Package 'TwinSimPublic'

March 6, 2020

Title Dizygotic twinning in humans

Version 0.9.1

Description This simulation of dizygotic twinning in humans uses data from Gambia, 1950 - 1975.

Depends R (>= 3.5.0)

License Private package

LazyData true

Imports binom

Suggests knitr (>= 0.1.0)

RoxygenNote 6.0.1

R topics documented:

---

paritiesinputs	*A simulation of dizygotic twinning in humans for up to 20 parities*

Description

This function was used to conduct the simualtions described in:

Wade N Hazel, Robert Black, Richard C Smock, Rebecca Sear, Joseph L Tomkins 2020. An age-dependent ovulatory strategy explains the evolution of dizygotic twinning in humans, Nature Ecology and Evolution 2020. DOI to be finalized.

This function estimates lifetime reproduction in a cohort of women by simulating ovulatory cycles that can result in successful live singleton or twin births depending on inputs controlling prenatal and postnatal mