infection. (b) When

we ask an infectee about possible infectors in the past, then the infectee will tend to mention persons with symptoms, who can be recognised because of the symptoms, and they might overlook infectious persons without symptoms (who convey an impression of health). Thus, already on those two counts, there is every reason to be cautious with data using symptoms. It tends to make little sense to test 100% of the units with microbiological tests when only 1% show symptoms. For SARS-CoV-2, doing such tests can be cheaper than a lockdown of the economy however.

PM 1. It would also be possible that latency would be longer than incubation, so that persons with symptoms still might not be infectious at the beginning. Perhaps this applies for a subgroup in the SARS-CoV-2 case too (just a comment from theory). In that case, the "average" will hide quite some differences in the population. However, there seem to be no reports on this and thus we neglect this possibility.

PM 2. For a model with hospitalisation, we would have to include a period when a person has cleared the infection but still suffers from symptoms related to the disease, for example when the pneumonia has taken hold, or when the immune system has started to attack the own body. It depends upon what one calls symptoms (being in a hospital might be a symptom). But we do not regard hospitalisation here.

10.3.2. Interarrival periods between two units

While it is *relatively* straightforward to measure the durations of symptomatic periods, the infection status is rather elusive. A strategy is to look at the *interarrival* periods (IAP) between *two* cases, i.e. the infector and the infectee, as opposed to the *arrival* periods and rates for single cases. The relevant terms are the **generation interval** for infections and the **serial interval** for symptoms.

- The generation interval is given by the period between the *moments of infection*, **from** the infection of the infector **to** when the infector transfers the virus to the infectee so that a new generation starts.

- Du et al. (2020): "The serial interval of COVID-19 is defined as the time duration between a primary case-patient (infector) having symptom onset and a secondary case-patient (infectee) having symptom onset."

In the following diagram, the infector and infectee both have a latency of 4 days. The infector has an incubation of 8 days: thus an asymptomatic period of 8 - 4 = 4 days. The infectee has an incubation of 10 days: thus an asymptomatic period of 10 - 4 = 6 days. The infection happens on day 7, thus with a *generation interval* of 7. Both units have a symptomatic period of 10 days. The *serial interval* is equal to the generation interval plus the difference of the asymptomatic periods, thus 7 + (6 - 4) = 9, as can be shown by the following formulas.

```
QuarantineColours["Period", 2]
```



```
Results[QuarantineColours, "Period", 2]
```
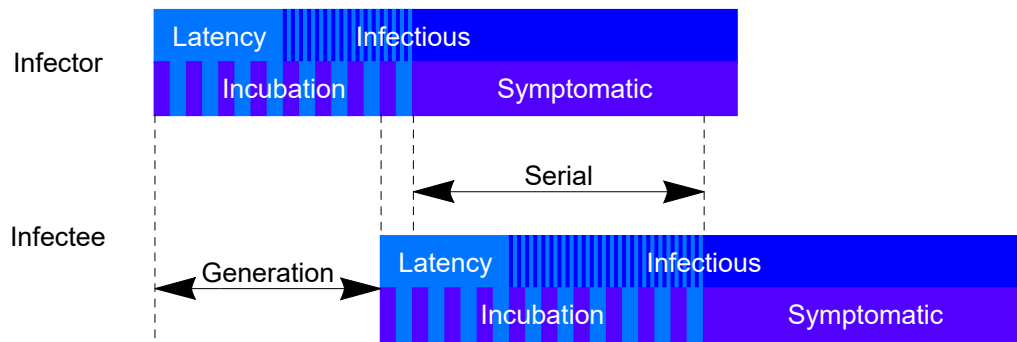
{GenerationInterval → 7, SerialInterval → 9}

Let us use the following variables:

(i) $inc_{1,2} = lat_{1,2} + a_{1,2}$ for the incubation, latency and asymptomatic periods of infector (1) and infectee (2)

(ii) $g = lat_1 + y$ for the generation interval, latency of the infector, and the period to the actual infection y

(iii) $s = (g - inc_1) + inc_2$ for the serial interval, correcting g for the incubation of the infector and the infectee

The formula for the serial interval then is:

$$s = (lat_1 + y - (lat_1 + a_1)) + (lat_2 + a_2) = y + lat_2 + (a_2 - a_1)$$

The assumption that $g = s$ is frustrated by differences in the periods of asymptomatic infectiousness. For example, the serial interval would be negative if the infector develops symptoms after the infectee.

10.3.3. Complication for the population averages

While above calculation was done at the micro level, it also applies to the macro level, and then the variables represent *means*. This has the special effect that y becomes the *mean infectious period,* so that g becomes the *transmission period,* or the whole length of the scheme (interpreting these as averages). Subsequently, we must assume *either* that the infectors and infectees come from the same population and thus have the same means, *or* that there is a structural difference. For population averages, we would tend to assume that both the infector and the infectee would have the characteristics of the averages, but this can also be an intellectual pitfall, e.g. when infectors would be structurally younger than infectees.

If the infector and infectee have the same aysmptomatic infectious period, then the generation period and the serial interval have the same duration. The following diagram displays the average situation, in which the infection happens at the *mean* moment, which is at the end of the shown bar. They both have the same asymptomatic periods. The incubation for both is 8 and the symptomatic period is 10, and thus the generation interval and serial interval are both 18.

```
QuarantineColours["Period", 2, 18, 4]
```



```
Results[QuarantineColours, "Period", 2]
```
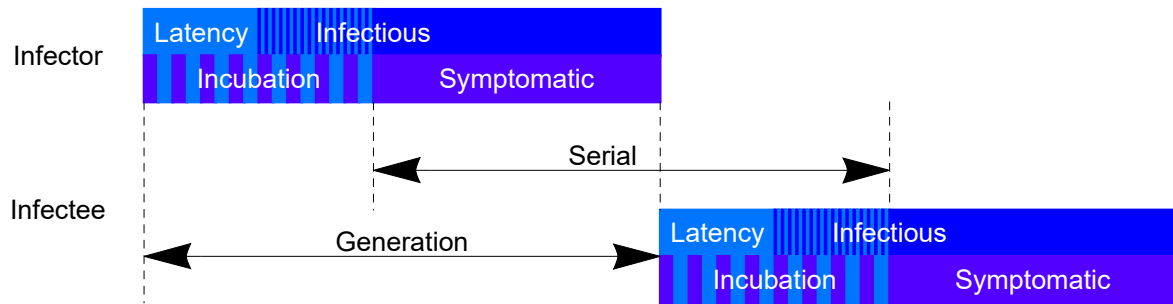
{GenerationInterval → 18, SerialInterval → 18}

The serial interval is officially an interarrival period because there are two persons involved, but in practice it is the mere addition of arrival periods, with the symptomatic phase of the infector and the incubation of the infectee, and using that the onset of symptoms is a rather reliable point of observation. The serial interval is primarily useful for the observations at the micro-level, where cases show variety. These display routines calculated these values from the input data but in practice researchers determine onsets of symptoms, try to match infectees with their infectors, and apply statistics to the observed serial intervals to determine the mean periods in the scheme. At the summation level, the serial interval either defines the transmission interval (the whole range) or must be made consistent with such other estimates.

Unless we model different compartments, there is no structural difference in the asymptomatic periods of infectors and infectees, and in practice we would expect $a_2 = a_1$, so that the effect of variation shows up in the variances only. However, if infectors are *structurally* healthier persons without symptoms and if infectees with symptoms are *structurally* the more vulnerable ones, then there can be a structural difference between the asymptomatic periods.

A problem with SARS-CoV-2 is that there are *many asymptomatic infectious.* The infectee might point to a possible infector who is not the relevant person. The SARS-CoV-2 literature shows authors reporting on the serial interval, because this is a commonly used statistic, but the same authors acknowledge that the statistic would not be so relevant without more information about the asymptomatic period. When it is reported that the serial interval is shorter than the incubation period, then this does not make sense in the second diagram.

Tindale et al. (2020): " Results: The mean incubation period was 7.1 (6.13, 8.25) days for Singapore and 9 (7.92, 10.2) days for Tianjin. Both datasets had shorter incubation periods for earlier-occurring cases. The mean serial interval was 4.56 (2.69, 6.42) days for Singapore and 4.22 (3.43, 5.01) for Tianjin. We inferred that early in the outbreaks, infection was transmitted on average 2.55 and 2.89 days before symptom onset (Singapore, Tianjin). The estimated basic reproduction number for Singapore was 1.97 (1.45, 2.48) secondary cases per infective; for Tianjin it was 1.87 (1.65, 2.09) secondary cases per infective. Conclusions: Estimated serial intervals are shorter than incubation periods in both Singapore and Tianjin, suggesting that pre-symptomatic transmission is occurring. Shorter serial intervals lead to lower estimates of R0, which suggest that half of all secondary infections should be prevented to control spread."

Du et al. (2020) report that 12.6% cases (59 of 468) have an asymptomatic origin: "Fifty-nine of the 468 reports indicate that the infectee had symptoms earlier than the infector. Thus, presymptomatic transmission might be occurring. Given these negative-valued serial intervals, COVID-19 serial intervals seem to resemble a normal distribution more than the commonly assumed gamma or Weibull distributions, which are limited to positive values [ref]." Du et al. (2020) indeed show a normal distribution for the serial interval, ranging between -10 and 20 and a mean at 4 days.

The average of 4 is remarkable. Anderson et al. (2020) have an incubation of 5-6 days, and Backer et al. (2020) (connected authors) have 6.4 days. Let us take as incubation for symptomatic disease: 6.4 days. For our own set of parameters we can calculate a generation interval of 11 days. This difference of 11 - 4 = 7 days suggests that *more* than 12.6% cases may have had an asymptomatic infection. At play is that we use γ = 1/10, which is not what these authors presume. The problem may lie in the incubation period. The incubation period was observed for the infectees and not for the infectors. When authors report an incubation period they should clarify for whom.

With $s = (g - inc_1) + inc_2$ we get 4 = (11 - inc_1) + 6.4 we find inc_1 = 13.4, meaning that the infectors need not develop *any* symptoms.

Single studies may not easily fit within a larger scheme though.

10.4. Additive equations in the package

For choosing values in the Options[SIA] the following equations may be handy.

We already had:

```
(Take[ SEYCD["R0[t]"], 5]) // MatrixForm
```

$$
\begin{pmatrix}
R0(t) = \frac{BetaSIA(t)}{GammaSIA()} \\
GammaSIA() = GammaSICD() + MuSICD() \\
Period(Infectious) = \frac{1}{GammaSIA()} \\
Period(Latency) = \frac{1}{AlphaSEYCD()} \\
Period(Transmission) = Period(Infectious) + Period(Latency)
\end{pmatrix}
$$

New concepts are the *interarrival* periods.

```
(Take[SEYCD["R0[t]"], -5]) // MatrixForm
```

$$
\begin{pmatrix}
Period(Generation) = InfectionInterval() + Period(Latency) \\
Period(SerialInterval) = Period(Transmission) \\
Period(SerialInterval) = Period(Incubation) + Period(Symptomatic) \\
Period(Asymptomatic\ infectious) = Period(Incubation) - Period(Latency) \\
Period(Symptomatic) = Period(At\ hospital) + Period(Delay\ to\ hospital)
\end{pmatrix}
$$

Our values for all equations are:

```
(eqs2 = SEYCD["R0[t]"] /. {R0[t] → 4., Period["Latency"] → 1, GammaSIA[] → .1,
    MuSICD[] → .0015, Period["Incubation"] → 6.4}) // MatrixForm
```

$$
\begin{pmatrix}
4. = 10. \, \text{BetaSIA}(t) \\
0.1 = \text{GammaSICD}() + 0.0015 \\
\text{Period(Infectious)} = 10. \\
1 = \dfrac{1}{\text{AlphaSEYCD}()} \\
\text{Period(Transmission)} = \text{Period(Infectious)} + 1 \\
\text{Period(Generation)} = \text{InfectionInterval}() + 1 \\
\text{Period(SerialInterval)} = \text{Period(Transmission)} \\
\text{Period(SerialInterval)} = \text{Period(Symptomatic)} + 6.4 \\
\text{Period(Asymptomatic infectious)} = 5.4 \\
\text{Period(Symptomatic)} = \text{Period(At hospital)} + \text{Period(Delay to hospital)}
\end{pmatrix}
$$

```
(List @@ Reduce[eqs2, {Period["SerialInterval"]}]) // MatrixForm
```

$$
\begin{pmatrix}
\text{BetaSIA}(t) = 0.4 \\
\text{GammaSICD}() = 0.0985 \\
\text{AlphaSEYCD}() = 1 \\
\text{InfectionInterval}() = \text{Period(Generation)} - 1 \\
\text{Period(Asymptomatic infectious)} = 5.4 \\
\text{Period(Transmission)} = 11. \\
\text{Period(Symptomatic)} = 4.6 \\
\text{Period(Infectious)} = 10. \\
\text{Period(At hospital)} = 4.6 - 1. \, \text{Period(Delay to hospital)} \\
\text{Period(SerialInterval)} = 11.
\end{pmatrix}
$$

10.5. Formulas for generation interval

10.5.1. Theory

Champredon at al. (2018): "The renewal [Euler] and ODE [SI(EY)A(CD)] approaches are based on different conceptualizations of dynamics. The renewal approach focuses on cohorts of infectious individuals, and how they spread infection through time, while the ODE approach focuses on counting individuals in different states. The renewal equation is less common than compartmental models in epidemiological applications, probably because the goal when modelling epidemics is often to identify optimal intervention strategies, which is facilitated by clearly distinguishing the various epidemiological states (e.g., susceptible, infectious, immune, vaccinated, quarantined, etc.) to act on. However, the simplicity of the renewal equation makes it particularly well adapted to estimate the effective reproduction factor from incidence time series [25] and to forecast epidemics [8]. As a notable example, it was used recently by the WHO Ebola Response Team to estimate the reproduction factor the Ebola epidemic [27]" (PM. Their manuscript is a preprint and "for review purposes only".)

Equally succinctly: "The generation interval [of infection between infector and infectee] is rarely observed, but through contact tracing it is possible to directly observe the serial interval (i.e., the interval of time between onset of symptoms for the infector and her/his infectee). Although different in theory, the serial interval distribution may be a good approximation to the generation interval distribution, especially for diseases for which the latent and incubation periods are similar (Appendix D and [14]). On the other hand, the latent and infectious periods —which are used to parametrize compartmental ODE models— can be observed only in clinical studies, which are more rare. Consequently, the [serial interval] distribution can be easier to obtain than the distributions of

latent and infectious periods, in which case a renewal equation might be easier to parameterize than an Erlang SEIR ODE model." (PM. Their manuscript is a preprint and "for review purposes only".)

Check Svensson (2005:9-10) and Champredon, Dushoff & Earn (2018:22) for their diagrams and compact notation, with also a statement on the variances of the relevant periods. Our own package is not so compact and uses "long names", see below.

10.5.2. Mean interarrival period (MIAP) for infections

The MIAP of the infection by the infector to the infectee is called the *generation interval.* In our case, the generation interval consists of the *sum* of two exponentially distributed variables, namely the exposed phase with mean $1/\alpha$ and the infectious phase with mean $1/\gamma$, of which the pdf are well-known. Svensson (2005, 2007) provides relevant formulas and a substitution for the some special cases, see below, like the gamma distribution. Both Breda et al. (2012) and Champredon et al. (2018) give this explicit formula for (intrinsic) generation pdf as we have been using (assuming $\alpha \neq \gamma$):

```
g[t] == Hold[α γ / (α - γ)] (Exp[-γ t] - Exp[-α t])
    (* stopping Mathematica on the fraction*)
```

$$g(t) = \left(e^{\gamma(-t)} - e^{\alpha(-t)}\right) \text{Hold}\left[\frac{\alpha\,\gamma}{\alpha - \gamma}\right]$$

For our parameters $\gamma = 1/10$ and $\alpha = 1$, we use our "SumTwoExp" function (and observe that it does not matter in what order the parameters are entered).

```
SumTwoExp[γ, α, t]
```

$$\frac{\alpha\,\gamma\left(e^{\gamma(-t)} - e^{\alpha(-t)}\right)}{\alpha - \gamma}$$

```
Plot [SumTwoExp[.1, 1, t], {t, 0, 20},
 AxesLabel → {Day, "PDF"}, PlotLabel → "Generation interval"]
```
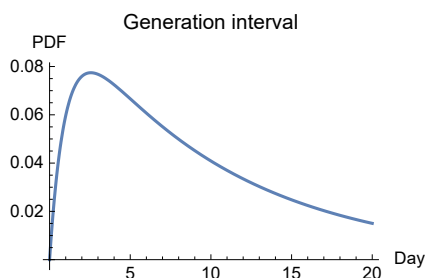

The expected value can be found by the usual integration, or substitution of the known outcome for the exponential distribution per component:

```
α γ / (α - γ) (1 / γ^2 - 1 / α^2) == 1 / α + 1 / γ
```

$$\frac{\alpha\,\gamma\left(\frac{1}{\gamma^2} - \frac{1}{\alpha^2}\right)}{\alpha - \gamma} = \frac{1}{\alpha} + \frac{1}{\gamma}$$

```
% // FullSimplify
```

True

PM. https://en.wikipedia.org/wiki/Exponential_distribution#Sum_of_two_independent_exponential_random_variables

```
PDF[ExponentialDistribution[λ], t]
```

$$\begin{cases} \lambda\, e^{\lambda\,(-t)} & t \geq 0 \\ 0 & \text{True} \end{cases}$$

```
Mean[ExponentialDistribution[λ]]
```

$$\frac{1}{\lambda}$$

10.5.3. Infections: Primary and secondary generation interval

Svensson (2005, 2007) provides us with the following for the "mean interarrival period" (MIAP).

In the following, we will put in the arrival period and find the derived interarrival period. In empirical testing, we will observe the interarrival period (e.g. the generation interval), and then must solve backwards to the arrival period that is relevant for the model.

An infection can be looked at from the position of the infector ("primary") or from the position of the infectee ("secondary").

(Pp) Infectors may have contacts that do not result into infections, which causes an "infinite interarrival period". Svensson's formula from this perspective on the generation interval takes the period of latency plus half of the mean infectious period (since half will be infected before and half after this mean). Thus *Pp = Platency + Pinfectious* / 2 (or *Pp = lat + inf* / 2 in above symbols).

(Ps) The secondary viewpoint, from the infectee, considers only contacts that result into infections. Apart from a general setup, there is a specific formula for a *constant infectious rate* with *random arrival period* with mean *mu* and standard deviation *sd*, following a period of latency with mean *Platency.*

```
miap[constant rate] == InfectionInterval[Platency, mu, sd]
```

$$\text{miap(constant rate)} = \frac{mu^2 + sd^2}{2\,mu} + \text{Platency}$$

If we allow a constant period as a particular case of a random period, then:

```
miap[constant period too] == InfectionInterval[Platency, Pinfectious, 0]
```

$$\text{miap(constant period too)} = \frac{\text{Pinfectious}}{2} + \text{Platency}$$

And this is the result the mean and sd for the exponential distribution.

```
miap[exponential] == InfectionInterval[1/α, 1/γ, 1/γ]
```

$$\text{miap(exponential)} = \frac{1}{\alpha} + \frac{1}{\gamma}$$

For the exponential case, relevant for SI(EY)A(CD), the transmission period is also the generation

interval (MIAP). We already derived this above but it is useful to draw attention to the other formula with mean and standard deviation.

There can be observational equivalence. Different models might explain the same phenomena. To discriminate situations, observations on the serial interval or MIAP would need to determine not only a mean infectious period but also how those are distributed over time (e.g. a constant interval with a uniform distribution of cases in it, or exponential decay). The assumption for exponential decay is that the virus may remain available forever but only doesn't become epidemic because of the low number of contacts.

10.6. Ganyani et al. (2020) on the generation interval

Writing in May 2020, with so *many* already infected in Holland, it is remarkable that studies refer to only a *few* data points also rather early in the pandemic in *Asia*, as if researchers in the West don't do source and contact tracing. Ganyani et al. (2020) appears to be relevant for us, since researchers of RIVM partook, and it matters for the comparison of findings about Holland. See section 2.9.3 on RIVM January 27.

Ganyani et al. (2020) refer to 91 confirmed SARS-CoV-2 cases in Singapore 2020-02-26, of which 54 cases provide adequate data, and which are bundled in four clusters. They also have 135 cases from Tianjin of 2020-02-27, of which 114 can be traced to one of 16 clusters. For some questions there are too few datapoints in Singapore. The authors: "quarantine and other containment measures were already in place at the time of data collection, which may inflate the proportion of infections from pre-symptomatic individuals."

Ganyani et al. (2020) assume that that infectors and infectees have the same average incubation periods. Subsequently, they fix the incubation period to a gamma distribution with mean 5.2 and sd of 2.8 days, referring to Zhang et al. (2020). Due to the variation, it is still possible to find a fraction of asymptomatic infections. A sensitivity analysis is performed with incubation averages of 4.8 (sd 2.6)and 6.4 (sd 2.3) days.

The authors acknowlegde (p7): "Our study does have some limitations. First, we rely on previous estimates for the incubation period. However, our sensitivity analyses showed that changing the incubation period distribution does not have a big impact on our estimates of the generation interval distribution." However, there is a distinction between putting in different values *under the assumption of sameness* for infector and infectee, and, alternatively, dropping that assumption and allowing that infectors have a different mean than infectees.

They find a mean "generation time" of 3.95 days for Tianjin and 5.2 days for Singapore. They find the same serial intervals, though with different standard deviations.

The authors (p4): "When allowing for negative serial intervals in the Tianjin data, the mean generation time decreased to 2.90 days (95% CrI: 1.85–4.12). The sensitivity analyses showed that the assumptions made about the incubation period have only moderate impact on the results. On the other hand, assumptions made about the underlying transmission network (e.g. acknowledging possibly negative serial intervals) had a large impact on our results. As expected, the proportion of pre-symptomatic transmission increased (...) from 62% (95% CrI: 50–76) to 77% (95% CrI: 65–87) for the Tianjin data."

"Hence, a large proportion of transmission appears to occur before symptom onset, which is an important point to consider when planning intervention strategies. It is worth noting that the outbreak data we used were collected in the presence of intervention measures such as case isolation and quarantining of identified contacts. This means that our estimates do not necessarily reflect the natural epidemiology of COVID-19, but instead reflect what is observed in the presence of these intervention measures. It is expected that these measures reduce the proportion of symptomatic transmission, which implies that a high proportion of infections is likely to have occurred before symptom onset because isolation prevents symptomatic transmission."

10.7. Kucharski et al. (2020) and Goh (2020) epidemic calculator on symptoms

The Kucharski et al. (2020) paper was used by Goh (2020) to select default parameters for his (recommendable) "epidemic calculator". Kucharski et al. (2020) present their results with reference to a SEIRD model, but with the special aspect that alongside the transmission from *Exposed* to *Infectious* there is a parallel *onset of symptoms*.

Their article, data and model actually appear to concern the *development of symptoms* rather than *infections.*

Thus their model adopts the structure of the SEIRD or SI(EY)A(CD) model but it is not the "pure" SEIRD in terms of infections. Perhaps there exist infections such that latency = incubation. I am not familiar with the history of S(E)IR(D) and it may be that this is quite conventional. For me it appeared a discovery, so to speak, to realise that the model, the data and results had to be interpreted differently than infections even though the model was claimed to be SEIRD.

Application to symptoms doesn't appear to change the functioning of the SEIRD or SI(EY)A(CD) model as such, or the usefulness of the Goh epidemic calculator. Below we will reproduce the core of that calculator (but not all of its features).

Points are:

- The Kucharski model (parameters used by Goh) has the structure of SEIR(D) or SI(EY)A(CD), but does not quite apply to the SARS-CoV-2 situation with the *infectious without symptoms*. They report: " Our results suggested there were around ten times more symptomatic cases in Wuhan in late January than were reported as confirmed cases (figure 2) ..." The problem however are also the asymptomatic cases that are not reported.

- Goh (2020) presents a "case fatality rate" (2%), which causes us to be alert about the distinction between the *Infection Fatality Factor* (IFF) and the symptomatic *Case Fatality Factor* (sCFF).

- At first it seems that the translation from Goh's webpage to our model is straightforward, since he provides the model and parameter values. However, he presents data for *incubation of symptoms,* which is not the *latency* which the SEIR(D) or SI(EY)A(CD) model is about. Thus, the translation to our model setup with infections is more involved. A practical solution is to merely run the model with their parameters, and disregard the content, to merely check the mechanics, and stop wondering about the meaning and policy implications.

- The Kucharski et al. (2020) diagram of the model (their figure 1) shows that symptoms develop between the Exposed and Infectious compartments. The Exposed are *non-symptomatic* as is stan-

dard. About the Infectious the legend explains: "*infectious (and symptomatic)*". It is correct to count the Exposed and latency period as part of *incubation,* but it is another issue to take them as equal.

On the *last page* of the Kucharski et al. (2020) paper we finally get clarity:

"We also assumed that the latent period is equal to the incubation period (ie, individuals become infectious and symptomatic at the same time) and all infected individuals will eventually become symptomatic. However, there is evidence that transmission of SARS-CoV-2 can occur with few reported symptoms. [ref] Therefore, we did a sensitivity analysis in which transmission could occur in the second half of the incubation period, but this did not change our overall conclusions of a decline in Rt from around 2·4 to almost 1 during the last 2 weeks of January."

The latter is an important conclusion for their paper, of course. It is not helpful for *our* purpose of reconstructing their model and finding the proper parameter values for latency and infectiousness. Subsequently, halving the incubation period generates (i) a part *exposed* (latency), (ii) a part *infectious and asymptomatic,* and (iii) a part *symptomatic and infectious*, which are three stages and is not our SEIR(D) or SI(EY)A(CD) model. Now, we could take "half of the incubation period" as the latency period, as Kurcharski et al. (2020) do for their sensitivity analysis, but based upon what data, and while it is not reproduced by Goh's epidemic calculator ?

It is not clear to me how the SEIR(D) or SEIA(CD) model was developed in history. This present notebook has not modeled symptoms and for us the following two models could apply:

- the Exposed are the units in latency, i.e. before being infectious (presentation in this notebook)

- the Exposed are units in latency = incubation, and afterwards are symptomatic = infectious (Kucharski et al.)

Under the assumption of the latter, we can reproduce the SEIR(D) or SI(EY)A(CD) part of Goh's epidemic calculator, but the relation with SARS-CoV-2 is limited, due to the asymptomatic infections. It is somewhat of a mathematical exercise, merely to show a reproduction of results. See section 11.4 for the run and plot.

This notebook does not look at symptomatic cases. The Kucharski et al. (2020) paper concerns these symptomatic cases, and the Goh epidemic calculator properly illuminates basic dynamics. Above comments are merely about the interpretation.

However, see below for the discussion of 10.8.4 Tao Liu et al. (2020), that was as the basis of the Kucharski et al. (2020) paper. With our assumption of $\alpha = 1$, and then *incubation - 1 / α = asymptomatic infectiousness*, we still can arrive at an indication what this study on symptoms may imply for our application to infections.

10.8. Other literature

Other papers in the literature mentioned in our references also tend to look at symptoms. They all rely upon verified SARS-CoV-2 cases, but such testing does not overcome the conventional problem of measuring the generation interval in a more direct manner, and thus the standard solution remains taking the route via the serial interval.

10.8.1. Anderson et al. (2020)

Anderson et al. (2020) is a review comment and no new empirical research. They suggest a serial interval of 4.4-7.5 days (paper) or 4-7 days (supplement 1). The term "serial interval"applies to symptomatic diseases, and it is more likely that they have this in mind, but in supplement 1 it is mentioned at the diagram for the infectious trajectory. They also mention an incubation period of 5-6 days. They correctly wonder that *means* of 5.5 for both *serial interval* and *incubation* result into a symptomatic period of mean length 0: "Combining this time with a similar length serial interval suggests there might be considerable presymptomatic infectiousness (appendix 1)."

If we assume that their serial interval is also the generation interval then we can apply Svensson's formula. The assumption of the SI(EY)A(CD) model is that the infectious period would be exponential, thus with *mu = sd.* The MIAP outcome of 5.5 days must be explained by these input parameters for the arrival times.

```
5.5 == InfectionInterval[1, 1/γ, 1/γ]
```

$$5.5 = \frac{1}{\gamma} + 1$$

An infectious period of 4.5 days seems rather short, not only in terms of the *Diamond Princess* but also common sense. A common cold may take 7 to 10 days, and those are corona viruses too.

Anderson et al. (2020) indeed comment: "The fourth uncertainty is the duration of the infectious period for COVID-19. The infectious period is typically short for influenza A, but it seems long for COVID-19 on the basis of the few available clinical virological studies, perhaps lasting for 10 days or more after the incubation period. [8]" The latter however is not an average.

A curious remark is this: "School closure, a major pillar of the response to pandemic influenza A [14] is unlikely to be effective given the apparent low rate of infection among children, although data are scarce." Given that children are infectious in general, the prior assumption should rather be that they are also infectious for SARS-CoV-2 except when proven differently. Perhaps children are less symptomatic but they might be infectious for their environment (like grandparents).

10.8.2. Flaxman et al. (2020)

A fair point of consideration is that if we take the *R*[0] from Flaxman et al. (2020) then we should also consider taking the *generation interval* from their study. NB. They call it a "serial interval" but describe it as "the time between when a person gets infected and when they subsequently infect another other people", and they actually use the letter "g" instead of "s". Perhaps they assume that all asymptomatic periods are the same, so that *numerically* there is no difference (so skip the difference in terms).

They also refer to the SIR model and not SEIR: "The renewal model is related to the Susceptible-Infected-Recovered model, except the renewal is not expressed in differential form."

Flaxman et al. (2020:18) use a gamma distribution with a mean of 6.5 and a coefficient of variation of 0.62. We can apply Svensson's formula again, albeit with a zero mean for the latency phase ($\alpha \rightarrow \infty$).

```
6.5 == InfectionInterval[0, 1/γ, 1/γ]
```

$6.5 = \dfrac{1}{\gamma}$

Alternatively, if we want to maintain our SEYCD structure, and use $\alpha = 1$ and still use the Flaxman et al. (2020) estimate of the generation interval, then we get:

```
6.5 == InfectionInterval[1, 1/γ, 1/γ]
```

$6.5 = \dfrac{1}{\gamma} + 1$

The latter would require a different distribution, namely the "SumTwoExp" discussed above, thus with the same mean but more dispersion over shorter and longer intervals. This notebook uses $\gamma = 1/10$, thus a third version.

PM. For the parameterisation of the gamma distribution in *Mathematica*, the Flaxman et al. format translates as follows.

```
Solve[{6.5 == Mean[GammaDistribution[alpha, beta]],
  0.62 == 1/6.5 * StandardDeviation[GammaDistribution[alpha, beta]]}, {alpha, beta}]
```

{{alpha → 2.60146, beta → 2.4986}}

The Economics Pack has a parameterisation of the gamma distribution with a mean and sd.

```
Plot[{Legended[PDF[GammaMuSigma[6.5, 6.5 * 0.62], x],
   "α = ∞, mean 6.5 (Flaxman et al.)"],
  Legended[SumTwoExp[1, 1/5.5, x], "α = 1, mean 6.5 (Flaxman et al.)"],
  Legended[SumTwoExp[1, 1/10, x], "α = 1, mean 11 (notebook)"]
 }, {x, 0, 20},
 ImageSize → 300, AxesLabel → {"Days\nbetween\ninfections", "PDF"}]
```
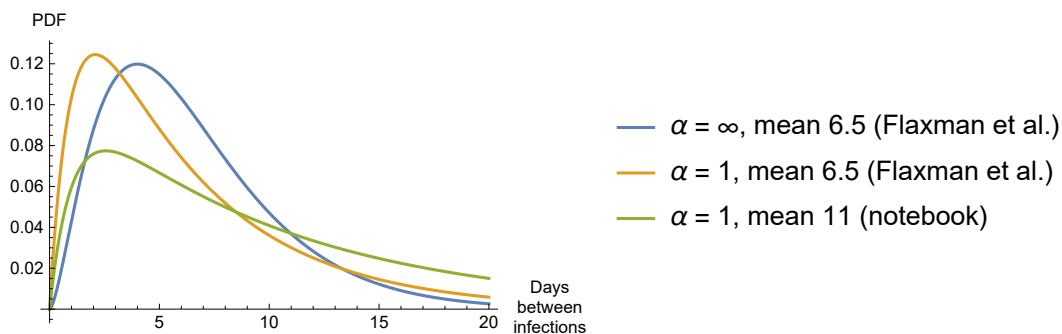

Though the gamma distribution is not symmetric, the above may still be read that roughly 50% of observations would occur before the mean. It would likely be those infections that drive the exponential growth.

PM. Potentially, we spent too much attention to $1/\gamma$, given the observation that it is mostly a normalisation factor.

10.8.3. Other cases mentioned by Svensson (2005, 2007)

The assumption of a constant infectious period (a limiting value of these assumptions) generates a value of $mu = 11$.

```
6.5 == InfectionInterval[1, mu, 0]
```

$$6.5 = \frac{mu}{2} + 1$$

As said, Svensson (2005, 2007) has also a formula for the *primary* interarrival period. This looks at the point of view of the infectors, of which a part may have no infections, with thus an "infinite interarrival period". This perspective seems less relevant. This formula gives $6.5 = 1/\alpha +$ Tinf $/ 2$, which gives Tinf = 11. However, this is less relevant.

10.8.4. Tao Liu et al. (2020)

Tao Liu et al. (2020) reported early in the year: "The average incubation period [...] was 4.8 days, ranging from 1 to 14 days, and the average period from onset to isolation was 2.9 days". This onset must refer to symptoms. With latency of 1 day, the asymptomatic infectious period was 4.8 - 1 = 3.8 days. The symptomatic period of 2.9 can be seen as infectious (otherwise there is no need for isolation). The infectious period is 3.8 + 2.9 - 1 = 6.7 days, of which 3.8 days asymptomatic. Their finding was adopted by Kucharski et al. (2020), who actually adapted to incubation of 5.2 days, to a total infectious period of 7.1 days. This was adopted by Goh (2020) in his "epidemic calculator", that looks at the symptomatic development rather than purely infections.

```
5.2 + 2.9 == InfectionInterval[1, 1/γ, 1/γ]
```

$$8.1 = \frac{1}{\gamma} + 1$$

10.8.5. Feretti et al. (2020)

Ferretti et al. (2020) recognise different subgroups of infectiousness, but still condition on above outcome of (Tao) Liu et al. (2020), using the early data from China, causing the question whether the Western nations haven't been able to do similar follow up studies.

10.8.6. Summing up

In our setup, $R_0 = 4$ has been given, and a longer $1/\gamma$ only means a lower β. The value of 10 days seems appropriate because of the aetiology of the process of clearing, while we are interested in the period before intervention. The SI(EY)A(CD) model comes with the assumption that this is also the infectious period. If R_0 had not been given, but β, then the longer the period the more infections, the higher R_0. In itself, a value of 4 seems required for the Dutch record of the Carnival at the end of February and the first death in the beginning of March. See the Chapters on intervention and estimation below.

11. Comparison with SEYCD studies by others

The literature has common names for variables S, E, I, R, D and parameters β and γ but there are differences in others. This notebook hopes to present uniform names, including the quantity q and proportion $p,$ and the rebaptising into the acquitted $A = C + D$ and the split of $I = E + Y.$ The following makes comparisons also by running the model and reproduce plots.

11.1. Yan and Chowell (2019) on SEIA

Yan & Chowell (2019:168-169) plot both SIA and SEIA (no CD) to show the same levels and delay.

11.1.1. Names

```
Our label  Their label  Value
   N[0]          m       10 000
    Sq           Sd
    Sp           x
    Aq          Rd, Cd
     I        P = E + I
   Seed       x₀ = 0.9999    1
   SpTop        m / R₀     4444.4
   Sq[∞]      ♯ escaped    1466
   Ap[∞]          η       0.8534
```

The numerical outcomes actually helped to verify this labeling. The same outcomes are reproduced.

11.1.2. Set options, time horizon, run and plot

```
NqSeed = 10^4;
SetOptions[SIA, Nqt0 → NqSeed, Seed → 1 (*not to forget*),
 BetaSIA[] → .75, GammaSIA[] → 1/3., AlphaSEYCD[] → 1/3.,
 PhiSICD[] → 0 (*SEYA, not deaths*), R0[] → Null (*not used *)]
```

{Aqt0 → 0, Cqt0 → 0, DataMold → {S, I, A}, Dqt0 → 0, Iqt0 → Null,

  Nqt0 → 10 000, Onset → 0, Seed → 1, Yqt0 → 0, AlphaSEYCD() → 0.333333,

  BetaSIA() → 0.75, GammaSIA() → 0.333333, PhiSICD() → 0, R0() → Null}

```
SetOptions[SetDatabank, Databank → SEYCD]; (* for Explain *)
```

```
f0 = BetaSIA[]/GammaSIA[] /. Options[SIA]  (*will be handy*)
```

2.25

```
SEYCD[Run, Pr, YC168, t, 150];
```

The variables are now available in levels and proportions.

Yan & Chowell (2019:169) plot with SIA and identification of the peak prevalence $I = E + Y$ with the values at $t = 55.6$.

```
TforTopIq[]
```

55.6089

```
SEYCD[Plot, 0, 120]
```
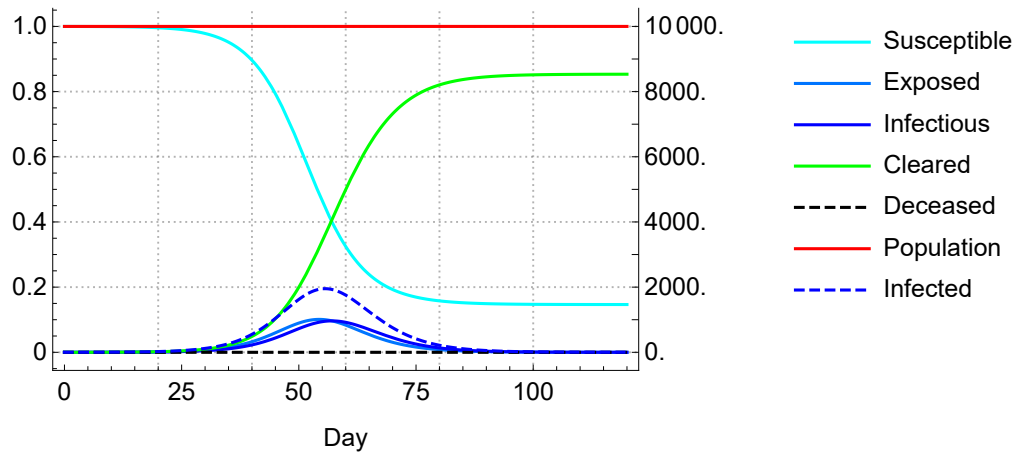

11.1.3. Limit values

Yan & Chowell (2019:168-169) have limit values {S, I, A}q[∞] = {1466, 0, 8534}, rounded up, as reproduced here. The Sq[∞] is called "the escaped".

```
SIA[Limit] (* selection from SEYCD *)
```

{0.146556, 0, 0.853444}

```
SEYCD[Limit]
```

{0.146556, 0, 0, 0.853444, 0., 1., 0, 0.853444}

```
TheFormer NqSeed // Explain
```

{S → 1465.56, E → 0, Y → 0, C → 8534.44, D → 0., N → 10 000., I → 0, A → 8534.44}

The above relies upon setting all options to their proper values. A direct check on the productlog relation is:

```
SIA[ProductLog, f0, 0.9999, 0]
```

0.146556

11.1.4. The top of Iq

Y & C identify Iq as the Peak Prevalence (of the infectious), with the top value of 1951.

```
SIA[Ip, Top, R₀] == SIA[Ip, Top, f0] (* starting value for A =  0 *)
```

$$\left\{ \frac{1}{R_0}, \frac{R_0 - \log(R_0) - 1}{R_0}, \frac{\log(R_0)}{R_0} \right\} = \{0.444444, 0.195142, 0.360413\}$$

The time of the peak prevalence fits their graph as well.

```
tTopI = TforTopIq[Automatic, 0, 100]
```

55.6089

```
pnt = SEYCD[tTopI] // Explain
```

{S → 4444.44, E → 999.813, Y → 952.053, C → 3603.69, D → 0., N → 10 000., I → 1951.87, A → 3603.69}

11.1.5. Reset options

```
ResetOptions[SIA] (* best to reset *)
```

{Aqt0 → 0, Cqt0 → 0, DataMold → {S, I, A}, Dqt0 → 0, Iqt0 → Null, Nqt0 → 1.74×10^7, Onset → 0, Seed → 100,

   Yqt0 → 0, AlphaSEYCD() → 1., BetaSIA() → 0.4, GammaSIA() → 0.1, PhiSICD() → 0.015, R0() → Null}

11.2. Lin et al. (2010) on SICD

Lin, Muthuraman and Lawley (2010) apply the SICD model with some variants. They also model a non-pharmaceutical intervention (NPI) with decision variable $u[t]$, not in our model (here). Their paper is interesting for the economic aspects too, not further discussed here. We now only focus on their Figure 3 on page 7, and only consider the outcomes without the NPI.

11.2.1. Names

```
Our label  Their label    Value
    Sp           s
    Ip           i
    Cp           r
    Dp           d
   n.a.         u[t]
    γ          γ + τ        0.3
    λ            γ          0.25
    μ            τ          0.05
    φ                     0.05 / 0.3
   N[0]       100 × %       100
   Seed      i₀ = 1 × %      1
```

They first assign $x[t] = \text{SICD}[t]$, but then redefine to $x[t] = \{Sp, Ip\}$. Subsequently, they consider variants $x[0] = \{.99, .01\}$, $\{.67, .33\}$ and $\{.50, .50\}$.

For us, it provides a useful comparison on the starting conditions. We only consider their models (a) and (c).

```
SetOptions[SetDatabank, Databank → SICD]; (* for Explain *)
```

11.2.2. Their model (a): Set options, time horizon, run and plot

```
NqSeed = 100;
SetOptions[SIA, Nqt0 → NqSeed, Seed → 1 (*not to forget*),
 BetaSIA[] → .40, GammaSIA[] → 0.3, AlphaSEYCD[] → Null,
 PhiSICD[] → 0.05 / 0.3, R0[] → Null (*not used*)]
```

{Aqt0 → 0, Cqt0 → 0, DataMold → {S, I, A}, Dqt0 → 0, Iqt0 → Null, Nqt0 → 100, Onset → 0, Seed → 1,
  Yqt0 → 0, AlphaSEYCD() → Null, BetaSIA() → 0.4, GammaSIA() → 0.3, PhiSICD() → 0.166667, R0() → Null}

```
f0 = BetaSIA[] / GammaSIA[] /. Options[SIA]  (*will be handy*)
```

1.33333

```
SICD[Run, Pr, LMLp7fig3, t, 150];
```

The variables are now available in levels and proportions.

Lin et al. (2010) only plot in a shorter range, and only plot Sp and Ip. Their plot also contains the NPI variant, that we neglect.

```
SICD[Plot, 0, 100, PlotRange → {0, 0.10}]
```
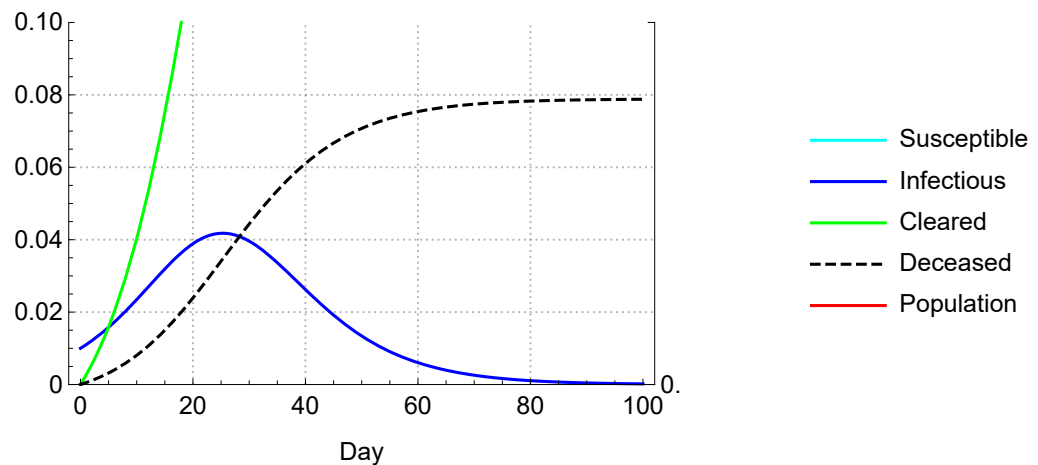

They do not provide other outcomes, but it may be useful to give our limit outcome here.

```
SICD[Limit]  // Explain
```

{S → 0.526713, I → 0, C → 0.394406, D → 0.0788812, N → 0.921119, A → 0.473287}

11.2.3. Their model (c): Set options, time horizon, run and plot

```
NqSeed = 100;
SetOptions[SIA, Nqt0 → NqSeed, Seed → 33 (*not to forget*),
 BetaSIA[] → .40, GammaSIA[] → 0.3, AlphaSEYCD[] → Null,
 PhiSICD[] → 0.05 / 0.3 (**), R0[] → Null (*not used*)]
```

{Aqt0 → 0, Cqt0 → 0, DataMold → {S, I, A}, Dqt0 → 0, Iqt0 → Null, Nqt0 → 100, Onset → 0, Seed → 33,
   Yqt0 → 0, AlphaSEYCD() → Null, BetaSIA() → 0.4, GammaSIA() → 0.3, PhiSICD() → 0.166667, R0() → Null}

```
f0 = BetaSIA[] / GammaSIA[] /. Options[SIA]  (*will be handy*)
```

1.33333

```
SICD[Run, Pr, LMLp7fig3, t, 150];
```

The variables are now available in levels and proportions.

Lin et al. (2010) only plot in a shorter range, and only plot Sp and Ip.

```
SICD[Plot, 0, 100, PlotRange → {0, 0.35}]
```
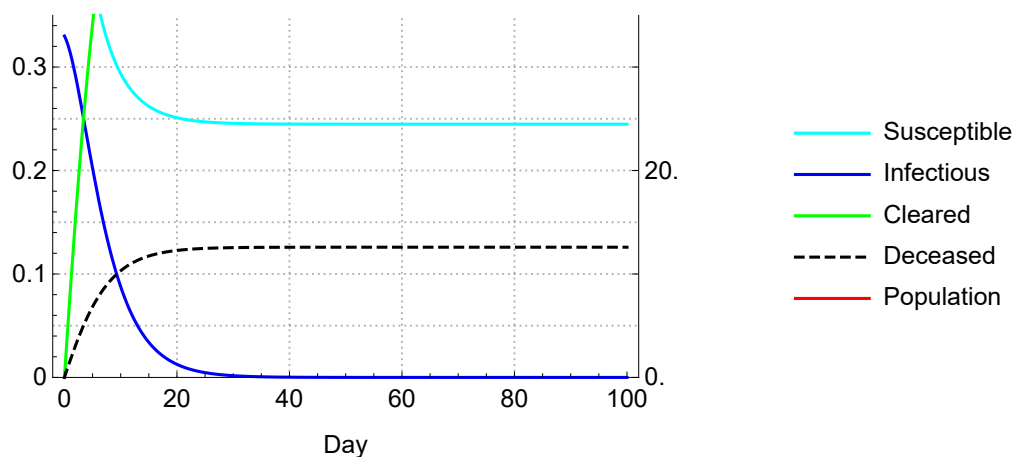

They do not provide other outcomes, but it may be useful to give our limit outcome here.

```
SICD[Limit] // Explain
```

{S → 0.244765, I → 0, C → 0.629363, D → 0.125873, N → 0.874127, A → 0.755235}

11.2.4. Reset options

```
ResetOptions[SIA] (* best to reset *)
```

{Aqt0 → 0, Cqt0 → 0, DataMold → {S, I, A}, Dqt0 → 0, Iqt0 → Null, Nqt0 → 1.74×10^7, Onset → 0, Seed → 100,
   Yqt0 → 0, AlphaSEYCD() → 1., BetaSIA() → 0.4, GammaSIA() → 0.1, PhiSICD() → 0.015, R0() → Null}

11.3. Maugeri et al. (2020) using SEYCD on SARS-CoV-2 in China

Maugeri et al. (2020) (published May 5) present both an example of the classic SEYCD model and then do a study on China. The first however runs into problems of interpretation of the parameters. Their main result on China is that 92.9% of infections were not reported. Their assumption is that reported deaths "are less likely to be affected by reporting bias", but, at least in Holland, many "normal" deaths that do not occur in a hospital are not tested on the virus and may still be caused by the virus (a phenomenon called "excess death" compared to earlier years).

NB. They also take α as "the infection rate and was assumed to be the inverse of the incubation period (i.e. the period from infection to the onset of symptoms)". Thus see our discussion above about the Kucharski et al. (2020) model. At first it seems that their model concerns symptoms but it appears that they concentrate on (unobserved) infections, estimated on death counts.

PlotLegendsSIA[SEYCD]

PM. A nice aspect of their model formulation is the use of $\mu = \gamma$ IFF and $\lambda = \gamma$ (1 - IFF), so that it is directly clear how to transform from the sCFF to the model (also when γ changes).

11.3.1. Their example: Set options, time horizon, run and plot

This is only an example. Still, their *example* IFF = 0.2 = 20% still is remarkably high. When we run their example, then the other variables are okay but we cannot get the dead to their level. It is more likely that their example used a figure closer to 10%.

```
Our label  Their label  Value
    α            σ         0.3
    β            β         0.8
    γ            γ         0.2
    φ            μ         0.2
  N[0]           N       10000
  Seed                     1
```

NqSeed = 10000;
SetOptions[SIA, Nqt0 → NqSeed, Seed → 1 (∗not to forget∗),
 BetaSIA[] → 0.8, GammaSIA[] → 0.2, AlphaSEYCD[] → 0.3, PhiSICD[] → 0.1]

{Aqt0 → 0, Cqt0 → 0, DataMold → {S, I, A}, Dqt0 → 0, Iqt0 → Null, Nqt0 → 10000, Onset → 0, Seed → 1,
   Yqt0 → 0, AlphaSEYCD() → 0.3, BetaSIA() → 0.8, GammaSIA() → 0.2, PhiSICD() → 0.1, R0() → Null}

SetOptions[SetDatabank, Databank → SEYCD]; (∗ for Explain ∗)

f0 = BetaSIA[] / GammaSIA[] /. Options[SIA] (∗will be handy∗)

4.

SEYCD[Run, Pr, Metalex, t, 100];

The variables are now available in levels and proportions.

```
SEYCD[Plot, 0, 100, PlotRange → All]
```
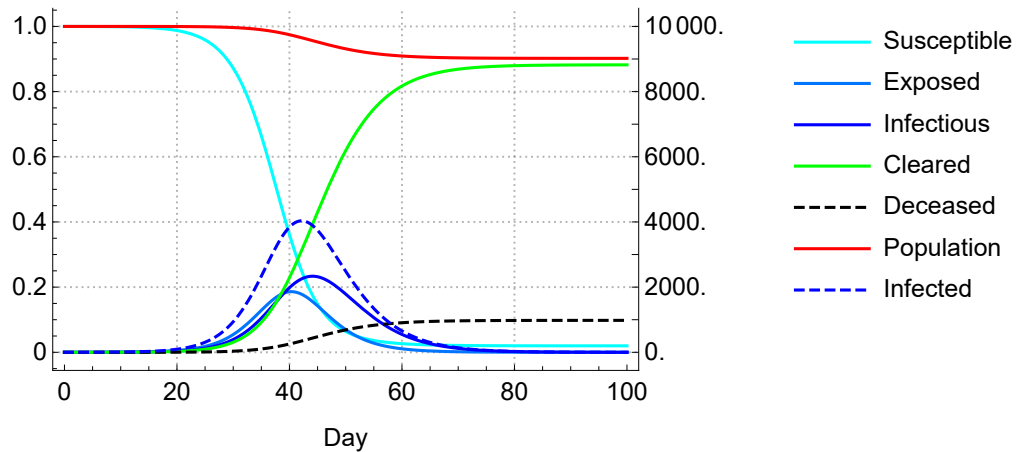

11.3.2. Their finding on China: Set options, time horizon, run and plot

The reported *R0* = 2.43*, β* = 0.73 and *γ* = 0.28 do not fit (and their reference needs to be checked). It appears that *γ* has been fixed as a parameter and thus there should be no problem in dividing the values. (If *β and γ* had been estimates, then the expectation of a ratio is not the same as the ratio of the expectations. This ratio is so important for the SI(EY)A(CD) model that it would be wise to include the restriction of sameness.)

There is the always awkward rounding of parameters. The authors give both *γ* and $1/γ$ values, but *γ* is rounded (and perhaps $1/γ$ originally was too). In practice it is useful to give six significant digits for all, and only round for the abstract or general discussion.

```
{{.19, 1/5.2}, {.28, 1/3.5}}
```
$$\begin{pmatrix} 0.19 & 0.192308 \\ 0.28 & 0.285714 \end{pmatrix}$$

Interestingly, the authors comment (p6) about a sensitivity analysis that when *γ* = 0.1 then R0 = 4.07 (which fits our choice of parameters in this notebook). They say that the results do not change much, but they do not mention the error or what "much" means. They also estimate an IFF of 1.5% (as in this notebook).

```
Our label  Their label           Value
   R0          R0                 2.43
   α           σ         0.19 (fixed as 1 / 5.2 days)
   β           β            0.73 (estimated)
   γ           γ         0.28 (fixed as 1 / 3.5 days)
   φ           μ            0.015 (estimated)
   λ        γ (1 - μ)      0.28 * (1 - 0.015)
  N[0]         N                  10^9
  Seed                            1
```

```
NqSeed = 10^9;
SetOptions[SIA, Nqt0 → NqSeed, Seed → 1 (*not to forget*),
 BetaSIA[] → .73, GammaSIA[] → 0.28, AlphaSEYCD[] → 0.19, PhiSICD[] → 0.015]
```

{Aqt0 → 0, Cqt0 → 0, DataMold → {S, I, A}, Dqt0 → 0, Iqt0 → Null,

   Nqt0 → 1 000 000 000, Onset → 0, Seed → 1, Yqt0 → 0, AlphaSEYCD() → 0.19,

   BetaSIA() → 0.73, GammaSIA() → 0.28, PhiSICD() → 0.015, R0() → Null}

```
f0 = BetaSIA[] / GammaSIA[] /. Options[SIA]  (*will be handy*)
```

2.60714

```
SEYCD[Run, Pr, Metalres, t, 300];
```

The variables are now available in levels and proportions. Let us make a table for some points along the path.

```
tab = Table[t, {t, 0, 300, 100}];
```

```
res = SEYCD /@ tab;
```

```
TableForm[res // PopulationForm, TableHeadings → {tab, SEYCD[DataMold]},
 TableSpacing → {1, 2}, TableAlignments → Right]
```

	S	E	Y	C	D	N	I	A
0	$1. \times 10^9$	1	0	0	0	$1. \times 10^9$	1	0
100	$998. \times 10^6$	680 847	308 088	606 897	9242	$1. \times 10^9$	988 935	616 139
200	95.4×10^6	1.62×10^6	1.82×10^6	$888. \times 10^6$	13.5×10^6	$986. \times 10^6$	3.43×10^6	$901. \times 10^6$
300	94.3×10^6	22	25	$892. \times 10^6$	13.6×10^6	$986. \times 10^6$	47	$906. \times 10^6$

Remarkably, the authors do not present a plot of the model with these parameters for China, but we can do so. It is sobering that the numbers on the right hand axis have a stepsize of 100 million persons.

```
SEYCD[Plot, 0, 300, PlotRange → All]
```
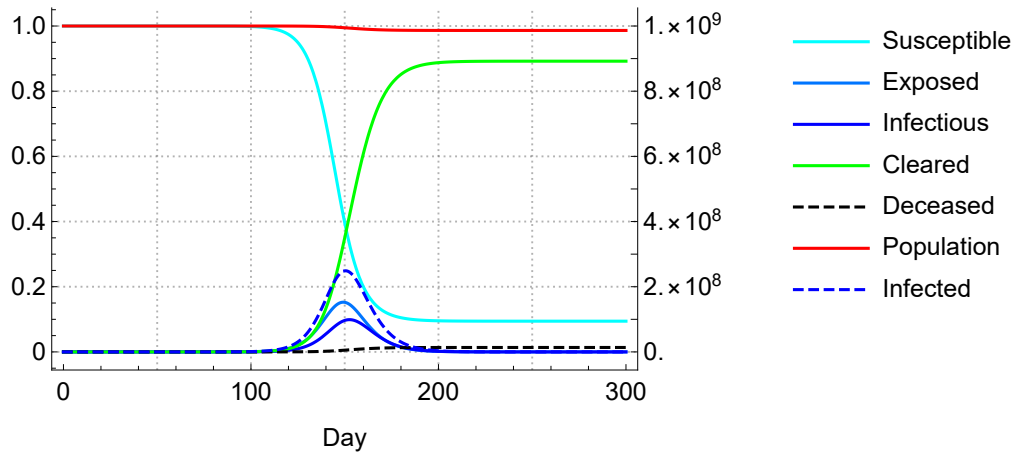

The authors do not provide other outcomes. The limiting values are rather clear from above figures and plot. If unhindered, there would have been 13.6 million deaths per billion inhabitants.

```
SEYCD[Limit] // Explain
```

{S → 0.0942981, E → 0, Y → 0, C → 0.892116, D → 0.0135855, N → 0.986414, I → 0, A → 0.905702}

```
(SEYCD[Limit] NqSeed) // Explain
```

$\{$S → 9.42981×10^7, E → 0, Y → 0, C → 8.92116×10^8,

  D → 1.35855×10^7, N → 9.86414×10^8, I → 0, A → $9.05702 \times 10^8\}$

The authors report in their figure 4 that February 2 had a death toll of 300, also right on the estimated model path. We can determine the time when this happened, and calculate back to the date of onset.

```
FindTime[Dq[t] == 300, {t, 10, 100}]
```

75.5335

This would be the moment when patient 0 occurred.

```
DayPlus[{2020, 2, 2}, -Round[TheFormer]]
```

Day: Mon 18 Nov 2019

11.3.3. Reset options

```
ResetOptions[SIA] (* best to reset *)
```

$\{$Aqt0 → 0, Cqt0 → 0, DataMold → {S, I, A}, Dqt0 → 0, Iqt0 → Null, Nqt0 → 1.74×10^7, Onset → 0, Seed → 100,

  Yqt0 → 0, AlphaSEYCD() → 1., BetaSIA() → 0.4, GammaSIA() → 0.1, PhiSICD() → 0.015, R0() → Null$\}$

11.4. Kucharski et al. (2020), with use in the "epidemic calculator" by Goh (2020)

Kucharski et al. (2020), now published in *The Lancet Infectious Diseases*, was used by Goh (2020) for basic values in his "epidemic calculator". The following tries to use the same parameters which Goh presents as defaults, May 14 2020. It is easy to just plug in the parameters, but it is a challenge to interpret them. The Kucharski et al. (2020) paper namely looks at symptomatic cases, which causes questions on the asymptomatic cases. This has been discussed above in section 10.7, and here we merely plug in the parameters and reproduce the results (with margins of error) without further discussion.

Let us set the plot colours and legends to *symptoms* lest we forget this later on.

```
PlotLegendsSIA["Symptoms"]
```

11.4.1. Goh's parameters on SEIRD or SI(EY)A(CD) itself

For us, the relevant parameters given by Goh are:

R_0 = 2.2
Tincubation = 5.2 days
Tinfectious = 2.9 days
Case fatality rate = 2% (actually the *symptomatic Case Fatality Factor*, sCFF)

11.4.2. Names

```
Our label  Their label   Value
  N[0]      Population   7 × 10^6
   R₀           R₀          2.2
   β                     R₀ * γ
   γ        1 / Tinf     1 / 2.9
   α        1 / Tinc     1 / 5.2
   φ           sCFF        0.02
  Seed     Initial Inf     1
```

11.4.3. Set options, time horizon, run and plot

```
NqSeed = 7 × 10^6;
SetOptions[SIA, Nqt0 → NqSeed, Seed → 1 (*not to forget*),
 BetaSIA[] → 2.2 / 2.9 ,
 GammaSIA[] → 1/2.9,
 AlphaSEYCD[] → 1/5.2,
 (* remarkably long: not latency but concerns incubation to symptoms *)
 PhiSICD[] → 0.02 ]
```

{Aqt0 → 0, Cqt0 → 0, DataMold → {S, I, A}, Dqt0 → 0, Iqt0 → Null,

  Nqt0 → 7 000 000, Onset → 0, Seed → 1, Yqt0 → 0, AlphaSEYCD() → 0.192308,

  BetaSIA() → 0.758621, GammaSIA() → 0.344828, PhiSICD() → 0.02, R0() → Null}

```
SetOptions[SetDatabank, Databank → SEYCD]; (* for Explain *)
```

```
f0 = BetaSIA[]/GammaSIA[] /. Options[SIA]  (*will be handy*)
```

2.2

```
SEYCD[Run, Pr, KetalandGoh, t, 250];
```

The variables are now available in levels and proportions.

The Goh calculator doesn't run beyond day 218 and restricts the vertical axis to about 2 million. We find the top around day 130, like in his plot. Observe that the high α causes that his "Exposed" are now higher than the "Infectious", or rather that the "Incubating" are higher than the "Symptomatic". Goh has the time of the first death at day 41. For us:

```
FindTime[1 == Dq[t], {t, 10, 50}]
```

35.3905

It would be wrong to reproduce Goh's legend (i.e. use SI(EY)A(CD) labels while it concerns symptoms).

```
SEYCD[Plot, 0, 218, PlotRange → {0, 2/7}]
```
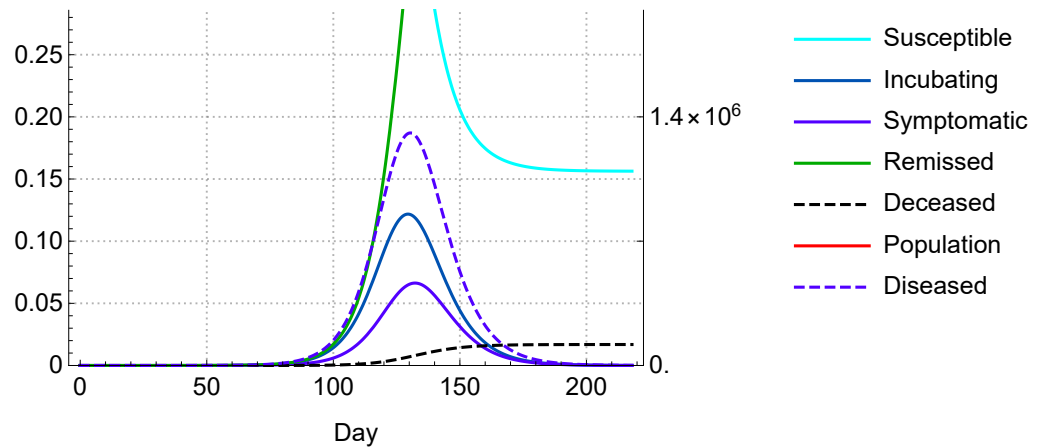

11.4.4. Numerical outcomes from reading the graph

Goh's numerical outcomes on day 218 can be read from the graph:

```
Goh218 = {Sq → 1095195, Eq → 101, Yq → 100, Cq → 5701049,
  Dq → 111316, Nq → 7 × 10^6 - 111316, Iq → 101 + 100, Aq → 5904604}
```

{Sq → 1 095 195, Eq → 101, Yq → 100, Cq → 5 701 049,
  Dq → 111 316, Nq → 6 888 684, Iq → 201, Aq → 5 904 604}

This compartments add up to the population.

```
{7 × 10^6, "=?=", Sq + Iq + Aq} /. Goh218
```

{7 000 000, =?=, 7 000 000}

There is the separate group of hospitalised.

```
{7 × 10^6, "=?=", Sq + Eq + Yq + Cq + Dq, "hospital" → 85033} /. Goh218
```

{7 000 000, =?=, 6 907 761, hospital → 85 033}

```
{Aq, "=?=", Cq + Dq, "hospital" → 85033} /. Goh218
```

{5 904 604, =?=, 5 812 365, hospital → 85 033}

Our outcomes are comparable though.

```
(res = SEYCD[218]) // Explain
```

{S → 1.09423 × 10^6, E → 429.572, Y → 336.57, C → 5.7869 × 10^6,
  D → 118 100., N → 6.8819 × 10^6, I → 766.141, A → 5.905 × 10^6}

For direct comparison:

```
SetDatabank[SEYCD[DataMold], {Last /@ Goh218, res}, anylabel] ;
ShowData[]
```

	1	2
S	1 095 195	1.09423×10^6
E	101	429.572
Y	100	336.57
C	5 701 049	5.7869×10^6
D	111 316	118 100.
N	6 888 684	6.8819×10^6
I	201	766.141
A	5 904 604	5.905×10^6

11.4.5. Goh's parameters on hospitalisation

Goh's model also contains a part on hospitalisation that is not immediately relevant here (but later would be). However, this also involves mortality. Let us quote his explanation on this part:

"Model Details. The clinical dynamics in this model are an elaboration on SEIR that simulates the disease's progression at a higher resolution, subdividing I, R into mild (patients who recover without the need for hospitalization), moderate (patients who require hospitalization but survive) and fatal (patients who require hospitalization and do not survive). Each of these variables follows its own trajectory to the final outcome, and the sum of these compartments add up to the values predicted by SEIR. Please refer to the source code for details. Note that we assume, for simplicity, that all fatalities come from hospitals, and that all fatal cases are admitted to hospitals immediately after the infectious period."

Thus there are mild (80%, with a recovery time of 11.1 days) vs hospital (20%). The parameter "time from end of incubation to death" (32 days) would apparently include the 2.9 days infectious period and the 5 days time to hospitalisation. Does this assume that people can be 2.1 days symptomatic without being infectious ? There are reports (Gill (2020)) that people clear the virus but that the immune system starts attacking the own body. There are moderate and fatal patients (the latter apparently stay 27 = 32 - 5 days). The average of 28.6 days in the hospital seems to require that the moderate cases require longer to recover (the fatal cases might end faster).

Overall, we reproduce the SI(EY)A(CD) (symptoms) part, and we did not look at the hospitalisation part.

11.4.6. Reset options

```
ResetOptions[SIA] (* best to reset *)
```

$\{$Aqt0 → 0, Cqt0 → 0, DataMold → {S, I, A}, Dqt0 → 0, Iqt0 → Null, Nqt0 → 1.74×10^7, Onset → 0, Seed → 100,

Yqt0 → 0, AlphaSEYCD() → 1., BetaSIA() → 0.4, GammaSIA() → 0.1, PhiSICD() → 0.015, R0() → Null$\}$

```
PlotLegendsSIA[SEYCD]
```

12. Intervention

Clear

```
SIA[Clear]

ResetOptions[SIA];

SetOptions[SetDatabank, Databank → SEYCDT]; (* for Explain *)
```

12.1. Introduction

A less challenging but tedious issue here concerns the *plot colours and legends*. When we plot two scenarios in the same graph, the colours and legends must allow for distinction. The routine PlotLegendsSIA allows control of such. The Base scenario is plotted in dot-dashing, so that it remains somewhat vague in the background. The Alternative scenario is plotted in drawn lines because it has our new interest.

12.1.1. Keeping $\gamma = 1/10$ fixed

The discussion in sections 1.5 and 16.13 shows that γ is a tricky parameter and rather should not be used for modeling interventions in our current setup. Thus we keep $\gamma = 0.1$ fixed in all scenario's.

12.2. Two methods

The Base scenario in SEYCD is *without* intervention and the Alternative scenario is *with* intervention. There are two ways to implement the Alternative model:

(1) By use of a SEYCD model too, but starting at another moment than 0 while using the new parameter values. This is discussed in **Chapter 13**.

(2) By use of the Mathematica *WhenEvent* feature to dynamically put the parameter change into the model. This method generates the **SEYCDT** "insert" variant for R_0 and **SEYCDB** for β. It appears that this method is useful for estimation. **Chapter 14** discusses the data. Estimation must deal with the death counts caused by parameters *before* and *after* the intervention, and it is useful to have these and the overlap available in one single run. Keeping γ fixed, then analytically there should be no difference in adapting β or R_0. However, there appears to be the practical issue of "conditioning" in numerical estimation. This notebook and package have had these phases: (i) first estimating both β and γ like in the SEYCD setup, and discovering both that the problem likely was ill-conditioned and that γ caused weird effects, (ii) keep γ fixed and switch to using R_0 because a larger sized parameter ought to improve the conditioning, but discovering that the numerical routines then converged on less interesting values than generated by the use of β, (iii) develop the graph shown in section 1.5.3 how estimates depend upon additional assumptions on the day of onset, (iv) decide to keep both approaches available for users, with the suggestion that the use of β, with γ fixed, might still be best anyhow. Thus there are now SEYCDT for R_0 and SEYCDB for β. **Chapter 15** and the *Introduction* section 1.6 present findings with SEYCDT (nicely plotting values for R_0), but the starting values for those estimation routines were originally found by using SEYCDB (**Chapter 16**). The routines use

FindMinimum (local) and NMinimize (wider) and not FindFit (yet). For our dataset, the use of confidence intervals is out of proportion, and our discussion is only an introduction into the notions of intervention and estimation.

12.3. Interpretation of the estimation result

When the onset would be at Carnival February 23 2020 indeed, then the death count till May 20 can be explained by $R_0 = 5.7$. We have been using $R_0 = 4$ throughout this notebook. This value of 4 appears to better fit the assumption that the onset wasn't at Carnival, February 23, but nine to ten days earlier, at Valentine's Day, February 14. The FindMinimum and NMinimize routines do not give confidence intervals. However, we can check that the difference in outcomes is not so relevant.

Without an intervention, the death toll for $R_0 = 4$ would be 255,000. For $R_0 = 5.7$ it would be 260,000.

```
"D" * NqSeed /. Explain[SICD[Limit, BetaSIA[] → 0.4], SICD]
```

255 825.

```
"D" * NqSeed /. Explain[SICD[Limit, BetaSIA[] → 0.57], SICD]
```

260 110.

The difference is remarkably small, and this has the theoretical reason already mentioned in the *Introduction* section 1.6, and shown with a graph: for values of R_0 higher than 4, the limit value of the share of the acquitted does not change so much anymore. The compartment of the deceased is only a proportion of the acquitted. Let us plot directly for the deceased.

```
"D" * NqSeed /. Explain[SICD[Limit, BetaSIA[] → r0 GammaSIA[]], SICD] // Simplify
```

$$261\,000.\frac{W\left(-0.999994\,e^{-1.\,r0}\,r0\right)}{r0} + 261\,000.$$

```
Plot[TheFormer, {r0, 1, 8}, AxesLabel → {"R0", "Deceased"},
 PlotStyle → {{Dashed, Black}}, AxesOrigin → {0, 0}]
```
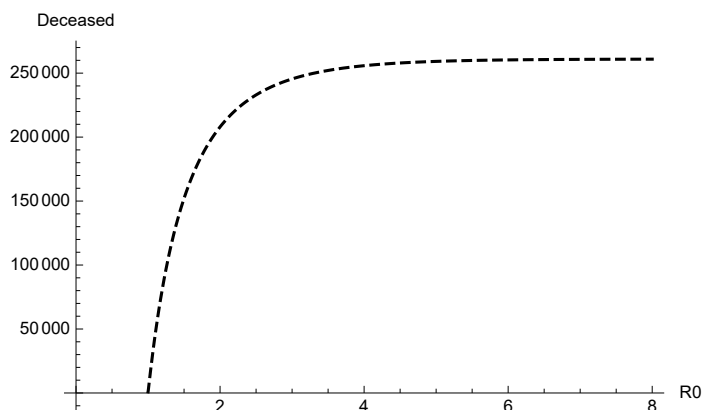

12.2. Principles for parameters

The intervention - (of March 12-15 of a partial lockdown, with an implementation delay of 3 days till March 18 (day 24)) - caused a change in the pandemic parameters. The modeling exercise helps us to better understand the meaning of those parameters.

This was discussed in the *Introduction,* section 1.4.6. Our base case has $\beta = R_0\,\gamma = 4 / 10 = 0.4$. Suppose that an intervention on (self-) quarantine reduces the "real infectious period" from 10 to 5

days. So we may want to use a new value of $\gamma = 1/5 = 0.2$. However, this means that the number of acquitted cases rises much faster, and certainly around the day of intervention, and thus also the number of deaths contained in them. A simulation causes the paradoxical result of increased death in the short run. It might make the estimate look good because it reaches a lower limit indeed, but the effect in the short run is illogical. The underlying reason is that SI(EY)A(CD) modeling concerns the assignment to compartments, and does not model actual moments of death. For us it makes more sense now to regard γ as a *normalisation parameter* that allows comparison over applications. It suffices to modify β or R_0. The choice is immaterial on content, but in practice estimation is influenced by the *conditioning* of the problem (sensitivity to the step size).

12.3. Principles for routines

Modeling of an intervention can be done in at least two ways. Let Base be the base scenario and Alternative be the intervention scenario. We assume that Base has a fully developed model and time trajectory.

12.3.1. Method 1. Using two full SEYCD models

The values of Base[$t_{\text{Intervention}}$] at the moment of intervention can be stored, and Alternative[t] can be run with Onset $\rightarrow t_{\text{Intervention}}$, with new parameters and such starting values. This has the advantage that the Options[SIA], when set to the new values of the intervention, also generate the appropriate top and limit values. This is the most elegant approach. The implementation below is straightforward, and actually only requires (i) a transfer from variable values of Base to starting values of Alternative, and (ii) the setting of the new parameter values and onset parameter.

However, this method also required the small computational trick, that the model specifies Nq[t0] = Nqt0 - Dqt0, slightly changing our standard interpretation of the option parameter Nqt0. The trick leaves unaffected that the ratio's are still defined w.r.t. Nqt0.

PM. The current model implementation has $D = \varphi A$. Other models elsewhere use interdependent parameters, e.g. μ and λ that actually depend upon γ. One should take care not to change the underlying IFF. For interdependent parameters, the Options[SIA] (in other formats) should also be corrected from the moment of intervention onwards.

12.3.2. Method 2. A fast track with SEYCDT and SEYCDB inserts

Mathematica has the feature *WhenEvent,* such that a model with differential equations can be solved with discrete variables that can be changed at a particular moment. See *?WhenEvent* and *?DiscreteVariables*. This gives a fast result without little administration at the moment of input. This setup seems useful for estimation of the effect on β and γ by an intervention. However, it causes additional administration when we want to determine the appropriate top and limit values in output, since those routines depend upon starting values. This approach is very elegant for estimation but would be more cumbersome on those other issues. Its implementation has been done as follows:

(2.1) The SEYCDT equations have an insert of SEYCD with time-dependent R0[t] and GammaSIA[t], with the starting values taken from Options[SIA]. (We do not have a reason for an intervention on

PhiSICD or AlphaSEYCD. For estimation, it appeared that our parameter φ includes the non-hospital deaths while the data concern hospital deaths. The estimate would not be affected by a level difference.)

(2.2) Values on the day of intervention and its parameters are stated in the Options[Intervention]. Thus those are not in Options[SIA].

(2.3) After a call of SEYCDT[Run, Pr, arbitraryInterventionModelName, t, tmax] we can use the environment of SEYCD as usual. As said, this comes with the warning that values for the top and limit still use Options[SIA] and not the updated values of the new situation. A quick check is provided by the routine TopIq.

(2.4) We can formulate SEYCDT["RMSE", …] expressions of the Root Mean Squared Error (RMSE) of the model *outcomes* on Dq[t] and infection fatality *statistics* Drec[τ] (with different definition and moment of observation). We do not need a separate estimation routine yet but can call Mathematica's FindMinimum and NMinimize (without extra's on confidence intervals etcetera).

NB. It appeared useful to also have a SEYCDB insert variant that uses BetaSIA[t] and GammaSIA[t]. It has the same structure. See Chapter 16.

12.3.3. Comparison of the methods

Both approaches require an administration of the values at the time of the intervention. Eventually we want to know about the top and limit values anyhow. Once we are used to method (1), it may tend to be as fast as (2).

However, when the intervention causes $R_0 < 1$, then the actual use of routines for SIA appears to be rather disappointing. *The infection peters out while most of our routines are focused on a developing infection.* It may well be that the second method is more fruitful, since it facilitates such estimation.

It remains an important insight that the lack of "interesting results" derives from the properties of the new *parameters*, and not from "some new model". When the old routines generate less interesting outcomes, the routines still work and still are valid for the problem at hand: only the parameters in the problem have changed. (Thus we need not be worried when there would be no interesting outcomes, or when there are many warning messages about out-of-range values.)

13. Intervention using two full SEYCD models

Chapter 12 introduced the first method: this models an intervention by having two full SEYCD models.

Clear

```
SIA[Clear]

ResetOptions[SIA];

SetOptions[SetDatabank, Databank → SEYCD]; (* for Explain *)
```

13.1. Base case without intervention

The base case will consist of the SEYCD model with the default parameters in the package, and discussed above. We repeat the calculation here without discussion. The only changes are: (a) A longer time horizon, to allow for later comparison. (b) A different legend to highlight that this is the base case. PM. The other runs will compare with this base case. Commonly, it suffices to compare with the plot. If one wishes to compare particular outcomes of variables, then it may be required to run the base case again. An advise is to remember this section number or copy the section to a separate "base" file.

```
PlotLegendsSIA[SEYCD, DotDashed] (* an alternative might be Dashed *)
```

```
ResetOptions[SIA]
```

$\{Aqt0 \rightarrow 0, Cqt0 \rightarrow 0, DataMold \rightarrow \{S, I, A\}, Dqt0 \rightarrow 0, Iqt0 \rightarrow Null, Nqt0 \rightarrow 1.74 \times 10^7, Onset \rightarrow 0, Seed \rightarrow 100,$
$Yqt0 \rightarrow 0, AlphaSEYCD() \rightarrow 1., BetaSIA() \rightarrow 0.4, GammaSIA() \rightarrow 0.1, PhiSICD() \rightarrow 0.015, R0() \rightarrow Null\}$

```
NqSeed = 17.4 × 10^6;
(* SetOptions[SIA, BetaSIA[] → .6, GammaSIA[] → .1] *)
```

```
SEYCD[Run, Pr, newModel, t, 250];
```

```
pbase = SEYCD[Plot, 0, 250]
```
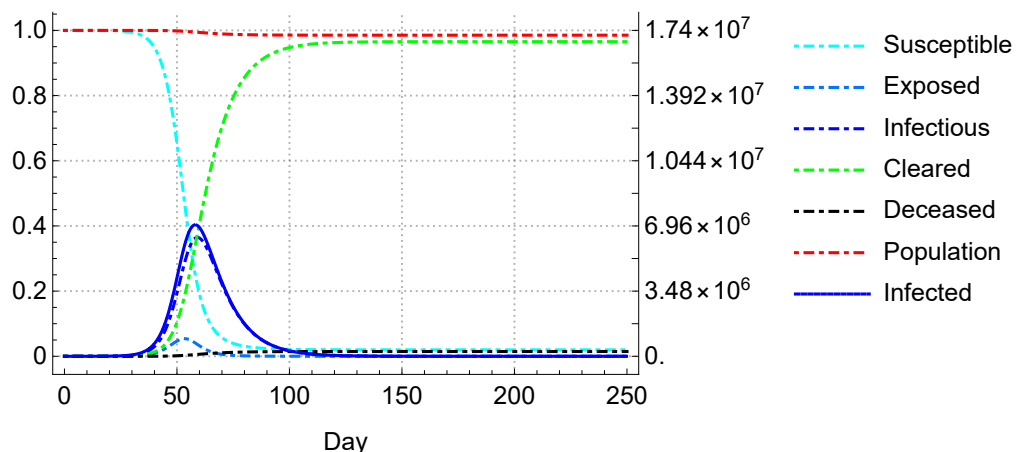

The values at day 24, the day of the intervention (including 3 days for implementation):

```
Explain[SEYCD[24], SEYCD]
```

$\{S \rightarrow 1.73741 \times 10^7, E \rightarrow 4777.44, Y \rightarrow 14\,660.4,$
$C \rightarrow 6355.58, D \rightarrow 96.7855, N \rightarrow 1.73999 \times 10^7, I \rightarrow 19\,437.8, A \rightarrow 6452.37\}$

13.2. Alternative case, with the intervention (SEYCD)

13.2.1. Choosing the intervention parameters

Below we will discuss that we will keep $\gamma = 0.1$ and select $R_0 = 0.99$, so that the new $\beta = 0.099$. This suffices for now, and our focus is upon stitching two models and their runs together: the first run is

with the old parameters till t = 24, and the other run is from t = 24 with new parameters.

13.2.2. The model values at t = 24

For the intervention on day 24, the values of the variables of the Base scenario are important as starting variables for the new process. They must be transported into the Options[SIA], with the new values of β and γ, and the information that t = 24 has become the new Onset.

SEYCDB[Intervention, 24, .099, .1]

{Seed → 4777.44, Yqt0 → 14 660.4, Cqt0 → 6355.58, Dqt0 → 96.7855,
  Aqt0 → 6452.37, BetaSIA() → 0.099, GammaSIA() → 0.1, Onset → 24}

SetOptions[SIA, TheFormer]

{Aqt0 → 6452.37, Cqt0 → 6355.58, DataMold → {S, I, A}, Dqt0 → 96.7855,
  Iqt0 → Null, Nqt0 → 1.74 × 10^7, Onset → 24, Seed → 4777.44, Yqt0 → 14 660.4,
  AlphaSEYCD() → 1., BetaSIA() → 0.099, GammaSIA() → 0.1, PhiSICD() → 0.015, R0() → Null}

Above routine on transferring the values of the variables uses the SEYCDB[DataMold], defined by using the names for the starting values. (Not used are Nqt0, Sqt0 and Iqt0.)

13.2.3. Running the Alternative model

The trick is that Options[SIA] has the Onset option. The default value is 0 so that the Base runs from t_0 = 0. This value has now been set to 24, and the new model will run from t_0 = 24.

The following uses standard SEYCD.

The model with the intervention uses the default plotting colours and legends (drawn lines).

```
PlotLegendsSIA[SEYCD]
(* intervention uses the default plotting colours and legends *)
```

We saved only the starting values of the variables and not the values of their derivatives. Mathematica calculates those anew, and gives a warning message.

We can now run the model. At the day 24 of the intervention, the cumulative death count was 96. In the days after the intervention, there was still a rise of the death count because of people "in the pipeline". By day 42, the statistically reported number of "deaths per day" apparently may have a peak value. However, the infection is not fully suppressed yet, with β = 0.099, and there will be new deaths also in the regime of (partial) lockdown.

SEYCD[Run, Pr, intModelFull, t, 250]; (* uses option Onset as t0 *)

Check that the model has the proper starting values and new parameters.

intModelFull[t]

Hold[{Sq′(t) = −5.68966 × 10^{-9} Sq(t) Yq(t), Eq′(t) = −1. Eq(t) − Sq′(t), Yq′(t) = 1. Eq(t) − Aq′(t),
  Aq′(t) = 0.1 Yq(t), Dq(t) = 0.015 Aq(t), Cq(t) = Aq(t) − Dq(t), Nq(t) = 1.74 × 10^7 − Dq(t),
  Iq(t) = Eq(t) + Yq(t), Nq(24) = 1.73999 × 10^7, Sq(24) = 1.73741 × 10^7, Eq(24) = 4777.44,
  Yq(24) = 14 660.4, Cq(24) = 6355.58, Dq(24) = 96.7855, Aq(24) = 6452.37, Iq(24) = 19 437.8}]

```
Explain[SEYCD[42], SEYCD] (* supposed top of "reported deaths per day" *)
```

$\{S \to 1.73432 \times 10^7, E \to 1713.13, Y \to 17\,338.6, C \to 37\,143.6,$

$D \to 565.638, N \to 1.73994 \times 10^7, I \to 19\,051.7, A \to 37\,709.2\}$

At first the following plot seems awkward, but, the intervention suppressed the infection, and thus there is hardly anything to see.

```
pintFull = SEYCD[Plot, 24, 250]
```
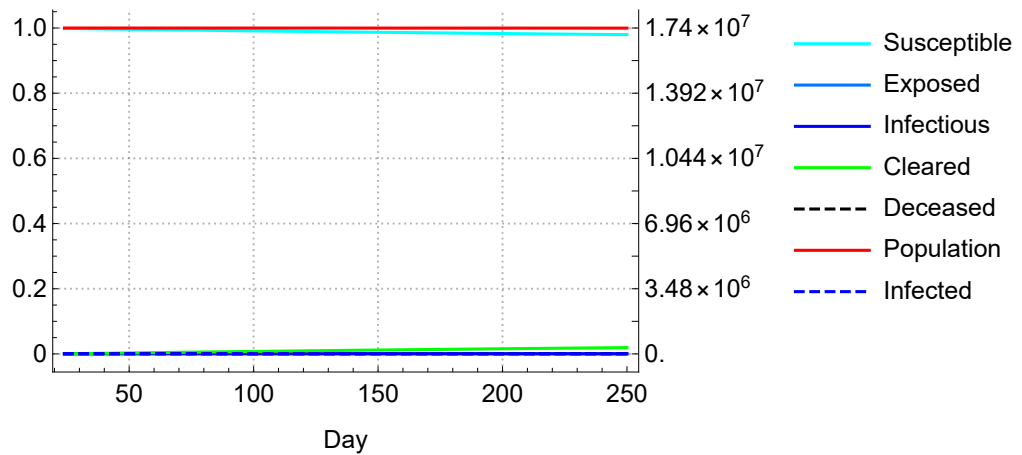

After 250 days, the death toll has developed into some 5100. However, the daily number has been dropping and the cumulative figure has been flattening for a long while.

```
Explain[SEYCD[250], SEYCD]
```

$\{S \to 1.70473 \times 10^7, E \to 1107.8, Y \to 11\,389.4, C \to 335\,057.,$

$D \to 5102.39, N \to 1.73949 \times 10^7, I \to 12\,497.2, A \to 340\,159.\}$

```
Show[pbase, pintFull]
```
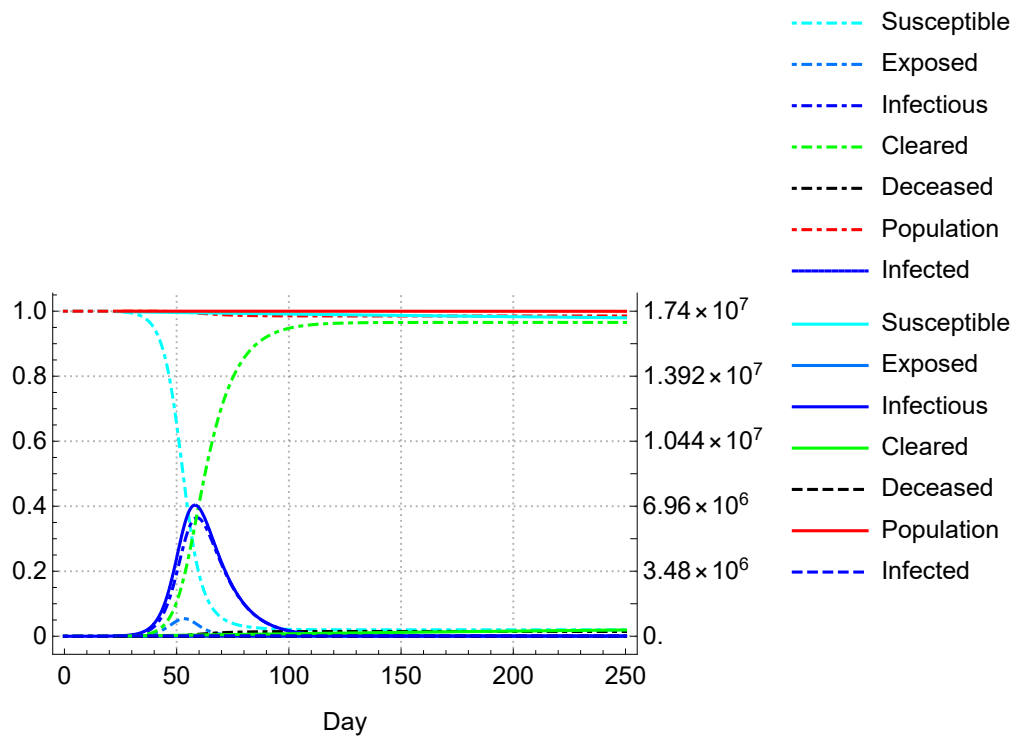

We see most activity around the day of the intervention. We can clearly see that two models have been stitched together. The period till day 24 only shows the base scenario with the dotted lines. From day 24 onwards there are also the default lines given by the intervention scenario. (The LHS plot range is so small that there are no vertical ticks.)

```
Show[pbase, pintFull, PlotRange → {{20, 30}, {0, .002}}]
```


–·–·–	Susceptible
–·–·–	Exposed
–·–·–	Infectious
–·–·–	Cleared
–·–·–	Deceased
–·–·–	Population
———	Infected
———	Susceptible
———	Exposed
———	Infectious
———	Cleared
– – –	Deceased
———	Population
– – –	Infected
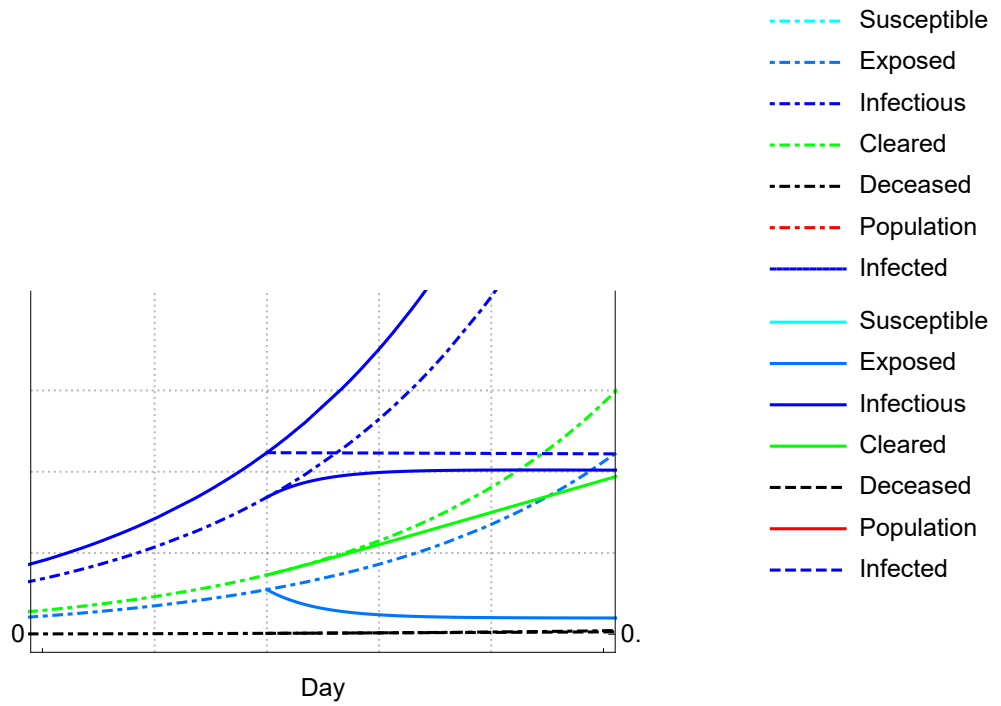
```
Show[pbase, pintFull, PlotRange → {{20, 150}, {0, .003}}]
```


–·–·–	Susceptible
–·–·–	Exposed
–·–·–	Infectious
–·–·–	Cleared
–·–·–	Deceased
–·–·–	Population
———	Infected
———	Susceptible
———	Exposed
———	Infectious
———	Cleared
– – –	Deceased
———	Population
– – –	Infected
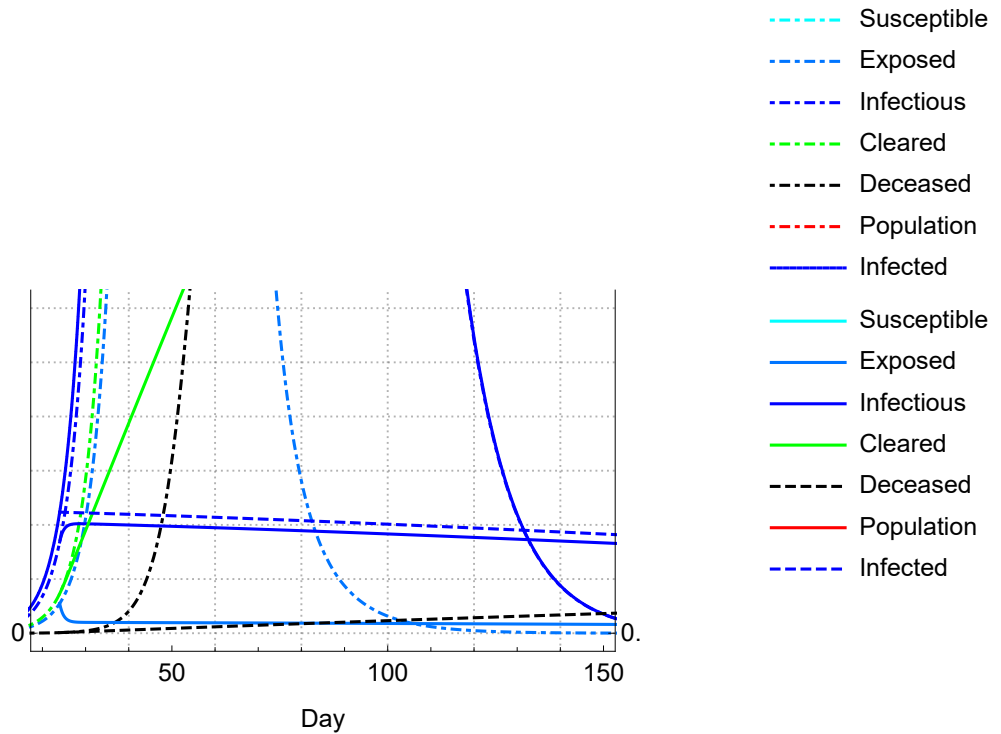
13.2.4. Use of default routines on top and limit

The advantage of a full (Alternative) model is that the routines on top and limit can now use Options[SIA].

However, with $R_0 < 1$, the infection peters out. The standard routines are mostly interesting for *spreading* infections instead, and for domains where the interpolating functions have been defined.

(1) For the limiting values, we see that new β would cause a death count of 4066 at the limit. However, we have to add the 96 dead from day 24 from the old regime.

```
Explain[SICD[Limit] * NqSeed, SICD]
```

$\{S \to 1.71289 \times 10^7, I \to 0., C \to 267051., D \to 4066.77, N \to 1.73959 \times 10^7, A \to 271118.\}$

(2) For the top, the standard formula is no use for $R_0 < 1$. We might insert the share of the acquitted compartment at $t = 24$, but this doesn't help much.

```
Ap[24]
```

0.000370826

```
SIA[Ip, Top, .99, Ap[24]]
```

SIA: The value of R0 = 0.99 must be above 1

{1.0101, −0.000319981, −0.00978103}

We might use NMaximize. However, the effect of the intervention is that the infections immediately start falling down from their value at $t = 24$.

```
NMaximize[Iq[t], t] // Quiet (* suppresses warning messages about ranges *)
```

$\{19437.8, \{t \to 24.\}\}$

Much of the same holds for the standard routine TforTopIq.

```
TforTopIq[] // Quiet
```

23.996

(3) For the 95% score of the limit values, there are two values, that both fall outside of our interpolation range. If we neglect the starting values, and assume that Sp = 1 and Ap = 0, then the routine looks for a value below t = 24. When we include the starting values from t = 24, then the routine looks for a value beyond t = 250.

```
TforAp95[Not, 0.99]  // Quiet (* exclude starting values *)
```

23.9742

```
TforAp95[0.99] // Quiet(* standardly include starting values *)
```

666.21

13.2.5. Conclusion on the comparison of the Base versus Alternative model

The Base and Alternative models provide for *models, runs, trajectories for the variables, plots, and routines for answering particular questions.* They seem a relevant environment for applications of SI(EY)A(CD) issues. While these aspects are feasible, section 13.2.4 has indicated what is *less* feasible when the intervention forces $R_0 < 1$.

Let us now proceed with the question how to choose the intervention parameters, with the subsequent issue of estimation.

14. Data

14.1. Recorded death data

There is a distinction between the Dq[t] in the model and the recorded death figures Drec[τ]. The Dq in the model concern the end of their infectious period. The statistical records mention the actual day of dying. This can include a longer period in the hospital, with effective quarantine. Thus the recorded days are not directly useful for measuring the end of the infectious period.

14.2. Back-tracing in linear manner

Assume that both t and τ are measured from the day of onset. The day of death τ follows the day t of the end of the infectious period with some constant delay plus some distribution for the duration of the death-bed. Currently we have no other indication for the latter duration other than the distance $\tau - t$ itself. In formula: $\tau = t$ + constant + slope ($\tau - t$) with constant 0 and slope 1. However, this inspires the approach to back-trace τ to a value t in the model as t = constant + slope τ, for some estimated constant and slope values. The constant would be negative (and cannot be more negative than the period of the first death). The slope would be below 1. We can try some values below, and can include the parameters in the estimation.

14.3. The data, and elimination of the presumed onset

We model the (partial) lockdown by the Dutch government on March 15 2020. Allowing for a delay of some days, the lockdown became effective on March 18, day 24. We can model this by different values for both β and γ.

By April 5 (Day 42) we see a top of the number of recorded deaths *per day* (not shown here). The cumulated death count then is 1766, and it continues to rise to 5748 by May 20 (Day 87), though in ever lower values of deaths per day. The deaths in the weeks immediately following day 24 have been caused by the infections before. If we assume that 3000 cases were in the pipeline under the old values of β and γ, then the other 3000 cases on May 20 must come from the new regime of the lockdown. Even if the new $R_0 < 1$, there will still be some deaths to reckon with, given the many (asymptomatic) infectious units still available in the population.

```
PlotLine[Take /. Options[SEYCDT],
 AxesLabel → {"Day\nsince\nonset", "Obs. Cum. Deceased"}]
```
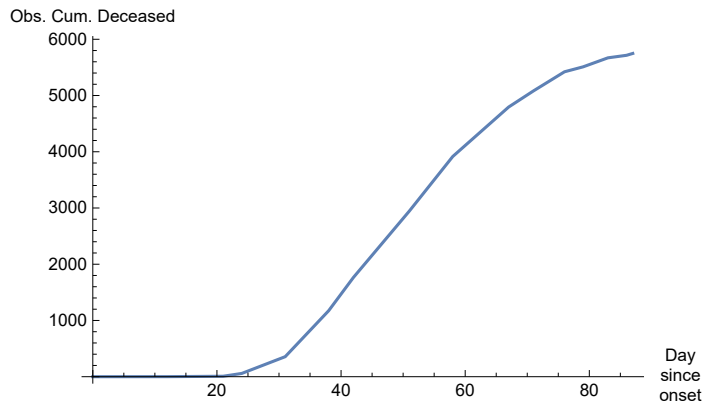

These are very noisy observations: (i) GP's and their patients have become much more reluctant to have patients go to the hospital and ICU, because they have learned more about survival options, and about the possibility that one could die in an unconscious state in an ICU bed without contact with family. (ii) There is the notion of "excess deaths" compared to the level in normal years. There is still no testing of deaths in general, and if the pandemic proceeds then the share of *untested infection related deaths* will first increase and then, when more testing is done, go down again. (iii) It is dubious to assume that we can neglect "normal deaths" when normal treatment has collapsed because of this pandemic, though it is correct that these do not fall under the *Infection Fatality Factor* as such.

Given this noise, and that estimation will be slower when we include all points, we do not take all available data, but select a few points that allow a rough reproduction of the shape of the curve. Some 16 points including onset appear to be enough.

Some example dates and values are the following.

```
{onset = {2020, 02, 23};
 DayCount[onset, {2020, 03, 18}],
 DayCount[onset, {2020, 03, 25}],
 DayCount[onset, {2020, 04, 05}],
 DayCount[onset, {2020, 05, 20}]}
```

{24, 31, 42, 87}

Counting from the presumed day of onset, the following are our data, consisting of #day and #Drec. The element with zero deaths will not be included in the estimation, and must be removed. We consider days before the recorded days of death, but we cannot look before 0, because the interpolation starts at 0.

```
Take /. Options[SEYCDT]
```

$$\begin{pmatrix} 0 & 0 \\ 12 & 1 \\ 21 & 8 \\ 24 & 58 \\ 31 & 356 \\ 38 & 1173 \\ 42 & 1766 \\ 51 & 2945 \\ 58 & 3916 \\ 67 & 4795 \\ 71 & 5082 \\ 76 & 5422 \\ 79 & 5510 \\ 83 & 5670 \\ 86 & 5715 \\ 87 & 5748 \end{pmatrix}$$

```
SEYCDT[Options, Update, Rest]; (* remove the zero date of onset *)

Length /. Options[SEYCDT] (* the number of our data points *)
```

15

14.4. Comparison with the data

Some 6000 deaths in May compares favourably to the 250,000 deaths that could have occurred. Let us compare the recorded deaths to what would have happened without an intervention.

PM. For comparison, we need to run the base scenario again. This is also useful for the later exercise on the ProductLog.

```
ResetOptions[SIA] ; SEYCD[Run, Pr, newModel, t, 250];
```

The following plot uses the routine SEYCDT[PlotLine, ...] which takes the *onset date* from Options[S-EYCDT], which is regarded as Day 0.

```
SEYCDT[PlotLine, PlotLabel → "Model without intervention"]
(* same blue line as above, max 6000 *)
```
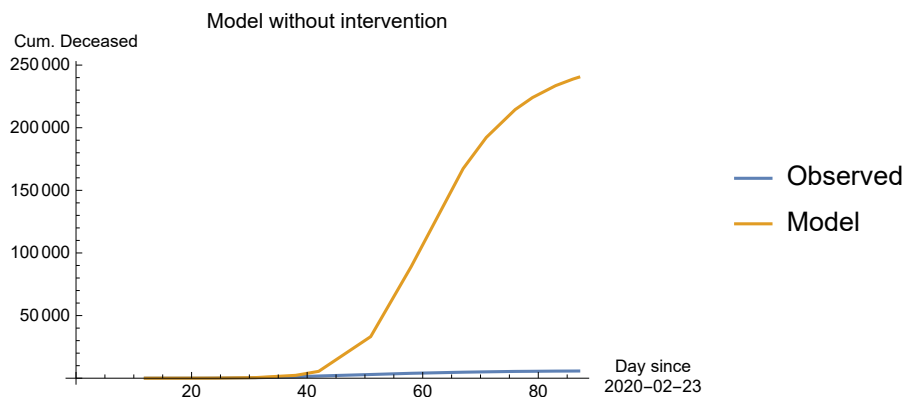
15. Intervention and estimation, using SEYCDT

15.1. Introduction

This chapter has the same structure as the later Chapter 16 on SEYCDB. The only difference is the practical problem of numerical estimation with an ill-conditioned problem. After considering both approaches, my impression for the current setup is that using β actually is a bit better than using R_0, but this impression may be caused by that I originally started with β.

In retrospect, we should better take account of the fact that the data are *discrete* and *defined per day*. Thus we could make a grid of days for the "actual onset before the chosen onset" and the "constant in the back-trace", and estimate the continuous parameters for the separate values in the grid. This ought to reduce the problems that we now have with the conditioning. Then, perhaps with more data and a better specification for the back-tracing, we might used FindFit to even consider confidence intervals. But we are oriented at didactics now and not at the best estimate (on dubious data).

15.2. Find the intervention parameters for SEYCDT

The recorded cumulated number of deaths gives us important information. The curve is bending towards some limiting value. This provides information about the implied R_0. Let us assume that the death toll for the remainder of the year moves gradually to 7000. Assume that 7000 - x deaths would be attributed to the new regime, assuming a value of $R_0 < 1$, and x are still accountable to the old regime. For the new limit value, the "starting values" at day 24 are important. We have the ProductLog rule for the limiting value. Let us substitute an unknown R_0 and the values of the variables at day 24 of the intervention, and assume that $\gamma = 1/10$ remains the same. (This requires that the Base model in section 13.1 has run.)

```
startvalues = SEYCDT[Intervention, 24, f0, .1]
```

{Seed → 4777.44, Yqt0 → 14 660.4, Cqt0 → 6355.58, Dqt0 → 96.7855,

  Aqt0 → 6452.37, R0() → f0, GammaSIA() → 0.1, BetaSIA() → 0.1 f0, Onset → 24}

The death toll that has already occurred or is in the pipeline is:

```
Dqt0 + PhiSICD[] (Seed + Yqt0) /. startvalues /. Options[SIA]
```

388.353

In the SICD[Limit] list, the Dp are in the 4th position. The limiting value applies to the excess deaths, given the startvalues. Let us include a number x that can also be attributed to the Base case, above to the latter pipeline (which changes above notion of x).

```
assign[x_] =
  (7000 - TheFormer - x) / NqSeed == SICD[Limit, Sequence @@ startvalues][[4]]
```

$5.74713 \times 10^{-8} (6611.65 - x) =$

$$\frac{1. \, W\left(-0.999355 \, e^{-0.999629 \, f0} \, f0\right)}{f0} + \frac{0.985 \log\left(-\frac{1.00065 \, W\left(-0.999355 \, e^{-0.999629 \, f0} \, f0\right)}{f0}\right)}{f0} + 0.999635$$

For $x = 0$, the R_0 of the new regime would be slightly above 1.

```
FindRoot[assign[0], {f0, 1}]
```

{f0 → 1.00207}

It appears that the idea of assigning a death toll to a regime is somewhat simplistic. A clearcut number is the recorded death toll and the IFF applied to the already infected. What happens after the intervention is not only influenced by the original situation but also by the intervention. We may make an arbitrary choice along the following contour. In the end, it are the practical measures that determine the β and γ of the new regime. For our exercise, a solution might be to say that $x/2$ is caused by the old regime and $x/2$ by the new regime. Or it might be fairer to say that most deaths still apply to the new regime because it will exist for a much longer period. At this stage, all theory is arbitrary, and it suffices to take $R_0 = .99$.

```
r0fromx[x_] := f0 /. FindRoot[assign[x], {f0, 1}]
```

```
newR0 == r0fromx[6612 / 2]
```

newR0 = 0.98423

```
Plot[r0fromx[x], {x, 0, 7000},
 AxesLabel → {"x deaths\ncaused by\nold regime", "New R0"}, AxesOrigin → {0, 0}]
```


A consideration may be that $1/\gamma$ would be shorter because of self-quarantine. However, a higher value of γ also means a faster development of the acquitted, and thus also a surge of the number of deaths in the period just after the intervention. This is illogical. Gamma is best understood as a normalisation constant, and we keep it the same. The whole burden of change falls onto β.

```
SetOptions[Intervention, Time[Intervention] → 24, R0[Intervention] → .99,
 BetaSIA[Intervention] → .99 * .1, GammaSIA[Intervention] → .1]
```

{Time(Intervention) → 24, R0(Intervention) → 0.99,

  BetaSIA(Intervention) → 0.099, GammaSIA(Intervention) → 0.1}

15.3. Intervention at t = 24, using SEYCDT

```
PlotLegendsSIA[SEYCD]
(* intervention uses the default plotting colours and legends *)
```

```
NqSeed = 17.4 × 10^6;
ResetOptions[SIA] (* not uninmportant *)
```

$\{Aqt0 \to 0, Cqt0 \to 0, DataMold \to \{S, I, A\}, Dqt0 \to 0, Iqt0 \to Null, Nqt0 \to 1.74 \times 10^7, Onset \to 0, Seed \to 100,$

$\quad Yqt0 \to 0, AlphaSEYCD() \to 1., BetaSIA() \to 0.4, GammaSIA() \to 0.1, PhiSICD() \to 0.015, R0() \to Null\}$

The equations for SEYCDT use R0[] instead of BetaSIA[]. The first reason is conceptually, that we are effectively estimating the new reproductive factor when we keep GammaSIA[] constant. The second reason is that the estimation relations appear to be a bit ill-conditioned, whence it seems better to use a parameter that is ten times larger.

```
SetOptions[SIA, R0[] → 4]; (* value before intervention *)
```

```
Options[Intervention]  (* set above *)
```

$\{Time(Intervention) \to 24, R0(Intervention) \to 0.99,$

$\quad BetaSIA(Intervention) \to 0.099, GammaSIA(Intervention) \to 0.1\}$

See above how this routine could be done in steps.

```
SEYCDT[Run, Pr, intModel, t, 250]; (* longer period,
but in estimation shorter is faster *)
```

```
intModel[t] (* notice the WhenEvent statement *)
```

$Hold\big[\{Sq'(t) = -5.74713 \times 10^{-8} GammaSIA(t) R0(t) Sq(t) Yq(t), Eq'(t) = -1. Eq(t) - Sq'(t),$

$\quad Yq'(t) = 1. Eq(t) - Aq'(t), Aq'(t) = GammaSIA(t) Yq(t), Dq(t) = 0.015 Aq(t),$

$\quad Cq(t) = Aq(t) - Dq(t), Nq(t) = 1.74 \times 10^7 - Dq(t), Iq(t) = Eq(t) + Yq(t), Nq(0) = 1.74 \times 10^7,$

$\quad Sq(0) = 1.73999 \times 10^7, Eq(0) = 100, Yq(0) = 0, Cq(0) = 0, Dq(0) = 0, Aq(0) = 0, Iq(0) = 100,$

$\quad R0(0) = 4, GammaSIA(0) = 0.1, WhenEvent[t > 24, \{R0(t) \to 0.99, GammaSIA(t) \to 0.1\}]\}\big]$

The supposed top of "recorded deaths per day" was on day 42, with a cumulated total of 1766 deaths. The run of the model of the intervention, see below, mentions 565 deaths. However, we must keep in mind that these death counts have different meanings. See the estimation below.

```
Explain[SEYCD[42], SEYCD] (* day 24: supposed top of "reported deaths per day" *)
```

$\{S \to 1.73432 \times 10^7, E \to 1713.13, Y \to 17\,338.6, C \to 37\,143.6,$

$\quad D \to 565.638, N \to 1.73994 \times 10^7, I \to 19\,051.7, A \to 37\,709.2\}$

The most useful plot now might be the one on ICU beds.

```
seicdplotoptionsICU = SEYCD[Plot, Options, ICU, PlotRange → {0, 10000}];
Plot @@ {SEYCD[ICU, 1500, t], {t, 0, 150}, seicdplotoptionsICU}
```


The values at the end of the 250 days are (some 5000 deaths and not 7000 yet):

```
Explain[SEYCD[250], SEYCD]
```

$\{S \to 1.70473 \times 10^7, E \to 1107.8, Y \to 11\,389.4, C \to 335\,057.,$

$\quad D \to 5102.39, N \to 1.73949 \times 10^7, I \to 12\,497.2, A \to 340\,159.\}$

```
pintv = SEYCD[Plot, 0, 250]
```


Combining the two plots of the base case and the intervention.

```
Show[pbase, pintv]
```


The following is around the day of the intervention, day 24, using a plotrange of 0.2% (no axes ticks). This is the same outcome as the separate running of the Base and Alternative scenario's, except that the WhenEvent model allows drawn lines from day 0 too.

```
Show[pbase, pintv, PlotRange → {{20, 30}, {0, .002}}]
```


15.4. Estimation: aim and result

For estimation we are interested in the new value of R_0. For estimation it however appears to be better to also estimate the old value and the day of onset. The number of infected at the day of the intervention is not only determined by the old R_0 but also by how long the infection has been raging.

The summary of our findings on estimation is as follows. The following graph gives the Root Mean Squared Error (RMSE) as a function of both the R_0 before the intervention (R_0-pre) and the possibility that the onset of infections happened some days earlier than Carnival, February 23 2020. It appears that our choice of $R_0 = 4$ agrees with 9-10 days earlier: Valentine's Day. With a higher R_0 then there is less need for an earlier onset. We can maintain the onset at Carnival, if R_0-pre would be 5.7, and then have the same RMSE level as in other cases. In all these cases we maintained $\gamma = 0.1$.
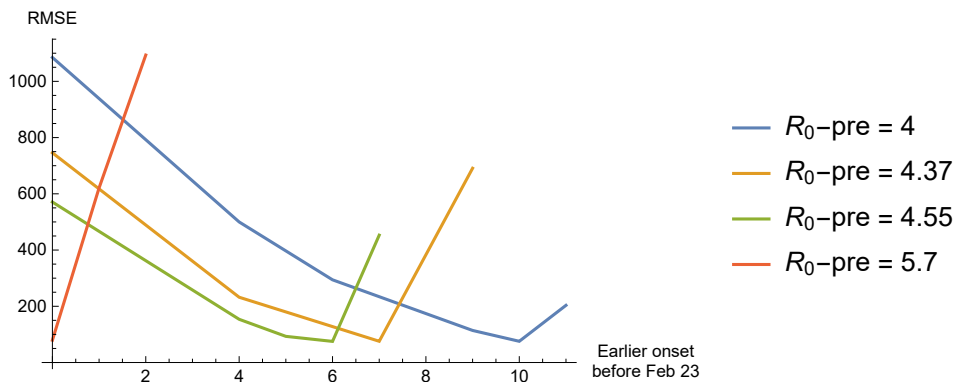
In all such minimum combinations, we get a graph like below, on the fit of the model estimate of

the Dq[t] and the recorded death count Drec[τ]. The following gives the combination of R_0-pre = 4 and the required earlier onset of about 10 days (stated by Plus → 9.847).
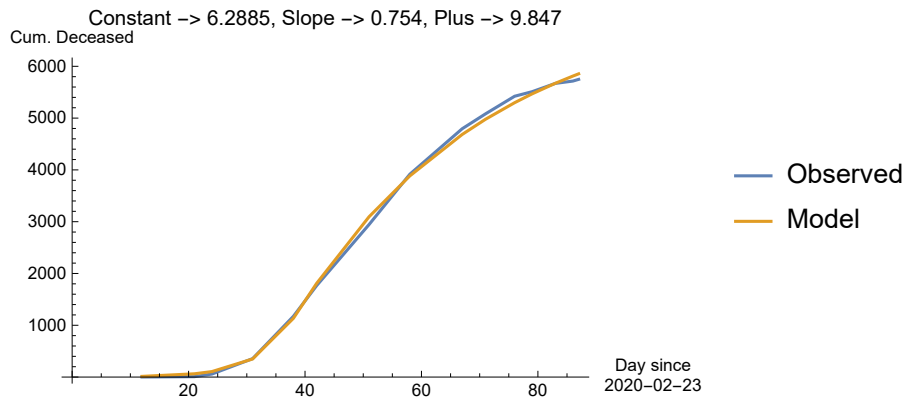
In all such minimum combinations, the R_0 after the intervention drops to about 0.5 (imposed lower boundary), and the value of deaths after 250 days is a bit above 7,000.

15.5. Major routine: the definition of the error

Because of estimation we may restrict the time horizon to the available data.

```
tEst = 100.;
```

```
NqSeed = 17.4 × 10^6;
ResetOptions[SIA] (* not uninmportant *)
```

{Aqt0 → 0, Cqt0 → 0, DataMold → {S, I, A}, Dqt0 → 0, Iqt0 → Null, Nqt0 → 1.74×10^7, Onset → 0, Seed → 100,
  Yqt0 → 0, AlphaSEYCD() → 1., BetaSIA() → 0.4, GammaSIA() → 0.1, PhiSICD() → 0.015, R0() → Null}

The equations for SEYCDT use R0[] instead of BetaSIA[]. The first reason is conceptually, that we are effectively estimating the new reproductive factor when we keep GammaSIA[] constant. The second reason is that the estimation relations appear to be a bit ill-conditioned, whence it seems better to use a parameter that is ten times larger.

```
SetOptions[SIA, R0[] → 4]; (* value before intervention *)
```

```
ResetOptions[SEYCDT];
```

```
SEYCDT[Options, Update, Rest]; (* remove the zero date of onset *)
```

```
SetOptions[Intervention, Time[Intervention] → 24, R0[Intervention] → .99,
 BetaSIA[Intervention] → .99 * .1, GammaSIA[Intervention] → .1]
```

{Time(Intervention) → 24, R0(Intervention) → 0.99,
  BetaSIA(Intervention) → 0.099, GammaSIA(Intervention) → 0.1}

The following is a major routine. The major step in estimation is the definition of the error. Thereafter it is only a matter of minimising the error. The following runs the model with the stated intervention parameters, at the day of intervention in Options[Intervention]. It also give the Root Mean Squared Error (RMSE) using the data in the Options[SEYCDT] (with the options Day and Count for the death counts).

The following sets R_0 = 0.99 and keeps γ = 0.1. We set it now, but it will later be done by the estimation routine.

```
SEYCDT["RMSE", intModel, .99, .1, {t, tEst}]
```

2869.6

```
intModel[t]
```

$\text{Hold}\big[\{Sq'(t) = -5.74713 \times 10^{-8}\ \text{GammaSIA}(t)\ R0(t)\ Sq(t)\ Yq(t),\ Eq'(t) = -1.\ Eq(t) - Sq'(t),$

$\qquad Yq'(t) = 1.\ Eq(t) - Aq'(t),\ Aq'(t) = \text{GammaSIA}(t)\ Yq(t),\ Dq(t) = 0.015\ Aq(t),$

$\qquad Cq(t) = Aq(t) - Dq(t),\ Nq(t) = 1.74 \times 10^{7} - Dq(t),\ Iq(t) = Eq(t) + Yq(t),\ Nq(0) = 1.74 \times 10^{7},$

$\qquad Sq(0) = 1.73999 \times 10^{7},\ Eq(0) = 100,\ Yq(0) = 0,\ Cq(0) = 0,\ Dq(0) = 0,\ Aq(0) = 0,\ Iq(0) = 100,$

$\qquad R0(0) = 4,\ \text{GammaSIA}(0) = 0.1,\ \text{WhenEvent}[t > 24, \{R0(t) \to 0.99,\ \text{GammaSIA}(t) \to 0.1\}]\}\big]$

The plots shows that the model fits the last weeks of March but it doesn't explain the observed deaths in April and May. A possibility is that the pre-intervention R_0 was much higher, but as stated, $R_0 = 4$ already is high. We are drawn into an evaluation of the date of onset and the meaning of the dates of the death counts.

PM. The PlotLabel informs us of the values of the Constant and Slope in the Options[SEYCDT] about the relation between the recorded day of death τ and the end of the infectious period t. The value of Plus gives the number of days that should be added to the horizontal axis if we would take onset so many days earlier.

```
SEYCDT[PlotLine]
```
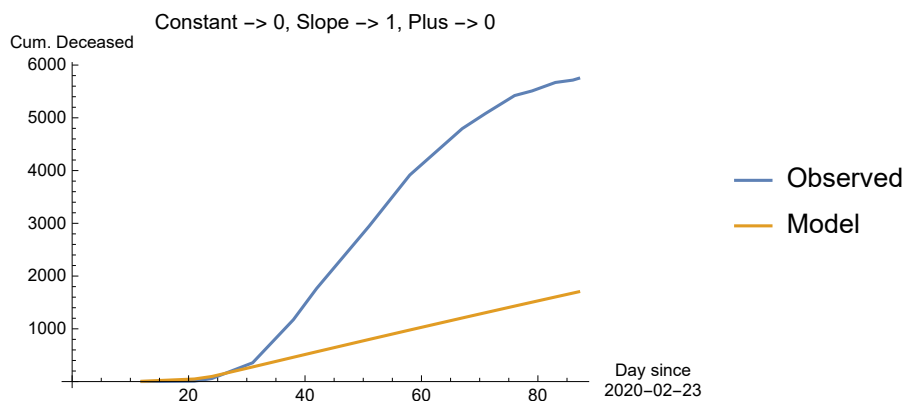

15.6. Minimising the RMSE

Given this high RMSE, we may try to minimise it. We might consider that the death counts from the early period, say to day 35, are determined by the original R_0 and γ, and then exclude these data, to estimate the new parameters. However, the SEYCDT model uses the *WhenEvent* feature and thus accurately manages the regime change. The problem does not reside with the Dq[t] that the model calculates. The true issue is that the observed death counts concern *actual days of death*, while the model requires the *end of the infectious period*.

Thus, we must back-trace the death counts to an earlier moment in time. It is an option to extend the model, say with $U' = u\,D$, with the hospital phase (u taken from ICU), like the Goh (2020) "epidemic calculator". At this point it seems more logical to work more directly with the data, if only to better grasp what the implications are.

The best approach would seem to be to allow the estimation routine to determine both model parameters and the time shift between Dq and the actually recorded deaths. It takes some care to

find the proper interpretation and minimum. We may consider to actually also estimate R_0 and γ for the period before the intervention. Also, the assumption of the day of onset is only an assumption.

For example, consider a fixed back-tracing of 7 days and a proportional reduction of 80%. The first recorded death on day 12 then actually ended its infectious period at day -7 + 0.8 * 12 ≈ 3. We drop the first observation, which is the day of onset (with no death count). In above chart, the model looked like a good fit for the first weeks of March, but taking account of back-tracing shows that it would actually be a bad fit for the whole period.

Let us see how the error looks when we impose these changes. The error is large and there is still a wide gap between the model and the data.

```
ResetOptions[SEYCDT];

SEYCDT[Options, Update, Rest]; (* remove the zero date of onset *)

SetOptions[SEYCDT, Constant → -7, Slope → .8];

SEYCDT["RMSE", intModel, .99, .1, {t, tEst}]
(* uses both model output and the data values of the options *)
```
3310.02

```
SEYCDT[PlotLine]
```
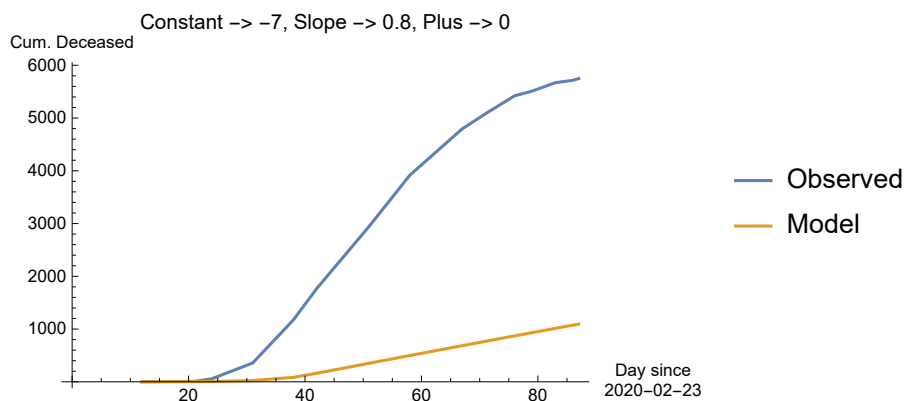

Taking advantage of some estimations not shown, it appears that the estimation routine has a tendency to take a *positive* constant in the back-tracing, instead of a negative one, as we actually required. However, our assumption of the onset may be wrong. Above graph shows that the model is still in the exponential phase and still far away from the slowing down towards peaking. Allowing a positive shift *in the model* actually means that the onset must have been earlier in the year.

Let us use a positive constant as a proxy for an earlier date of onset. (Setting the value of Plus has no effect, since this option is controlled by the estimation routine on the onset, and we don't do an estimation on onset yet.)

```
SetOptions[SEYCDT, Constant → 14];

SEYCDT["RMSE", intModel, .99, .1, {t, tEst}]
```
2883.9

This has a lower RMSE but the growth of infections in the model remains low due to the fixed

intervention date at "day 24". When the onset was earlier in the year, then also the day number of intervention shifts ...

SEYCDT[PlotLine]
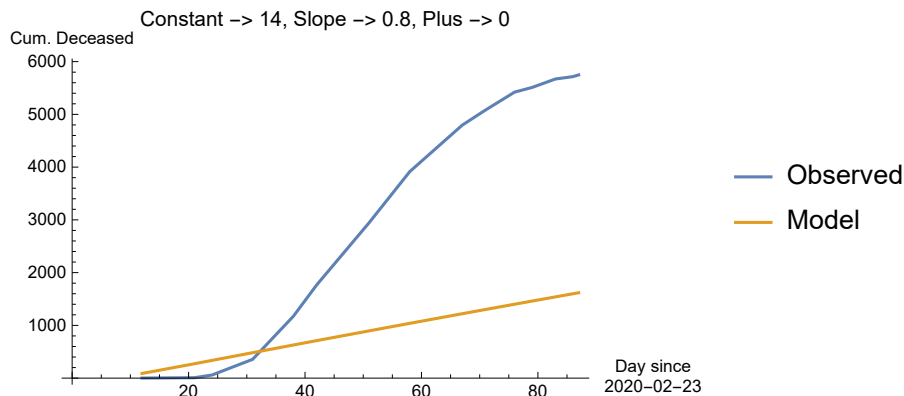
15.7. Combined estimate of new onset, pre-intervention R_0, intervention R_0, and back-tracing

This subsection forms the core of the chapter. The following routine combines above separate statements. When we have a RMSE then we can apply FindMinimum and NMinimize to it.

15.7.1. Calculating RMSE given numerical parameter values

The following only calculates a RMSE given numerical parameter values. Let us assume that onset happened 14 days earlier. Let the pre-intervention $R_0 = 6$ and the intervention value $R_0 = 0.99$.

SEYCDT["RMSE", 14, −7, 0.8, intModel, R0, 6, 0.99, {t, tEst}]

157 155.

This overshoots, which gives confidence that an estimation should be able to find values in-between.

Check that the model indeed has R0[0] == 6 and that the WhenEvent is at day 24 + 14 = 38.

intModel[t]

$\text{Hold}\big[\{Sq'(t) = -5.74713 \times 10^{-8} \text{ GammaSIA}(t) \text{ R0}(t) \text{ Sq}(t) \text{ Yq}(t),\ Eq'(t) = -1.\ Eq(t) - Sq'(t),$
$Yq'(t) = 1.\ Eq(t) - Aq'(t),\ Aq'(t) = \text{GammaSIA}(t)\ Yq(t),\ Dq(t) = 0.015\ Aq(t),$
$Cq(t) = Aq(t) - Dq(t),\ Nq(t) = 1.74 \times 10^7 - Dq(t),\ Iq(t) = Eq(t) + Yq(t),\ Nq(0) = 1.74 \times 10^7,$
$Sq(0) = 1.73999 \times 10^7,\ Eq(0) = 100,\ Yq(0) = 0,\ Cq(0) = 0,\ Dq(0) = 0,\ Aq(0) = 0,\ Iq(0) = 100,$
$R0(0) = 6,\ \text{GammaSIA}(0) = 0.1,\ \text{WhenEvent}[t > 38,\ \{R0(t) \to 0.99,\ \text{GammaSIA}(t) \to 0.1\}]\}\big]$

15.7.2. FindMinimum on the RMSE

The following is somewhat involved: (i) we shift the onset and intervention, and (ii) we estimate R_0 before and after the intervention, and (iii) we relate the death counts to the infectious periods. Apparently it works nicely, but we lack information about confidence intervals. PM. Estimation time can be reduced by taking tmax = 100, which is the interval for which there are data. But it can be useful to project to t = 250.

Unfortunately, FindMinimum is sensitive to starting values. The following takes advantage of earlier runs.

```
FindMinimum[SEYCDT["RMSE", onsetplus, c, s, intModel, R0, r0pre, r0, {t, 250}],
 {{onsetplus, 7, 0, 21},
  {c, -3, -4, 0}, {s, .75, .1, 1}, {r0pre, 4.37, 1., 6.}, {r0, 0.55, 0.5, 4.}}]
```

... FindMinimum: The line search decreased the step size to within the tolerance specified by AccuracyGoal and PrecisionGoal but was unable to find a sufficient decrease in the function. You may need more than MachinePrecision digits of working precision to meet these tolerances.

$\{75.1128, \{onsetplus \to 7.33732, c \to -1.27676, s \to 0.682958, r0pre \to 4.36302, r0 \to 0.509851\}\}$

Observe that the model now uses the latter parameters.

```
intModel[t]
```

$\text{Hold}\big[\{Sq'(t) = -5.74713 \times 10^{-8} \text{ GammaSIA}(t) \text{ R0}(t) \text{ Sq}(t) \text{ Yq}(t), Eq'(t) = -1. Eq(t) - Sq'(t),$
$Yq'(t) = 1. Eq(t) - Aq'(t), Aq'(t) = \text{GammaSIA}(t) Yq(t), Dq(t) = 0.015 Aq(t), Cq(t) = Aq(t) - Dq(t),$
$Nq(t) = 1.74 \times 10^7 - Dq(t), Iq(t) = Eq(t) + Yq(t), Nq(0) = 1.74 \times 10^7, Sq(0) = 1.73999 \times 10^7,$
$Eq(0) = 100, Yq(0) = 0, Cq(0) = 0, Dq(0) = 0, Aq(0) = 0, Iq(0) = 100, R0(0) = 4.36302,$
$\text{GammaSIA}(0) = 0.1, \text{WhenEvent}[t > 31.3373, \{R0(t) \to 0.509851, \text{GammaSIA}(t) \to 0.1\}]\}\big]$

The death toll will rise to some 7000 over the course of the year.

```
Explain[SEYCD[250], SEYCD]
```

$\{S \to 1.693 \times 10^7, E \to 0.293306, Y \to 5.62931, C \to 462\,981.,$
$D \to 7050.47, N \to 1.73929 \times 10^7, I \to 5.92262, A \to 470\,031.\}$

Reading below chart requires the following explanation.

(a) The time axis below specifies that February 23 is day 0. This applies to the recorded death counts.

(b) The value of option Plus gives above estimate that the day of onset was actually a week earlier. Thus the model started a week before t = 0 in the plot, and also the intervention was later.

(c) The back-tracing of the death counts to the end of the infectious period has used the formula t = constant + slope τ, with the stated values, with t the value in the model (using the plus value) and τ the recorded day since 2020-02-23 (thus without the plus value). Thus the Constant in below plot (6.06) is the value of Plus (7.33) + the value of the constant in the estimate (-1.27).

```
SEYCDT[PlotLine]
```
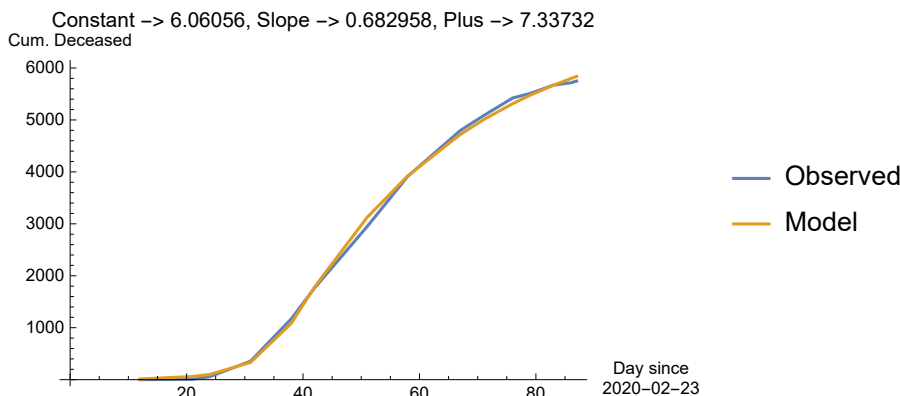
15.7.3. Indication of lack of sensitivity to small changes

Above estimate of pre-intervention R_0 = 4.4 gives us some worry about our choice of R_0 = 4. In another run, not shown, a value was fixed with R_0 = 4. For these values, the RMSE hardly changes and the plot and limit values remain quite the same.

```
SEYCDT["RMSE", 9.847, -3.5585, 0.754, intModel, R0, 4, 0.595, {t, 250}]
```

78.9544

```
SEYCDT[PlotLine]
```
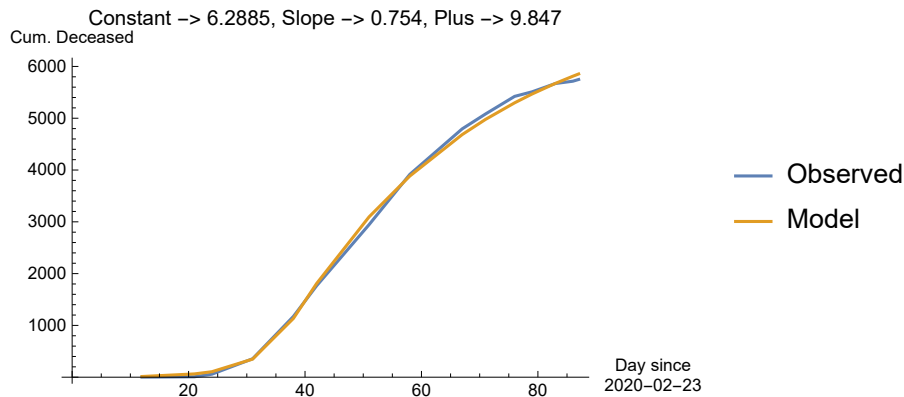

The implications towards later in the year:

```
Explain[SEYCD[250], SEYCD]
```

$\{S \rightarrow 1.69144 \times 10^7, E \rightarrow 1.92206, Y \rightarrow 31.9096, C \rightarrow 478\,297.,$
$\quad D \rightarrow 7283.71, N \rightarrow 1.73927 \times 10^7, I \rightarrow 33.8316, A \rightarrow 485\,581.\}$

The conclusion seems warranted that onset in Holland was some 9-10 days before Carnival February 23, our earlier choice of the onset. Ah, yes, Valentine's day, February 14. Potentially people returned to Holland to celebrate Valentine's day.

15.7.4. FindMinimum with a condition

Let us fix onsetplus = 0, so that the onset was really at Carnival, and we want to know what R_0 would have been needed. For this, we use the possibility of a condition.

```
FindMinimum[{SEYCDT["RMSE", onsetplus, c, s, intModel, R0, r0pre, r0, {t, tEst}],
  {onsetplus == 0} },
 {{onsetplus, 7, 0, 21},
  {c, -3, -4, 0}, {s, .75, .1, 1}, {r0pre, 5.7, 1, 6}, {r0, .55, .5, 4}}]
```

⋯ FindMinimum: Failed to converge to the requested accuracy or precision within 500 iterations.

$\{79.1202, \{onsetplus \rightarrow 0., c \rightarrow -0.0000777267, s \rightarrow 0.675223, r0pre \rightarrow 5.68902, r0 \rightarrow 0.517493\}\}$

```
SEYCDT[PlotLine]
```

Constant –> –0.0000777267, Slope –> 0.675223, Plus –> 0.
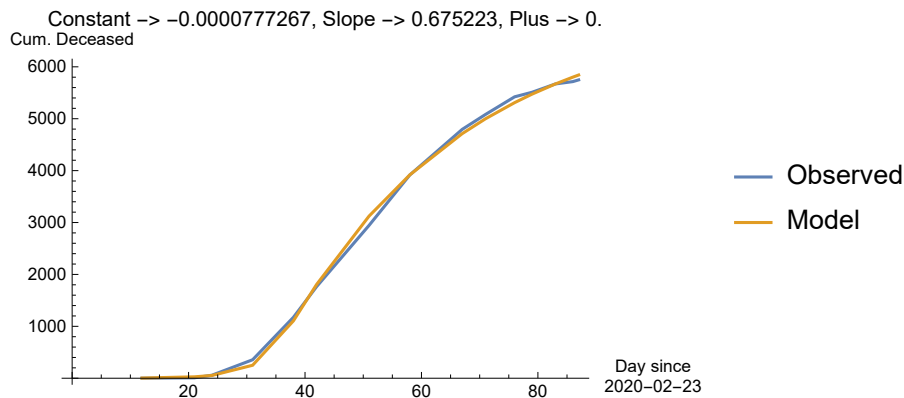
The implications towards later in the year are yet unclear since we used tEst = 100.

```
Explain[SEYCD[tEst], SEYCD]
```

$\{S \to 1.69319 \times 10^7, E \to 302.136, Y \to 5716.88, C \to 455\,187.,$
$\quad D \to 6931.79, N \to 1.73931 \times 10^7, I \to 6019.02, A \to 462\,119.\}$

15.8. Variant of startvalues from 1 to 6

Mathematica's FindMinimum has a known sensitivity to starting values. We might also use NMinimize, that is more robust in this respect, but in this case it is helpful to see the various outcomes in fairly quick manner.

Check the sensitivity of FindMinimum to starting values, and that our choice based upon the literature on $R_0 \approx 4$ and $1/\gamma \approx 10$ was not unwise. A consideration is that the linear scheme for back-tracing might influence the outcome too much, but this can only be tested when there would be relevant other schemes.

```
Table[
 FindMinimum[SEYCDT["RMSE", onsetplus, c, s, intModel, R0, r0pre, r0, {t, tEst}],
  {{onsetplus, 7, 0, 21},
   {c, -1, -4, 0}, {s, .75, .1, 1}, {r0pre, test, 1, 8}, {r0, 0.55, 0.5, 4}}],
 {test, 1, 6, 1}]
```

... **FindMinimum**: The point {7.01156, –0.996157, 1., 1.98066, 0.765377} is at the edge of the search region {0.1, 1.} in coordinate 3 and the computed search direction points outside the region.

... **FindMinimum**: The point {7.04332, –0.996187, 1., 2.6822, 0.751079} is at the edge of the search region {0.1, 1.} in coordinate 3 and the computed search direction points outside the region.

... **FindMinimum**: The point {7.04081, –0.996208, 1., 3.51968, 0.766233} is at the edge of the search region {0.1, 1.} in coordinate 3 and the computed search direction points outside the region.

... **General**: Further output of FindMinimum::reged will be suppressed during this calculation.

... **FindMinimum**: The line search decreased the step size to within the tolerance specified by AccuracyGoal and PrecisionGoal but was unable to find a sufficient decrease in the function. You may need more than MachinePrecision digits of working precision to meet these tolerances.

$$
\begin{pmatrix}
3905.02 & \{onsetplus \to 7.01156,\ c \to -0.996157,\ s \to 1.,\ r0pre \to 1.98066,\ r0 \to 0.765377\} \\
3720.64 & \{onsetplus \to 7.04332,\ c \to -0.996187,\ s \to 1.,\ r0pre \to 2.6822,\ r0 \to 0.751079\} \\
2613.12 & \{onsetplus \to 7.04081,\ c \to -0.996208,\ s \to 1.,\ r0pre \to 3.51968,\ r0 \to 0.766233\} \\
100.349 & \{onsetplus \to 7.0481,\ c \to -1.00749,\ s \to 0.707694,\ r0pre \to 4.3179,\ r0 \to 0.646773\} \\
5250.47 & \{onsetplus \to 6.98213,\ c \to -1.00105,\ s \to 0.68387,\ r0pre \to 4.89226,\ r0 \to 0.5\} \\
39\,281.4 & \{onsetplus \to 6.97927,\ c \to -1.00131,\ s \to 0.670628,\ r0pre \to 5.91791,\ r0 \to 0.5\}
\end{pmatrix}
$$

15.9. NMinimize, but the problem is not so well-behaved

NMinimize would tend to give us a global minimum for well-behaved problems. The cell is locked because of the execution time. The parameter values and the outcome of the RMSE make us wonder whether the problem is well-behaved.

```
ResetOptions[SEYCDT];

SEYCDT[Options, Update, Rest]; (* remove the zero date of onset *)

tEst = 100;

NMinimize[{
  SEYCDT["RMSE", onsetplus, c, s, intModel, R0, r0pre, r0, {t, tEst}],
  {c <= 0}},
 {{onsetplus, 0, 21},
  {c, -4, 0}, {s, .1, 1}, {r0pre, 1, 6}, {r0, .5, 4}}]
```

{81.825, {onsetplus → 8.56545, c → –1.86302, s → 0.710459, r0pre → 4.16219, r0 → 0.585631}}

Using the latter value of R_0 before the intervention, we actually get a slightly lower outcome by using FindMinimum again.

```
FindMinimum[{SEYCDT["RMSE", onsetplus, c, s, intModel, R0, r0pre, r0, {t, tEst}],
  {r0pre == 4.16, c <= 0}},
 {{onsetplus, 8.56, 0, 21},
  {c, -1.86, -4, 0}, {s, .71, .1, 1}, {r0pre, 4.16, 1, 6}, {r0, .586, .5, 4}}]
```

... **FindMinimum**: Failed to converge to the requested accuracy or precision within 500 iterations.

{75.2685, {onsetplus → 8.76736, c → –2.77884, s → 0.722734, r0pre → 4.16, r0 → 0.537232}}

15.10. The lowest value found

In the various runs, the following value was found to be the lowest. Given the other properties we cannot attach much value to this.

```
NMinimize[{SEYCDT["RMSE", onsetplus, c, s, intModel, R0, r0pre, r0, {t, 120}],
  {c <= 0}},
 {{onsetplus, 6.06359, 21},
  {c, -.5141, 0}, {s, .665523, 1}, {r0pre, 4.55, 6}, {r0, .5, 4}}]
```

{74.9765, {onsetplus → 5.52779, c → −0.211321, s → 0.658037, r0pre → 4.63632, r0 → 0.492405}}

15.11. Contours

Let us plot some contours of combinations of the earlier onset and values of R_0 before the intervention. The accuracy and numbers of iterations can be reduced because we only wish to have a rough indication.

```
contour[rb_, op_] := contour[rb, op] =
  Results[contour, rb, op] =
   FindMinimum[{SEYCDT["RMSE", onsetplus, c, s, intModel, R0, r0pre, r0, {t, tEst}],
     {r0pre == rb, onsetplus == op, c <= 0}},
    {{onsetplus, 5, 0, 21},
     {c, -3, -4, 0}, {s, .8, .1, 1},
     {r0pre, 4, 1, 6}, {r0, 0.55, .5, 4}}, AccuracyGoal → 4,
   MaxIterations → 100]

pnt = 4.;
listab[pnt] = {0, 4, 6,  9, 10, 11};

tab[pnt] = contour[pnt, #] & /@ listab[pnt]
```

⋯ FindMinimum: Failed to converge to the requested accuracy or precision within 100 iterations.

⋯ FindMinimum: Failed to converge to the requested accuracy or precision within 100 iterations.

⋯ FindMinimum: Failed to converge to the requested accuracy or precision within 100 iterations.

⋯ General: Further output of FindMinimum::cvmit will be suppressed during this calculation.

$$
\begin{pmatrix}
1084.92 & \{\text{onsetplus} \to 0., c \to -0.0000194853, s \to 1., \text{r0pre} \to 4., \text{r0} \to 1.45255\} \\
499.929 & \{\text{onsetplus} \to 4., c \to -0.00116596, s \to 0.999976, \text{r0pre} \to 4., \text{r0} \to 1.15234\} \\
294.125 & \{\text{onsetplus} \to 6., c \to -2.83313, s \to 1., \text{r0pre} \to 4., \text{r0} \to 0.987921\} \\
114.08 & \{\text{onsetplus} \to 9., c \to -3.99999, s \to 0.813062, \text{r0pre} \to 4., \text{r0} \to 0.719199\} \\
75.4312 & \{\text{onsetplus} \to 10., c \to -3.99984, s \to 0.756323, \text{r0pre} \to 4., \text{r0} \to 0.563462\} \\
203.372 & \{\text{onsetplus} \to 11., c \to -2.726 \times 10^{-7}, s \to 0.593483, \text{r0pre} \to 4., \text{r0} \to 0.5\}
\end{pmatrix}
$$

```
pnt = 4.37;
listab[pnt] = {0, 4,  7,  9};

tab[pnt] = contour[pnt, #] & /@ listab[pnt]
```

⋯ FindMinimum: Failed to converge to the requested accuracy or precision within 100 iterations.

$$
\begin{pmatrix}
745.806 & \{\text{onsetplus} \to 0., c \to -0.0000335252, s \to 0.999999, \text{r0pre} \to 4.37, \text{r0} \to 1.27888\} \\
232.16 & \{\text{onsetplus} \to 4., c \to -4., s \to 0.981113, \text{r0pre} \to 4.37, \text{r0} \to 0.918376\} \\
75.5744 & \{\text{onsetplus} \to 7., c \to -2.85448, s \to 0.733904, \text{r0pre} \to 4.37, \text{r0} \to 0.55227\} \\
690.91 & \{\text{onsetplus} \to 9., c \to -2.1302 \times 10^{-7}, s \to 0.489079, \text{r0pre} \to 4.37, \text{r0} \to 0.5\}
\end{pmatrix}
$$

```
pnt = 4.55;
listab[pnt] = {0, 4, 5, 6, 7};

tab[pnt] = contour[pnt, #] & /@ listab[pnt]
```

⋯ FindMinimum: Failed to converge to the requested accuracy or precision within 100 iterations.

⋯ FindMinimum: Failed to converge to the requested accuracy or precision within 100 iterations.

$$
\begin{pmatrix}
570.787 & \{\text{onsetplus} \to 0., c \to -0.0000481244, s \to 0.999999, \text{r0pre} \to 4.55, \text{r0} \to 1.18818\} \\
153.719 & \{\text{onsetplus} \to 4., c \to -3.99959, s \to 0.881768, \text{r0pre} \to 4.55, \text{r0} \to 0.808408\} \\
93.0472 & \{\text{onsetplus} \to 5., c \to -3.99999, s \to 0.807496, \text{r0pre} \to 4.55, \text{r0} \to 0.672654\} \\
75.1005 & \{\text{onsetplus} \to 6., c \to -0.884901, s \to 0.677291, \text{r0pre} \to 4.55, \text{r0} \to 0.510108\} \\
452.897 & \{\text{onsetplus} \to 7., c \to -2.36983 \times 10^{-7}, s \to 0.547174, \text{r0pre} \to 4.55, \text{r0} \to 0.5\}
\end{pmatrix}
$$

```
pnt = 5.7;
listab[pnt] = {0, 1, 2};

tab[pnt] = contour[pnt, #] & /@ listab[pnt]
```

$$
\begin{pmatrix}
79.3334 & \{\text{onsetplus} \to 0., c \to -0.0000392834, s \to 0.672531, \text{r0pre} \to 5.7, \text{r0} \to 0.504915\} \\
620.949 & \{\text{onsetplus} \to 1., c \to -2.36358 \times 10^{-7}, s \to 0.530617, \text{r0pre} \to 5.7, \text{r0} \to 0.5\} \\
1093.48 & \{\text{onsetplus} \to 2., c \to -2.06263 \times 10^{-7}, s \to 0.439888, \text{r0pre} \to 5.7, \text{r0} \to 0.5\}
\end{pmatrix}
$$

```
ListLinePlot[{Legended[Transpose[{ listab[4.], First /@ tab[4.]}], "R₀-pre = 4"],
  Legended[Transpose[{ listab[4.37], First /@ tab[4.37]}], "R₀-pre = 4.37"],
  Legended[Transpose[{ listab[4.55] , First /@ tab[4.55]}], "R₀-pre = 4.55"],
  Legended[Transpose[{ listab[5.7] , First /@ tab[5.7]}], "R₀-pre = 5.7"]
}, AxesLabel → {"Earlier onset\nbefore Feb 23", "RMSE"}]
```
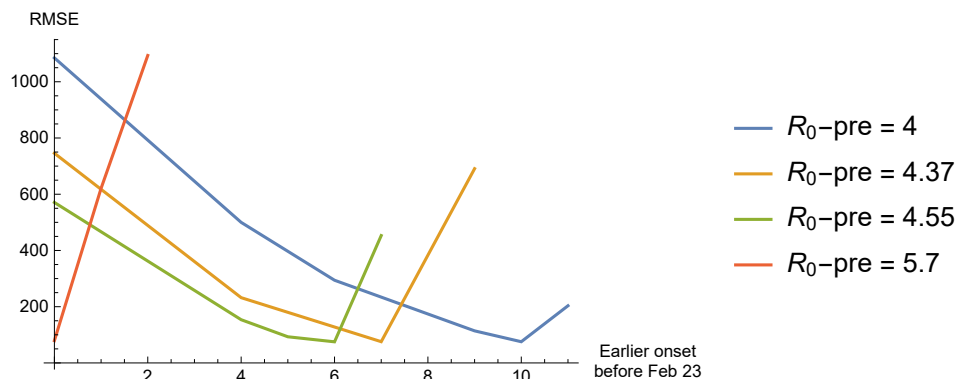

16. Intervention and estimation using SEYCDB

16.1. Introduction

This chapter has the same structure as Chapter 15 on SEYCDT. The only difference is the practical problem of numerical estimation with an ill-conditioned problem. After considering both approaches, my impression for the current setup is that using β actually is a bit better than using R_0, but this impression may be caused by that I originally started with β.

In retrospect, we should better take account of the fact that the data are *discrete* and *defined per day*. Thus we could make a grid of days for the "actual onset before the chosen onset" and the "constant in the back-trace", and estimate the continuous parameters for the separate values in the grid. This ought to reduce the problems that we now have with the conditioning. Then, perhaps

with more data and a better specification for the back-tracing, we might used FindFit to even consider confidence intervals. But we are oriented at didactics now and not at the best estimate (on dubious data).

16.2. Find the intervention parameters for SEYCDB

The recorded cumulated number of deaths gives us important information. The curve is bending towards some limiting value. This provides information about the implied R_0. Let us assume that the death toll for the remainder of the year moves gradually to 7000. Assume that 7000 - x deaths would be attributed to the new regime, assuming a value of $R_0 < 1$, and x are still accountable to the old regime. For the new limit value, the "starting values" at day 24 are important. We have the ProductLog rule for the limiting value. Let us substitute an unknown R_0 and the values of the variables at day 24 of the intervention, and assume that $\gamma = 1/10$ remains the same. (This requires that the Base model in section 13.1 has run.)

`startvalues = SEYCDB[Intervention, 24, f0 * .1, .1]`

{Seed → 4777.44, Yqt0 → 14 660.4, Cqt0 → 6355.58, Dqt0 → 96.7855,

Aqt0 → 6452.37, BetaSIA() → 0.1 f0, GammaSIA() → 0.1, Onset → 24}

The death toll that has already occurred or is in the pipeline is:

`Dqt0 + PhiSICD[]` (`Seed + Yqt0`) `/. startvalues /. Options[SIA]`

388.353

In the SICD[Limit] list, the Dp are in the 4th position. The limiting value applies to the excess deaths, given the startvalues. Let us include a number x that can also be attributed to the Base case, above to the latter pipeline (which changes above notion of x).

`assign[x_] =`
(`7000 – TheFormer – x`) `/ NqSeed == SICD[Limit, Sequence @@ startvalues][[4]]`

$5.74713 \times 10^{-8} (6611.65 - x) =$

$$\frac{1. \, W\!\left(-0.999355 \, e^{-0.999629 \, \text{f0}} \, \text{f0}\right)}{\text{f0}} + \frac{0.985 \log\!\left(-\dfrac{1.00065 \, W\!\left(-0.999355 \, e^{-0.999629 \, \text{f0}} \, \text{f0}\right)}{\text{f0}}\right)}{\text{f0}} + 0.999635$$

For $x = 0$, the R_0 of the new regime would be slightly above 1.

`FindRoot[assign[0], {f0, 1}]`

{f0 → 1.00207}

It appears that the idea of assigning a death toll to a regime is somewhat simplistic. A clearcut number is the recorded death toll and the IFF applied to the already infected. What happens after the intervention is not only influenced by the original situation but also by the intervention. We may make an arbitrary choice along the following contour. In the end, it are the practical measures that determine the β and γ of the new regime. For our exercise, a solution might be to say that $x / 2$ is caused by the old regime and $x / 2$ by the new regime. Or it might be fairer to say that most deaths still apply to the new regime because it will exist for a much longer period. At this stage, all theory is arbitrary, and it suffices to take $R_0 = .99$.

```
r0fromx[x_] := f0 /. FindRoot[assign[x], {f0, 1}]
```

```
newR0 = r0fromx[6612 / 2]
```

newR0 = 0.98423

```
Plot[r0fromx[x], {x, 0, 7000},
 AxesLabel → {"x deaths\ncaused by\nold regime", "New R0"}, AxesOrigin → {0, 0}]
```


A consideration may be that $1/\gamma$ would be shorter because of self-quarantine. However, a higher value of γ also means a faster development of the acquitted, and thus also a surge of the number of deaths in the period just after the intervention. This is illogical. Gamma is best understood as a normalisation constant, and we keep it the same. The whole burden of change falls onto β.

```
SetOptions[Intervention, Time[Intervention] → 24, R0[Intervention] → .99,
 BetaSIA[Intervention] → .99 * .1, GammaSIA[Intervention] → .1]
```

{Time(Intervention) → 24, R0(Intervention) → 0.99,

    BetaSIA(Intervention) → 0.099, GammaSIA(Intervention) → 0.1}

16.3. Intervention at t = 24, using SEYCDB

For comparison to the base scenario, the latter plot ("pbase") must have been produced in the above.

```
PlotLegendsSIA[SEYCD]
(* intervention uses the default plotting colours and legends *)
```

```
NqSeed = 17.4 × 10^6;
ResetOptions[SIA] (* not uninmportant *)
```

{Aqt0 → 0, Cqt0 → 0, DataMold → {S, I, A}, Dqt0 → 0, Iqt0 → Null, Nqt0 → 1.74×10^7, Onset → 0, Seed → 100,

    Yqt0 → 0, AlphaSEYCD() → 1., BetaSIA() → 0.4, GammaSIA() → 0.1, PhiSICD() → 0.015, R0() → Null}

The equations for SEYCDB use BetaSIA[]. The pre-intervention β is the default provided in the options. The intervention parameters are entered in Options[Intervention] now.

```
Options[Intervention]  (* set above *)
```

{Time(Intervention) → 24, R0(Intervention) → 0.99,

    BetaSIA(Intervention) → 0.099, GammaSIA(Intervention) → 0.1}

See above how this routine could be done in steps.

```
SEYCDB[Run, Pr, intModel, t, 250];
```

```
intModel[t] (* notice the WhenEvent statement *)
```

$\text{Hold}\big[\{Sq'(t) = -5.74713 \times 10^{-8} \, \text{BetaSIA}(t) \, Sq(t) \, Yq(t), \, Eq'(t) = -1. \, Eq(t) - Sq'(t),$

$Yq'(t) = 1. \, Eq(t) - Aq'(t), \, Aq'(t) = \text{GammaSIA}(t) \, Yq(t), \, Dq(t) = 0.015 \, Aq(t), \, Cq(t) = Aq(t) - Dq(t),$

$Nq(t) = 1.74 \times 10^7 - Dq(t), \, Iq(t) = Eq(t) + Yq(t), \, Nq(0) = 1.74 \times 10^7, \, Sq(0) = 1.73999 \times 10^7,$

$Eq(0) = 100, \, Yq(0) = 0, \, Cq(0) = 0, \, Dq(0) = 0, \, Aq(0) = 0, \, Iq(0) = 100, \, \text{BetaSIA}(0) = 0.4,$

$\text{GammaSIA}(0) = 0.1, \, \text{WhenEvent}[t > 24, \{\text{BetaSIA}(t) \to 0.099, \, \text{GammaSIA}(t) \to 0.1\}]\}\big]$

The values are now available as SEYCD[t] (and not as SEYCDB[t], that only includes the parameters).

The supposed top of "recorded deaths per day" was on day 42, with a cumulated total of 1766 deaths. The run of the model of the intervention, see below, mentions 565 deaths. However, we must keep in mind that these death counts have different meanings. See the estimation below.

```
Explain[SEYCD[42], SEYCD] (* day 24: supposed top of "reported deaths per day" *)
```

$\{S \to 1.73432 \times 10^7, \, E \to 1713.13, \, Y \to 17\,338.6, \, C \to 37\,143.6,$

$D \to 565.638, \, N \to 1.73994 \times 10^7, \, I \to 19\,051.7, \, A \to 37\,709.2\}$

The most useful plot now might be the one on ICU beds.

```
seicdplotoptionsICU = SEYCD[Plot, Options, ICU, PlotRange → {0, 10000}];
Plot @@ {SEYCD[ICU, 1500, t], {t, 0, 150}, seicdplotoptionsICU}
```


The values at the end of the 250 days are (some 5000 deaths and not 7000 yet):

```
Explain[SEYCD[250], SEYCD]
```

$\{S \to 1.70473 \times 10^7, \, E \to 1107.8, \, Y \to 11\,389.4, \, C \to 335\,057.,$

$D \to 5102.39, \, N \to 1.73949 \times 10^7, \, I \to 12\,497.2, \, A \to 340\,159.\}$

```
pintv = SEYCD[Plot, 0, 250]
```


Combining the two plots of the base case and the intervention.

```
Show[pbase, pintv]
```


The following is around the day of the intervention, day 24, using a plotrange of 0.2% (no axes ticks). This is the same outcome as the separate running of the Base and Alternative scenario's, except that the WhenEvent model allows drawn lines from day 0 too.

```
Show[pbase, pintv, PlotRange → {{20, 30}, {0, .002}}]
```


16.4. Estimation: aim and result

For estimation we are interested in the new value of β. This value however depends upon the old value, while the old value depends upon our choice of the day of onset.

The summary of our findings on estimation is as follows. The following graph gives the Root Mean Squared Error (RMSE) as a function of both the β before the intervention (β-pre) and the possibility that the onset of infections happened some days earlier than Carnival, February 23 2020. It appears that our choice of β = 0.4 agrees with 9-10 days earlier, Valentine's Day. With a higher β then there is less need for an earlier onset. We can maintain the onset at Carnival, if β-pre would be 0.57, and then have the same RMSE level as in other cases. In all these cases we maintained γ = 0.1.
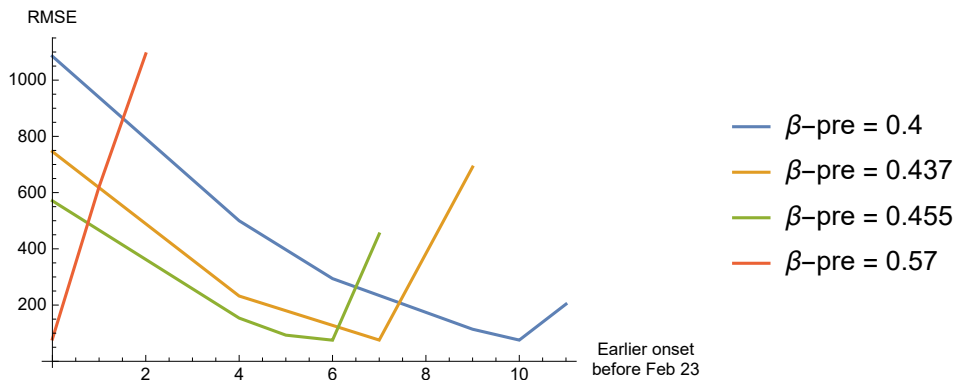
In all such minimum combinations, we get a graph like below, on the fit of the model estimate of the Dq[t] and the recorded death count Drec[τ]. The following gives the combination of β-pre = 0.4 and the required earlier onset of about 9-10 days.

Constant –> 6.2885, Slope –> 0.754, Plus –> 9.847
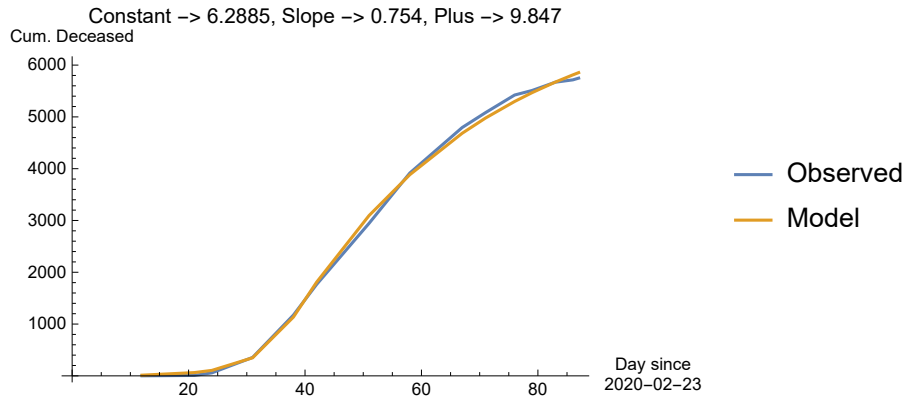
In all such minimum combinations, the β after the intervention drops to about 0.05 (imposed lower boundary), and the value of deaths after 250 days is a bit above 7,000.

16.5. Major routine: the definition of the error

Because of estimation we may restrict the time horizon to the available data.

```
tEst = 100.;
```

```
NqSeed = 17.4 × 10^6;
ResetOptions[SIA] (* not uninmportant *)
```

$\{$Aqt0 → 0, Cqt0 → 0, DataMold → {S, I, A}, Dqt0 → 0, Iqt0 → Null, Nqt0 → 1.74×10^7, Onset → 0, Seed → 100,

   Yqt0 → 0, AlphaSEYCD() → 1., BetaSIA() → 0.4, GammaSIA() → 0.1, PhiSICD() → 0.015, R0() → Null$\}$

```
ResetOptions[SEYCDB]; (* difference with SEYCDT is only the datamold on BetaSIA[] *)
```

```
SEYCDB[Options, Update, Rest]; (* remove the zero date of onset *)
```

```
SetOptions[Intervention, Time[Intervention] → 24, R0[Intervention] → .99,
 BetaSIA[Intervention] → .99 * .1, GammaSIA[Intervention] → .1]
```

{Time(Intervention) → 24, R0(Intervention) → 0.99,

   BetaSIA(Intervention) → 0.099, GammaSIA(Intervention) → 0.1}

The following is a major routine. The major step in estimation is the definition of the error. Thereafter it is only a matter of minimising the error. The following runs the model with the stated intervention parameters, at the day of intervention in Options[Intervention]. It also give the Root Mean Squared Error (RMSE) using the data in the Options[SEYCDB] (with the options Day and Count for the death counts).

The following sets $\beta = 0.099$ and keeps $\gamma = .1$.

```
SEYCDB["RMSE", intModel, .099, .1, {t, tEst}]
```

2869.6

The plots shows that the model fits the last weeks of March but it doesn't explain the observed deaths in April and May.

PM. The PlotLabel informs us of the values of the Constant and Slope in the Options[SEYCDB] about the relation between the recorded day of death τ and the end of the infectious period t. The value of Plus gives the number of days that should be added to the horizontal axis if we would take onset so

many days earlier.

SEYCDB[PlotLine]

Constant –> 0, Slope –> 1, Plus –> 0
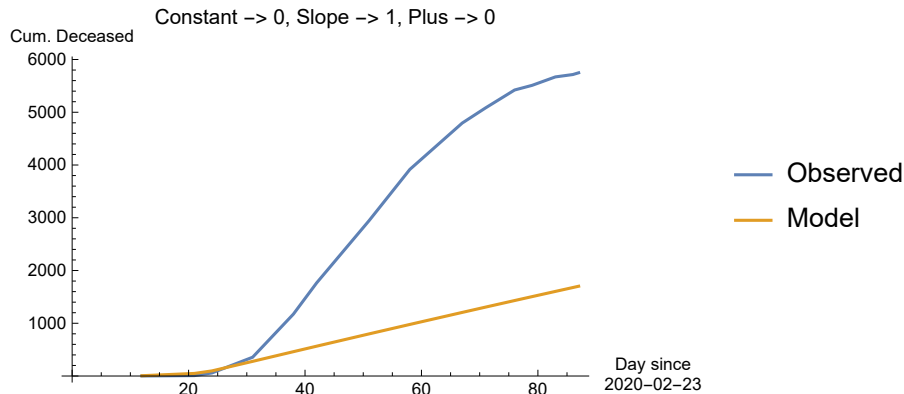
16.6. Minimising the RMSE

Given this high RMSE, we may try to minimise it. We might consider that the death counts from the early period are determined by the original β and γ, and then drop these data, to estimate the new parameters. However, the SEYCDB model uses the *WhenEvent* feature and thus accurately manages the regime change. The problem does not reside with the Dq[t] that the model calculates. The true issue is that the observed death counts concern *actual days of death*, while the model requires the *end of the infectious period*.

Thus, we must back-trace the death counts to an earlier moment in time. It is an option to extend the model, say with $U' = u\,D,$ with the hospital phase (u taken from ICU), like the Goh (2020) "epidemic calculator". At this point it seems more logical to work directly with the data, if only to better grasp what the implications are.

The best approach would seem to be to allow the estimation routine to determine both model parameters and the time shift between Dq and the actually recorded deaths. It takes some care to find the proper interpretation and minimum. We may consider to actually also estimate β and γ for the period before the intervention. Also, the assumption of the day of onset is only an assumption.

For example, consider a fixed back-tracing of 7 days and a proportional reduction of 80%. The first recorded death on day 12 then actually ended its infectious period at day -7 + 0.8 * 12 ≈ 3. We drop the first observation, which is the day of onset (with no death count). In above chart, the model looked like a good fit for the first weeks of March, but taking account of back-tracing shows that it would actually be a bad fit for the whole period.

```
ResetOptions[SEYCDB];

SEYCDB[Options, Update, Rest]; (* remove the zero date of onset *)

SetOptions[SEYCDB, Constant → –7, Slope → .8];

SEYCDB["RMSE", intModel, .099, .1, {t, tEst}]
(* uses both model output and the data values of the options *)
```

3310.02

SEYCDB[PlotLine]
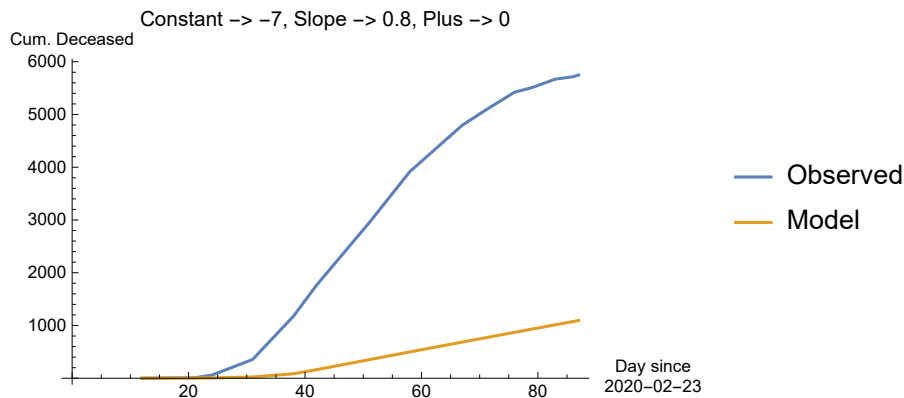
Constant –> –7, Slope –> 0.8, Plus –> 0

Taking advantage of some estimations not shown, it appears that the estimation routine has a tendency to take a *positive* constant in the back-tracing, instead of a negative one, as we actually required. However, our assumption of the onset may be wrong. Above graph shows that the model is still in the exponential phase and still far away from the slowing down towards peaking. Allowing a positive shift *in the model* actually means that the onset must have been earlier in the year.

Let us use a positive constant as a proxy for an earlier date of onset. (Setting the value of Plus has no effect, since this option is controlled by the estimation routine, and we do not do an estimation on onset yet.)

SetOptions[SEYCDB, Constant → 14];

SEYCDB["RMSE", intModel, .099, .1, {t, tEst}]

2883.9

This has a lower RMSE but the growth of infections in the model remains low due to the fixed intervention date. When the onset was earlier in the year, then also the day number of intervention shifts.

SEYCDB[PlotLine]

Constant –> 14, Slope –> 0.8, Plus –> 0
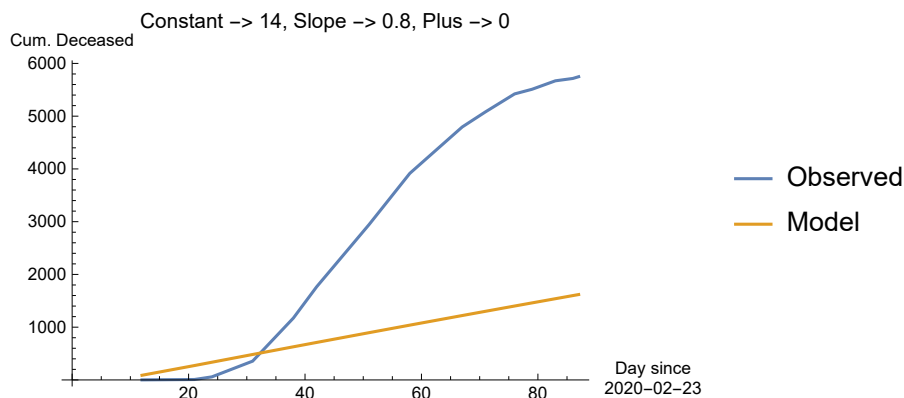
16.7. Combined estimate of new onset, pre-intervention beta, intervention beta, and back-tracing

This subsection forms the core of the chapter. The following routine combines above separate

statements. When we have a RMSE then we can apply FindMinimum and NMinimize to it.

16.7.1. Calculating RMSE given numerical parameter values

The following only calculates a RMSE given numerical parameter values. Let us assume that onset happened 14 days earlier. Let the pre-intervention β be 0.6, thus $F_0 = 6$.

```
SEYCDB["RMSE", 14, -7, 0.8, intModel, BetaSIA, 0.6, 0.099, {t, tEst}]
```

157 155.

This overshoots, which gives confidence that an estimation should be able to find values in-between.

Check that the model indeed has BetaSIA[0] == 0.6 and that the WhenEvent is at day 24 + 14 = 38.

```
intModel[t]
```

$\text{Hold}\big[\{Sq'(t) = -5.74713 \times 10^{-8}\, BetaSIA(t)\, Sq(t)\, Yq(t),\ Eq'(t) = -1.\, Eq(t) - Sq'(t),$
$\quad Yq'(t) = 1.\, Eq(t) - Aq'(t),\ Aq'(t) = GammaSIA(t)\, Yq(t),\ Dq(t) = 0.015\, Aq(t),\ Cq(t) = Aq(t) - Dq(t),$
$\quad Nq(t) = 1.74 \times 10^7 - Dq(t),\ Iq(t) = Eq(t) + Yq(t),\ Nq(0) = 1.74 \times 10^7,\ Sq(0) = 1.73999 \times 10^7,$
$\quad Eq(0) = 100,\ Yq(0) = 0,\ Cq(0) = 0,\ Dq(0) = 0,\ Aq(0) = 0,\ Iq(0) = 100,\ BetaSIA(0) = 0.6,$
$\quad GammaSIA(0) = 0.1,\ WhenEvent[t > 38,\ \{BetaSIA(t) \to 0.099,\ GammaSIA(t) \to 0.1\}]\}\big]$

16.7.2. FindMinimum on the RMSE

The following is somewhat involved: (i) we shift the onset and intervention, and (ii) we estimate β before and after the intervention, and (iii) we relate the death counts to the infectious periods. Apparently it works nicely, but we lack information about confidence intervals. PM. Estimation time can be reduced by taking tmax = 100, which is the interval for which there are data. But it can be useful to project to t = 250.

Unfortunately, FindMinimum is sensitive to starting values. The following takes advantage of earlier runs.

```
FindMinimum[SEYCDB["RMSE", onsetplus, c, s, intModel, BetaSIA, bpre, b, {t, 250}],
 {{onsetplus, 7, 0, 21},
  {c, -3, -4, 0}, {s, .8, .1, 1}, {bpre, .4, .1, .6}, {b, .1, .05, .4}}]
```

⋯ FindMinimum: The line search decreased the step size to within the tolerance specified by AccuracyGoal and PrecisionGoal but was unable to find a sufficient decrease in the function. You may need more than MachinePrecision digits of working precision to meet these tolerances.

$\{75.6131,\ \{onsetplus \to 6.99788,\ c \to -3.00216,\ s \to 0.737404,\ bpre \to 0.436991,\ b \to 0.0551216\}\}$

Observe that the model now uses the latter parameters.

```
intModel[t]
```

$\text{Hold}\big[\{Sq'(t) = -5.74713 \times 10^{-8}\, BetaSIA(t)\, Sq(t)\, Yq(t),\ Eq'(t) = -1.\, Eq(t) - Sq'(t),$
$\quad Yq'(t) = 1.\, Eq(t) - Aq'(t),\ Aq'(t) = GammaSIA(t)\, Yq(t),\ Dq(t) = 0.015\, Aq(t),\ Cq(t) = Aq(t) - Dq(t),$
$\quad Nq(t) = 1.74 \times 10^7 - Dq(t),\ Iq(t) = Eq(t) + Yq(t),\ Nq(0) = 1.74 \times 10^7,\ Sq(0) = 1.73999 \times 10^7,$
$\quad Eq(0) = 100,\ Yq(0) = 0,\ Cq(0) = 0,\ Dq(0) = 0,\ Aq(0) = 0,\ Iq(0) = 100,\ BetaSIA(0) = 0.436991,$
$\quad GammaSIA(0) = 0.1,\ WhenEvent[t > 30.9979,\ \{BetaSIA(t) \to 0.0551216,\ GammaSIA(t) \to 0.1\}]\}\big]$

The death toll will rise to some 7000 over the course of the year.

Explain[SEYCD[250], SEYCD]

$\{S \rightarrow 1.69283 \times 10^7, E \rightarrow 0.692471, Y \rightarrow 12.3457,$

$\quad C \rightarrow 464\,615., D \rightarrow 7075.36, N \rightarrow 1.73929 \times 10^7, I \rightarrow 13.0381, A \rightarrow 471\,691.\}$

Reading below chart requires the following explanation.

(a) The time axis below specifies that February 23 is day 0. This applies to the recorded death counts.

(b) The above estimate is that the day of onset was actually a week earlier, which is the day 0 in the model. This has been specified by the option Plus.

(c) The back-tracing of the death counts to the end of the infectious period has used the formula $t =$ constant + slope τ, with the stated values, with t the value in the model (using the plus value) and τ the recorded day since 2020-02-23 (thus without the plus value). Thus the Constant in below plot is Plus + the c value in the estimate.

(c) The back-tracing of the death counts to the end of the infectious period has used the formula $t =$ constant + slope τ, with the stated values, with t the value in the model (using the plus value) and τ the recorded day since 2020-02-23 (thus without the plus value). Thus the Constant in below plot (3.99) is the value of Plus (6.99) + the value of the constant in the estimate (-3.00).

SEYCDB[PlotLine]
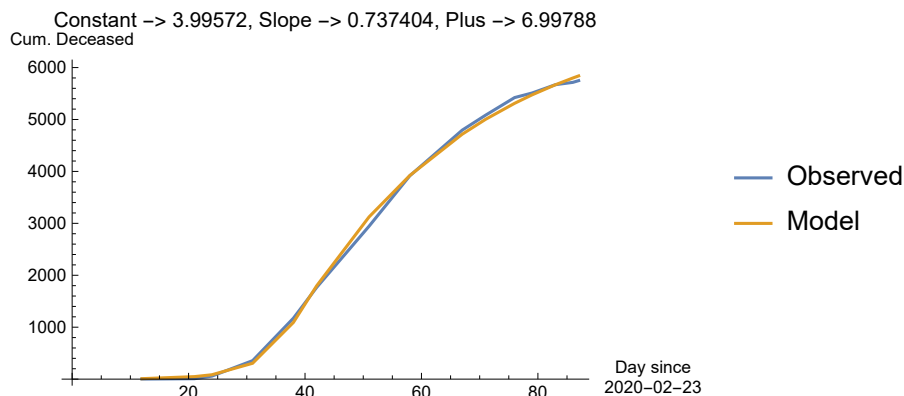
16.7.3. Indication of lack of sensitivity to small changes

Above estimate of pre-intervention $\beta = 0.44$ gives us some worry about our choice of $\beta = 0.4$. In another run, not shown, a value was fixed with $\beta = 0.4$. For these values, the RMSE hardly changes and the plot and limit values remain quite the same.

SEYCDB["RMSE", 9.847, -3.5585, 0.754, intModel, BetaSIA, 0.4, 0.0595, {t, 250}]

78.9544

`SEYCDB[PlotLine]`
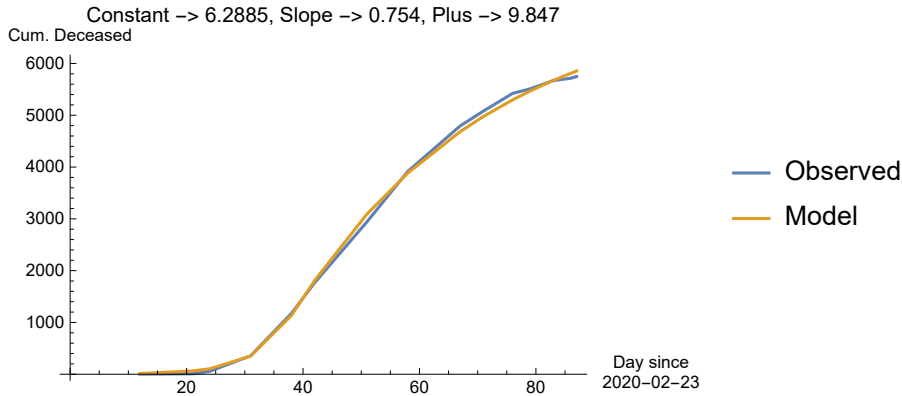
Constant –> 6.2885, Slope –> 0.754, Plus –> 9.847

The implications towards later in the year:

`Explain[SEYCD[250], SEYCD]`

$\{S \to 1.69144 \times 10^7, E \to 1.92206, Y \to 31.9095, C \to 478\,297.,$
$\quad D \to 7283.71, N \to 1.73927 \times 10^7, I \to 33.8316, A \to 485\,581.\}$

The conclusion seems warranted that onset in Holland was some 7-10 days before Carnival February 23, our earlier choice of the onset. Ah, yes, Valentine's day, February 14. Potentially people returned to Holland to celebrate Valentine's day.

16.7.4. FindMinimum with a condition

Let us fix onsetplus = 0, so that the onset was really at Carnival, and we want to know what β would have been needed. For this, we use the possibility of a condition.

```
FindMinimum[{SEYCDB["RMSE", onsetplus, c, s, intModel, BetaSIA, bpre, b, {t, tEst}],
  {onsetplus == 0} },
 {{onsetplus, 7, 0, 21},
  {c, -3, -4, 0}, {s, .8, .1, 1}, {bpre, .4, .1, 1}, {b, .1, .05, .4}}]
```

⋯ FindMinimum: Failed to converge to the requested accuracy or precision within 500 iterations.

$\{79.1203, \{\text{onsetplus} \to 0., c \to -0.00015925, s \to 0.675232, \text{bpre} \to 0.568898, b \to 0.0517535\}\}$

`SEYCDB[PlotLine]`

Constant –> –0.00015925, Slope –> 0.675232, Plus –> 0.
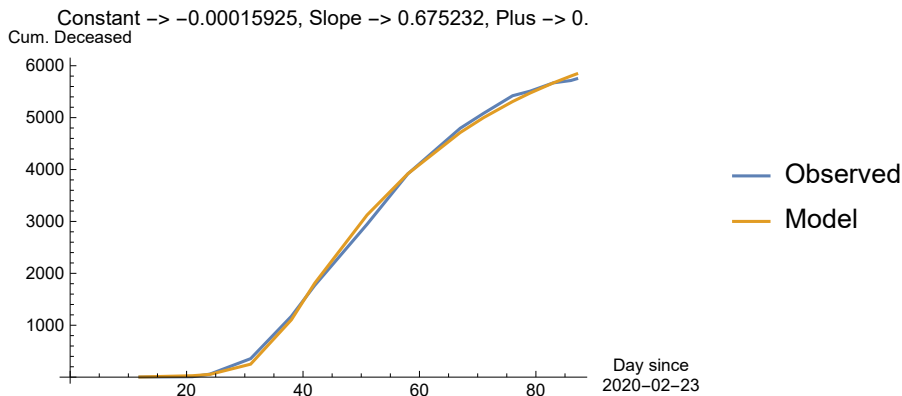
The implications towards later in the year:

```
Explain[SEYCD[ tEst], SEYCD]
```

$\{S \to 1.69319 \times 10^7, E \to 302.238, Y \to 5718.36,$
$\quad C \to 455\,197., D \to 6931.93, N \to 1.73931 \times 10^7, I \to 6020.6, A \to 462\,129.\}$

16.8. Variant of startvalues from 0.1 to 0.6

Mathematica's FindMinimum has a known sensitivity to starting values. We might also use NMinimize, that is more robust in this respect, but in this case it is helpful to see the various outcomes in fairly quick manner.

https://reference.wolfram.com/language/tutorial/ConstrainedOptimizationOverview.html

Check the sensitivity of FindMinimum to starting values, and that our choice of β around 0.40 based upon the literature on $R_0 \approx 4$ and $1/\gamma \approx 10$ was not unwise. A consideration is that the linear scheme for back-tracing might influence the outcome too much, but this can only be tested when there would be relevant other schemes.

```
Table[
 FindMinimum[SEYCDB["RMSE", onsetplus, c, s, intModel, BetaSIA, bpre, b, {t, tEst}],
  {{onsetplus, 7, 0, 21},
   {c, -3, -4, 0}, {s, .8, .1, 1}, {bpre, test, .1, .8}, {b, .1, .05, .4}}],
 {test, .1, .6, .1}]
```

... FindMinimum: The point {6.99857, 0., 0.722138, 0.34779, 0.170131} is at the edge of the search region {–4., 0.} in coordinate 2 and the computed search direction points outside the region.

... FindMinimum: The point {7.15878, –4., 0.913819, 0.403589, 0.0876309} is at the edge of the search region {–4., 0.} in coordinate 2 and the computed search direction points outside the region.

... FindMinimum: The point {6.99538, –3.00005, 0.77177, 0.434057, 0.05} is at the edge of the search region {0.05, 0.4} in coordinate 5 and the computed search direction points outside the region.

... General: Further output of FindMinimum::reged will be suppressed during this calculation.

... FindMinimum: The line search decreased the step size to within the tolerance specified by AccuracyGoal and PrecisionGoal but was unable to find a sufficient decrease in the function. You may need more than MachinePrecision digits of working precision to meet these tolerances.

... FindMinimum: The line search decreased the step size to within the tolerance specified by AccuracyGoal and PrecisionGoal but was unable to find a sufficient decrease in the function. You may need more than MachinePrecision digits of working precision to meet these tolerances.

... FindMinimum: The line search decreased the step size to within the tolerance specified by AccuracyGoal and PrecisionGoal but was unable to find a sufficient decrease in the function. You may need more than MachinePrecision digits of working precision to meet these tolerances.

... General: Further output of FindMinimum::lstol will be suppressed during this calculation.

$\left(\begin{array}{ll} 1278.92 & \{onsetplus \to 6.99857, c \to 0., s \to 0.722138, bpre \to 0.34779, b \to 0.170131\} \\ 204.389 & \{onsetplus \to 7.15878, c \to -4., s \to 0.913819, bpre \to 0.403589, b \to 0.0876309\} \\ 280.552 & \{onsetplus \to 6.99538, c \to -3.00005, s \to 0.77177, bpre \to 0.434057, b \to 0.05\} \\ 75.6135 & \{onsetplus \to 6.99697, c \to -3.00226, s \to 0.737417, bpre \to 0.437001, b \to 0.0551245\} \\ 75.5908 & \{onsetplus \to 7.06619, c \to -3.0101, s \to 0.737263, bpre \to 0.436116, b \to 0.0550712\} \\ 1159.75 & \{onsetplus \to 6.96573, c \to -3.00235, s \to 0.558729, bpre \to 0.424127, b \to 0.146957\} \end{array}\right)$

16.9. NMinimize, but the problem is not so well-behaved

NMinimize would tend to give us a global minimum for well-behaved problems. The cell is locked

because of the execution time. The parameter values and the outcome of the RMSE make us won-der whether the problem is well-behaved.

```
ResetOptions[SEYCDB];

SEYCDB[Options, Update, Rest]; (* remove the zero date of onset *)

tEst = 100;

NMinimize[SEYCDB["RMSE", onsetplus, c, s, intModel, BetaSIA, bpre, b, {t, tEst}],
  {{onsetplus, 0, 21},
   {c, -4, 0}, {s, .1, 1}, {bpre, .1, .6}, {b, .05, .4}}]
```

$\{75.3063, \{\text{onsetplus} \rightarrow 5.89458, c \rightarrow -1.42644, s \rightarrow 0.693891, \text{bpre} \rightarrow 0.455377, b \rightarrow 0.0521679\}\}$

Using the latter value of β before the intervention, we actually get a slightly lower outcome by using FindMinimum again.

```
FindMinimum[{SEYCDB["RMSE", onsetplus, c, s, intModel, BetaSIA, bpre, b, {t, tEst}],
   {bpre == 0.455} },
  {{onsetplus, 7, 0, 21},
   {c, -3, -4, 0}, {s, .8, .1, 1}, {bpre, .4, .1, .6}, {b, .1, .05, .4}}]
```

$\{75.0092, \{\text{onsetplus} \rightarrow 6.06359, c \rightarrow -0.5141, s \rightarrow 0.665523, \text{bpre} \rightarrow 0.455, b \rightarrow 0.0500001\}\}$

16.10. The lowest value found

In the various runs, the overall lowest value was found by using SEYCDT. Given the other properties we cannot attach much value to this.

When we use the SEYCDT point but only using some first six digits, then the error is still a bit larger than said minimum.

```
SEYCDB["RMSE", 5.52779, -0.21132, .658057,
  intModel, BetaSIA, .46363, .04944, {t, tEst}]
```

75.2173

However, NMinimize on SEYCDB using the latter starting point does not find the minimum as SEYCDT does (though the starting points are not the same since we now took SEYCDT's somewhat rounded outcome).

```
NMinimize[{SEYCDB["RMSE", onsetplus, c, s, intModel, BetaSIA, bpre, b, {t, tEst}],
   {c <= 0}},
 {{onsetplus, 5.52779, 21},
  {c, -0.21132, 0}, {s, .658057, 1}, {bpre, .46363, .6}, {b, .04944, .4}}]
(*locked*)
```

- ... **InterpolatingFunction**: Input value {102.829} lies outside the range of data in the interpolating function. Extrapolation will be used.

- ... **InterpolatingFunction**: Input value {105.803} lies outside the range of data in the interpolating function. Extrapolation will be used.

- ... **InterpolatingFunction**: Input value {106.794} lies outside the range of data in the interpolating function. Extrapolation will be used.

- ... **General**: Further output of InterpolatingFunction::dmval will be suppressed during this calculation.

{75.045, {onsetplus → 5.12303, c → −0.221474, s → 0.660199, bpre → 0.469651, b → 0.0495565}}

16.11. Contours

Let us plot some contours of combinations of the earlier onset and values of β before the intervention. The accuracy and numbers of iterations can be reduced because we only wish to have a rough indication.

```
contour[bp_, op_] := contour[bp, op] =
  Results[contour, bp, op] = FindMinimum[
    {SEYCDB["RMSE", onsetplus, c, s, intModel, BetaSIA, bpre, b, {t, tEst}],
     {bpre == bp, onsetplus == op}},
    {{onsetplus, 5, 0, 21},
     {c, -3, -4, 0}, {s, .8, .1, 1},
     {bpre, .4, .1, .6}, {b, .1, .05, .4}}, AccuracyGoal → .4,
    MaxIterations → 100]
```

```
listab[0.4] = {0, 4, 6,  9, 10, 11};
```

```
tab[0.4] = contour[.4, #] & /@ listab[0.4]
```

- ... **FindMinimum**: Failed to converge to the requested accuracy or precision within 100 iterations.

- ... **FindMinimum**: Failed to converge to the requested accuracy or precision within 100 iterations.

- ... **FindMinimum**: Failed to converge to the requested accuracy or precision within 100 iterations.

- ... **General**: Further output of FindMinimum::cvmit will be suppressed during this calculation.

$$\begin{pmatrix} 1084.95 & \{\text{onsetplus} \to 0., c \to -0.000444557, s \to 0.999989, \text{bpre} \to 0.4, b \to 0.145257\} \\ 499.869 & \{\text{onsetplus} \to 4., c \to -3.83539 \times 10^{-7}, s \to 1., \text{bpre} \to 0.4, b \to 0.115231\} \\ 294.126 & \{\text{onsetplus} \to 6., c \to -2.8331, s \to 0.999998, \text{bpre} \to 0.4, b \to 0.0987922\} \\ 114.08 & \{\text{onsetplus} \to 9., c \to -4., s \to 0.813064, \text{bpre} \to 0.4, b \to 0.0719196\} \\ 75.4328 & \{\text{onsetplus} \to 10., c \to -3.99708, s \to 0.756262, \text{bpre} \to 0.4, b \to 0.0563488\} \\ 203.372 & \{\text{onsetplus} \to 11., c \to -2.73954 \times 10^{-7}, s \to 0.593483, \text{bpre} \to 0.4, b \to 0.05\} \end{pmatrix}$$

```
listab[0.437] = {0, 4,  7, 9};
```

```
tab[0.437] = contour[.437, #] & /@ listab[0.437]
```

··· FindMinimum: Failed to converge to the requested accuracy or precision within 100 iterations.

$$\begin{pmatrix} 745.804 & \{\text{onsetplus} \to 0., c \to -9.25962 \times 10^{-6}, s \to 1., \text{bpre} \to 0.437, b \to 0.127888\} \\ 232.16 & \{\text{onsetplus} \to 4., c \to -4., s \to 0.981114, \text{bpre} \to 0.437, b \to 0.0918376\} \\ 75.5744 & \{\text{onsetplus} \to 7., c \to -2.85448, s \to 0.733904, \text{bpre} \to 0.437, b \to 0.055227\} \\ 690.91 & \{\text{onsetplus} \to 9., c \to -2.13943 \times 10^{-7}, s \to 0.489079, \text{bpre} \to 0.437, b \to 0.05\} \end{pmatrix}$$

```
listab[0.455] = {0, 4, 5, 6, 7};
```

```
tab[0.455] = contour[.455, #] & /@ listab[0.455]
```

··· FindMinimum: Failed to converge to the requested accuracy or precision within 100 iterations.

$$\begin{pmatrix} 570.784 & \{\text{onsetplus} \to 0., c \to -0.0000131116, s \to 1., \text{bpre} \to 0.455, b \to 0.118818\} \\ 153.715 & \{\text{onsetplus} \to 4., c \to -4., s \to 0.881781, \text{bpre} \to 0.455, b \to 0.0808401\} \\ 93.0471 & \{\text{onsetplus} \to 5., c \to -4., s \to 0.807496, \text{bpre} \to 0.455, b \to 0.0672654\} \\ 75.1005 & \{\text{onsetplus} \to 6., c \to -0.884902, s \to 0.677291, \text{bpre} \to 0.455, b \to 0.0510108\} \\ 452.897 & \{\text{onsetplus} \to 7., c \to -2.37777 \times 10^{-7}, s \to 0.547174, \text{bpre} \to 0.455, b \to 0.05\} \end{pmatrix}$$

```
listab[0.57] = {0, 1, 2};
```

```
tab[0.57] = contour[.57, #] & /@ listab[0.57]
```

··· InterpolatingFunction: Input value {−1.92631} lies outside the range of data in the interpolating function. Extrapolation will be used.

··· InterpolatingFunction: Input value {−0.954094} lies outside the range of data in the interpolating function. Extrapolation will be used.

··· InterpolatingFunction: Input value {−0.630022} lies outside the range of data in the interpolating function. Extrapolation will be used.

··· General: Further output of InterpolatingFunction::dmval will be suppressed during this calculation.

··· FindMinimum: Failed to converge to the requested accuracy or precision within 100 iterations.

$$\begin{pmatrix} 79.3352 & \{\text{onsetplus} \to 0., c \to -0.000763437, s \to 0.672546, \text{bpre} \to 0.57, b \to 0.0504912\} \\ 620.949 & \{\text{onsetplus} \to 1., c \to -2.3739 \times 10^{-7}, s \to 0.530617, \text{bpre} \to 0.57, b \to 0.05\} \\ 1093.48 & \{\text{onsetplus} \to 2., c \to -2.07552 \times 10^{-7}, s \to 0.439888, \text{bpre} \to 0.57, b \to 0.05\} \end{pmatrix}$$

```
ListLinePlot[{Legended[Transpose[{ listab[0.4], First /@ tab[0.4]}], "β-pre = 0.4"],
  Legended[Transpose[{ listab[0.437], First /@ tab[0.437]}], "β-pre = 0.437"],
  Legended[Transpose[{ listab[0.455] , First /@ tab[0.455]}], "β-pre = 0.455"],
  Legended[Transpose[{ listab[0.57] , First /@ tab[0.57]}], "β-pre = 0.57"]
 }, AxesLabel → {"Earlier onset\nbefore Feb 23", "RMSE"}]
```
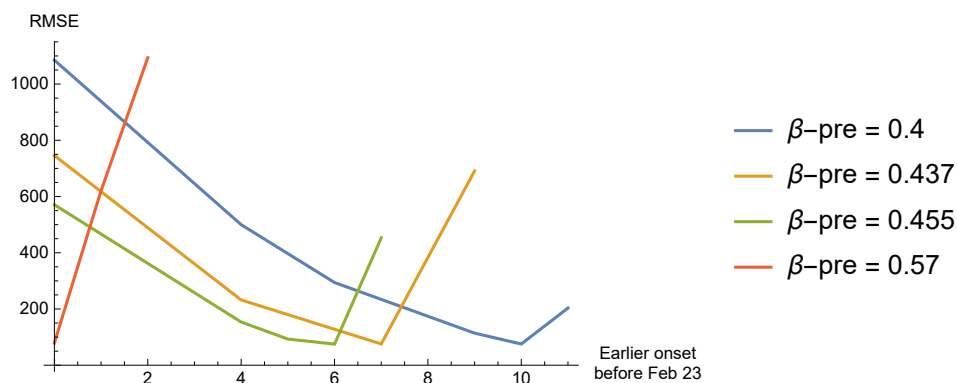

16.12. Earlier finding on the need to perhaps move the date of onset

Above analysis has a clear summary in the first subsection of this chapter. Before we got there, there were some preliminary results, that caused us to think into that direction. It may be useful to maintain some of these earlier results.

The following two estimation runs do not move the day of onset, but they allow the constant in the back-tracing to have positive rather than negative values, which is an indication of earlier onset. The drawback is that the day of intervention remains fixed at 24. The error term appears to be ill-conditioned. In the first run, the constant jumps from starting value 0 to the edge 20. In the second run, the constant hardly moves from the given startvalue 14. The infection still is in its exponential phase, even after the intervention. The new β still is relatively high, in order to meet with the number of deaths at the end of the observations.

```
ResetOptions[SEYCDB]; SEYCDB[Options, Update, Rest];
```

```
FindMinimum[SEYCDB["RMSE", c, s, intModel, BetaSIA, b, {t, tEst}],
 {{c, 0, -4, 20}, {s, .5, .1, 1}, {b, .05, .001, .4}}]
```

... FindMinimum: The point {20., 0.433442, 0.238955} is at the edge of the search region {-4., 20.} in coordinate 1 and the computed search direction points outside the region.

$\{1502.82, \{c \to 20., s \to 0.433442, b \to 0.238955\}\}$

```
SEYCDB[PlotLine]
```
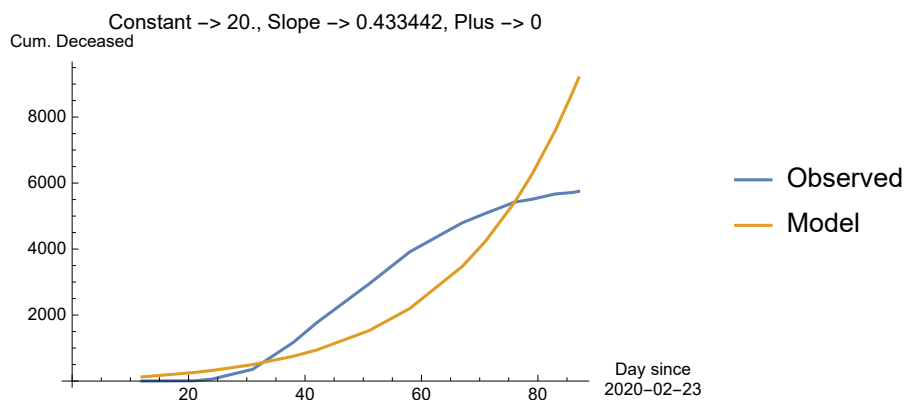


```
FindMinimum[SEYCDB["RMSE", c, s, intModel, BetaSIA, b, {t, tEst}],
 {{c, 14, -4, 20}, {s, .5, .1, 1}, {b, .05, .001, .4}}]
```

$\{725.838, \{c \to 14.0754, s \to 1., b \to 0.132167\}\}$

```
SEYCDB[PlotLine]
```
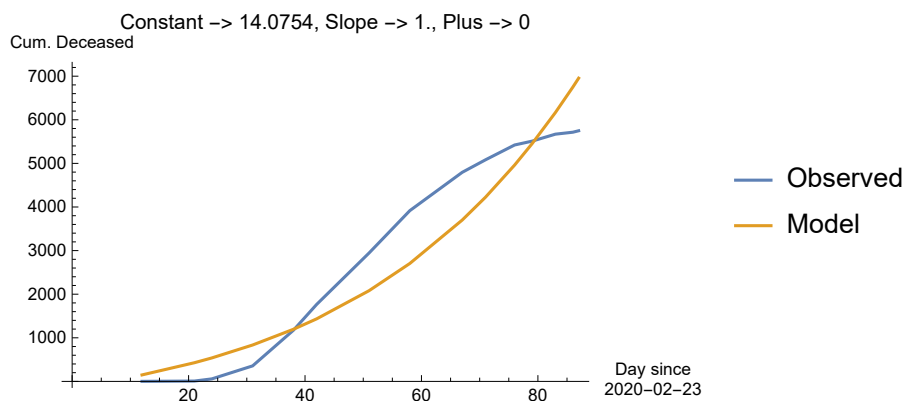
16.13. Post mortem on γ: intervention at t = 24 - variant

16.13.1. Why changing γ is tricky

In the *Introduction* section 1.5 it was argued that changing γ can be problematic. This also holds when estimating a new value following the intervention. The following shows why this is.

We first set a base, now with $\beta = 0.6$ and $\gamma = 0.1$. Thus $R_0 = 6$.

Subsequently we do an intervention with $\beta = 0.4$ and $\gamma = 2$. Thus $R_0 = 0.2$.

16.13.2. Choosing base values with $\beta = 0.6$ and $\gamma = 0.1$

The SEYCD model discussed above will be the base case. We repeat the calculation here without discussion. The only changes are: (a) A longer time horizon, to allow later comparison. (b) A different legend to highlight that this is the base case.

```
PlotLegendsSIA[SEYCD, DotDashed]
```

```
ResetOptions[SIA];
```

```
NqSeed = 17.4 × 10^6;
SetOptions[SIA, BetaSIA[] → .6, GammaSIA[] → .1]
```

$\{$Aqt0 → 0, Cqt0 → 0, DataMold → $\{$S, I, A$\}$, Dqt0 → 0, Iqt0 → Null, Nqt0 → 1.74×10^7, Onset → 0, Seed → 100, Yqt0 → 0, AlphaSEYCD() → 1., BetaSIA() → 0.6, GammaSIA() → 0.1, PhiSICD() → 0.015, R0() → Null$\}$

```
SEYCD[Run, Pr, newModel, t, 250];
```

```
pbase = SEYCD[Plot, 0, 250]
```
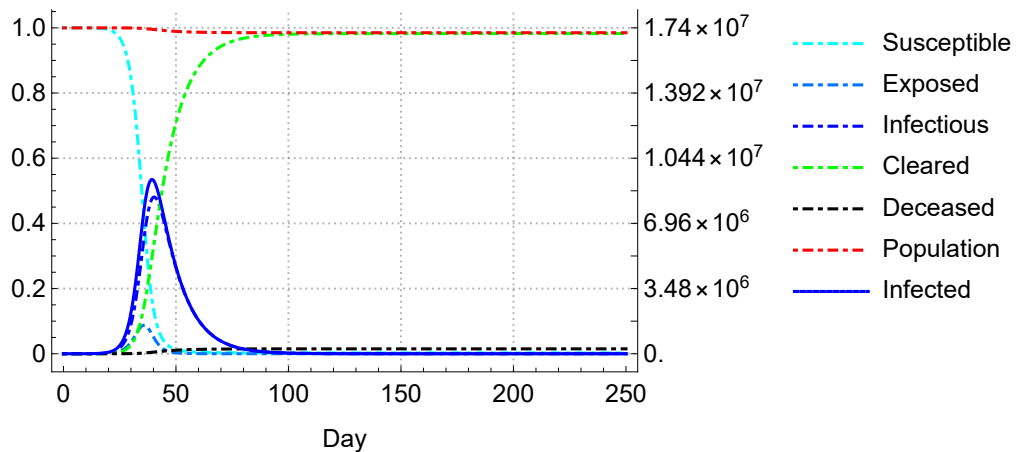

16.13.3. Intervention

```
PlotLegendsSIA[SEYCD]
(* intervention uses the default plotting colours and legends *)
```

```
NqSeed = 17.4 × 10^6;
SetOptions[SIA, Nqt0 → NqSeed]
```

{Aqt0 → 0, Cqt0 → 0, DataMold → {S, I, A}, Dqt0 → 0, Iqt0 → Null, Nqt0 → 1.74×10^7, Onset → 0, Seed → 100,

   Yqt0 → 0, AlphaSEYCD() → 1., BetaSIA() → 0.6, GammaSIA() → 0.1, PhiSICD() → 0.015, R0() → Null}

```
SetOptions[Intervention, Time[Intervention] → 24,
 BetaSIA[Intervention] → .4, GammaSIA[Intervention] → 2]
```

{Time(Intervention) → 24, R0(Intervention) → 0.99,

   BetaSIA(Intervention) → 0.4, GammaSIA(Intervention) → 2}

```
SEYCDB[Run, Pr, intModel, t, 250];
```

```
intModel[t]
```

Hold$\left[\{Sq'(t) = -5.74713 \times 10^{-8}\right.$ BetaSIA(t) Sq(t) Yq(t), Eq'(t) = −1. Eq(t) − Sq'(t),

   Yq'(t) = 1. Eq(t) − Aq'(t), Aq'(t) = GammaSIA(t) Yq(t), Dq(t) = 0.015 Aq(t), Cq(t) = Aq(t) − Dq(t),

   Nq(t) = 1.74×10^7 − Dq(t), Iq(t) = Eq(t) + Yq(t), Nq(0) = 1.74×10^7, Sq(0) = 1.73999×10^7,

   Eq(0) = 100, Yq(0) = 0, Cq(0) = 0, Dq(0) = 0, Aq(0) = 0, Iq(0) = 100, BetaSIA(0) = 0.6,

   GammaSIA(0) = 0.1, WhenEvent[t > 24, {BetaSIA(t) → 0.4, GammaSIA(t) → 2}]$\}\big]$

```
Explain[SEYCD[42], SEYCD] (* supposed top of "reported deaths per day" *)
```

{S → 1.69412×10^7, E → 0.444803, Y → 0.342246,

   C → 451 909., D → 6881.86, N → 1.73931×10^7, I → 0.787048, A → 458 791.}

The most useful plot now might be the one on ICU beds.

```
seicdplotoptionsICU = SEYCD[Plot, Options, ICU, PlotRange → {0, 10000}];
Plot @@ {SEYCD[ICU, 1500, t], {t, 0, 150}, seicdplotoptionsICU}
```
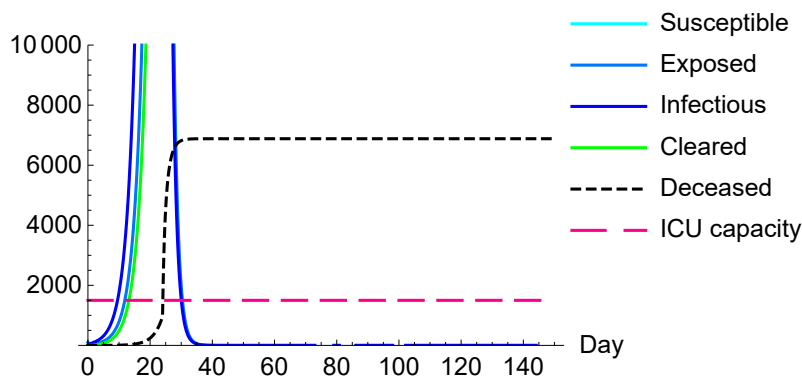

The values at the end of the 250 days are:

```
Explain[SEYCD[250], SEYCD]
```

{S → 1.69412×10^7, E → -3.20474×10^{-13}, Y → -2.46583×10^{-13},

   C → 451 910., D → 6881.88, N → 1.73931×10^7, I → -5.67057×10^{-13}, A → 458 792.}

```
pintv = SEYCD[Plot, 0, 250]
```
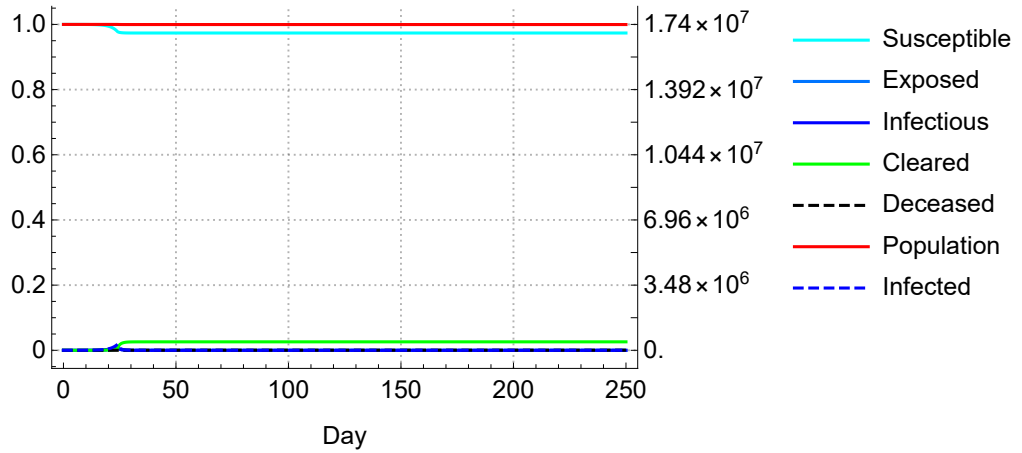

Combining the two plots.

```
Show[pbase, pintv]
```
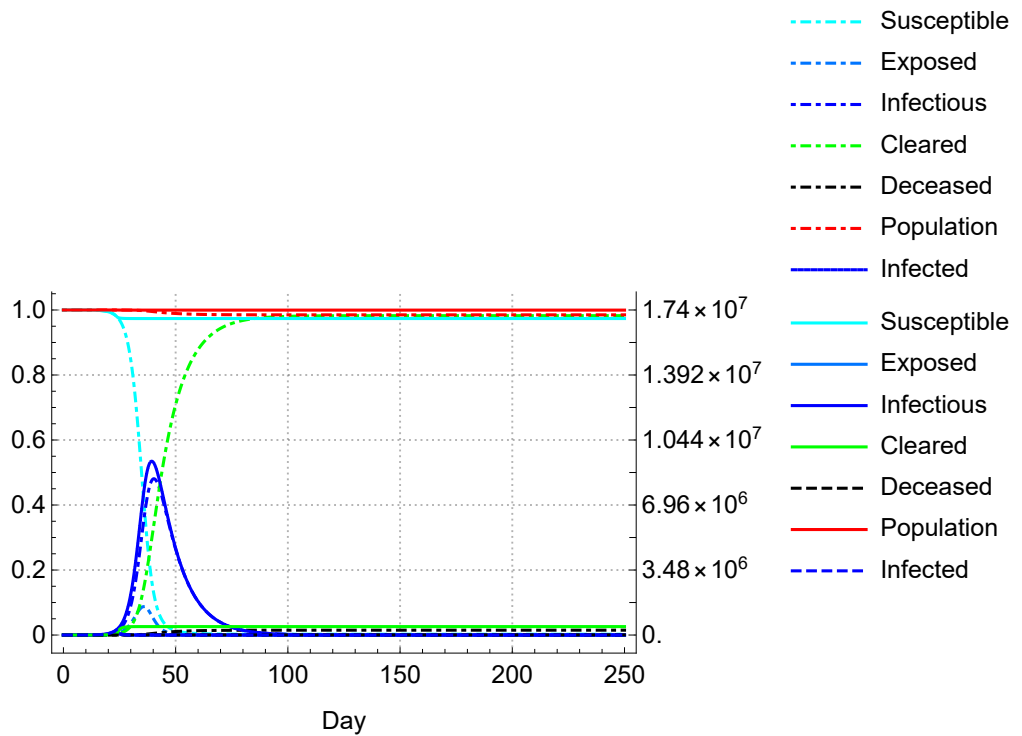

This is around the day of the intervention, day 24, using a plotrange of 0.2% (no axes ticks).

The following to graphs show the problem. Because of $\gamma = 2$, there is a large wave from the infectious Y into the acquitted A, and a fixed proportion of those are the deaths D.

Thus, while the intervention is supposed to reduce the R_0 from 6 to 0.2 and while the infectious period has been reduced from 10 days to 1/2 day, we see a surge of infections and deaths in the few

days after the intervention. However, what happens is that when the new infectious period has ended, the units are allocated to their respective compartments, and for the deaths there still is a longer personal path towards the actual day of dying. Thus we should keep in mind what the definitions are. Also, our estimate would be much influenced by how we have formulated the path for Dq[t] to Drec[τ] now. For out setup it is best to keep γ constant.

```
Show[pbase, pintv, PlotRange → {{20, 30}, {0, .002}}]
```
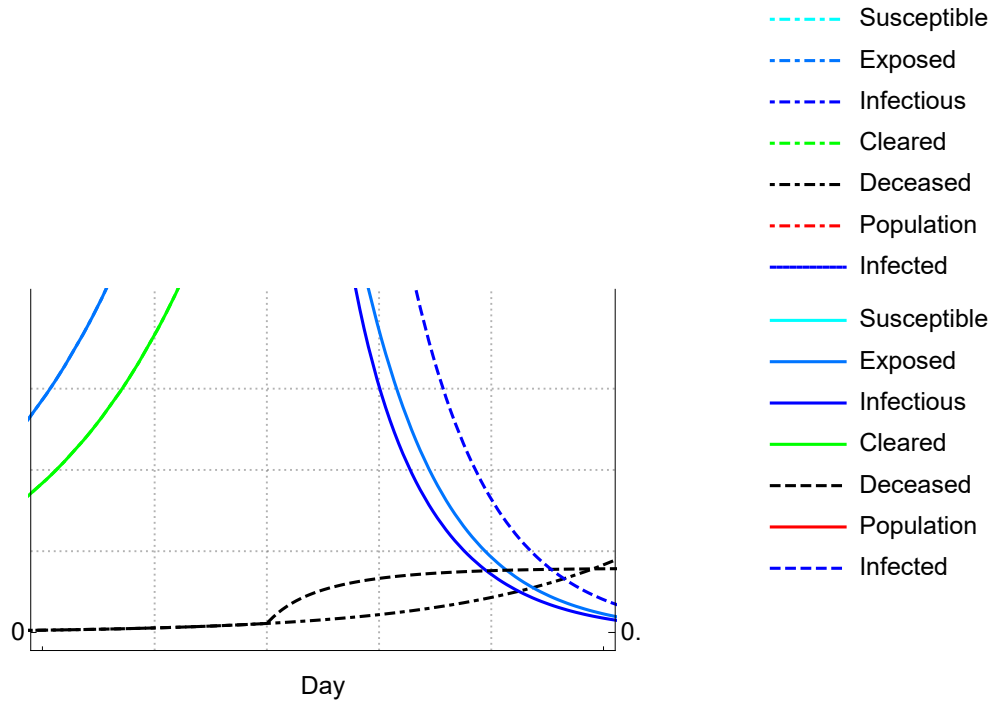
```
Show[pbase, pintv, PlotRange → {{20, 30}, {0, .008}}]
```


Day
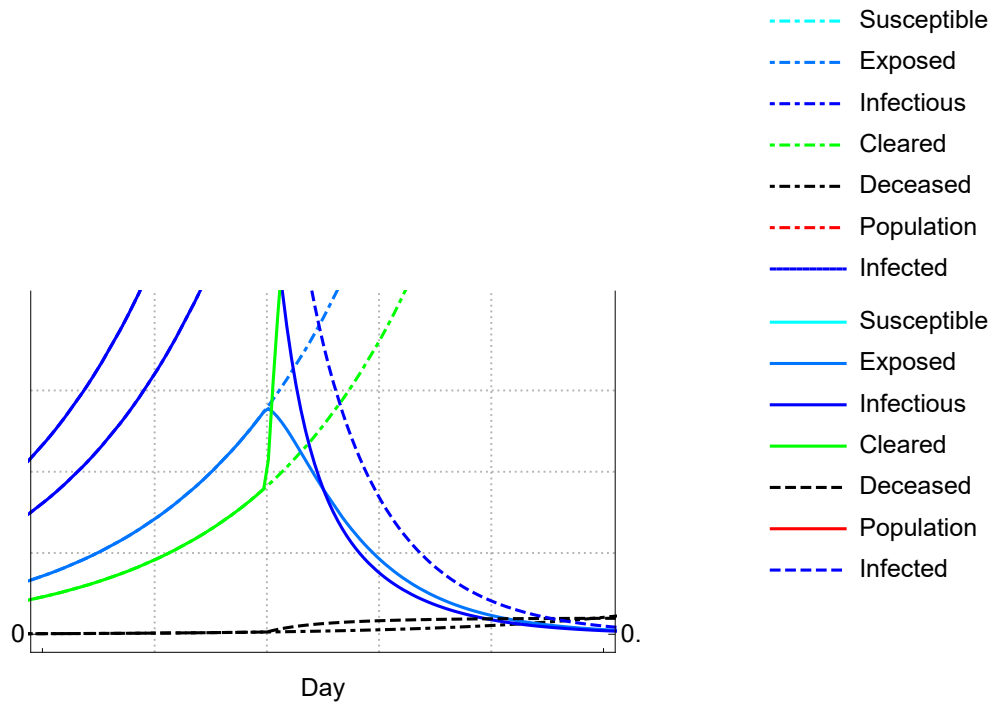
16.13.4. One of the earlier estimation efforts

The following includes γ and a condition that $R_0 = 0.99$ as above, and we do not accept a β that is larger than in the pre-intervention period. This cell has been locked since it takes more than 500 iterations.

```
FindMinimum[{SEYCDB["RMSE", c, s, intModel, b, g, {t, 250}],
  {b < .99 g}}, {{c, -3, -4, 0}, {s, .5, .1, 1}, {b, .3, .001, .4}, {g, .2, .05, .4}}]
```

⋯ FindMinimum: Failed to converge to the requested accuracy or precision within 500 iterations.

$\{981.421, \{c \to -2.4521 \times 10^{-6}, s \to 1., b \to 0.396, g \to 0.4\}\}$

Thus it seemed a good idea to include back-tracing into the estimation, and the idea paid-off since we are now better aware of complications. It seems that we would also have to include the β from before the invention, and the date of onset, before we can properly explain the shape of the observed death count. We actually did so in the earlier section.

Appendix A. Recall exponential decay, reliability, survival

A.1 Introduction

It will be useful to give a short review of survival, reliability analysis and event analyses. Analyses of these aspects are quite similar, but there are differences, for example w.r.t. (1) durations (ages) versus calendar time, (2) homogeneous or heterogeneous populations (which matters for the

denominator). One might argue that these aspects differ so much from SI(EY)A(CD) modeling that an effort at comparison might be confusing. However, the meaning of an assumption would be better understood by looking at cases when the assumption does not apply. (Though, it is not argued here, that the best understanding of SI(EY)A(CD) is achieved by looking at all cases where its assumptions do not apply.) This appendix only discusses the Success vs Failure model in short.

A.2 Half-life and doubling period

For any growth process, it is useful to have this rule of thumb available.

Consider growth over a period p, for which the outcome at the beginning is a and and the outcome at the end is b. Let us assume that the growth process is (approximately) exponential. This means that we assume that a and b are on a curve $f[t] = f[0]$ Exp[$g\ t$]. From this, we can find the starting value and the average growth rate. With this estimated growth rate, we can also determine the *half-life* or *doubling period,* depending upon whether there is decay or growth. The values of *double* or *half* are only convenient. We might also take 10 times as much, or whatever. For example, with observation levels 23 and 56 that are 5 units of time apart:

HalfLife[23, 56., 5] (* default factor 2 *)

$$\left\{56. = 23\ e^{0.177971\ time},\ Coefficient \to 0.177971,\ Factor \to \left\{2, \frac{1}{2}\right\},\ Period \to 3.89471\right\}$$

HalfLife[23, 56., 5, 10] (* a different factor, here 10 *)

$$\left\{56. = 23\ e^{0.177971\ time},\ Coefficient \to 0.177971,\ Factor \to \left\{10, \frac{1}{10}\right\},\ Period \to 12.9379\right\}$$

In fact, for any factor h:

If $f[t] = f[0]$ Exp[$g\ t$] then $f[t] = h\ f[0]$ iff $t = Ln[h] / g$.

For $h = 2$ there is the doubling time for Ln[2] / g.

The half-life has Ln[1/2] / -g = Ln[2] / g too.

Ln[2] ≈ 0.70 and for percentage growth rates $p\%$ the rule 70 / p is used.

The above can be done for any factor, e.g. 10 or 1/10 times as much.

? HalfLife

Symbol

HalfLife[a, b, t, factor:2] assumes that b/a = Exp[g t], or g = Log[b/a] / t. Given g, we can find the

    halftime (g < 0) or doubling time (g > 0), or this actually for any factor as t = Log[factor] / g

A.3 Two compartments, one with exponential decay. Duration vs calendar time (stocks and flows)

Consider a Population with compartments *Success* versus *Failure*. Success has exponential *decay.* Failure thus grows, but its growth cannot be exponential growth as such. While this model is dynamic, there is a key difference between use of *calendar time* (decay or mortality by incidents

over time) or *ages* (e.g. mortality per age or by duration of disease). We will use the same symbol *t* whenever there is a fundamental similarity, but it is best to switch to age *a* or duration *d* or period *p* when the use of *t* might cause confusion.

In all cases we might work with *factors* or *rates.* Success can be modeled with factor $0 \leq S \leq 1$ as $s[t] = S^t = (1 - r)^t$, for the share in the population at time *t* and *r* the perunage decay rate. Thus also $s[t] = \text{Exp}[\text{Log}[S]\, t] = \text{Exp}[\text{Log}[1 - r]\, t] = \text{Exp}[-\lambda\, t]$, with exponential *failure rate* $\lambda = -\text{Log}[S] \approx r \geq 0$ in which the negative sign is used since $0 \leq S \leq 1$. Recall that $\text{Log}[1 - r] \approx -r$ for small values of *r*, or Limit$[\text{Log}[1+r] / r, r \to 0] = 1$. Thus λ and *r* hardly differ but better be distinguished for accuracy. The growth rate of $s[t]$ is $d\text{Log}[S^t] / dt = d(t\, \text{Log}[S]) / dt = \text{Log}[S] = -\lambda$, actually the rate of decay.

Failure has no constant factor like Success. Failure is just the opposite $f[t] = 1 - s[t] = 1 - \text{Exp}[-\lambda\, t]$, so that $s[t] + f[t] = 1$. Observe that $s[0] = 1$ and $f[0] = 0$, and that Failure thus cannot grow exponentially from 0. A property is that $f'[t] = -s'[t] = \lambda\, s[t]$.

The expected value of a moment of decay or the "mean time to failure" (MTTF) is $\mu = 1/\lambda$, see below.

Success and Failure always intersect at the value $s[t] = f[t] = 1/2$. The involved *t* is the *half-life*, and regarded sometimes as more informative than the rate. We have $S^{t_{1/2}} = 1/2$. Using logarithms we find $t_{1/2} = \text{Log}[2] / \text{Log}[S] = \text{Log}[2] / \lambda = \mu\, \text{Log}[2]$. This allows for quick calculation. With $\text{Log}[2] \approx 0.7$, half-life is about 70% of the mean, or the mean is about 1.4 of half-life. Failure doesn't have a constant doubling time since it is not exponential.

The following plots for $\lambda = -\text{Log}[S] = 0.25$, so that $\mu = 1/\lambda = 4$. There is little need to explain this, since it shows in the graph.

SuccessFailurePlot[.25 , E → False]
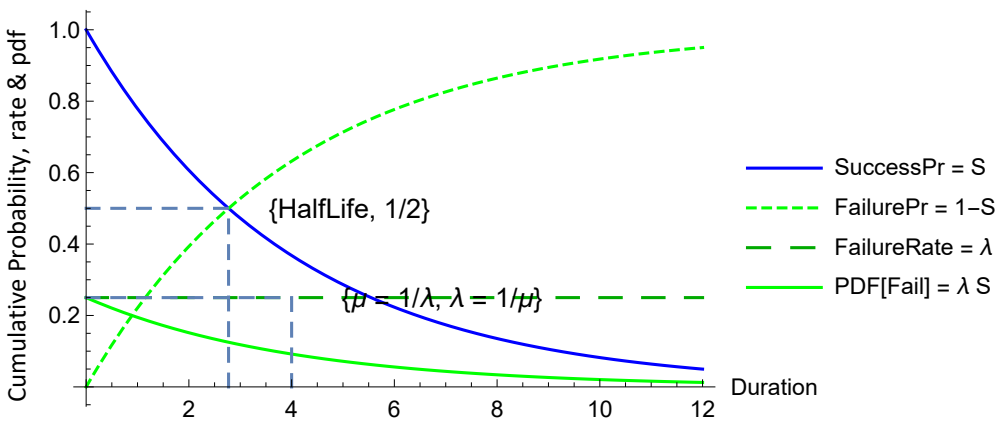
A.4. Factors and rates

The reason to work with rates is twofold. The minor one is that it reduces clutter in derivatives. The main one is that the focus is on *decay of success* and *growth of failure*, which processes may be better grasped by rates per time. However, the use of factors cannot be avoided. Rates are not observed directly. We only *count events* and *measure lengths and durations*, then turn these into factors, and then turn those into rates. Working with factors can sometimes be illuminating, also because they remind us where the rates come from.

For example, when we have two processes in sequence, then the *joint success* requires the two of them, which satisfies the logical "*and*" notion, which means multiplication. The joint success is $s = s_1 s_2$ with $\lambda = \lambda_1 + \lambda_2$ and $f = 1 - s = 1 - (1 - f_1)(1 - f_2)$.

When we have two processes in parallel, then the whole system fails when both fail. Thus *joint failure* satisfies the logical "and" notion, so that $f = f_1 f_2$ and $s = 1 - (1 - s_1)(1 - s_2)$. The relation for λ is more involved. We can find this by looking at the mean durations. For example (1), with n units in parallel and each with same MTTF $1/\lambda$, then the whole has a MTTF of $1/\lambda \sum_{i=1}^{n} (1/i) \approx \text{Log}[n] / \lambda$. If we were to increase n, then the rate of increase of the MTTF is $1/(n\lambda)$. For example (2), with two systems in parallel with different service times, then the joint MTTF is $1/\lambda_1 + 1/\lambda_2 - 1/(\lambda_1 + \lambda_2)$.

```
FailurePr["ParallelMTTF", Identity, m, 2] == FailurePr["ParallelMTTF", m1, m2]
```

$$\frac{3\,m}{2} = -\frac{1}{\frac{1}{m1} + \frac{1}{m2}} + m1 + m2$$

```
TheFormer /. {m1 → m, m2 → m}
```

True

See https://reference.wolfram.com/language/ref/ReliabilityDistribution.html. PM. Within this approach it is advisable to speak about the *rate of decay of success* and the *rate of growth of failure*. Sometimes λ is called the failure rate, which is okay when there is no confusion, but it is only the *rate of decay of success*. The growth of failure is $d\text{Log}[f[t]] / dt = f'[t] / f[t] = \lambda\, s[t] / f[t]$.

To translate SIA to the Success & Failure model, we regard *I* in SIA as the success that is decaying, and *A* as the failure that collects the decay. The rate of decay λ in the Success and Failure model links to the γ in SIA as the rate of decay of *I*.

For SICD, we have $D = \varphi\,A$. Thus here $D = \varphi\,f[t] = \varphi\,(1 - s[t])$. Thus $D' = -\varphi\,s'[t] = -\varphi\,\text{Log}[S]\,s[t] = \varphi\,\lambda\,s[t]$. The latter translates back to SIA again as $D' = \varphi\,\gamma\,I$, which is the formulation in the conventional format of SIA.

$$
\begin{array}{c|c|c|c}
\begin{array}{c} \text{SIA} \\ I \\ \\ \\ \\ D' = \varphi\,\gamma I \end{array} &
\begin{array}{c} \text{SIA} \\ A \\ A' = \gamma\,I \\ D = \varphi A \\ \end{array} &
\begin{array}{c} \text{S \& F} \\ s[t] \\ \\ \\ \\ \end{array} &
\begin{array}{c} \text{S \& F} \\ f[t] \\ f'[t] = -s'[t] = \lambda s[t] \\ D = \varphi f[t] \\ D' = -\varphi s'[t] = \varphi\,\lambda s[t] \end{array}
\end{array}
$$

PM. The addition in SIA of $A = C + D$ with $A' = C' + D' = ((1 - \varphi)\,\gamma + \varphi\,\gamma)\,I = \gamma\,I$ applies to the (*additive*) proportions in failure, and must not be confused with a *multiplicative* split of success $s[t]$ and addition $\lambda = \lambda_1 + \lambda_2$ in the joint success. It is also possible to split *I* in proportions $\varphi\,I$ and $(1 - \varphi)\,I$ and have each decay with γ. (Some readers might not see this as acceptable without the first explanation that $D' = -\varphi\,s'[t]$ translates into $D' = \varphi\,\gamma\,I$.)

For SEYCD, the exposed *E* and infectious *Y* are in time sequence but *not* as a product of decay. Here, the model of queueing theory with two sequential service units applies (we add the periods as $1/\alpha + 1/\gamma$).

? SumTwoExp

Symbol

SumTwoExp[gamma, alpha, t] gives the PDF of the sum of one exponentially

distributed phase with parameter gamma (mean 1/gamma) and another such phase

with parameter alpha (mean 1/alpha). The expected value of the sum of both outcomes

is 1/gamma + 1/alpha. If gamma == alpha, then this is PDF[ErlangDistribution[2, alpha], t]

A.5. One point of observation suffices

The setup is determined by only one observation different from 0. If we observe $\{T, S_{Obs}\}$ for T the time of observation, then $S_{Obs} = s[T] = S^T$ so that $S = S_{Obs}^{1/T}$ gives a direct estimate of the unknown factor S per period. Alternatively $S_{Obs} = \text{Exp}[-\lambda\, T]$, and $\lambda = -\text{Log}[S_{Obs}] / T$. In the above we actually used $T = 1$.

Having more points of observation causes the idea that the last observation may be the most relevant. A possibility is to transform the data to Log[s[t]] and apply linear regression, if one assumes that the error around the path of decay is lognormal. There is also the consideration that the *mean duration* of failure times better be accurate. Curve-fitting then is replaced by finding the best estimate for μ.

(In a SIA model, the infection grows by $r = \beta - \gamma$ in the first phase. If we find a prevalence of 100 in one week, and indeed started from 1 person, then this has a rate of $100 = (1 + r)^7$, or $r = 100^{1/7} - 1 \approx$ -Log[100]/7 = 34% per day. If the infectious period is 3 days, then $\beta = .342 + 1/3 = .675$ and $R_0 = \beta / \gamma$ = 3 * .675 = 2, which suggests a serious outbreak.)

A.6. Dynamic growth versus probability. CDF, PDF, and properties

For T the random variable for the time to an event of decay, the CDF of the exponential distribution is $P[T < t] = 1 - \text{Exp}[-\lambda\, t]$ while $t \geq 0$ and undefined elsewhere. This is exactly the same form as *above growth process for Failure*. This means that we can sometimes use interpretations and analogies, while keeping in mind that a (perhaps even deterministic) *growth process* is not the same as a *probability distribution*. However, the useful interpretation is that T is the *period to failure*, and we may observe that it has an exponential distribution (though in reality we don't check for eternity).

The expectation or mean μ of the exponential distribution has the value $\mu = 1 / \lambda$. As said, we can now interpret it as the "mean time to failure" or MTTF. It may sometimes make for better understanding of the CDF by writing $P[T < t] = 1 - \text{Exp}[-t / \mu]$, and $s[t] = \text{Exp}[-t / \mu]$. However, it seems that the human inclination to count is stronger than the inclination to measure, so that many might regard "6 items per minute" as having more impact than "10 seconds per item". It obviously depends upon context whether the count or the period is in focus.

Observe that $s[\mu] = \text{Exp}[-1] = 1 / e \approx 0.37$ and $f[\mu] \approx 0.63$. Thus μ on the horizontal axis can be identified not only by the half-life but also by using these values on the vertical axis.

```
SuccessFailurePlot[.25 , HalfLife → False]
```
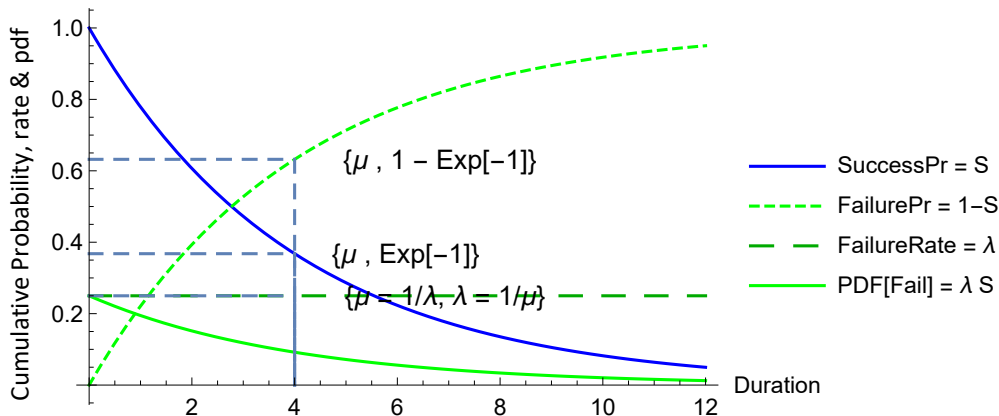

A.7. Subject areas: duration or calendar time

Let us now consider calendar time versus two types of durations.

(1.a) Periods of a homogeneous population, only information about periods

This concerns phenomena like ages, life-spans, or clearing, recovery and acquittal periods, where the time axis concerns durations or the age of the items. We assume that there is no cut-off or censoring point, and that the longest observation has indeed been made. The *n* items are considered comparable and only the durations differ. The main parameter is the *mean duration* $\mu = \Sigma$ durations $/ n = D / n$. This parameter suffices for the exponential distribution. The assumption for this distribution would be based upon the histogram of the observations and other considerations. However, there is no other information why the durations differ, they just differ. We may say that the durations are characterised by the *event* of their ending, but then there still is a difference with a heterogeneous population.

(1.b) Periods of a heterogeneous population, also with information on events

This concerns phenomena like incidents in a population, in which we relate *counts* to *item-durations,* e.g. getting a disease divided by person-years. There are at least two groups here. We are modeling a process of separation from a common source. With *N* the total population or sample size, there are *n* incidents with their *period to an incident*, and *N - n* unaffected. This distinction is missing in the above (having *n = N*).

There is the issue of censoring. For now we assume a cut-off at the latest incident, and all inferences are conditional to this. The *item-durations* of the unaffected are also included in the denominator. The main parameter is the mean period to an incident $\mu = \Sigma$ durations $/ n = D / n$. Again we have $\lambda = 1 / \mu$. There is an implication for the prevalence $f = n / N,$ but this is less in focus. Often *N* is only the sample size, and the focus is on properly measuring the durations, so that the sample outcomes can be confronted with another study on the actual prevalence.

Alternatively, we split the durations of the Success (unaffected) and Failure (diseased) groups. Then $\mu_F = D_F / n$ would be the estimate under (1.a) for said homogeneous group who all contacted the

disease. Then $\mu_S = D_S / (N - n)$ is the mean period for those without an incidence, in some cases merely the duration of the study. Thus $\mu = \mu_F + \mu_S (1 - f) / f$. It would seem that prevalence cannot be neglected, but that would be another topic of discussion.

(2) Calendar time

In this case Failure also stands for the prevalence of a disease, when a population starts unaffected, and meets with exponential decay because of the event, e.g. a disease. The notion of a prevalence can be relevant for cases (1ab) too, but the issue of prevalence does not really occur in the discussion above yet. However, prevalence has direct relevance for the discussion for calendar time. Namely, when $f = n / N$ is the prevalence at time T, then we may have *a single observation*, that allows the estimate of exponential decay of Success.

For example, when it is observed that a disease has spread to 10% of the population in a period of 5 years, then linear increments of 2% per year might underestimate developments in the first 5 years and overestimate developments after 5 years. If we assume exponential decay, $s[T] = S^5 = 90\%$ then the failure rate $\lambda = \text{Log}[S] = -\text{Log}[1 - 10\%] / 5 = 2.1\%$ could give a more reliable estimate, not quite because of its value but because of its modeling of exponential growth.

Thus in this case prevalence information is used, and there are only implied assumptions for mean duration and person-years. A rough estimate of the person-years involved are $D = T N$. The linear mean rate $\lambda_{\text{linear}} = n / D = f N / (T N) = f / T = 10\% / 5 = 2\%$ differs hardly from the exponential estimate, but, as said, linear and exponential interpolation and extrapolation have quite different outcomes. In actuality we might use the model $(1 - 0.02)^t$ as a rough estimate so that the switch to Exp is accurate but perhaps less clear about its actual meaning. The actual factor $0.9^{1/5} \approx 0.979148$ is accurate though.

The routine allows us to plug in a moment of observation, when the rate of decay was estimated from. We have been using the failure rate of 0.25 above. There the prevalence of 50% is already reached in the half-life of around three periods. The following supposes that the rate was determined by an observation at time 6. The routine does not give a warning when the data of rate, Tobs and Fobs do not match: the user might have a reason for wanting no match.

```
SuccessFailurePlot[.25, Level → {6, 1 - Exp[- 0.25 * 6]},
 PDF → False, HalfLife → False, E → False]
```
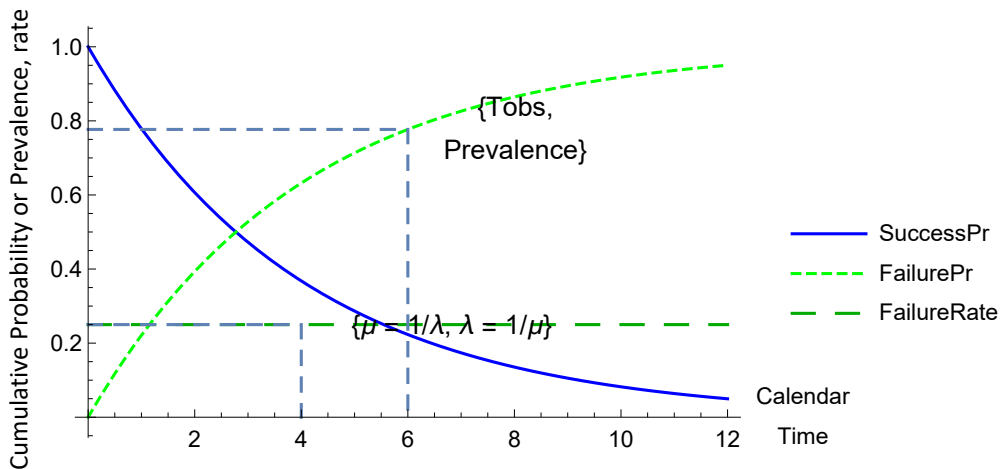


```
? SuccessFailurePlot
```

```
? SuccessPr
```

A.8. Comparison with SICD and SEYCD

The Success & Failure model is important to understand the basics of exponential decay, and for the consequences of sequential or parallel processes (also when those do *not* apply). For comparison to the SICD and SEYCD models, there remains a pitfall.

(1) For SICD, the distinction between *C* and *D* seems "parallel", but the setup is different. To translate SIA to the Success & Failure model, we regard *I* in SIA as the success that is decaying, and *A* as the failure that collects the decay. The rate of decay λ in the Success and Failure model links to the γ in SIA as the rate of decay of *I*. For SICD, we have $D = \varphi A$. Thus here $D = \varphi f[t] = \varphi (1 - s[t])$. Thus $D' = -\varphi s'[t] = -\varphi \text{Log}[S] s[t] = \varphi \lambda s[t]$. The latter translates back to SIA again as $D' = \varphi \gamma I$, which is the formulation in the conventional format of SIA. The addition in SIA of $A = C + D$ with $A' = C' + D' = ((1 - \varphi) \gamma + \varphi \gamma) I = \gamma I$ applies to the proportions in failure, and must not be confused with the addition $\lambda = \lambda_1 + \lambda_2$ above in the joint success.

(2) For SEYCD, the exposed *E* and infectious *Y* are in time sequence but *not* as a product of decay. Here, the model of queueing theory with two sequential service units applies (we add the periods as $1/\alpha + 1/\gamma$).

```
? SumTwoExp
```

> Symbol
>
> SumTwoExp[gamma, alpha, t] gives the PDF of the sum of one exponentially
>
> distributed phase with parameter gamma (mean 1/gamma) and another such phase
>
> with parameter alpha (mean 1/alpha). The expected value of the sum of both outcomes
>
> is 1/gamma + 1/alpha. If gamma == alpha, then this is PDF[ErlangDistribution[2, alpha], t]
>
> ⌄

Appendix B. Recall Poisson, Erlang and exponential

B.1. Toolkit

This Appendix is not intended to *directly* contribute to the analysis of the SI(EY)A(CD) model, but only concerns the *toolkit* level. This Appendix identifies the general analytical environment on the exponential distribution, with links to routines available in *Mathematica.*

My background is from queueing theory. SI(EY)A(CD) models look like queueing. The *renewal equation* concerns exponential population growth with birth and death. The interpretation in contacts and possible infections could be a Poisson process in the *virginal period before* the sigmoid takes over. When writing this notebook it seemed useful to recall some of this queueing theory, simply because it was my background.

Hillier & Lieberman (1967, 1972) discuss queueing theory and underlying birth-death processes. The basic case is when *counts* follow the Poisson process, or equivalently when the *interarrival periods* (IAP) are distributed by the exponential distribution. For example, when cars run in a street, we count the cars and may find the arrival rate of AR = 100 cars per day (always a mean), and then the most practical random variable concerns the IAP with a (mean) MIAP = 24 hours / (100 cars). A recent online restatement is by Hossein Pishro-Nik: https://www.probabilitycourse.com/chapter11/11_1_2_basic_concepts_of_the_poisson_process.php

Comparing SIA and queueing theory, my main finding is:

- There is indeed the same toolkit. *Rate* γ and *period* $1 / \gamma$ are related by definition (inviting the exponential distribution).

- There might be some application of queueing theory for some non-infectious disease with Poisson arrivals. The queueing modeling requires that the service rate γ is larger than the arrival rate β, or that the utilisation rate $\rho = \beta / \gamma < 1$, otherwise the queue explodes. This reminds of $R_0 < 1$. However, there is a clear distinction between infectious and non-infectious disease. For infections, there is no "joint server with a queue", since all infections "have their own server system" (natural immunity or quarantine).

- The modeling of $E + Y$ is the same as the sequential server model, with the sum of two exponential distributions with total mean $1/\alpha + 1/\gamma$, which for equal values collapses into the Erlang 2 distribution. The expected time in the system W is the sum of the time in the queue W_q plus this mean service time, $W = W_q + (1/\alpha + 1/\gamma)$, but for infections $W_q = 0$.

Thus, the modeling issues of queueing theory and SI(EY)A(CD) are different. Key summary statistics on queueing, like the number of units in the system (line length) as $L = \beta W$ (using exponential arrival β in the initial phase) require a steady state, and there is a clear conflict between "initial phase" and "steady state". An effort at comparison tends to be distracting. The models on queueing and SI(EY)A(CD) use the same toolkit but build something else. Thus this appendix only concerns the toolkit.

B.2. Basics

When we follow convention that μ is the symbol for the expected value of a distribution (like in the Normal Distribution with mean μ and spread σ), and when we model the distribution of the IAPs, then the mean IAP is μ = MIAP and the arrival rate (always a mean) is AR = λ = 1 / μ.

For example, for cars running in a street, you wait till the first car, but you do not include this in the count, because you have no observation on its IAP. When the first car has been spotted, you start the stopwatch, count cars, and when car n has been counted then you stop the watch, giving time T. Then you have n interarrival periods, and the mean interarrival period is MIAP = μ = T / n (all excluding the first (censored) car). The (always mean) arrival rate AR λ = 1 / μ = n / T. Indeed, you also spotted the first car, but it does not not count as "a car within the time window of observation", since it only indicated the *beginning* of the counting process.

Hillier & Lieberman (1967, 1972) p302 have another parameterisation of the Erlang distribution than in *Mathematica.* Below reproduces their form and plots for μ = MIAP = 10, comparable with the *mean acquittal period* of say μ = 10 days for SARS-CoV-2.

For the Hillier & Lieberman parameterisation it holds: If T_1, ..., T_k are i.i.d. with exponential distribution with mean μ = 1/(λ k), then the sum T has an Erlang (Gamma) distribution with parameters λ and k. For k = 1 the Erlang distribution gives the exponential distribution.

For the sum of a Poisson process *for any k* it holds: for a particular arrival rate AR = λ we have the *expectation* or *mean* interarrival *period* MIAP = μ = 1 / λ.

NB. It makes more sense to parameterise these functions in such manner that the stated parameter is the mean, like for the normal distribution. In that case it is clearer what variable is being modeled: IAPs and not counts. This is discussed in the next subsection. Now it suffices that we have a clear distinction about the meanings of μ and λ. As a general rule: do not use μ as a parameter in these functions unless it really is the expected value.

```
ErlangHillierLieberman[k, λ]
```

ErlangDistribution[*k, k λ*]

```
ehl[k_, λ_, t_] = ErlangHillierLieberman[PDF, k, λ, t]
```

$$\begin{cases} \frac{k^k \, \lambda^k \, t^{k-1} \, e^{-k\lambda t}}{\Gamma(k)} & t > 0 \\ 0 & \text{True} \end{cases}$$

```
Expectation[x, x ≈ ErlangHillierLieberman[k, λ]]
```

$$\frac{1}{\lambda}$$

```
StandardDeviation[ErlangHillierLieberman[k, λ]]
```

$$\frac{1}{\sqrt{k} \, \lambda}$$

For k = 1 we get the exponential distribution.

```
ehl[1, λ, t]
```

$$\begin{cases} \lambda \, e^{\lambda \, (-t)} & t > 0 \\ 0 & \text{True} \end{cases}$$

For the following MIAP = μ = 10.

```
mu = 10; labda = 1 / mu;
```

```
plotoptshere = Sequence[ImageSize → Medium, BaseStyle → {FontSize → 13},
    PlotRange → All, AxesOrigin → {0, 0}, AxesLabel → {"Time", "Prob. density"}];
```

The conventional parameterisation has the advantage that the exponential distribution starts for t = 0 at the value of the parameter. The expected value of the IAPs can be found as MIAP = $\mu = 1 / \lambda$. For example, for $\mu = 10$ and $k = 3$, there would be three servers with each a service time of 1 / (λ 3) = 1 / .3 = 3.33 so that the mean (expected) service time is 10.

```
ErlangHillierLieberman[Plot, labda, 3, plotoptshere]
```

Prob. density
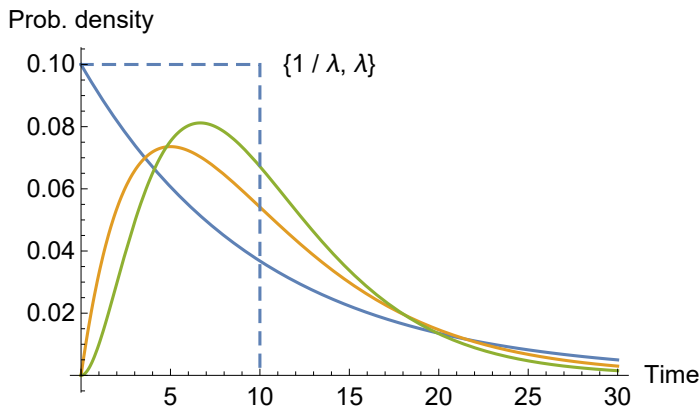
? ErlangHillierLieberman

Symbol

ErlangHillierLieberman[k, labda] is the parameterisation of the Erlang distribution as used by

Hiller and Lieberman (1967), "Introduction to Operations Research", Holden–Day, p302

ErlangHillierLieberman[PDF, k, labda, t] gives the PDF for t

ErlangHillierLieberman[Plot, labda, x:3] reproduces their plot for k in {1, 2, 3} for (t, 0, x * 1 / labda}

B.3. Parameterisation such that μ = expected value

Awkwardly, the standard functions in the literature and programming for the Poisson process are parameterised with λ while the expectation of the *interarrival periods* is given as $1 / \lambda$. For our purposes it would be more convenient if the parameter were μ and the expectation were μ, like we have for the Normal Distribution.

For who considers these issues for the first time, the question is whether λ must be taken as MIAP or AR. Given how these functions are conventionally denoted in the literature and programmed in computer codes, we must input the AR, to force that the expected value, which we want to see, becomes the MIAP.

In the parameterisation by Hillier & Lieberman, it are the IAP that have a distribution and expectation, and their expected value is MIAP = $(1/\lambda) / k$, but not λ / k.

It is clearer to define the distribution such that the mean becomes the parameter. Since *Mathematia* has a pre-defined ErlangDistribution with a stated parameterisation, we make a small routine

that translates from the mean to their conventional "inverted" use.

```
ErlangMean[k, μ]
```

$\text{ErlangDistribution}\Big[k,\ \dfrac{k}{\mu}\Big]$

```
ErlangMean[PDF, k, μ, t]
```

$$\begin{cases} \dfrac{k^k\,\mu^{-k}\,t^{k-1}\,e^{-\frac{kt}{\mu}}}{\Gamma(k)} & t>0 \\ 0 & \text{True} \end{cases}$$

```
Expectation[x, x ≈ ErlangMean[k, μ]]
```

μ

```
StandardDeviation[ErlangMean[k, μ]]
```

$\dfrac{\mu}{\sqrt{k}}$

We now can now use MIAP as the parameter of the function, and see how the random draws of the IAPs are distributed with this expectation.

```
mu = 10; (* mean interarrival period *)
```

```
ErlangMean[Plot, mu, 3, plotoptshere]
```
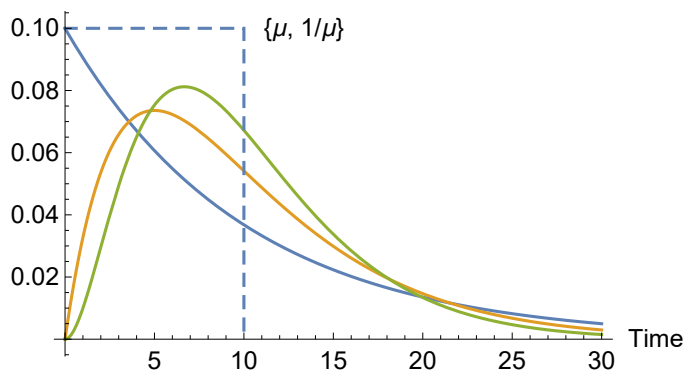

Once you have noticed the (awkward) convention in the literature, it can become an automated habit though. One may simply follow convention and put in the AR in order to model the IAPs with their MIAP.

B.4. The exponential distribution

The *Mathematica* documentation provides an example of a Poisson process, stating: "The times between events are independent and follow $\text{ExponentialDistribution}[\mu]$."

The *Mathematica* parameterisation of the exponential distribution generates the same plot as we have seen above, with the same inversion of AR and MIAP. Potentially it is more convenient to follow the literature and standard programming.

```
PDF[ExponentialDistribution[λ], t]
```

$$\begin{cases} \lambda\, e^{\lambda\,(-t)} & t \geq 0 \\ 0 & \text{True} \end{cases}$$

```
Expectation[x, x ≈ ExponentialDistribution[λ]]
```

$$\frac{1}{\lambda}$$

```
StandardDeviation[ExponentialDistribution[λ]]
```

$$\frac{1}{\lambda}$$

```
mu = 10;
```

```
Plot[PDF[ExponentialDistribution[1 / mu], t], {t, 0, 3 mu}, Evaluate[plotoptshere]]
```
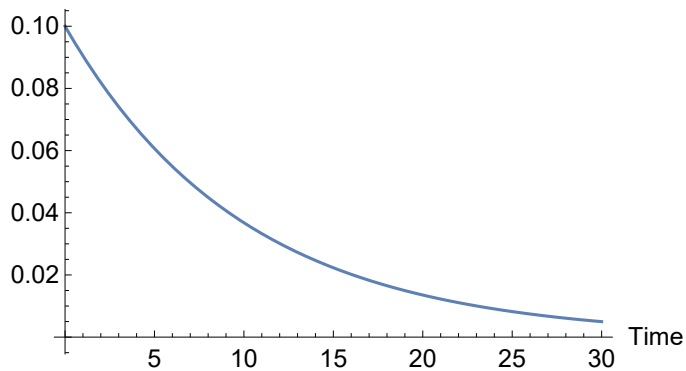

B.5. A random simulation with a Poisson process

The *Mathematica* documentation provides an example of a Poisson process, stating:

(a) "PoissonProcess at time t is the number of events in the interval 0 to t."

(b) "The times between events are independent and follow ExponentialDistribution$[\mu]$."

Their μ is not the expected value and they better write λ.

The *Mathematica* specification of the Poisson process requires us to specify a time window. How do the random draws fit within that window ? For this simulation, the stopwatch starts properly at 0, and thus the first count is made after a proper interarrival period. However, the *last* observation is tricky. A random draw might generate an interarrival period that extends over the time horizon for the simulation. It appears for this particular routine that the final observation is truncated to fit the simulation period. It actually is no observation. Thus we must drop the last element.

When we use the Poisson process with our μ = MIAP = 10, then we get the following.

```
mu = 10; labda = 1 / mu;
```

A short run

```
data = RandomFunction[PoissonProcess[labda], {0, 5 mu}];
```

The simulated interarrival periods are the following. The option False means that we do not elimi-nate the last element yet.

```
iaps = InterArrivalPeriod[Data, data, False]
```

{8.53136, 1.8443, 13.3736, 9.53722, 4.1882, 8.86907, 3.65626}

```
"Observation period" → Total[iaps]
```

Observation period → 50.

In *Mathematica*'s ListStepPlot *a count is a vertical line upwards*. The plot doesn't show the last (truncated) observation *anyway.* The plot extends the horizontal timeline of the *penultimat*e obser-vation, which is a bit confusing (as if there is a long last iap), but after some thought we grow aware that the programmers might have had some problem in cutting off the plot at exactly the last line upwards.

```
ListStepPlot[data, Filling → Axis,
 AxesLabel → {"Time", "Count"}, BaseStyle → {FontSize → 13}]
```
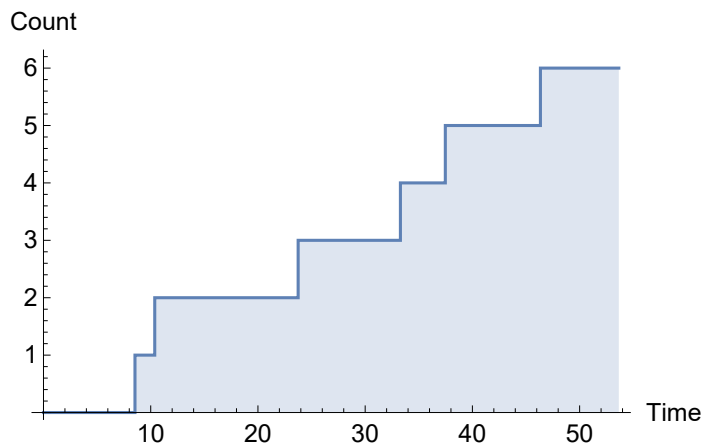


```
obs = Drop[iaps, -1];
```

$\{N \rightarrow (\text{nn = Length[obs]}), \text{Time} \rightarrow (\text{tt = Total[obs]}), \text{MIAP} \rightarrow \text{tt} / \text{nn}, \text{AR} \rightarrow \text{nn} / \text{tt}\}$

{*N* → 6, Time → 46.3437, MIAP → 7.72396, AR → 0.129467}

```
{Mean[obs], StandardDeviation[obs]}
```

{7.72396, 4.10467}

This is an alternative plot. The *iap per count* are light grey blocks. The time axis now stops at the final count.

```
InterArrivalPeriod[Plot, obs]
```

Count
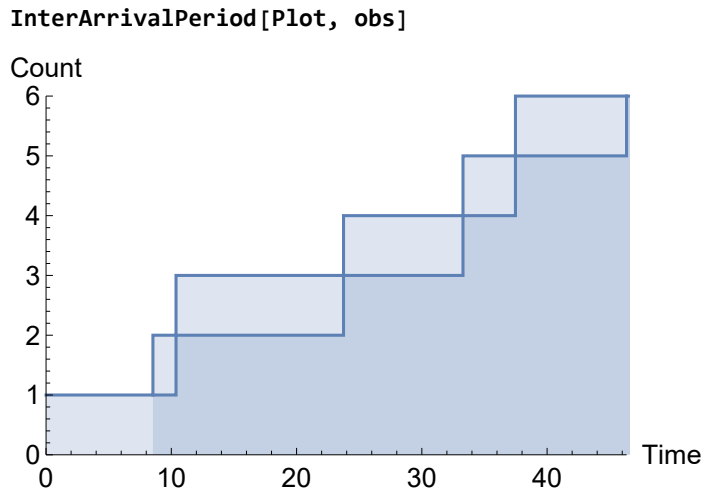
A long run

We must run the simulation for a long period to get a smoother overall view.

```
horizon = (3 mu) ^2
```

900

Let us also plot $n[t] = t / 10 = t / \mu = \lambda t$.

```
plabda = Plot[labda * t , {t, 0, horizon}, PlotRange → All,
    AxesLabel → {"Time t", "λ t"}, BaseStyle → {FontSize → 13}];

data = RandomFunction[PoissonProcess[labda], {0, horizon}];

p0 = ListStepPlot[data, Filling → Axis,
    AxesLabel → {"Time", "Count"}, BaseStyle → {FontSize → 13}];

Show[p0, plabda]
```
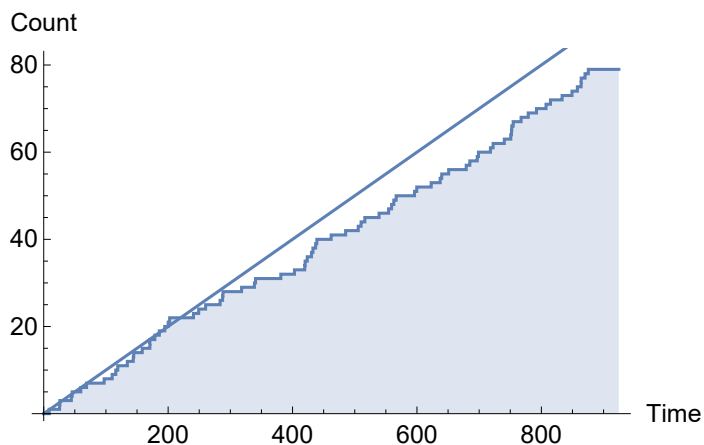
Count

Since this is random, we can only hope for roughly μ = MIAP = 900 / 90. The graph will tend to be almost linear, with the formula for the cumulated count $n[t] = t / 10 = t / \mu = \lambda t$. In the long run, in theory, the estimate of AR $\lambda = n / T$ is fair for large values. However, there can be random walks such that the estimate is off for long periods of time (and taking long to get back to the expected value).

```
iaps = obs = InterArrivalPeriod[Data, data]; (* now drops the last element *)
```

```
{N → (nn = Length[obs]), Time → (tt = Total[obs]), MIAP → tt / nn, AR → nn / tt}
```

{$N \to 79$, Time $\to 875.48$, MIAP $\to 11.082$, AR $\to 0.0902362$}

```
{Mean[obs], StandardDeviation[obs]}
```

{11.082, 9.23728}

The histogram of the interarrival periods generally looks like an exponential density.

```
Histogram[iaps, AxesLabel → {"Period\nLengths", "Count"},
 BaseStyle → {FontSize → 13}]
```
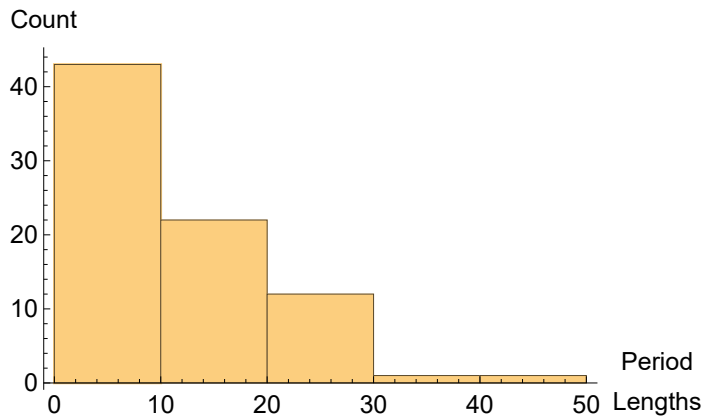

B.6. Poisson process (exponential IAP) versus exponential process

The *Poisson process on counts* (with exponentially distributed periods) should not be confused with an *exponential growth process on counts*. Also, SI(EY)A(CD) has exponential phases at the beginning and the end but not in the middle.

Comparison with the former longer run simulation

The Poisson count process is almost linear. If we want to approximate it by an exponential count process, then we know that this cannot work. At the end of the period we would get Exp[c T] = n, so that we have an estimate for the value of this growth coefficient c.

```
Clear[coef]; Exp[coef horizon] ⩵ nn
```

$e^{900\,coef} = 79$

```
coef = Log[nn] / horizon // N
```

0.00485494

```
p1 = Plot[Exp[coef t], {t, 0, horizon}];
```

```
Show[p0, p1, plabda]
```

Count
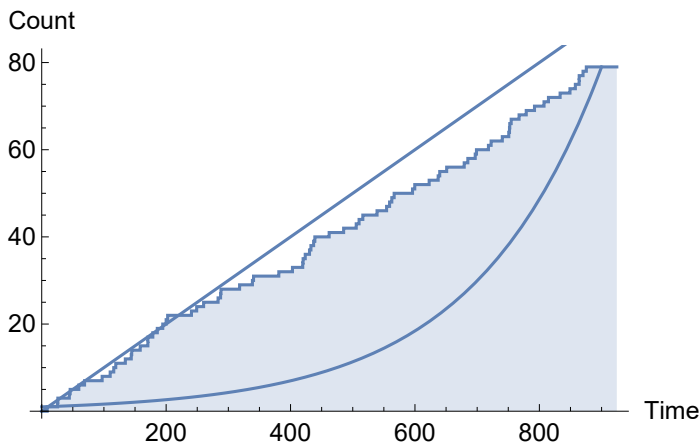
Normalising to another period

If we run Exp[λ t] till $t = 100$, then we get a sizeable number.

```
mu = 10; labda = 1 / mu;
```

```
{(num = Exp[labda * horizon]) , N[num] }
```

$\{e^{90}, 1.2204 \times 10^{39}\}$

We can somewhat normalise to the value at the horizon.

```
Plot[Exp[labda * t] / Exp[labda * horizon ], {t, 0, horizon}, PlotRange → All,
 AxesLabel → {"Time t", "E[labda t] / Last value"}, BaseStyle → {FontSize → 13}]
```

E[labda t] / Last value
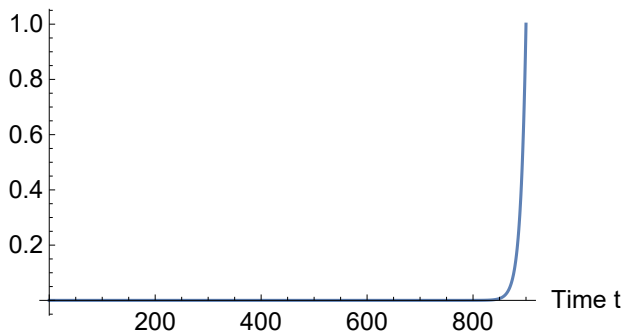
While we take μ as a mean acquittal period of 10 days, we might also make it 10 / 30 of a month. Then λ becomes 30 / 10 = 3 acquittals per month. We may now normalise to the value at the end of the month.

```
coef = 30;
Plot[Exp[coef labda * t] / Exp[coef labda], {t, 0, 1}, PlotRange → All,
 AxesLabel → {"Time t", "E[labda t] / Last value"}, BaseStyle → {FontSize → 13}]
```

E[labda t] / Last value
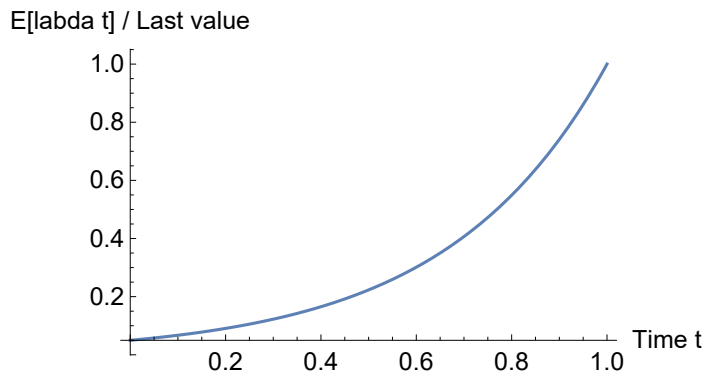
However, the main point is that we should not confuse the *exponential distribution for the interar-rival times* with an *exponential growth process for the number of cases*. The number of cases in the SI(EY)A(CD) models are generated by another process, though there are exponential (-like) phases.

B.7. Package on queueing

Colignatus (1995, 2020e) "The Economics Pack", has routines for queueing theory, based upon Hillier & Lieberman. See e.g. the online User Guide pdf.

```
Economics[Queue]
```

⌄ **Cool`Queue`**

AdjustedRates	MeanNQueue	MultipleServersContours	QueueSymbols
BerthOccupancyPlot	MeanNSystem	MultipleServersModel	*R*
FiniteSourceSS	MeanProcessTime	NBeingServed	SingleServer
L	MeanQueueWait	NQueue	SingleServerModel
Lq	MeanSacrificeRatio	NServers	ToQueueSymbols
Ls	MeanServiceRate	NSystem	*W*
MeanArrivalRate	MeanServiceTime	QueueBasics	Wq
MeanArrivalTime	MeanUtilisationRate	QueueDataAnalysis	λ
MeanNBeingServed	MultipleServers	QueueDataToGantt	ρ

Superficially there are ("SIR") compartments, namely the general population (S), the customers in the queue and at a service station (I), and the customers served (A). However, an effort at comparison can be rather distracting, since the questions differ about what is actually being modeled.

Appendix C. Recall Survival analysis

The Economics Pack, see Colignatus (1995, 2020e), has some packages about epidemiology and survival analysis. They are documented by notebooks in the Pack. The functionality may be seen as somewhat comparable to David Kleinbaum (2015) "ActiveEpi", but then for the environment of *Mathematica,* and then again extended for my purposes at the time. I much appreciated the teach-

ing by Kleinbaum in 2002-2004.

The intention was to document these packages in a separate book, like the other books ALOE, COTP and VTFD, but there was no time. If there is time, I intend to see whether the following packages could still be used for an educational book.

Over time, WRI has included much survival analysis by default within *Mathematica,* see *https://reference.wolfram.com/language/guide/SurvivalAnalysis.html.* The following does not yet take advantage of this wealth. Indeed, there is a learning curve, and it would not be unwise to follow it, to first find out what already is available in *Mathematica* before one writes a routine oneself. It is likely that there will be an increasing overlap of what *Mathematica* and these packages in The Economics Pack provide. In the mean time I myself regard them as having been useful at one time, and perhaps I will find time to see whether eventually there will be such book that uses them, or not.

C.1. Load packages

```
ToFileName[{$TEPDirectory[], "Applications",
  "Economics", "Documentation", "English", "LifeSciences"}]
```

C:\Users\Eigenaar\AppData\Roaming\Mathematica\Applications\Economics\Documentation\English\
    LifeSciences\

```
Economics[Survival]
```

Symbol

SurvivalPackages[] gives an overview of the packages in survival statistics and epidemiology.

These packages aren't supported with palettes yet.

Mathematica's I is the imaginary number (complex operator), but the String "I" may be used for Incidence

⌄

? Sp

Symbol

Sp is a Symbol for specificity

   *** AddedUsage by Cool`Survival`SIA`: ***

Sp[t] is the proportion of S susceptible units at time t, see Economics[Survival`SIA]

Sp is a Symbol for specificity, see Economics[Survival`Epidemiology]

⌄

C.2. Some spare examples

Example 1

`Economics[Survival`Common]`

⌄ **Cool`Survival`Common`**

Age	**DeathDis**	**HazardRate**	**L$**	**PeriodHazard**
AgeGroupLabels	**DeathOC**	**h$**	*m*	**Population**
Alive	**DegreeOfLiving**	**Intervention**	**Male**	**PopulationForm**
b	**Disability**	*l*	**MidSurvival**	*q*
Birth	**DiscLifeExp**	*L*	**Mortality**	**QALE**
b$	**DiscQALE**	**LastLifeExp**	**MortalityQuotient**	**QALY**
Cohort	**d$**	**LExample**	**MortalityRate**	**QualityOfLife**
CohortEnd	*e*	**LifeExpectancy**	**m$**	**q$**
CohortNPeriod	**e$**	**LifeTable**	*n*	**SetSurvival**
CohortPeriod	**Female**	**LifeTableSymbols**	**n$**	**ToLifeTableSymbols**
d	*h*	**LifeYears**	**Onset**	
DALE	*H*	**Living**	**Options$Cohort**	
DALY	**Hazard**	**l$**	**Options$Intervention**	

`ListOfSymbols[LifeTable]`

$$
\begin{pmatrix}
\text{Birth} & b \\
\text{Death} & d \\
\text{Hazard} & H \\
\text{LifeExpectancy} & e \\
\text{LifeYears} & L \\
\text{Living} & l \\
\text{MortalityQuotient} & q \\
\text{MortalityRate} & m \\
\text{Period} & n \\
\text{PeriodHazard} & h \\
\text{Survival} & S
\end{pmatrix}
$$

Example 2

SetSurvival["Explain"]

The formulas for the SetSurvival model are:

H[t]   = Integrate[μ[τ], {τ, 0, t}]

S[t]   = – Log[H[t]]

d[t, n] = S[t] – S[t+n]

q[t, n] = d[t, n] / S[t] = 1 – S[t+n]/S[t]

h[t, n] = H[t+n] – H[t] = – Log[1 – q[t, n]]

L[t, n] = Integrate[S[τ], {τ, t, t+n}]

e[t]   = L[t, ∞] / Survival[t]

Default option is Function –> Simplify. Otherwise use Identity

Mortality[t_] = .5

0.5

SetSurvival[t]

{Mortality → 0.5, Survival → $e^{-0.5\,t}$, LifeExpectancy → 2.}

The function Mortality in this package may also be called MortalityRate.

MortalityRate[t]

0.5

Plot[Survival[t], {t, 0, 10}]
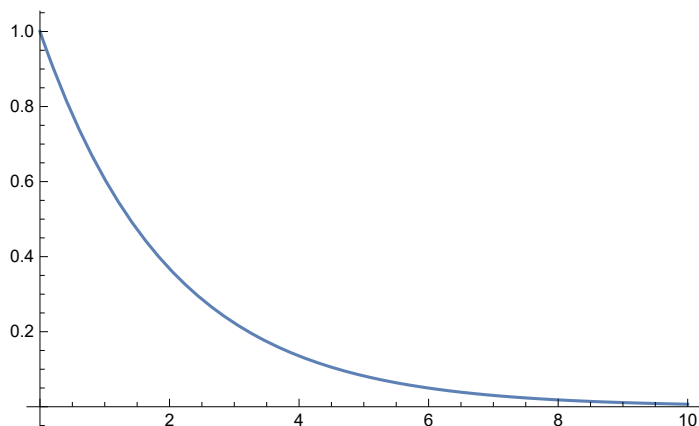
? SetSurvival

Symbol

SetSurvival[t, opts] sets continuous function Survival[t] and its implied functions Death, MortalityQuotient,

MortalityRate, LifeYears, Hazard, PeriodHazard, LifeExpectancy, all as functions of a (predefined)

Mortality[t]. For example, first defining Mortality[t_] = μ gives exponential death, with constant

life expectancy 1 / μ. The option Clear –> True (default False) first clears these functions (use

this in doubt only: clear also concerns lists Survival[] etc.!). The option Function may be used

to remove possible "If" clauses and such, e.g. use (Re[μ] > 0) –> True. Default Function –>

Simplify simplifies. Use Identity if you want to trace how the expressions relate to each other.

Note: For display purposes, the routine runs in the Global` context and does Clear[tt, nn, T,

τ]. Note: Survival[0] = 1; for Living[t] values: multiply all results by Birth[]. Output only shows

the list {Mortality[t], Survival[t], LifeExpectancy[t]} though the other functions are available.

SetSurvival[Clear] clears the parameters; advisable to do when you have used SetSurvival[t] and then

want to use LifeTableEquations[]. Note: the equations use t, n and τ in the global context, and

these will be cleared too. Note, this is also called first when SetSurvival[…, Clear –> True] is used.

See SetSurvival["Explain"] for the formula overview

⌄

Appendix D. Example of *not* using the package

D.1. Needless learning curve for experienced users

This present notebook and package have the purposes of *improved didactics, accessibility and userfriendliness*. The target is a novice level of entry. The notebook and package provide a path towards better understanding of modeling of infections with SI(EY)A(CD) and computable writing on this. A major task is learning how to use the package itself.

However, when one already *understands such modeling and programming,* then this "userfriendliness" could very well be a barrier, with needless extra learning of what already is known.

The following are examples of quick results, adapted from Antonov (2020) on Stack Exchange. See Smith & Moore (2001, 2004) on the modeling. This was actually my starting point. It were the other considerations on didactics that caused this notebook and package.

Smith, David, and Lang Moore (2001, 2004), "https://www.maa.org/press/periodicals/loci/joma/the-sir-model-for-spread-of-disease-the-differential-equation-model

Antonov, Anton (2019), https://mathematica.stackexchange.com/questions/208357/solving-and-plotting-an-sir-epidemiology-model

D.2. A model

```
ClearAll[Sus, Inf, Recov, beta, gamma, tmax, totN, suscol, infcol, reccol, soln, p2, p4]
```

```
suscol = Cyan;
infcol = Blue;
reccol = Green;
```

```
tmax = 20;
soln = First@NDSolve[{
    Sus'[t] == (-beta) * Sus[t] * Inf[t],
    Inf'[t] == beta * Sus[t] * Inf[t] - gamma * Inf[t],
    Recov'[t] == gamma * Inf[t],
    Sus[0] == 761 / 762,
    Inf[0] == 1 / 762,
    Recov[0] == 0, beta == 762 × 0.00218, gamma == 0.44036},
   {Sus, Inf, Recov}, {t, 0, tmax}]
```

{Sus → InterpolatingFunction[➕ 〽 Domain: (0. 20.) Output: scalar],

Inf → InterpolatingFunction[➕ 〽 Domain: (0. 20.) Output: scalar],

Recov → InterpolatingFunction[➕ ⌡ Domain: (0. 20.) Output: scalar]}

```
Plot[{Sus[t] /. soln, Inf[t] /. soln, Recov[t] /. soln},
 {t, 0, tmax}, PlotRange → All, PlotLegends → {"Sus", "Inf", "Recov"},
 ImageSize → Medium, PlotStyle → {suscol, infcol, reccol}]
```
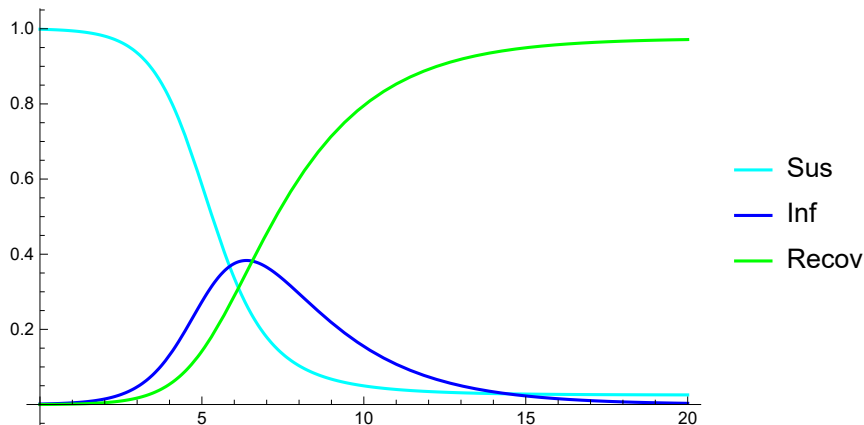

D.3. A model with an intervention

```
ClearAll[Sus, Inf, Recov, beta, gamma, tmax]
```

```
tmax = 150;
totN = 17 × 10^6;
gamma = 1 / 12;
soln = First@NDSolve[{
    Sus'[t] == (-beta[t]) * Sus[t] * Inf[t] / totN,
    Inf'[t] == beta[t] * Sus[t] * Inf[t] / totN - gamma * Inf[t],
    Recov'[t] == gamma * Inf[t],
    Sus[0] == totN - 50,
    Inf[0] == 50,
    Recov[0] == 0,
    beta[0] == 4 gamma,
    WhenEvent[t > 30, beta[t] → 2 gamma] },
   {Sus, Inf, Recov, beta}, {t, 0, tmax}, DiscreteVariables → {beta}]
```

{Sus → InterpolatingFunction[⊞ ◸ Domain: (0. 150.) Output: scalar],
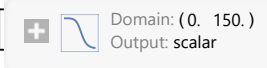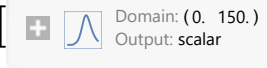
 Inf → InterpolatingFunction[⊞ ◹ Domain: (0. 150.) Output: scalar],

 Recov → InterpolatingFunction[⊞ ◿ Domain: (0. 150.) Output: scalar],

 beta → InterpolatingFunction[⊞ ◺ Domain: (0. 150.) Output: scalar]}
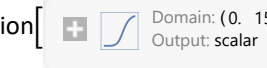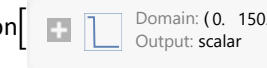
```
sus[t_] = Sus[t] /. soln;
inf[t_] = Inf[t] /. soln;
rec[t_] = Recov[t] /. soln;

p2 = Plot[{sus[t], inf[t], rec[t]}, {t, 0, tmax},
  PlotRange → All, PlotLegends → {"Sus SD", "Inf SD", "Rec SD"},
  ImageSize → Medium, PlotStyle → {suscol, infcol, reccol}]
```
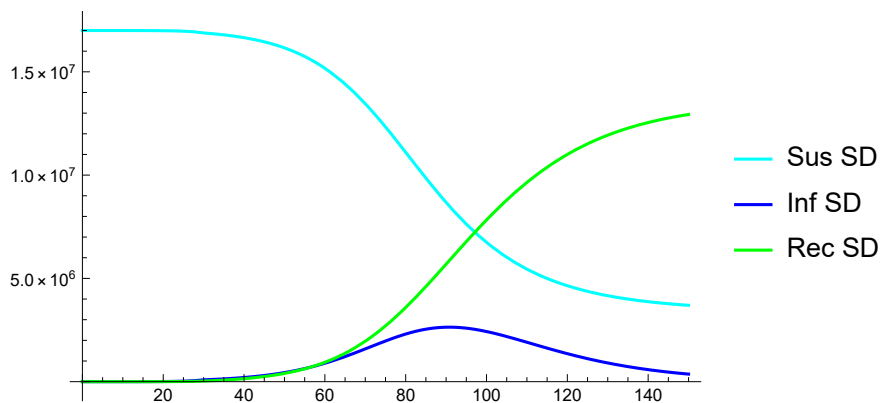
```mathematica
Plot[{inf[t], rec[t]}, {t, 0, 60}, PlotRange → All, PlotLegends → {"Inf SD", "Rec SD"},
 ImageSize → Medium,  PlotStyle → {infcol, reccol }] (* detail only *)
```
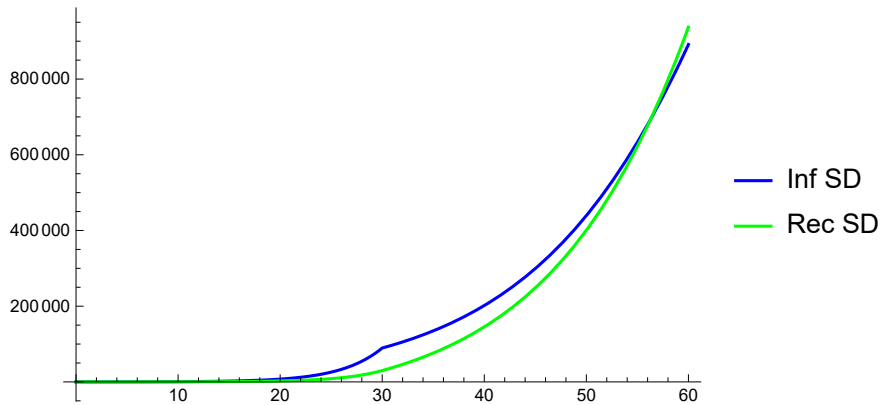


```mathematica
tmax = 150;
totN = 17 × 10^6;
gamma = 1 / 12;
soln = First@NDSolve[{
    Sus'[t] == (-beta) * Sus[t] * Inf[t] / totN,
    Inf'[t] == beta * Sus[t] * Inf[t] / totN - gamma * Inf[t],
    Recov'[t] == gamma * Inf[t],
    Sus[0] == totN - 50,
    Inf[0] == 50,
    Recov[0] == 0,
    beta == 4 gamma }, {Sus, Inf, Recov}, {t, 0, tmax}]
```

$\{$Sus → InterpolatingFunction[▦ Domain: (0. 150.) Output: scalar],

Inf → InterpolatingFunction[▦ Domain: (0. 150.) Output: scalar],

Recov → InterpolatingFunction[▦ Domain: (0. 150.) Output: scalar]$\}$

```mathematica
sus[t_] = Sus[t] /. soln;
inf[t_] = Inf[t] /. soln;
rec[t_] = Recov[t] /. soln;
```

```
p4 = Plot[{sus[t], inf[t], rec[t]}, {t, 0, tmax}, PlotRange → All,
   PlotLegends → {"Sus", "Inf", "Recov"}, ImageSize → Medium,
   PlotStyle → {{Dashed, suscol}, {Dashed, infcol}, {Dashed, reccol}}]
```
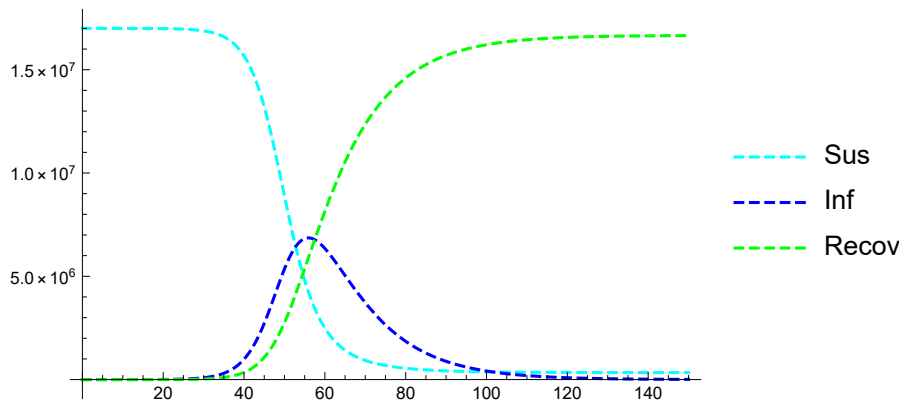
```
Show[p2, p4]
```
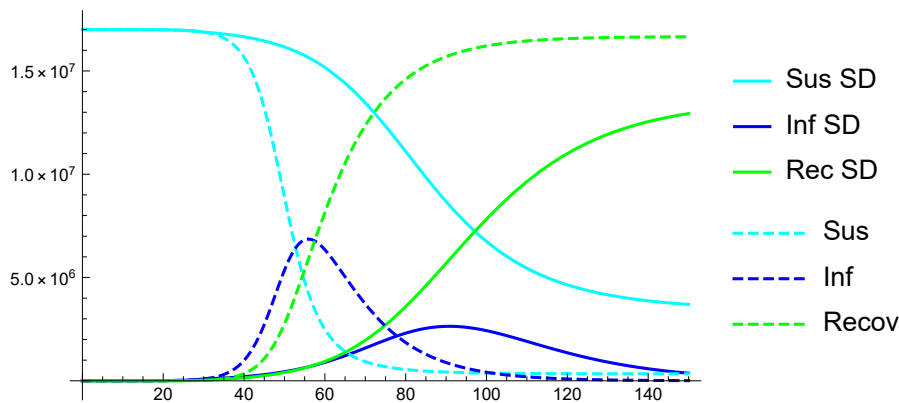
```
ClearAll[Sus, Inf, Recov, beta, gamma, tmax, totN, suscol, infcol, reccol, soln, p2, p4]
```

Appendix E. Life table and SARS-CoV-2

E.1. Introduction

While there is an abundance of books and software for the life table, the following gives our own presentation, namely for the purpose of application within the environment of the SI(EY)A(CD) models and within *Mathematica.*

Colignatus (1995, 2020e) & (2002-2020) *The Economics Pack* has some notebooks and packages for the life sciences. The routines have been used in particular settings but now have a didactic purpose only. The routines can skip details that are important within official medical or actuarial application, and thus should not be mistaken for such. For example, official application for life tables requires accuracy and then will tend to use 1-year age groups, with (institutionally selected) rules for the birth year and final age group. For other applications, 5-year age groups are often used. It depends upon the setting how such are handled, and for the present purposes of the Pack it suffices to reduce the application to didactics.

The following reproduces the *Introduction* of the 2004 *LifeTableConstruction.nb*, and proceeds with

some elements of the 2020 *LifeTableExample.nb* that refers to the SARS-CoV-2 pandemic. The routine packages have been loaded in Chapter 3 above.

E.2. Introduction of LifeTableConstruction.nb

E.2.1. Cohort, age groups and survival

The best introduction to the life table is to simply show it.

Consider the World Health Organisation (WHO) World Standard Population for 2000-2025, based upon the forecast by the UN Population Division 1998. The population is assumed to be stationary so that it also represents a cohort. The following cohort object gives age groups i = 1, ..., 18. For each age group we have a dataset {i, age-range label for i, $p(i)$}}. The age-range label is a text expression such as "0-4", and $p(i)$ is the population in that range. The total population sums to 100,000.

```
c = Cohort["WHO 2000", 18, 5]
```

$$\begin{pmatrix} 1 & 0-4 & 8800 \\ 2 & 5-9 & 8700 \\ 3 & 10-14 & 8600 \\ 4 & 15-19 & 8500 \\ 5 & 20-24 & 8200 \\ 6 & 25-29 & 7900 \\ 7 & 30-34 & 7600 \\ 8 & 35-39 & 7200 \\ 9 & 40-44 & 6600 \\ 10 & 45-49 & 6000 \\ 11 & 50-54 & 5400 \\ 12 & 55-59 & 4600 \\ 13 & 60-64 & 3700 \\ 14 & 65-69 & 3000 \\ 15 & 70-74 & 2200 \\ 16 & 75-79 & 1500 \\ 17 & 80-84 & 900 \\ 18 & 85+ & 600 \end{pmatrix}$$

The WHO 2000 Standard Population gives a snapshot of a current population, consisting of different cohorts born at different moments in time. Generally, this means that each cohort will have a separate life table of itself. However, when we assume a stationary population, in such manner that every 5 years 8800 are born, and such that they age in above pattern, then the last column can be interpreted as survival that provides a life table. The survival factors follow from dividing the last column by its first element.

The following plots these survival factors. The plot contains a reference survival, given by a dashed line, which by default is the European Standard Population interpreted as survival again.

```
{c1, c2, c3} = Transpose[c];
```

```
SurvivalListPlot[c3, 5]
```
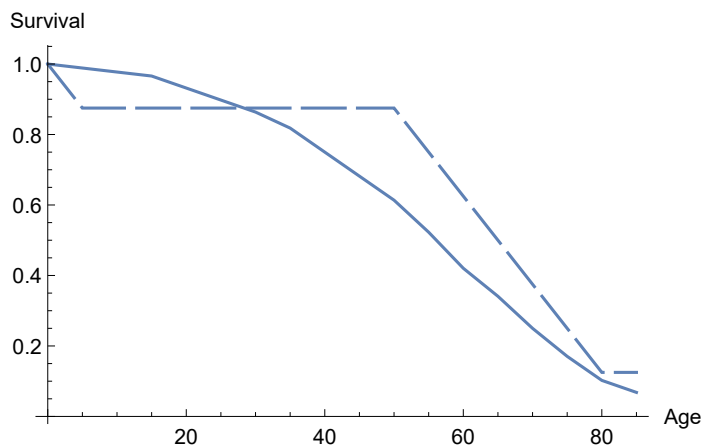

The survival function $S(t)$ gives the probability of surviving from 0 till period t. Survival is generally measured discretely as S_i at $i = 0, 1, 2, 3, ..., T$ periods, often periods of 5 years, which fits the choice of the age groups. The probabilities in the periods beyond T are less reliable, so that they are taken together in a final 'open' period $T+$.

The life years can be found as the area under the survival function. The life expectancy consists of the life years that can be expected to be lived from moment t till the maximal possible lifetime - normalised by current survival.

E.2.2. Life table calculation and display

There are two main routines:

(1) FromCohort to calculate the (implied) life table and store results under a label of choice.

(2) LifeTable to display the output. There are two tables: (2a) the basic table, and (2b) a "results" table with a Disability Adjusted Life Expectancy (DALE) and a Quality of Life Adjusted Life Expectancy (QALE). Each table has a column where its life expectancy is discounted by the rate of discount that has been given in the options. Each table also has a column where the UnitSqrt criterion is applied (see below).

First reset the options to the default values.

```
ResetOptions[Cohort]
```

{Birth → 100 000, BirthLifeYearFactor → False, CohortEnd → 105, CohortNPeriod → 100,
   CohortPeriod → 1, ContinuousDiscounting → True, Function → Min, Label → ,
   LastLifeExp → Automatic, Method → Line, RateOfDiscount → 0.03, Round → False, Transpose → True}

Let us include criteria for *disability* and *quality of life* in the calculation. The values are fictitious, and we now take qol = Sqrt[1 - disability] to include a degree of accommodation to the disability. The routine keeps track of the label (name) of a cohort, and it will be possible to compare different results by reference to those labels. Current cohort variables can also be accessed without label.

```
StoreCohort[] = {};
```

```
FromCohort[Cohort["WHO 2000", 18, 5],
QualityOfLife["Example", 18, 5],
Disability["Example", 18, 5], Round → True]
```

{WHO 2000}

The following lists have been created now: Cohort[], AgeGroupLabels[], Survival[], Living[], Death[], LifeYears[], LifeExpectancy[], MortalityRate[], MortalityQuotient[], PeriodHazard[], Disability[], DALY[], DALE[], QualityOfLife[], QALY[], QALE[], DiscLifeExp[], UnitSqrtLE[], QAAge[], UnitSqrtQALE[].

The routine LifeTable gives a TableForm presentation in a standard layout. First consider the basic table:

```
LifeTable[TableAlignments → ".", TableSpacing → {1, 1}] /.
 {MortalityRate → ".MortalityRate", "LifeExp" → "Life.Exp",
  "DiscLE" → "Disc.LE", UnitSqrtLE → "Unit.SqrtLE"}
```

	Living	q	Death	LifeYears	.MortalityRate	Life.Exp	Disc.LE	Unit.SqrtLE
0–4	100 000.	0.01136	1136.	497 160.	0.00228498	54.8	27.1	1.
5–9	98 864.	0.0115006	1137.	491 478.	0.00231343	50.4	26.2	0.953833
10–14	97 727.	0.0116242	1136.	485 795.	0.00233843	46.	25.1	0.906307
15–19	96 591.	0.0352931	3409.	474 433.	0.00718543	41.5	23.9	0.857037
20–24	93 182.	0.0365843	3409.	457 388.	0.0074532	37.9	22.8	0.809158
25–29	89 773.	0.0379736	3409.	440 343.	0.0077417	34.3	21.5	0.760405
30–34	86 364.	0.0526377	4546.	420 455.	0.0108121	30.5	20.1	0.710174
35–39	81 818.	0.0833313	6818.	392 045.	0.0173909	27.1	18.6	0.660486
40–44	75 000.	0.0909067	6818.	357 955.	0.0190471	24.3	17.3	0.614891
45–49	68 182.	0.0999971	6818.	323 865.	0.021052	21.5	15.9	0.568601
50–54	61 364.	0.148149	9091.	284 093.	0.0320001	18.6	14.3	0.520819
55–59	52 273.	0.195665	10 228.	235 795.	0.0433767	16.4	13.	0.479406
60–64	42 045.	0.189178	7954.	190 340.	0.0417884	14.8	12.	0.444783
65–69	34 091.	0.266669	9091.	147 728.	0.061539	12.7	10.6	0.403842
70–74	25 000.	0.3182	7955.	105 113.	0.0756808	11.4	9.7	0.373714
75–79	17 045.	0.4	6818.	68 180.	0.1	10.5	9.	0.350438
80–84	10 227.	0.333333	3409.	42 612.5	0.08	10.8	9.3	0.345349
85+	6818.	1.	6818.	68 180.	0.1	10.	8.7	0.324443

The concepts and symbols are:

```
ListOfSymbols[LifeTable]
```

$$\begin{pmatrix} \text{Birth} & b \\ \text{Death} & d \\ \text{Hazard} & H \\ \text{LifeExpectancy} & e \\ \text{LifeYears} & L \\ \text{Living} & l \\ \text{MortalityQuotient} & q \\ \text{MortalityRate} & m \\ \text{Period} & n \\ \text{PeriodHazard} & h \\ \text{Survival} & S \end{pmatrix}$$

```
LifeTableEquations[] /. ToLifeTableSymbols // MatrixForm
```

$$
\begin{pmatrix}
l(t) = b(0)\, S(t) \\
d(t, n) = l(t) - l(n + t) \\
q(t, n) = \frac{d(t,n)}{l(t)} \\
h(t, n) = -\log(1 - q(t, n)) \\
L(t, n) = \int_t^{n+t} l(\tau)\, d\tau \\
L(\text{Line})(t, n) = n\left(l(t) - \frac{1}{2}\, d(t, n)\right) \\
m(t, n) = \frac{d(t,n)}{L(t,n)} \\
e(t) = \frac{L(t,\infty)}{l(t)}
\end{pmatrix}
$$

LifeTable[Results] prints the results of adjustment for disability and quality of life.

Using DALY[] = LifeYears[] (1 - Disability[]) - not shown - we can compute the DALE[].

Using QALY[] = LifeYears[] QualityOfLife[] - not shown - we can compute the QALE[].

The table includes the input values for Disability[] and QualityOfLife[] (QoL).

The QoL adjustment seems more comprehensive in utility evaluation of the overall state of health, and thus is subjected to additional calculations. First, there is the discounting, with default discount rate 3% in Options[Cohort]. Secondly, the standard table, see above, includes UnitSqrtLE[] = UnitSqrt[age, 0, le], which assumes an intervention from a risk of instant death with recovery towards full life expectancy. For application to QALE[], the calendar age forms an index, and the QALYs are cumulated over that index. UnitsSqrtQALE[] = UnitSqrt[QAAge[], 0, QALE[]]. QAAge[] cumulates the QALYs = 5 QoL for a person.

```
LifeTable[Results]
```

	Living	LifeExp	Disability	DALE	QoL	QALE	DiscQALE	QAAge	UnitSqrtQALE
0–4	100 000.	54.8	0.046	51.1	0.977	52.8	26.7	0.	1.
5–9	98 864.	50.4	0.036	46.9	0.982	48.5	25.8	4.9	0.953187
10–14	97 727.	46.	0.034	42.6	0.983	44.2	24.7	9.8	0.904696
15–19	96 591.	41.5	0.04	38.2	0.98	39.7	23.4	14.7	0.854305
20–24	93 182.	37.9	0.04	34.7	0.98	36.2	22.2	19.6	0.805387
25–29	89 773.	34.3	0.044	31.2	0.978	32.6	20.9	24.5	0.755456
30–34	86 364.	30.5	0.05	27.5	0.975	28.9	19.4	29.4	0.703975
35–39	81 818.	27.1	0.05	24.2	0.975	25.5	17.9	34.3	0.653002
40–44	75 000.	24.3	0.054	21.4	0.973	22.7	16.5	39.1	0.605827
45–49	68 182.	21.5	0.06	18.6	0.97	19.9	15.	44.	0.557639
50–54	61 364.	18.6	0.068	15.7	0.965	16.9	13.3	48.9	0.507516
55–59	52 273.	16.4	0.084	13.4	0.957	14.6	11.9	53.7	0.463027
60–64	42 045.	14.8	0.102	11.5	0.948	12.8	10.7	58.5	0.424421
65–69	34 091.	12.7	0.132	9.1	0.932	10.6	9.1	63.2	0.378235
70–74	25 000.	11.4	0.19	7.3	0.9	8.9	7.8	67.9	0.340218
75–79	17 045.	10.5	0.27	5.7	0.854	7.5	6.7	72.4	0.306072
80–84	10 227.	10.8	0.36	4.7	0.8	6.8	6.1	76.6	0.284906
85+	6818.	10.	0.695	3.	0.516	5.2	4.8	80.6	0.245193

The LifeGain measures have been introduced for comparison. The following compares the first and last values. In terms of life expectancy, a baby or toddler is 5.48 more valuable than a person of 85+. The UnitSqrt of this life expectancy, reduces the value ratio to 3.08. The more elaborate discounted QALEs turns the ratio to 5.59 but the UnitSqrt equalises to 4.08 again.

```
lab = "WHO 2000";
```

```
First[#] / Last[#] & /@ (#[lab] & /@{LifeExpectancy, DiscLifeExp, UnitSqrtLE}) // N
```
{5.48296, 3.1349, 3.08221}

```
First[#] / Last[#] & /@ (#[lab] & /@{DALE, QALE, DiscQALE, UnitSqrtQALE})
```
{16.7564, 10.2439, 5.58865, 4.07843}

Comparison of the first and last elements can be too simple. When age groups are compared, like <60 and 60+, then we would have to weigh the scores by appropriate weights. For life expectancy the weights would be TLE = Living & LifeExp (since LifeExp has been calculated as TLE / Living), and for the UnitSqrt, that regards each life as 100%, the population weights (Living) are appropriate.

E.3. Effect of SARS-CoV-2 using LifeTableExample.nb

E.3.1. Read csv file with the published table

Before we can run a variant for 2020, we first need a base table.

The following uses the 2018 Dutch life table, published by CBS Statistics Netherlands, with 1-year age groups. It is a period table, i.e. with mortality quotients of the observation period, and it is not a "proper" cohort table with developed forecasts for the age groups.

https://opendata.cbs.nl/statline/portal.html?_la=nl&_catalog=CBS&tableId=37360ned&_theme=76

https://opendata.cbs.nl/statline/#/CBS/nl/dataset/37360ned/table?ts=1597566945306

The CBS Statistics Netherlands csv file has been reworked a bit and included in The Economics Pack.

```
dir = ToFileName[{$TEPDirectory[], "Applications", "Economics", "Data"}]
```
C:\Users\Eigenaar\AppData\Roaming\Mathematica\Applications\Economics\Data\

```
lis = CSVToMatrix[dir, "CBS-Dutch-Life-Expectancy-2018.csv" ];
Take[lis, 10] /. "Women" → "Wmn" (* fit on page *)
```

2018 age	All qx	All lx	All dx	All ex	Men qx	Men lx	Men dx	Men ex	Wmn qx	Wmn lx	Wmn dx	Wmn ex
0	0.0031	100 000	310	81.78	0.00346	100 000	346	80.16	0.00272	100 000	272	83.33
1	0.00043	99 690	43	81.53	0.00046	99 654	46	79.93	0.00041	99 728	41	83.06
2	0.00018	99 647	18	80.57	0.00022	99 608	22	78.97	0.00014	99 687	14	82.09
3	0.0001	99 629	10	79.58	0.00009	99 586	9	77.99	0.00011	99 673	11	81.11
4	0.00013	99 619	13	78.59	0.00014	99 577	14	76.99	0.00013	99 662	13	80.11
5	0.00007	99 606	7	77.6	0.00004	99 563	4	76.01	0.00009	99 649	9	79.12
6	0.00011	99 599	11	76.6	0.0001	99 559	10	75.01	0.00011	99 640	11	78.13
7	0.00008	99 588	8	75.61	0.00007	99 549	7	74.02	0.00009	99 629	9	77.14

E.3.2. Variables given in the csv file

```
Clear[vars, age, qx, lx, dx, ex]
```

```
vars = Prepend[
   Thread[#1[{all, men, women}]] & /@ {qx, lx, dx, ex} // Transpose // Flatten, age];
```

```
Evaluate[vars] = Transpose[Drop[lis, 2]]
```

0	1	2	3	4	5	6	7	8	9	10
0.0031	0.00043	0.00018	0.0001	0.00013	0.00007	0.00011	0.00008	0.00005	0.00007	0.00005
100 000	99 690	99 647	99 629	99 619	99 606	99 599	99 588	99 580	99 575	99 568
310	43	18	10	13	7	11	8	5	7	5
81.78	81.53	80.57	79.58	78.59	77.6	76.6	75.61	74.62	73.62	72.63
0.00346	0.00046	0.00022	0.00009	0.00014	0.00004	0.0001	0.00007	0.00008	0.00006	0.00006
100 000	99 654	99 608	99 586	99 577	99 563	99 559	99 549	99 542	99 534	99 528
346	46	22	9	14	4	10	7	8	6	6
80.16	79.93	78.97	77.99	76.99	76.01	75.01	74.02	73.02	72.03	71.03
0.00272	0.00041	0.00014	0.00011	0.00013	0.00009	0.00011	0.00009	0.00001	0.00009	0.00003
100 000	99 728	99 687	99 673	99 662	99 649	99 640	99 629	99 620	99 619	99 610
272	41	14	11	13	9	11	9	1	9	3
83.33	83.06	82.09	81.11	80.11	79.12	78.13	77.14	76.15	75.15	74.15

```
age // FullForm
```

List[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27,

   28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52,

   53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76,

   77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, "plus99"]

The life table with a birth cohort of 100,000 can also be seen as a stationary population. It so happens that the implied population in this present table is about half the actual Dutch population of 17.4 million in 2020. The number of deaths caused by SARS-CoV-2 on this table may be about half of actual deaths (in estimation).

```
lx[all] // Add
```

8 273 401

E.3.3. Values for case = all

The routine FromCohort (re-) uses the cohort Living[] vector, and calculates the other life table variables from this. The routine LifeTable only displays.

Let us select the case for the whole Dutch population. We can (re-) construct the (implied) variables for the LifeTable[] routine without calling FromCohort[…]. The call of LifeTable[] then shows the official data without interference by our own choice of methods.

```
case = all;
```

```
{MortalityQuotient[], Living[], Death[], LifeExpectancy[]} =
 (#[case] & /@ {qx, lx, dx, ex})
```

0.0031	0.00043	0.00018	0.0001	0.00013	0.00007	0.00011	0.00008	0.00005	0.00007	0.00005
100 000	99 690	99 647	99 629	99 619	99 606	99 599	99 588	99 580	99 575	99 568
310	43	18	10	13	7	11	8	5	7	5
81.78	81.53	80.57	79.58	78.59	77.6	76.6	75.61	74.62	73.62	72.63

```
CohortPeriod[] = 1;
```

```
AgeGroupLabels[] = age /. "plus99" → "99+";
```

```
Birth[] = First[Living[]];
```

```
Survival[] = Living[] / Birth[];
```

```
le = LifeExpectancy[]; (*for later comparison*)
```

```
tle = LifeYearsSum[] = Living[] × LifeExpectancy[];
```

```
LifeYears[] = Append[- Differences[tle], Last[tle]];
```

```
MortalityRate[] = Death[] / LifeYears[];
```

```
DiscLifeExp[] = ContinuousPV[1, #, .03] & /@ LifeExpectancy[];
```

```
UnitSqrtLE[] = UnitSqrtLE[0, age /. "plus99" → 99];
```

PM. In the official data, the deaths do not add up to the cohort birth. The mortality in the 99+ group should be larger than stated.

```
totd = Death[] // Add
```

98 395

PM. The average age of death is another estimate for the life expectancy at birth.

```
Add[Death[] (age /. "plus99" -> 99.)] / totd
```

81.4524

CBS Statistics Netherlands has a particular method to account for the life years of the birth cohort. See the Lexis Plot for the explanation. We can recover a "birth life-year factor", so that we can call the routine FromCohort with this factor, and then reproduce the official outcome (without knowing what their method precisely is).

```
blyf = BirthLifeYearFactor[Living[], LifeExpectancy[]]
```

0.503523

```
StoreCohort["Holland 2018 all"]
```

{WHO 2000, Holland 2018 all}

E.3.4. Reproduced life table for All

This abridged table selects the values of Living, LifeYearsSum, (discounted) LifeExpectancy and UnitSqrtLE at the 5 year age intervals, so that there remains consistency in LifeExpectancy = LifeYearsSum / Living. The values of MortalityQuotient, LifeYears and MortalityRate are derived from those. The consequence of this method (merely taking lines for the key variables) is that the relation between the survival curve and the life expectancy is more complex than the linear or exponential assumptions of survival analysis. When we compare a base and a variant (for example move from a period to a cohort life table with a forecast for the future), then we rather should use tables that have a uniform method for both, e.g. linear or exponential methods for the relation between survival and life expectancy, so that differences in outcomes only derive from the variant and not from the different manner of how the base has been calculated.

```
Abridge["Holland 2018 all", "Holland 2018 abridged", AgeGroupLabels[21]]
```

{WHO 2000, Holland 2018 all, Holland 2018 abridged}

```
LifeTable[]
```

	Living	q	Death	LifeYears	MortalityRate	LifeExp	DiscLE	UnitSqrtL
0–4	100 000.	0.00394	394.	448 574.	0.000878338	81.8	30.8	1.
5–9	99 606.	0.000381503	38.	497 802.	0.0000763356	77.6	30.4	0.969261
10–14	99 568.	0.000311345	31.	497 946.	0.0000622558	72.6	29.9	0.937539
15–19	99 537.	0.000924279	92.	497 482.	0.000184931	67.6	29.3	0.904716
20–24	99 445.	0.00133742	133.	496 956.	0.00026763	62.7	28.5	0.870742
25–29	99 312.	0.0016715	166.	496 400.	0.000334408	57.8	27.7	0.835483
30–34	99 146.	0.00212817	211.	494 950.	0.000426306	52.9	26.7	0.798768
35–39	98 935.	0.00289079	286.	494 146.	0.000578777	48.	25.6	0.760436
40–44	98 649.	0.00443998	438.	492 264.	0.000889767	43.1	24.4	0.720256
45–49	98 211.	0.00733115	720.	489 683.	0.00147034	38.3	22.9	0.678074
50–54	97 491.	0.0121037	1180.	484 558.	0.00243521	33.6	21.3	0.633741
55–59	96 311.	0.0210672	2029.	476 389.	0.00425913	28.9	19.4	0.587171
60–64	94 282.	0.0347999	3281.	464 442.	0.0070644	24.5	17.4	0.53854
65–69	91 001.	0.0537247	4889.	442 785.	0.0110415	20.3	15.3	0.487744
70–74	86 112.	0.0867823	7473.	413 561.	0.0180699	16.3	12.9	0.434599
75–79	78 639.	0.141151	11 100.	367 355.	0.030216	12.6	10.5	0.379128
80–84	67 539.	0.245399	16 574.	298 572.	0.0555109	9.2	8.1	0.321465
85–89	50 965.	0.430433	21 937.	201 059.	0.109107	6.4	5.8	0.263846
90–94	29 028.	0.65268	18 946.	94 546.7	0.200388	4.2	4.	0.212112
95–99	10 082.	0.747173	7533.	22 924.3	0.328604	2.8	2.7	0.170082
100+	2549.	1.	2549.	5607.8	0.454545	2.2	2.1	0.147442

E.3.5. All cause survival

We calculated Survival[] = Living[] / First[Living[]] above.

```
porg = PlotLine[Survival["Holland 2018 all"]]
```
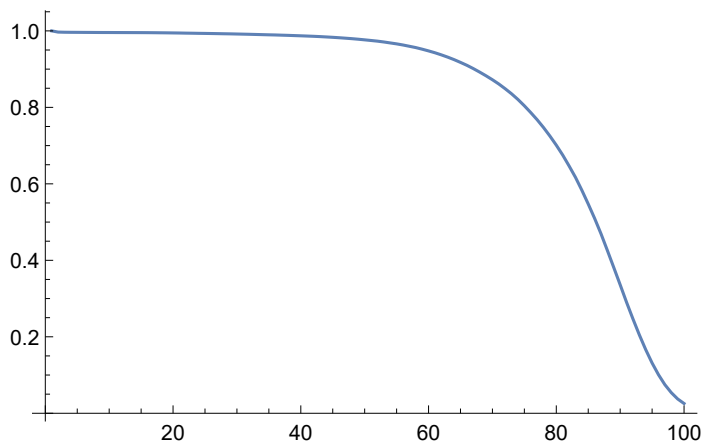

E.3.6. Cohort object

Let us create a Cohort object. If we want to re-generate the official life table by applying the routine FromCohort to this Cohort object, then we must adjust the factor for the life-years at birth.

```
Cohort["Holland 2018", 100, 1] = {Range[0, 99],
    ToAgeGroupLabels[0, 100, 1], Living["Holland 2018 all"] } // Transpose;

SetOptions[Cohort, BirthLifeYearFactor -> blyf]
```

{Birth → 100 000, BirthLifeYearFactor → 0.503523, CohortEnd → 105, CohortNPeriod → 100,

  CohortPeriod → 1, ContinuousDiscounting → True, Function → Min, Label → ,

  LastLifeExp → Automatic, Method → Line, RateOfDiscount → 0.03, Round → False, Transpose → True}

E.3.7. Using FromCohort

Let us now use FromCohort and display with LifeTable, and check the new LifeExpectancy. FromCohort relies upon the Living[] vector in the Cohort[....] data. From the Lexis Plot, we however know that the birth year must be treated differently. The CBS data on file apparently use the linear method for standard ages but indeed have (now unknown) methods for birth and end. Life expectancy at birth is off by 0.5 years if we follow the standard linear method as for the other age groups. This issue can be resolved by using above BirthLifeYearFactor. We do not adjust for the life expectancy at the end since we want that base and variant have the same method.

The advantage of using FromCohort is that the variables are calculated with known consistency and that variations will be calculated in the same manner, so that differences in outcome derive from assumptions for the variations and not from some hidden cause in the (official) data.

StoreCohort[] = {};

FromCohort[Cohort["Holland 2018", 100, 1]]

{Holland 2018}

The routine made sure that the sum of deaths is equal to the births in the cohort.

Death["Holland 2018"] // Add

100 000

There is no need to run LifeTable[] (it only shows and does no calculation).

CBS Statistics Netherlands has a life expectancy for the 95+ group of 2.2 years, but FromCohort arrives at 2.6 years.

ListLinePlot[le – LifeExpectancy[], PlotRange → All]
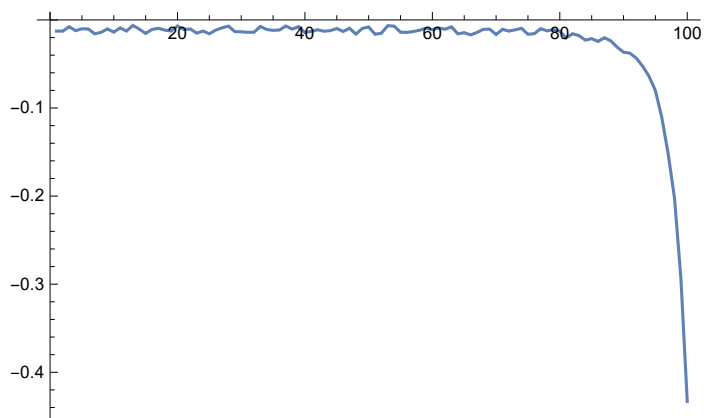
First[LifeExpectancy[]] // N

81.7927

Last[LifeExpectancy[]] // N

2.63305

For this table we do not have an adjustment for disability and quality of life yet. Given SARS-CoV-2 and the distinction between <60 and 60+ age groups, it still seems relevant to determine the

weighted averages of (discounted) life expectancy and UnitSqrtLE for these two groups. Life expectancy concerns years, and the weights will be the LifeYearsSum[]. The UnitSqrt[] measure takes each life as 100%, and the weights will be the Living[]. Currently we will use cohort weights but there is a good argument that the actual population is used. Let us use a small routine that relies upon some particulars of above cohort.

```
lab = "Holland 2018";

res1 = Weighing[LifeExpectancy[lab], LifeExpectancy[lab] * Living[lab], {60}] // N;
res2 = Weighing[DiscLifeExp[lab], DiscLifeExp[lab] * Living[lab], {60}];
res3 = Weighing[UnitSqrtLE[lab], Living[lab], {60}];
res = {res1, res2, res3} // Transpose;

TableForm[NRoundAt[Append[res, res[[1]] / res[[2]]] , 2],
 TableHeadings → { {"age < 60", "age ≥ 60", All, Ratio},
   {LifeExpectancy, DiscLifeExp, UnitSqrtLE}}]
```

	LifeExpectancy	DiscLifeExp	UnitSqrtLE
age < 60	58.98	26.58	0.79
age ≥ 60	16.97	12.79	0.39
All	54.98	24.56	0.68
Ratio	3.48	2.08	2.02

E.4. Using SARS-CoV-2 infection fatality factors for a life table cohort

E.4.1. Introduction

In a life table, the number of deaths remains the same, equal to the number of births, here 100,000. Deaths only shift up or down the ages. All reports indicate that SARS-CoV-2 works via comorbidity. Its victims thus can no longer die from the normal course of comorbidity later on. The virus has a "speeding up" effect on all-cause mortality. Other-cause mortality in later years will be reduced however because of the earlier deaths because of the virus.

Assume this scenario: there is no retention of immunity after a year, a *prevalence p is infected each year* and they suffer the disease-specific fatality, *year after year*. The mortality quotient then is $q = p$ IFF. The worst scenario is that $p = 1$. SARS-CoV-2 also is a nasty virus. For survivors it leaves scars and can negatively effect future survival prospects. Let us assume a 10% fraction that enhances the IFF in the subsequent year and 10% in the second year.

In the first half of 2020, Holland managed to keep the infection with SARS-CoV-2 below 10%. By comparison, the common cold corona viruses may have a prevalence of some 25% for the preceding two months, see Worrall (2011). In the first half of 2020, Holland managed to keep the SARS-CoV-2 prevalence below 10%. Let us now consider a scenario in which discipline becomes lax and the prevalence rises to 15% per annum. Subsequently we will look at results as a function of *p.*

E.4.2. Infection fatality factors

SARS-CoV-2 infection fatality factors for age groups 0-9, 10-19, …, 90+ are the following.

```
iff = {0.0, .0000364, .0000996, .000208, .0004203, .0024187, .011367, .051, .093, .093}
```

{0., 0.0000364, 0.0000996, 0.000208, 0.0004203, 0.0024187, 0.011367, 0.051, 0.093, 0.093}

We assume that these factors are constant within the age groups.

```
iffall = (Table[#, {10}] & /@ iff) // Flatten ;
```

The cohort size is pop = 8.27 million. If the prevalence is 100% then the IFF weighted by the cohort is 1.38%, while the steady state mortality is 100,000 / pop = 1.2%. Our routines will check that no more people can die than 100,000 in the cohort.

```
{toll = iffall * Living["Holland 2018"] // Add,
  pop = Living["Holland 2018"] // Add,
  toll / pop , 100000 / pop } // N
```

$\{114\,394., 8.2734 \times 10^6, 0.0138267, 0.0120869\}$

```
piff = PlotLine[ToSurvival[iffall] ];
```

A comparison of the earlier survival curve and the survival implied by the infection fatality factors, with $p = 1$, gives:

```
Show[porg, piff]
```


E.4.3. Effect of early death on subsequent mortality quotients

If we assume that covid-19 deaths in one year (age group) immediately reduce mortality from other causes in the next year, then the net effect of "speeding up" is as follows.

```
net = iffall – Drop[Prepend[iffall, 0], –1]
```

{0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.0000364, 0., 0., 0., 0., 0., 0., 0., 0., 0.0000632, 0., 0., 0., 0., 0.,
  0., 0., 0., 0., 0.0001084, 0., 0., 0., 0., 0., 0., 0., 0., 0.0002123, 0., 0., 0., 0., 0., 0., 0., 0., 0.,
  0.0019984, 0., 0., 0., 0., 0., 0., 0., 0., 0.0089483, 0., 0., 0., 0., 0., 0., 0., 0., 0.039633, 0.,
  0., 0., 0., 0., 0., 0., 0., 0.042, 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.}

This net effect tends to compare to the "excess death" due to SARS-CoV-2 in the first half of 2020.

```
net * Living["Holland 2018"] // Add
```

7329.6

However, it is likelier that the speeding up will be larger. Let us assume a prevalence of 15% and a delay with additionally 10% in each of the subsequent two years.

```
delay = Drop[Prepend[iffall, 0], –1];
```

```
iffin = iffall + 0.1 delay + 0.1 Drop[Prepend[delay, 0], -1];
```

```
iffp = 0.15 iffin; (* p = 0.15 *)
```

The annual death toll due to SARS-CoV-2 then would be, for the Cohort of 100,000, more than double the "excess death" from the first half of 2020.

```
covid19 = iffp Living["Holland 2018"]  // Round
```

{0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 4, 4, 4, 4, 4, 4,
    4, 4, 4, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 37, 39, 42, 42, 42, 42, 42, 42, 41, 41, 168, 179, 191, 189,
    188, 186, 184, 183, 181, 178, 688, 729, 767, 753, 738, 722, 705, 686, 666, 644, 1046, 1042,
    1033, 977, 916, 853, 786, 712, 638, 563, 486, 413, 343, 279, 220, 169, 125, 90, 63, 43}

```
covid19 // Add
```

20 271

E.4.4. Maintaining a constant total of 100,000 deaths which is the birth cohort

The other-cause mortality must be reduced with the same number. We adapt the other-cause death numbers proportionally from the age 10 onwards.

```
age09 = Take[Death["Holland 2018"], 10];
age10plus = Drop[Death["Holland 2018"], 10];
othercause = age09 ~Join~ Round[age10plus * (1 - Add[covid19] / Add[age10plus])]
```

{310, 43, 18, 10, 13, 7, 11, 8, 5, 7, 4, 4, 5, 6, 6, 10, 14, 14, 18, 18, 25, 19, 23, 18, 22, 24, 25, 23,
    29, 30, 30, 36, 37, 33, 32, 42, 40, 47, 47, 52, 52, 60, 76, 74, 88, 96, 99, 116, 130, 133, 138,
    172, 199, 195, 236, 266, 272, 334, 357, 386, 440, 471, 522, 574, 606, 679, 735, 772, 823, 885,
    992, 1055, 1182, 1309, 1414, 1487, 1631, 1771, 1900, 2051, 2271, 2372, 2689, 2873, 2995,
    3213, 3481, 3543, 3586, 3648, 3463, 3318, 3084, 2790, 2434, 2076, 1653, 1319, 952, 2030}

```
Total /@ {othercause, covid19, othercause + covid19}
```

{79 733, 20 271, 100 004}

Let us correct the total sum by using the last position of the other-cause.

```
othercause = Append[Drop[othercause, -1],
    Last[othercause] + 100 000 - Total[othercause + covid19]];
```

The new cohort thus has these total deaths.

```
Total /@ {othercause, covid19, othercause + covid19}
```

{79 729, 20 271, 100 000}

Thus we have these total deaths.

```
Death["Holland SARS-CoV-2"] = othercause + covid19;
```

Subsequently we have a revised survival curve.

```
Living["Holland SARS-CoV-2"] =
  Reverse[Cumulate[Reverse[Death["Holland SARS-CoV-2"]]]];
```

```
% // First
```

100 000

Since the cohort has been changed, the outcome of iffp * Living["Holland SARS-CoV-2"] will no longer be the same as iffp * Living["Holland 2018"], but there should not be a large difference with what we started out with.

E.4.5. Competing survival curves

We now can construct the incremental ("competing") survival curves for the virus and the other causes of death.

```
idss = IncrementalDeathsToSs[
    Living["Holland SARS-CoV-2"], {othercause, covid19} // Transpose] // N
```

$$\begin{pmatrix} 1. & 0.9969 & 0.99647 & 0.99629 & 0.99619 & 0.99606 & 0.99599 & 0.99588 & 0.9958 & 0.99575 & 0.99568 & 0.9956\ldots \\ 1. & 1. & 1. & 1. & 1. & 1. & 1. & 1. & 1. & 1. & 1. & 0.99999\ldots \end{pmatrix}$$

```
pi = ListLinePlot[idss, PlotStyle -> {Orange, Red}]
```
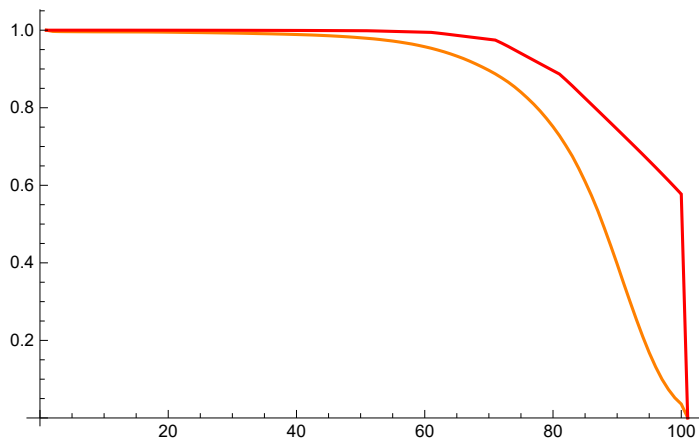

All-cause survival is given by Living["Holland SARS-CoV-2"] / 100,000 but also by multiplication of the incremental curves. Let us plot the (joint) all-cause curve (including SARS-CoV-2) in red, and the other-cause curve in orange.

```
combidss = {First[idss], Times @@ idss}
```

$$\begin{pmatrix} 1. & 0.9969 & 0.99647 & 0.99629 & 0.99619 & 0.99606 & 0.99599 & 0.99588 & 0.9958 & 0.99575 & 0.99568 & 0.9956\ldots \\ 1. & 0.9969 & 0.99647 & 0.99629 & 0.99619 & 0.99606 & 0.99599 & 0.99588 & 0.9958 & 0.99575 & 0.99568 & 0.9956\ldots \end{pmatrix}$$
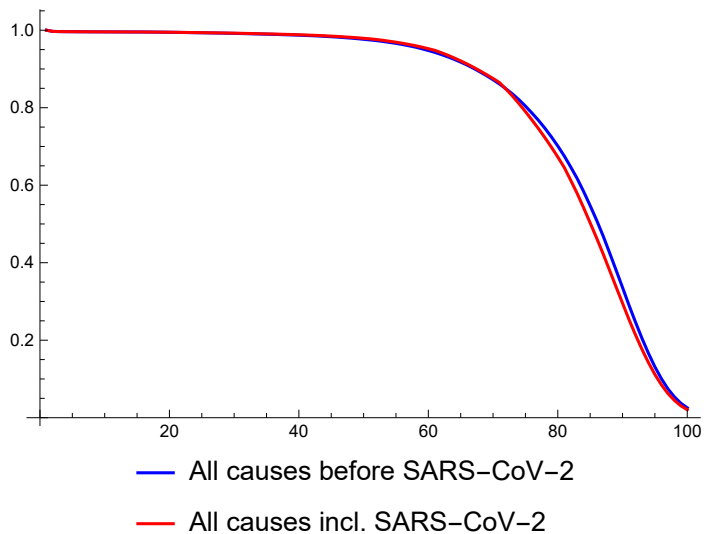
```
pcom = ListLinePlot[combidss, PlotStyle -> {Orange, Red},
  PlotLegends -> Placed[{"Other causes", "All causes incl. SARS-CoV-2"}, Below]]
```
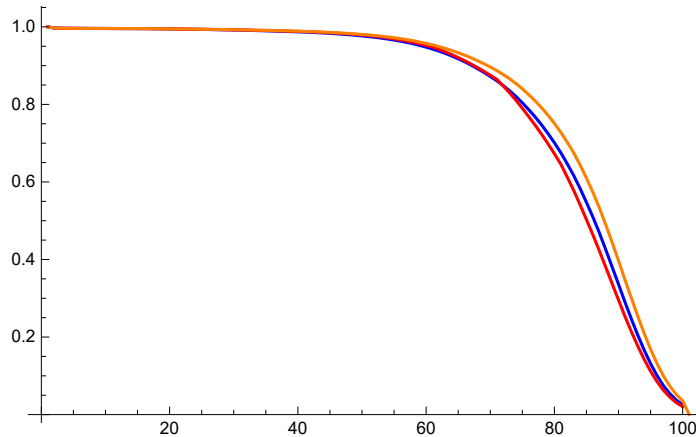

— Other causes — All causes incl. SARS–CoV–2
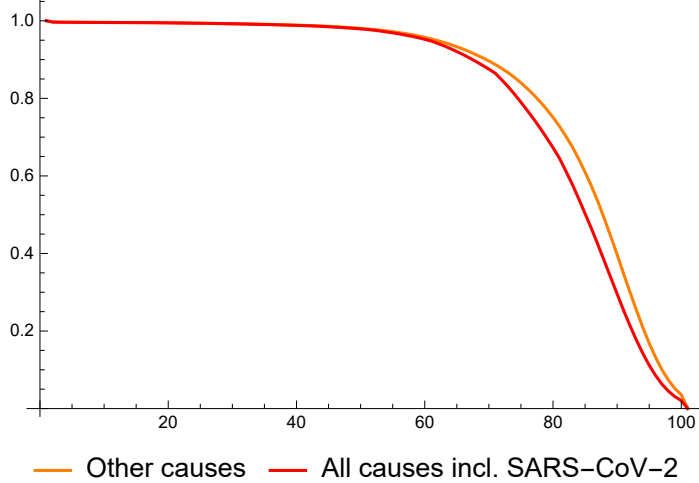
```
(* Check: gr = ListLinePlot[Living["Holland SARS-CoV-2"]/100000]; Show[pcom, gr] *)
```

The following compares the survival curves of 2018 and the effect of the virus, for $p = 0.15$.

```
ListLinePlot[{Living["Holland 2018"], Living["Holland SARS-CoV-2"]} / 100 000 ,
 PlotStyle -> {Blue, Red}, PlotLegends -> Placed[{"All causes before SARS-CoV-2",
    "All causes incl. SARS-CoV-2"}, Below]]
```


— All causes before SARS–CoV–2

— All causes incl. SARS–CoV–2

```
ListLinePlot[{Living["Holland 2018"] / 100 000, Living["Holland SARS-CoV-2"] / 100 000 ,
  First[idss]}, PlotStyle -> {Blue, Red, Orange},
 PlotLegends -> Placed[{"All causes before SARS-CoV-2",
    "All causes incl. SARS-CoV-2", "Other causes excl. SARS-CoV-2"}, Below]]
```


— All causes before SARS–CoV–2 — All causes incl. SARS–CoV–2

— Other causes excl. SARS–CoV–2

E.4.6. Effect on life expectancy

We create a Cohort object with a name and the proper age group labels, and then call FromCohort to determine the derived variables, like life expectancy as depending upon *p*.

```
Cohort["Holland SARS-CoV-2", 100, 1] =
   {Range[0, 99], ToAgeGroupLabels[0, 100, 1],
    Living["Holland SARS-CoV-2"]} // Transpose;
```

```
FromCohort[Cohort["Holland SARS-CoV-2", 100, 1]]
```

{Holland 2018, Holland SARS–CoV–2}

PM. It is not necessary to display the LifeTable[].

At birth:

```
ag = 0;
LEwas = LifeExpectancy["Holland 2018"][[ag + 1]];
LEsars = LifeExpectancy["Holland SARS-CoV-2"][[ag + 1]];
{LEwas, LEsars, LEwas - LEsars} // N
```

{81.7927, 81.132, 0.660697}

At age 50:

```
ag = 50;
LEwas = LifeExpectancy["Holland 2018"][[ag + 1]];
LEsars = LifeExpectancy["Holland SARS-CoV-2"][[ag + 1]];
{LEwas, LEsars, LEwas - LEsars} // N
```

{33.5763, 32.7736, 0.802654}

## E.4.7. Effect on UnitSqrt

The UnitSqrt[a, 0, LE] has been determined by FromCohort. We can compare the two outcomes. There appears to be little difference. Choices between the age-groups are not affected much by the change in the overall pattern of life-expectancy.

Since the UnitSqrt takes each life as 100%, the age-specific outcomes can be weighted by the age-groups. The overall average doesn't change much either, and the situation without SARS-CoV-2 is only marginally better.

```
UnitSqrtLE[#] . Living[#] / Total[Living[#] ] & /@
 {"Holland 2018", "Holland SARS-CoV-2"}
```

{0.680082, 0.679954}

The age-specific outcomes.

```
ListLinePlot[{UnitSqrtLE["Holland 2018"], UnitSqrtLE["Holland SARS-CoV-2"]},
 AxesLabel → {"Age", "UnitSqrtLE"},
 BaseStyle → {FontSize → 13}, PlotStyle → {Blue, Red}]
```
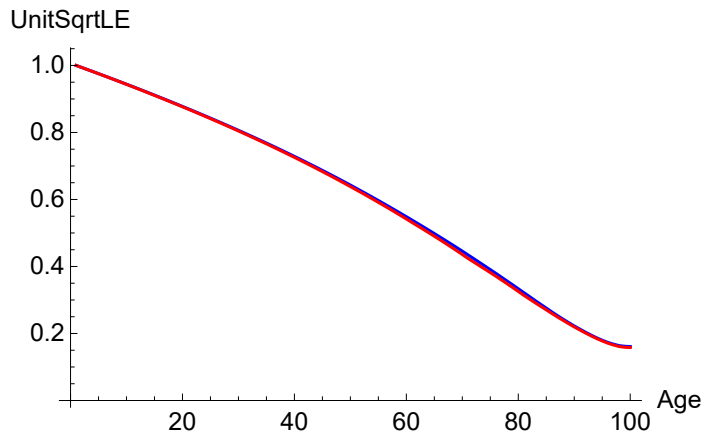

E.4.8. Routine

The above recalculation of the new survival can be done by this routine:

```
iffin = iffall + 0.1 delay + 0.1 Drop[Prepend[delay, 0], -1];
```

```
lis = SurvivalShift[Living["Holland 2018"], iffin, 0.15, Split -> 10];
```

```
lis === Living["Holland SARS-CoV-2"]
```

True

The above result can be turned into a Cohort object, and used for the life expectancy. The Results[-SurvivalShift] have the lists of Living, (all-cause) Death,"Specific" and "Other" deaths.

```
newidss =
 Hold[IncrementalDeathsToSs[Living, {"Other", "Specific"} // Transpose] // N]
```

$\text{Hold}\left[N\left[\text{IncrementalDeathsToSs}(\text{Living}, \{\text{Other}, \text{Specific}\}^{\text{T}})\right]\right]$

```
newidss = (newidss /. Results[SurvivalShift]) // ReleaseHold;
```

```
idss == newidss
```

True

E.4.9. Application with a symbolic parameter

Life expectancy can be seen as a function of prevalence *p*.

```
livfunc[p_] := SurvivalShift[Living["Holland 2018"], iffin, p, Split → 10];
```

```
newliv = SurvivalShift[Living["Holland 2018"], iffin, p, Split → 10];
(* p hardcoded *)
```

```
Cohort["Holland SARS-CoV-2 with p", 100, 1] =
   {Range[0, 99], ToAgeGroupLabels[0, 100, 1],  newliv} // Transpose;
```

```
FromCohort[Cohort["Holland SARS-CoV-2 with p", 100, 1] ]
```

{Holland 2018, Holland SARS–CoV–2, Holland SARS–CoV–2 with p}

PM. It is not necessary to display the LifeTable[].

Check that we have the same values.

At birth:

```
ag = 0;
LEwas = LifeExpectancy["Holland 2018"][[ag + 1]] ;
LEsars = LifeExpectancy["Holland SARS-CoV-2 with p"][[ag + 1]]  /. p -> 0.15;
{LEwas, LEsars, LEwas - LEsars} // N
```

{81.7927, 81.132, 0.660697}

At age 50:

```
ag = 50;
LEwas = LifeExpectancy["Holland 2018"][[ag + 1]] ;
LEsars = LifeExpectancy["Holland SARS-CoV-2 with p"][[ag + 1]]  /. p -> 0.15;
{LEwas, LEsars, LEwas - LEsars} // N
```

{33.5763, 32.7736, 0.802654}

E.4.10. Plot for number of disease specific deaths

The disease specific deaths cannot be larger than the cohort size.

```
lis = Table[livfunc[p];
  {p, Total[ "Specific" /. Results[SurvivalShift, p] ]}, {p, 0, 1, 0.1}]
```

$$
\begin{pmatrix}
0. & 0 \\
0.1 & 13\,500 \\
0.2 & 27\,018 \\
0.3 & 40\,524 \\
0.4 & 54\,045 \\
0.5 & 67\,548 \\
0.6 & 81\,063 \\
0.7 & 94\,560 \\
0.8 & 100\,000 \\
0.9 & 100\,000 \\
1. & 100\,000
\end{pmatrix}
$$

```
ListLinePlot[lis, AxesLabel → {"Prevalence\nof SARS-CoV-2", Deaths},
 PlotLabel → "Disease specific, cohort 100000", AxesOrigin -> {0, 0}]
```
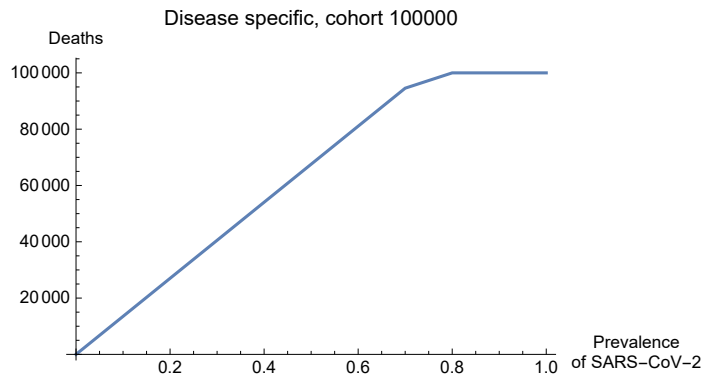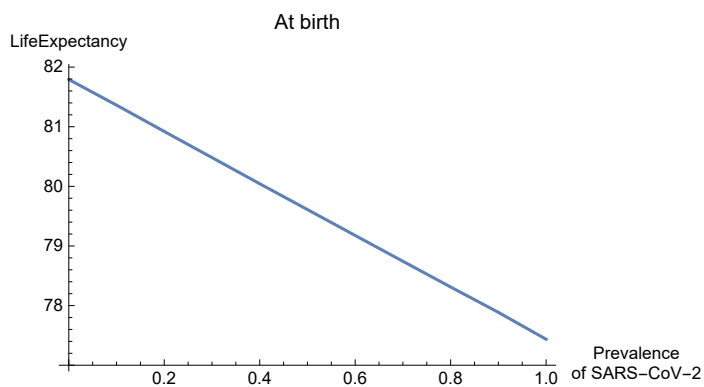

E.4.11. Plot for life expectancy

```
lis = Table[ag = 0;
  {p, LifeExpectancy["Holland SARS-CoV-2 with p"][[ag + 1]] }, {p, 0, 1, 0.1}]
```

$$\begin{pmatrix} 0. & 81.7927 \\ 0.1 & 81.3621 \\ 0.2 & 80.9209 \\ 0.3 & 80.4825 \\ 0.4 & 80.0435 \\ 0.5 & 79.6106 \\ 0.6 & 79.1763 \\ 0.7 & 78.7434 \\ 0.8 & 78.3118 \\ 0.9 & 77.8867 \\ 1. & 77.4354 \end{pmatrix}$$

```
ListLinePlot[lis, AxesLabel → {"Prevalence\nof SARS-CoV-2", LifeExpectancy},
 PlotLabel → "At birth", AxesOrigin -> {0, 77}]
```

```
lis = Table[ag = 50;
  {p, LifeExpectancy["Holland SARS-CoV-2 with p"][[ag + 1]]  // N}, {p, 0, 1, 0.1}]
```

$$\begin{pmatrix} 0. & 33.5763 \\ 0.1 & 33.0416 \\ 0.2 & 32.5095 \\ 0.3 & 31.9766 \\ 0.4 & 31.448 \\ 0.5 & 30.9218 \\ 0.6 & 30.4002 \\ 0.7 & 29.8784 \\ 0.8 & 29.363 \\ 0.9 & 28.8597 \\ 1. & 28.3273 \end{pmatrix}$$

```
ListLinePlot[lis, AxesLabel → {"Prevalence\nof SARS-CoV-2", LifeExpectancy},
  PlotLabel → "At age 50", AxesOrigin -> {0, 28}]
```
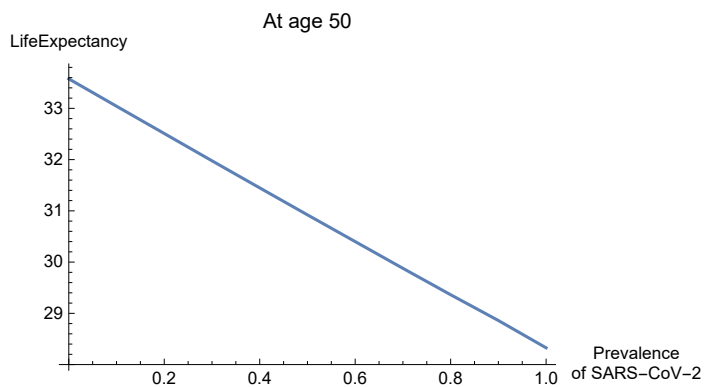

E.4.12. UnitSqrt with parameter

The following gives the age group weigthed outcome of UnitSqrtLE as a function of *p.* Apparently the higher age groups require some protection that some variables should remain nonnegative. I leave this for the moment. The overall effect is small, in the third digit.

```
lis = Table[hlp =  Living["Holland SARS-CoV-2 with p"] ; shr = hlp / Total[hlp];
  {p, UnitSqrtLE["Holland SARS-CoV-2 with p"] . shr }, {p, 0, 1, 0.1}]
```

$$\begin{pmatrix} 0. & 0.680082 \\ 0.1 & 0.680005 \\ 0.2 & 0.679883 \\ 0.3 & 0.679698 \\ 0.4 & 0.679455 \\ 0.5 & 0.679139 \\ 0.6 & 0.67876 \\ 0.7 & 0.678297 \\ 0.8 & 0.677776 \\ 0.9 & 0.677304 - 0.0000108308\,i \\ 1. & 0.676244 + 0.0000683604\,i \end{pmatrix}$$

```
ListLinePlot[lis /. x_?NumberQ :> Re[x],
 AxesLabel → {"Prevalence\nof SARS-CoV-2", UnitSqrtLE},
 PlotLabel → "Cohort average", AxesOrigin -> {0, .67}]
```


Cohort average
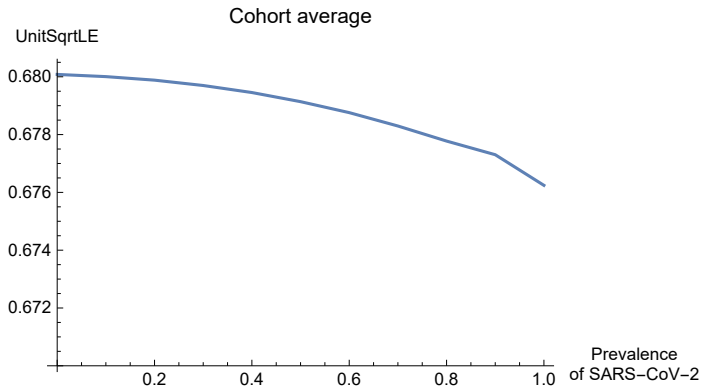
```
lis = Table[UnitSqrtLE["Holland SARS-CoV-2 with p"], {p, 0, 1, 0.1}]
```

1.	0.993924	0.987816	0.98167	0.975486	0.969265	0.963003	0.956702	0.950359	0.943973	0.93
1.	0.993892	0.987751	0.981573	0.975356	0.969101	0.962805	0.956469	0.950091	0.94367	0.93
1.	0.993859	0.987685	0.981472	0.975221	0.968931	0.962599	0.956228	0.949814	0.943355	0.93
1.	0.993825	0.987618	0.981371	0.975085	0.96876	0.962393	0.955985	0.949535	0.94304	0.93
1.	0.993792	0.98755	0.981269	0.974947	0.968587	0.962184	0.95574	0.949252	0.94272	0.93
1.	0.993758	0.987482	0.981166	0.97481	0.968414	0.961975	0.955495	0.948971	0.942401	0.93
1.	0.993724	0.987413	0.981063	0.974671	0.968239	0.961764	0.955247	0.948685	0.942077	0.93
1.	0.99369	0.987344	0.980958	0.974531	0.968063	0.961551	0.954997	0.948397	0.941751	0.93
1.	0.993655	0.987274	0.980853	0.974389	0.967885	0.961336	0.954744	0.948107	0.941422	0.93
1.	0.99362	0.987205	0.980748	0.974249	0.967708	0.961122	0.954493	0.947818	0.941094	0.93
1.	0.993583	0.98713	0.980636	0.974098	0.967518	0.960892	0.954223	0.947507	0.940742	0.9

```
ListLinePlot[Drop[lis, -1], AxesLabel → {"Age", UnitSqrtLE},
 PlotLabel → "Prevalence in steps of 0.1", AxesOrigin -> {0, 0}]
```


Prevalence in steps of 0.1
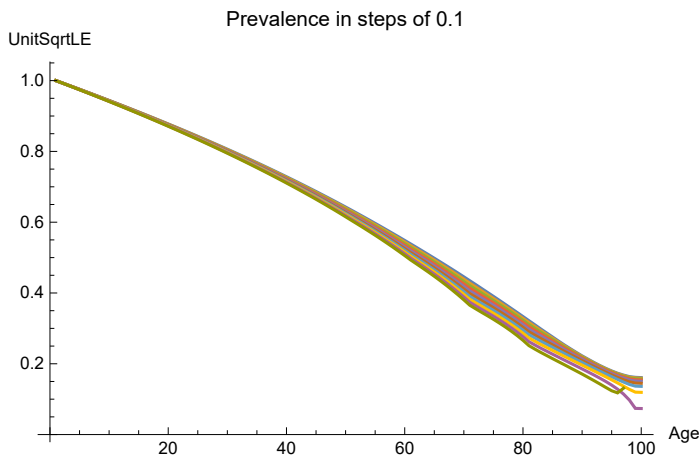
```
lis = Table[ag = 50;
  {p, UnitSqrtLE["Holland SARS-CoV-2 with p"][[ag + 1]]  // N}, {p, 0, 1, 0.1}]
```

$$
\begin{pmatrix}
0. & 0.633833 \\
0.1 & 0.630787 \\
0.2 & 0.627701 \\
0.3 & 0.624556 \\
0.4 & 0.621378 \\
0.5 & 0.618158 \\
0.6 & 0.614907 \\
0.7 & 0.611595 \\
0.8 & 0.608262 \\
0.9 & 0.604949 \\
1. & 0.601376
\end{pmatrix}
$$

```
ListLinePlot[lis , AxesLabel → {"Prevalence\nof SARS-CoV-2", UnitSqrtLE},
 PlotLabel → "Age 50", AxesOrigin -> {0, .60}]
```
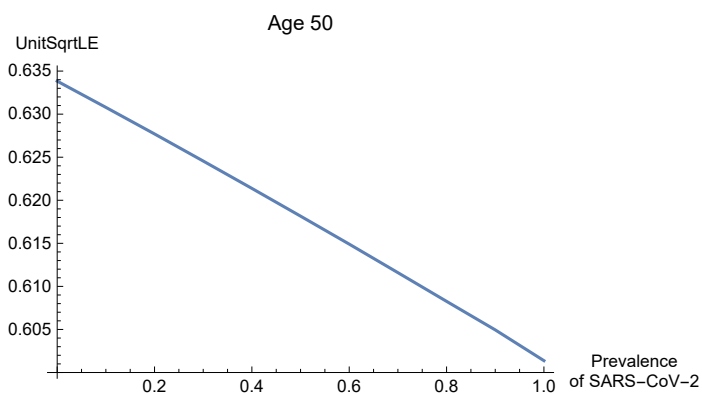


```
lis = Table[ag = 75;
  {p, UnitSqrtLE["Holland SARS-CoV-2 with p"][[ag + 1]]  // N}, {p, 0, 1, 0.1}]
```

$$
\begin{pmatrix}
0. & 0.379328 \\
0.1 & 0.373781 \\
0.2 & 0.367907 \\
0.3 & 0.361663 \\
0.4 & 0.355059 \\
0.5 & 0.348033 \\
0.6 & 0.340566 \\
0.7 & 0.332553 \\
0.8 & 0.324033 \\
0.9 & 0.315147 \\
1. & 0.304665
\end{pmatrix}
$$

```
ListLinePlot[lis /. x_?NumberQ :> Re[x],
 AxesLabel → {"Prevalence\nof SARS-CoV-2", UnitSqrtLE},
 PlotLabel → "Age 75", AxesOrigin -> {0, .30}]
```
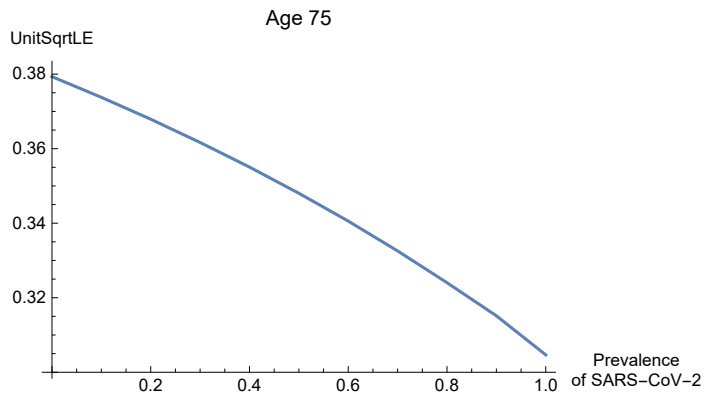

References

R.1. Timeline

Wikipedia is a portal and no source. The wikipedia timeline on the "corona crisis" in Holland 2020 is: https://en.wikipedia.org/wiki/2020_coronavirus_pandemic_in_the_Netherlands#Timeline

R.2. On the gap between society and science

Colignatus, Thomas (2020a), "Forum Theory & A National Assembly of Science and Learning", mijnbestseller.nl, http://thomascool.eu/Papers/FT&NASL/Index.html (1st edition February 10, 2nd edition, July 5)

R.3. Own research in public health economics before 2020

Colignatus, Thomas (2002-2020), Notebooks and packages within *The Economics Pack* w.r.t. the life sciences, as part of Colignatus (1995-2020e):

```
ToFileName[{$TEPDirectory[], "Applications",
  "Economics", "Documentation", "English", "LifeSciences"}]
```

Colignatus, Thomas (2003, 2020), "On the value of life", original 2003 https://ideas.repec.org/p/w-pa/wuwppe/0310003.html, and update 2020 https://mpra.ub.uni-muenchen.de/102535

Colignatus, Thomas (2004), "Modifying behaviour on STI including HIV and oncogenic HPV: a draft protocol for a registry open to the registered subject, generating the concept of a 'STI passport'", https://ideas.repec.org/p/wpa/wuwphe/0412001.html

Colignatus, Thomas (2019), "A numerical exercise on climate change and family planning: World population might reduce from 11 to 8 billion in 2100 if women of age 15-29 wait and have their first child at age 30+", https://mpra.ub.uni-muenchen.de/97447

R.4. Standard discussions about SI(EY)A(CD) or epidemiology in general

https://www.idmod.org/docs/hiv/model-seir.html

3Blue1Brown (2020), "Simulating an epidemic", https://www.youtube.com/watch?v=gxAaO2rsdIs

Barendregt, Jan, Gerrit van Oortmarssen, Theo Vos, and Christopher Murray (2003), "A generic model for the assessment of disease epidemiology: the computational basis of DisMod II", Popul Health Metrics 1, 4, https://doi.org/10.1186/1478-7954-1-4. See the software at https://www.who.int/healthinfo/global_burden_disease/tools_software/en/ and the In Memoriam http://www.epigear.com/Obituary/obituary.html

Barton, Jeffrey (2016), "Models for Life: An Introduction to Discrete Mathematical Modeling with Microsoft Office Excel", Wiley

Baylor College of Medicine (BCM) (2020), "Introduction to Infectious Diseases", website https://www.bcm.edu/departments/molecular-virology-and-microbiology/emerging-infections-and-biodefense/introduction-to-infectious-diseases

Bedford, Juliet, Jeremy Farrar, Chikwe Ihekweazu, Gagandeep Kang, Marion Koopmans, John Nkengasong (2019), "A new twenty-first century science for effective epidemic response", Nature 575, 130–136 (2019), https://doi.org/10.1038/s41586-019-1717-y

Breda, D., O. Diekmann, W. F. de Graaf, A. Pugliese, R. Vermiglio (2012), "On the formulation of epidemic models (an appraisal of Kermack and McKendrick)", Journal ofBiological Dynamics, 6:sup2, 103-117, DOI: 10.1080/17513758.2012.716454

Brauer, Fred, Pauline van den Driessche, and Jianhong Wu (eds) (2008), "Mathematical epidemiology", Springer, https://link.springer.com/book/10.1007/978-3-540-78911-6, https://www.researchgate.net/publication/266998673_Mathematical_Epidemiology, https://www.fields.utoronto.ca/programs/scientific/10-11/drugresistance/emergence/fred1.pdf

Cassini, Alessandro, et al. (2018), "Impact of infectious diseases on population health using incidence-based disability-adjusted life years (DALYs): results from the Burden of Communicable Diseases in Europe study, European Union and European Economic Area countries, 2009 to 2013", Eurosurveillance Volume 23, Issue 16, 19/Apr/2018, https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2018.23.16.17-00454?crawler=true

Champredon, David, Jonathan Dushoff, David J.D. Earn (2018), "Equivalence of the Erlang Seir Epidemic Model and the Renewal Equation", bioRxiv 319574; doi: https://doi.org/10.1101/319574

Delamater, Paul, et al. (2019), "Complexity of the Basic Reproduction Number (R0)", Emerging infectious diseases, Vol 25, No 1, January, https://wwwnc.cdc.gov/eid/article/25/1/17-1901_article

Diekmann, O., J.A.P. Heesterbeek, J.A.J. Metz (1990), "On the definition and the computation of the basic reproduction ratio R0 in models for infectious diseases in heterogeneous populations", Journal of Mathematical Biology28(4), http://dspace.library.uu.nl/handle/1874/8051

Dunne, Edward (2020), "Mathematics and epidemiology", March 15, https://blogs.ams.org/beyondreviews/2020/03/15/mathematics-and-epidemiology/

Fine, Paul (2003), "The Interval Between Successive Cases of an Infectious Disease", Am J Epidemiol, Dec 1;158(11):1039-47. PMID: 14630599 doi: 10.1093/aje/kwg251

Fine, Paul, Ken Eames, and David L. Heymann (2011), ""'Herd Immunity'': A Rough Guide", VAC-CINES d CID 2011:52 (1 April) d 911, DOI: 10.1093/cid/cir007

Flaxman, Abraham, James Collins (2019), https://github.com/ihmeuw/dismod_mr

Fox, J.P., L. Elveback, W. Scott, L. Gatewood, E. Ackerman (1971), "Herd immunity, basic concept and relevance for public health immunization practices", Am. J. of Epidemiology, September, Vol. 94 No 3, p179-189

Garnett, Geoffrey (2005), "Role of Herd Immunity in Determining the Effect of Vaccines against Sexually Transmitted Disease", The Journal of Infectious Diseases 2005; 191(Suppl 1):S97–106, https://doi.org/10.1086/425271

Glawell, Malcolm (2000, "The tipping point", Little, Brown

Kleinbaum, David (2015), "ActivEpi", http://www.activepi.com/

Lin, Feng, and Kumar Muthuraman and Mark Lawley (2010), "An optimal control theory approach to
non-pharmaceutical interventions", BMC Infectious Diseases 2010, 10:32
http://www.biomedcentral.com/1471-2334/10/32

Heesterbeek, J.A.P. (2002), "A brief history of R0 and a recipe for its calculation", Acta biotheoretica, volume 50, issue 3, pp. 189 - 204

Heesterbeek, J.A.P. et al. (2015), "Modeling infectious disease dynamics in the complex landscape of global health", Science, Mar 13; 347(6227): aaa4339, PMC4445966

Hillier, Frederick, and Gerald Lieberman (1967, 1972), "Introduction to Operations Research", Holden-Day

Jones, James Holland (2007), "Notes on R0", https://web.stanford.edu/~jhj1/teachingdocs/Jones-on-R0.pdf

Jones, James Holland (2008), "Models of Infectious Disease", https://web.stanford.edu/~jhj1/teachingdocs/Jones-Epidemics050308.pdf

Mangen, M.-J.J., D. Plass, M.E.E. Kretzschmar (2014), "Estimating the current and future burden of communicable diseases in the European Union and EEA/EFTA", RIVM report 210474001/2014, https://rivm.openrepository.com/bitstream/handle/10029/317045/210474001.pdf

Manton, K.G., and E. Stallard (1988), "Chronic Disease Modelling : Measurement and Evaluation of the Risks of Chronic Disease Processes", Oxford

Nishiura, N, C Castillo-Chavez , M Safan , G Chowell (2009), "Transmission potential of the new influenza A(H1N1) virus and its age-specificity in Japan", Eurosurveillance Volume 14, Issue 22, 04/Jun/2009

Obadia, T., Haneef, R. & Boëlle, P. (2012), "The R0 package: a toolbox to estimate reproduction numbers for epidemic outbreaks", BMC Med Inform Decis Mak 12, 147 (2012). https://-doi.org/10.1186/1472-6947-12-147

Oostrom, Sandra H. van, H. Susan J. Picavet, Boukje M. van Gelder, Lidwien C. Lemmens, Nancy Hoeymans, Robert A. Verheij, François G. Schellevis, Caroline A. Baan (2011), "Multimorbiditeit en comorbiditeit in de Nederlandse bevolking – gegevens van huisartsenpraktijken", Ntvg,

2011-05-30;155:A3193

Piot, Peter (2018), "Are we ready for the next pandemic?", October, https://www.youtube.com/watch?v=en06PYwvpbI

Roberts, M.G. & J. A. P. Heesterbeek (2007), "Model-consistent estimation of the basic reproduction number from the incidence of an emerging infection", Journal of Mathematical Biology volume 55, Article number: 803 (2007)

Smith, David, and Lang Moore (2001, 2004), "The SIA Model for Spread of Disease", MAA / JOMA, *Convergence.* The various chapters begin here: https://www.maa.org/press/periodicals/loci/joma/the-sir-model-for-spread-of-disease-introduction

Svensson Ake, (2006), "A note on generation times in epidemic models", Math Biosci. 2007 Jul;208(1):300-11. Epub 2006 Nov 9

Vink, Margaretha Annelie, Martinus Christoffel Jozef Bootsma, Jacco Wallinga (2014), "Serial Intervals of Respiratory Infectious Diseases: A Systematic Review and Analysis", American Journal of Epidemiology, Volume 180, Issue 9, 1 November 2014, Pages 865–875, https://doi.org/10.1093/aje/kwu209

Vos, Theo, et al. (2012), "Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010", Lancet. 2012 Dec 15; 380(9859): 2163–2196, PMCID: PMC6350784

Wallinga, Jacco, and Marc Lipsitch (2006, 2007), "How Generation Intervals Shape the Relationship Between Growth Rates and Reproductive Numbers", Proc Biol Sci. 2007 Feb 22;274(1609):599-604. doi: 10.1098/rspb.2006.3754

Wearing, H.J., Rohani, P., Keeling, M.J. (2005), "Appropriate models for the management of infectious diseases", PLoS Med. 7, 621–627 (2005)

Weisstein, Eric (2020a), "Kermack-McKendrick Model", https://mathworld.wolfram.com/Kermack-McKendrickModel.html

Wetering, EJ van de, and Stolk EA, van Exel NJ, Brouwer WB (2013), "Balancing equity and efficiency in the Dutch basic benefits package using the principle of proportional shortfall", Eur J Health Econ. 2013;14(1):107-115. doi:10.1007/s10198-011-0346-7

WHO GPEI & David Heymann (2020), "Poliovirus vs smallpox containment: An interview with David Heymann", http://polioeradication.org/news-post/poliovirus-vs-smallpox-containment-an-interview-with-david-heymann/

Wit, E. de, N. van Doremalen, D. Falzarano & V. J. Munster (2016), "SARS and MERS: recent insights into emerging coronaviruses", Nature Reviews Microbiology volume 14, pages 523–534, https://www.nature.com/articles/nrmicro.2016.81

Wilder-Smith, Annelies, Monica D. Teleman, Bee H. Heng, Arul Earnest, Ai E. Ling, and Yee S. Leo (2005), "Asymptomatic SARS Coronavirus Infection among Healthcare Workers, Singapore", Emerg Infect Dis. 2005 Jul; 11(7): 1142–1145, PMC3371799

Wilder-Smith, Annelies, Calvin Chiew, Vernon Lee (2020), "Can we contain the COVID-19 outbreak with the same measures as for SARS?", Lancet Infect Dis, Published Online March 5, https://doi.org/10.1016/S1473-3099(20)30129-8

Williams, A. (1997), "Intergenerational equity: an exploration of the 'Fair Innings' argument", Health Economics 6, 117–32

Worrall, Graham (2011), "Common Cold", Can Fam Physician, 57(11) 1289-1290

Yan, Ping, and Gerardo Chowell (2019), "Quantitative Methods for Investigating Infectious Disease Outbreaks", Springer

Zorginstituut Nederland (ZIN) (2018a), "Ziektelast in de praktijk. De theorie en praktijk van het berekenen van ziektelast bij pakketbeoordelingen, https://www.zorginstituutnederland.nl/publicaties/rapport/2018/05/07/ziektelast-in-de-praktijk

Zorginstituut Nederland (ZIN) (2018b), "Bijlage bij rapport 'Ziektelast in de praktijk': Samenvatting reacties externe partijen op concept rapport", https://www.zorginstituutnederland.nl/publicaties/rapport/2018/05/07/ziektelast-in-de-praktijk

R.5. Specifically about SARS-CoV-2 (Covid-19)

For Holland: https://www.rivm.nl/en/novel-coronavirus-covid-19/calculations/calculation-models-are-public-and-accessible

For the USA: https://datascience.nih.gov/covid-19-open-access-resources

Adamik, Barbara, et al. (2020), "Mitigation and herd immunity strategy for COVID-19 is likely to fail", MOCOS International research group, Wroclaw / Kaiserslautern / Trier / Koblenz / Warsaw, ResearchGate April 2

Anastassopoulou, Cleo, Lucia Russo, Athanasios Tsakris, Constantinos Siettos (2020), "Data-based analysis, modelling and forecasting of the COVID-19 outbreak", PLOS ONE, https://doi.org/10.1371/journal.pone.0230405 March 31, 2020 (received February 11)

Anderson, Roy, Hans Heesterbeek, Don Klinkenberg, T Deirdre Hollingsworth (2020), " How will country-based mitigation measures influence the course of the COVID-19 epidemic?", The Lancet, Vol 395 March 21 (online March 6), 2020, p931-934

Backer, Jantien, Don Klinkenberg, Jacco Wallinga (2020), "Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20–28 January 2020", Euro Surveill, v.25(5), 2000062, Feb 6, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7014672/

BBC Horizon (2020), "Coronavirus: A Horizon special - Part 2", Tuesday May 26, BBC 2, https://www.bbc.co.uk/programmes/m000jbhj

Bonten, Marc, et al. (2020), weblog texts starting in January 2020 at https://reflectionsipc.com/2020/01/24/novel-coronavirus-outbreak-an-update/

Borger, Peter (2020), "A SARS-like Coronavirus was Expected, but nothing was done to be Prepared", American Journal of Biomedical Science & Research 8(5), AJBSR.MS.ID.001312. DOI:10.34297/AJBSR.2020.08.001312 (April 29)

Britton, T., F. Ball, P. Trapman (2020a), "The disease-induced herd immunity level for Covid-19 is substantially lower than the classical herd immunity level", preprint, https://doi.org/10.1101/2020.05.06.20093336 (May 14)

Britton, T., F. Ball, P. Trapman (2020b), "A mathematical model reveals the influence of population

heterogeneity on herd immunity to SARS-CoV-2", Science 23 Jun 2020: eabc6810, DOI: 10.1126/-science.abc6810

Bunnik, Bram van, et al. (2020), "Segmentation and shielding of the most vulnerable members of the population as elements of an exit strategy from COVID-19 lockdown", May 5, https://www.ed.ac.uk/covid-19-response/latest-news/two-tier-approach-could-begin-lockdown-end

Chan, Jasper Fuk-Woo, and Kin-Hang Kok, Zheng Zhu, Hin Chu, Kelvin Kai-Wang To, Shuofeng Yuan & Kwok-Yung Yuen (2020), "Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan", Emerging Microbes & Infections, 9:1, 221-236, DOI: 10.1080/22221751.2020.1719902, January

Cheng PK, Wong DA, Tong LK, et al. (2004), "Viral shedding patterns of coronavirus in patients with probable severe acute respiratory syndrome" Lancet (London, England). 2004 May;363(9422):1699-1700. DOI: 10.1016/s0140-6736(04)16255-7.

Colignatus, Thomas (2020b), "Covid-19 and the value of life", https://boycottholland.wordpress.com/2020/03/31/covid-19-and-the-value-of-life

Colignatus, Thomas (2020c), "Colour-coding of Covid-19 quarantines", https://boycottholland.wordpress.com/2020/04/05/colour-coding-of-covid-19-quarantines

Colignatus, Thomas (2020d), "Two weeks – and 2000 or 4000 more deaths – later", https://boycottholland.wordpress.com/2020/04/19/two-weeks-and-2000-or-4000-more-deaths-later

Colignatus, Thomas (2020e), "SARS-CoV-2 and aerosols and ventilation", https://boycottholland.wordpress.com/2020/06/25/sars-cov-2-and-aerosols-and-ventilation/

Collins, Francis (2020), "The Prime Cellular Targets for the Novel Coronavirus", https://directorsblog.nih.gov/2020/05/05/the-prime-cellular-targets-for-the-novel-coronavirus

Draulans, Dirk, and Peter Piot (2020), "'Finally, a virus got me.' Scientist who fought Ebola and HIV reflects on facing death from COVID-19. Interview with Peter Piot", May 8, Science, https://www.sciencemag.org/news/2020/05/finally-virus-got-me-scientist-who-fought-ebola-and-hiv-reflects-facing-death-covid-19

Du, Zhanwei, Xiaoke Xu, Ye Wu, Lin Wang, Benjamin J. Cowling, Lauren Ancel Meyers (2020) "Serial Interval of COVID-19 among Publicly Reported Confirmed Cases", CDC Early Release, Emerging Infectious Diseases, Volume 26, Number 6—June 2020, https://wwwnc.cdc.gov/eid/article/26/6/20-0357_article

Eichenberger, Reiner, Rainer Hegselmann, David Savage, David Stadelmann & Benno Torgler (2020), "Certified Corona-Immunity as a Resource and Strategy to Cope with Pandemic Costs", March 25, 2020, CREMA Working Paper Series 2020-03, Center for Research in Economics, Management and the Arts (CREMA), https://ideas.repec.org/p/cra/wpaper/2020-03.html, paper at the Frankfurt School of Finance & Management, https://www.frankfurt-school.de/en/home/newsroom/news/2020/Maerz/corona-immune-ressource

Enserink, M. (2020), "Update: 'A bit chaotic.' Christening of new coronavirus and its disease name create confusion", Science, Feb. 12, https://www.sciencemag.org/news/2020/02/bit-chaotic-christening-new-coronavirus-and-its-disease-name-create-confusion

Ferguson, Neil, et al. (2020), "Impact of non-pharmaceutical interventions (NPIs) to reduce

COVID-19 mortality and healthcare demand", Report 9, 16 March 2020, Imperial CollegeCOVID-19 Response TeamDOI: https://doi.org/10.25561/77482

Ferretti, Luca, Chris Wymant, Michelle Kendall, Lele Zha1, Anel Nurta1, Lucie Abeler-Dörner, Michael Parker, David Bonsall, Christophe Fraser (2020), "Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing", Science 08 May 2020: Vol. 368, Issue 6491, eabb6936, DOI: 10.1126/science.abb6936

Flaxman, Seth, et al. (2020), "Report 13 - Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries", https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-13-europe-npi-impact

Frijters, Paul (2020), "The corona dilemma", https://clubtroppo.com.au/2020/03/21/the-corona-dilemma

Ganyani, Tapiwa , Cecile Kremer, Dongxuan Chen, Andrea Torneri, Christel Faes, Jacco Wallinga, Niel Hens (2020), "Estimating the generation interval for COVID-19 based on symptom onset data", Eurosurveillance Volume 25, Issue 17, 30/Apr/2020, https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2020.25.17.2000257 (MedRXiv, https://doi.org/10.1101/2020.03.05.200318150)

Gill, R.D. (2020), " Corona statistics, Corona uncertainty", presentation May 7 2020, https://www.math.leidenuniv.nl/~gill/Corona.pdf

Goh, Gabriel (2020), "Epidemic calculator", https://gabgoh.github.io/COVID/index.html

Gollier, Christian (2020), "Optimal Group Testing to Exit the Covid Confinement", https://econfip.org/policy-brief/optimal-group-testing-to-exit-the-covid-confinement/#

Grant, Alastair (2020), "Dynamics of COVID-19 epidemics: SEIR models underestimate peak infection rates and overestimate
epidemic duration", https://www.medrxiv.org/content/10.1101/2020.04.02.20050674v2

Hanage, William (2020), "I'm an epidemiologist. When I heard about Britain's 'herd immunity' coronavirus plan, I thought it was satire", Guardian, 15 Mar 2020

Höhle, Michael (2020), "Effective reproduction number estimation", weblog, https://staff.math.su.se/hoehle/blog/2020/04/15/effectiveR0.html

Ioannidis, John (2020, "Coronavirus disease 2019: The harms of exaggerated information and non-evidence-based measures", https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7163529/

Irribando, F. Javier, et al. (2020), "Rapid Decay of Anti–SARS-CoV-2 Antibodies in Persons with Mild Covid-19", https://www.nejm.org/doi/full/10.1056/NEJMc2025179 (July 21)

Jones, David (2020), "History in a Crisis — Lessons for Covid-19", NJEM 1681-1683 DOI:10.1056/NEJMp2004361 (April 30)

Jones, Terry, et al. (Christian Drosten) (2020), "An analysis of SARS-CoV-2 viral load by patient age", https://zoonosen.charite.de/fileadmin/user_upload/microsites/m_cc05/virologie-ccm/dateien_upload/Weitere_Dateien/analysis-of-SARS-CoV-2-viral-load-by-patient-age.pdf

Kimball, Anne, et al. (2020), "Asymptomatic and Presymptomatic SARS-CoV-2 Infections in Residents of a Long-Term Care Skilled Nursing Facility — King County, Washington, March 2020", CDC

Morbidity and Mortality Weekly Report (MMWR), Weekly / April 3, 2020 / 69(13);377–381, https://www.cdc.gov/mmwr/volumes/69/wr/mm6913e1.htm?s_cid=mm6913e1_w

Kirkcaldy, R.D., B.A. King, J.T. Brooks (2020), "COVID-19 and Postinfection ImmunityLimited Evidence, Many Remaining Questions", Viewpoint | COVID-19: Beyond Tomorrow, May 11, JAMA. 2020;323(22):2245-2246. doi:10.1001/jama.2020.7869

Kissler, Stephen, et al. (2020), "Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period", Science April 14, https://science.sciencemag.org/content/early/2020/04/14/-science.abb5793

Kretzschmar,Mirjam , Ganna Rozhnova, Martin Bootsma, Michiel van Boven, Janneke van de Wijgert, Marc Bonten (2020), "Time is of the essence: impact of delays on effectiveness of contact tracing4for COVID-19", MedRXiv, https://doi.org/10.1101/2020.05.09.20096289

Kucharski, Adam, et al. (2020), "Early dynamics of transmission and control of COVID-19: a mathematical modelling study", The Lancet Infectious Diseases, March 11, https://doi.org/10.1016/S1473-3099(20)30144-4

Li, Qun, Get al. (2020), "Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia", (January 29 & 31 online, print March 26), N Engl J Med, https://www.nejm.org/doi/full/10.1056/NEJMoa2001316

Liu, Tao, et al. (2020), "Transmission dynamics of 2019 novel coronavirus (2019-nCoV)", https://www.biorxiv.org/content/10.1101/2020.01.25.919787v1 update https://www.biorxiv.org/content/10.1101/2020.01.25.919787v2

Liu, Ying, Albert A Gayle, Annelies Wilder-Smith, and Joacim Rocklöv (2020), "The reproductive number of COVID-19 is higher compared to SARS coronavirus", J Travel Med. 2020 Mar; 27(2): taaa021, (online Feb 13), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7074654/

Lloyd-Smith, J. O., S. J. Schreiber,P. E. Kopp, and W. M. Getz (2005), "Superspreading and the effect of individual variation on disease emergence", Nature. 2005; 438(7066): 355–359, doi: 10.1038/nature04153, PMCID: PMC7094981, PMID: 16292310

MacLean, Oscar A, Richard J Orton, Joshua B Singer, David L Robertson (2020), "No evidence for distinct types in the evolution of SARS-CoV-2", Virus Evolution, Volume 6, Issue 1, January 2020, veaa034, https://doi.org/10.1093/ve/veaa034

Marois, Guillaume, Raya Muttarak, Sergei Scherbov (2020), "Assessing the potential impact of COVID-19 on life expectancy", IIASA WP-20-005

Maugeri, Andrea, Martina Barchitta, Sebastiano Battiato and Antonella Agodi (2020), "Estimation of Unreported Novel Coronavirus (SARS-CoV-2) Infections from Reported Deaths: a Susceptible Exposed Infectious Recovered Dead Model, https://www.preprints.org/manuscript/202004.0052/v1, published in the J. Clin. Med. 2020, 9(5), 1350, https://www.mdpi.com/2077-0383/9/5/1350 (May 5)

McKie, Robin (2020), "UK scientists want to infect volunteers with Covid-19 in race to find vaccine. Trials could be speeded up by using risky strategy of deliberately introducing the virus", Guardian, Sunday May 24

Papachristodoulou, E., L. Kakoullis, K. Parperis, G. Panos (2020), "Long-term and herd immunity against SARS-CoV-2: implications from current and past knowledge", Pathogens and Disease,

Volume 78, Issue 3, April 2020, ftaa025, https://doi.org/10.1093/femspd/ftaa025

Park, Sang Woo, David Champredon, Joshua S. Weitz, Jonathan Dushoff (2019), "A practical genera-tion-interval-based approach to inferring the strength of epidemics from their speed", Epidemics, Volume 27, June 2019, Pages 12-18

Petersen E, Koopmans M, Go U, et al. (2020), "Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics" [published online ahead of print, 2020 Jul 3]. Lancet Infect Dis. 2020;S1473-3099(20)30484-9. doi:10.1016/S1473-3099(20)30484-9

Pueyo, Tomas (2020), "Coronavirus: The Hammer and the Dance", https://medium.com/@-tomaspueyo/coronavirus-the-hammer-and-the-dance-be9337092b56

Richardson, Sylvia, and David Spiegelhalter (2020), "Coronavirus statistics: what can we trust and what should we ignore? ", The Guardian, April 12

Rocklöv, J, H Sjödin, A Wilder-Smith (2020), "COVID-19 outbreak on the Diamond Princess cruise ship: estimating the epidemic potential and effectiveness of public health countermeasures", Journal of Travel Medicine, taaa030, https://doi.org/10.1093/jtm/taaa030

Saito, T. (2020), "Contact-tracing and peer pressure: how Japan has controlled coronavirus", June 6, https://www.theguardian.com/commentisfree/2020/jun/06/contact-tracing-japancoronavirus-covid-19-patients-social-etiquette

Sample, Ian (2020a), "Why we might not get a coronavirus vaccine", Guardian, May 22, https://www.theguardian.com/world/2020/may/22/why-we-might-not-get-a-coronavirus-vaccine

Sample, Ian (2020b), "Immunity to Covid-19 could be lost in months, UK study suggests", https://www.theguardian.com/world/2020/jul/12/immunity-to-covid-19-could-be-lost-in-months-uk-study-suggests

Sample, Ian, and Rowena Mason (2020), "UK could relax lockdown for millions if over-70s are shielded, say scientists. Plan involves beefing up protection for vulnerable people while easing controls on others", Guardian, Tue 5 May 2020

Seow, J. et al. (2020), "Longitudinal evaluation and decline of antibody responses in SARS-CoV-2 infection", https://www.medrxiv.org/content/10.1101/2020.07.09.20148429v1

Spinney, Laura (2020), "Covid-19 expert Karl Friston: 'Germany may have more immunological "dark matter"'. The neuroscientist who advises Independent Sage on Covid-19 discusses the predictive power of his mathematical modelling and the risk of a second wave", Guardian, Sunday May 31

Shridhar, Devi, and Yasmin Rafiei (2020), "The problem with 'shielding' people from coronavirus? It's almost impossible", Guardian, Fri 29 May, modified Mon 1 June

Silver, Nate (2020), "Coronavirus Case Counts Are Meaningless.* (*Unless you know something about testing. And even then, it gets complicated.)", https://fivethirtyeight.com/features/coron-avirus-case-counts-are-meaningless/

Thorp, H., V. Vinson, C. Ash (2020), "Modeling herd immunity", Editor's blog, June 23, https://blogs.-sciencemag.org/editors-blog/2020/06/23/modeling-herd-immunity/

Tindale, Lauren, et al. (2020), "Transmission interval estimates suggest pre-symptomatic spread of COVID-19", MedRxiv https://doi.org/10.1101/2020.03.03.20029983 (March 6)

Verity, Robert, et al. (2020), "Estimates of the severity of coronavirus disease 2019: a model-based analysis", Lancet, March 30, https://doi.org/10.1016/S1473-3099(20)30243-7

Visser, Ellen de (2020), "De slooptocht van het coronavirus in het lichaam: bijna geen orgaan blijft onberoerd", Volkskrant May 17

Vlas, Sake de, Luc Coffeng (2020), "A phased lift of control: a practical strategy to achieve herd immunity against Covid-19 at the country level", April 12, https://www.medrxiv.org/content/10.1101/2020.03.29.20046011v2

Wu JT, Leung K, Leung GM. (2020), "Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study" (January 31), Lancet 395: 689–97, https://doi.org/10.1016/S0140-6736(20)30260-9

Wu, Zunyou, and Jennifer McGoogan (2020), "Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China. Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention", JAMA. 2020;323(13):1239-1242. doi:10.1001/jama.2020.2648

Yu H, Wang XC, Li J, et al. Zhonghua Yu Fang Yi Xue Za Zhi (2020) [Genomic analysis of a 2019-novel coronavirus (2019-nCoV) strain in the first COVID-19 patient found in Hangzhou] ;54(5):486-490. doi:10.3760/cma.j.cn112150-20200217-00128

Zhang J, Litvinova M, Wang W, Wang Y, Deng X, Chen X, et al. (2020), "Evolving epidemiology of novel coronavirus diseases 2019 and possible interruption of local transmission outside Hubei Province in China: a descriptive and modeling study", medRxiv.2020;20026328: (Preprint). Available from: https://doi.org/10.1101/2020.02.21.20026328

R.6. With Mathematica code

See also https://reference.wolfram.com/language/guide/SurvivalAnalysis.html

See also https://demonstrations.wolfram.com/search.html?query=Epidemics

See also https://search.wolfram.com/?query=Epidemics

e.g. https://www.wolfram.com/system-modeler/examples/more/life-sciences/an-influenza-epidemic

Less easy to search older files in https://library.wolfram.com/infocenter/MathSource/

Antonov, Anton (2019), "Answer to Solving and plotting an SIR epidemiology model", https://mathematica.stackexchange.com/questions/208357/solving-and-plotting-an-sir-epidemiology-model

Buzing, Arnoud (2020), "[NB] The SIR Model for Spread of Disease", https://community.wolfram.com/groups/-/m/t/1903289

Chandler, Seth (2011), "A Parameterized Multistate Life Table", https://demonstrations.wolfram.com/AParameterizedMultistateLifeTable/

Colignatus, Thomas (1995, 2020e), "The Economics Pack. Applications of Mathematica", http://thomascool.eu/TheEconomicsPack/index.html

Cool, Thomas (1999, 2001), "The Economics Pack. User Guide", ISBN 90-804774-1-9, ca 555 pp.

(updated edition). JEL reference number is JEL 1999-0820. (Journal of Economic Literature, volume 37, no. 3, September 1999). PDF at http://thomascool.eu/TheEconomicsPack/index.html

Hurt, Jan, and Daniel Hlubinka (2006), "The Kermack-McKendrick Model of Epidemics and Its Extensions", https://library.wolfram.com/infocenter/Conferences/6462/

Kaurov, Vitaliy (2014), "Modeling a Pandemic like Ebola with the Wolfram Language", https://blog.wolfram.com/2014/11/04/modeling-a-pandemic-like-ebola-with-the-wolfram-language/

Kaurov, Vitaliy (2020), "[UPDATES] Resources For Novel Coronavirus COVID-19 ", https://community.wolfram.com/groups/-/m/t/1872608 (with various contributions by others)

Nachbar, Robert (2020), "[Notebook] Epidemiological Models for Influenza and COVID-19", https://community.wolfram.com/groups/-/m/t/1896178

Pegg, Ed (2003), "Superflu modeling", http://www.mathpuzzle.com/MAA/09-Superflu%20Modeling/mathgames_12_22_03.html and https://library.wolfram.com/infocenter/MathSource/4906/

Smith, David, and Lang Moore (2004), "The SIR Model for Spread of Disease", MAA, *Convergence.* The various chapters begin here: https://www.maa.org/press/periodicals/loci/joma/the-sir-model-for-spread-of-disease-introduction

Strain, Steven (2011), "SIR Epidemic Dynamics", https://demonstrations.wolfram.com/SIREpidemicDynamics/

Weisstein, Eric W. (2020b), "Life Expectancy." From MathWorld--A Wolfram Web Resource. https://mathworld.wolfram.com/LifeExpectancy.html

Wolfram Language & System Documentation Center (2020), "Actuarial computation", https://reference.wolfram.com/language/guide/ActuarialComputation.html and https://reference.wolfram.com/language/ref/MortalityData.html

R.7. The Dutch case

https://www.rivm.nl/en/novel-coronavirus-covid-19/calculations/calculation-models-are-public-and-accessible

https://www.ntvg.nl/covid-19

1V EenVandaag (2020), "Het gelijk van microbioloog Alex Friedrich: 'Testen, testen, testen' krijgt nu landelijk navolging", April 17, https://eenvandaag.avrotros.nl/item/het-gelijk-van-microbioloog-alex-friedrich-testen-testen-testen-krijgt-nu-landelijk-navolging/

Bessems, Kustaw (2020), "Luisteren we in de coronacrisis niet naar een te klein groepje deskundigen?", April 9, https://www.volkskrant.nl/cultuur-media/luisteren-we-in-de-coronacrisis-niet-naar-een-te-klein-groepje-deskundigen~bc117875/?referer=https%3A%2F%2Fwww.google.com%2F

Blok, Michael, and Jaap Stronks (2020), "Inkeer doet pijn", May 6, https://joop.bnnvara.nl/opinies/inkeer-doet-pijn (woordvoerders van de actiegroep Containment Nu! die ook de petitie Testen Testen Testen! organiseerde)

Breuker, Remco (2020), "Zuid-Koreaans collectivisme in aanpak coronavirus is een hersenspinsel. Nemen we de Zuid-Koreaanse aanpak wel serieus?", https://www.scienceguide.nl/2020/03/zuid-koreaans-collectivisme-in-aanpak-coronavirus-is-een-hersenspinsel/

Brinkhorst, L.J. and Pieter van Geel (2004), "Kamerbrief 29200-XI-125",
https://zoek.officielebekendmakingen.nl/kst-29200-XI-125.html

Bruin, Jonathan de (2020), "CoronaWatchNL. Dataset: COVID-19 case counts in The Netherlands",
https://github.com/J535D165/CoronaWatchNL/blob/master/README.md

CBS (2020), "Sterfte in coronatijd", 29 mei, https://www.cbs.nl/nl-nl/nieuws/2020/22/sterfte-in-coronatijd#fd35eac6-fa0b-422e-ae86-47526e8252d9

Erp, Pepijn van (2020), "De denkfouten van Willem Engel over het coronavirus", https://klopt-datwel.nl/2020/06/06/de-denkfouten-van-willem-engel-het-coronavirus/, zie ook https://www.maurice.nl/2020/05/30/kijk-naar-dit-interview/

FMS & KNMG (2020), "Triage op basis van niet-medische overwegingen voor IC-opname ten tijde van Fase 3 stap C in de COVID-19 pandemie", https://www.demedischspecialist.nl/nieuws/publicatie-draaiboek-keuzes-ic-opname-bij-schaarste with a link to the pdf

Goudsmit, Jaap (2020), "Zijn wij niet meer dan kuddedieren? Hoe de groepsimmuniteit kwam en ging", https://jaapgoudsmit-over-corona.nl/nl/blog/2020/03/21/zijn-wij-niet-meer-dan-kud-dedieren-hoe-de-groepsimmuniteit-kwam-en-ging/

Gupta (2020), "In de slipstream van corona: een secundaire crisis in de zorg", March 23, https://gupta-strategists.nl/studies/in-de-slipstream-van-corona-een-secundaire-crisis-in-de-zorg

Hendrickx, Frank, and Huib Modderkolk (2020), "Februari: de verloren maand in de strijd tegen het coronavirus", April 11, https://www.volkskrant.nl/nieuws-achtergrond/februari-de-verloren-maand-in-de-strijd-tegen-het-coronavirus~b09e4c7a8/

Husby, Trond, and Lenny Stoeldraijer and Hans Visser (2020), "Statistische Trends Oversterfte tijdens de corona-epidemie: toepassing van een dynamisch regressiemodel", CBS, Statistische Trends,

Jaspers, Arnout (2020), "Relativerend rekenen aan Covid-19", March 20, https://www.nemokennislink.nl/publicaties/relativerend-rekenen-aan-covid-19/

KAG, Koninklijk Actuarieel Genootschap (2020), "AG-Commissie Sterfte Onderzoek: effecten levensverwachting Covid-19 op korte en lange termijn", https://www.ag-ai.nl/nieuws.php?action=view&Nieuws_Id=909

Keulemans, Maarten (2020), "Interview met Jaap van Dissel (RIVM)", March 12, https://www.volkskrant.nl/nieuws-achtergrond/doen-we-wel-genoeg-rivm-baas-van-dissel-zodra-iets-op-besmetting-in-het-gezin-wijst-isolatie~bc5d1b14/

KNAW (2020), "Videoregistratie, bijeenkomst het nieuwe coronavirus", https://www.knaw.nl/nl/actueel/nieuws/videoregistratie-bijeenkomst-het-nieuwe-coronavirus

NVIC & Diederik Gommers (2020), "Covid-19 - bericht van de voorzitter 9", April 2020, https://nvic.nl/sites/nvic.nl/files/20200426%20bericht%20van%20de%20voorzitter%209.pdf

NVIC (2020), "IC-capaciteit | Vraag naar IC capaciteit zal door COVID IC 639 bedden hoger zijn de komende 3 jaar", April 24, NVIC

RIVM (2020a), "Advies n.a.v. OMT 2019-nCoV (Wuhan)", January 27, https://zoek.officielebekendmakingen.nl/blg-921175.pdf

RIVM (2020b), "Epidemiologische situatie COVID-19 in Nederland", COVID-19_WebSite_rapport_20200520_1022.pdf

RIVM, Jaap van Dissel (2020a), "Powerpoint presentation for the Dutch House of Parliament", March 25,
https://www.tweedekamer.nl/sites/default/files/atoms/files/20200325_briefing_coronavirus_tweede_kamer_presentatie_rivm.pdf

RIVM, Jaap van Dissel (2020b), "Powerpoint presentation for the Dutch House of Parliament", April 8, https://www.tweedekamer.nl/kamerstukken/detail?id=2020D13392&did=2020D13392

RIVM, Jaap van Dissel (2020c), "Powerpoint presentation for the Dutch House of Parliament", April 22, https://www.tweedekamer.nl/sites/default/files/atoms/files/20200422_technische_briefing_-jaap_van_dissel_rivm_22_april.pdf

RIVM, Jaap van Dissel (2020d), "Powerpoint presentation for the Dutch House of Parliament", May 7, https://www.tweedekamer.nl/downloads/document?id=e5444e8a-886d-4efa-9dda-c8fc9b70188c&title=PowerPoint%20presentatie%20dhr.%20Van%20Dissel%20-%20RIVM.pdf

RIVM, Jaap van Dissel (2020e), "Powerpoint presentation for the Dutch House of Parliament", May 20, https://www.tweedekamer.nl/sites/default/files/atoms/files/presentatie_jaap_van_dissel_-_technische_briefing_20_mei_2020.pdf

RIVM, Jaap van Dissel (2020f), "Powerpoint presentation for the Dutch House of Parliament", May 25, https://www.tweedekamer.nl/sites/default/files/atoms/files/20200525_technische_briefing_l-nv_presentatie_jaap_van_dissel.pdf

RIVM, Outbreak Management Team (OMT) (2020), "Inhoudelijke onderbouwing met betrekking tot a-, pre- en vroegsymptomatische transmissie SARS-CoV-2", (May 6) https://lci.rivm.nl/covid-19/bijlage/onderbouwing-a-pre-vroegsymptomatische-transmissie

Rek, Wilma de, and Jaap Goudsmit (2020), "Jaap Goudsmit over de oorlog (die hij in getallen vatte) en de coronacrisis (waarin hij gegevens mist)", de Volkskrant, May 1

Rutte, Mark (2020), Rijksoverheid, "TV-toespraak van minister-president Mark Rutte", https://www.rijksoverheid.nl/documenten/toespraken/2020/03/16/tv-toespraak-van-minister-president-mark-rutte

Slot, Ed, et al. (2020), "Herd immunity is not a realistic exit strategy during a COVID-19 outbreak", April 29, Research Square, In Review, https://www.researchsquare.com/article/rs-25862/v1

Stokmans, Derk, en Mark Lievisse Adriaanse (2020), "Hoe Nederland de controle verloor De corona-uitbraak van dag tot dag", NRC-Handelsblad, 19 juni 2020

Tweebeeke, Marielle, and Jaap van Dissel (2020), "Hoe werkt groepsimmuniteit?", broadcast March 16, https://www.npostart.nl/hoe-werkt-groepsimmuniteit-tegen-corona/16-03-2020/POM-S_NOS_16034089

Tweebeeke, Marielle, and Siebe Sietsma (2020), "Duizenden coronatesten bleven op de plank liggen", June 10, https://www.npostart.nl/nieuwsuur/09-06-2020/VPWON_1310822

R.8. Economics about the pandemic

Acemoglu, Daron, Victor Chernozhukov, Iván Werning, Michael D. Whinston (2020), "Optimal Targeted Lockdowns in a Multi-Group SIR Model", NBER Working Paper No. 27102, https://www.nber.org/papers/w27102

Baldwin, Richard (2020a), "It's not exponential: An economist's view of the epidemiological curve", VoxEU, March 12, https://voxeu.org/article/it-s-not-exponential-economist-s-view-epidemiological-curve

Baldwin, Richard (2020b), "VoxEU readership in the COVID-19 crisis ", VoxEU, May 7, https://voxeu.org/article/voxeu-readership-covid-19-crisis

Baldwin, Richard, and Beatrice Weder di Mauro (eds) (2020), "Economics in the Time of COVID-19", A VoxEU.org Book, CEPR Press

Berger, David, Kyle Herkendorff, Simon Mongey (2020), "An SEIR Infectious Disease Model with Testing and Conditional Quarantine", Federal Reserve Bank of Minneapolis, Staff report 597, April, https://doi.org/10.21034/sr.597. The Python code is at https://sites.google.com/site/kyleherkenhoff/research

Bodenstein, Martin, Giancarlo Corsetti, and Luca Guerrieri (2020), "Social distancing and supply disruptions in a pandemic", Cambridge working papers in economics 2031, http://www.econ.cam.ac.uk/research-files/repec/cam/pdf/cwpe2031.pdf

Colignatus, Thomas (2020a), "Forum Theory & A National Assembly of Science and Learning", MijnBestseller.nl, https://mpra.ub.uni-muenchen.de/98568, http://thomascool.eu/Papers/FT&NASL/Index.html (2nd edition July 5)

Colignatus, Thomas (2020f), "The Tinbergen & Hueting Approach in the Economics of National Accounts and Ecological Survival", draft, http://thomascool.eu/Papers/THAENAES/Index.html

Colignatus, Thomas (2020g), "National accounts in the anthropocene: Hueting's environmental functions and environmentally sustainable national income: Translation and relevance for ecosystem services", Ecosystem Services, Volume 43, June 2020, 101094, https://www.sciencedirect.com/science/article/abs/pii/S221204162030036X

Rowthorn, Robert, and Flavio Toxvaerd (2013, 2020), "The Optimal Control of Infectious Diseases via Prevention and Treatment", Cambridge wp2013, https://www.inet.econ.cam.ac.uk/research-papers/wp-abstracts?wp=2013

Tabarrok, Alex (2020), "A Solution if We Act", March 30, https://marginalrevolution.com/marginalrevolution/2020/03/a-solution-if-we-act.html

Thurstrom, Linda, et al. (2020), "The Benefits and Costs of Using Social Distancing to Flatten the Curve for COVID-19", Forthcoming Journal of Benefit-Cost Analysis, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3561934

Wren-Lewis, Simon (2020), "Following the Science or a National Scandal", https://mainlymacro.blogspot.com/2020/03/following-science-or-national-scandal.html

R.9. Other redesign of didactics

Colignatus, Thomas (2009, 2015), "Elegance with Substance", Samuel van Houten Genootschap, https://zenodo.org/record/291974

Colignatus, Thomas (2018a), "A child wants nice and no mean numbers", Samuel van Houten Genootschap, MijnBestseller.nl, https://zenodo.org/record/1434693

Colignatus, Thomas (2018b), "Arithmetic with H = -1: subtraction, negative numbers, division, rationals and mixed numbers", https://zenodo.org/record/1251687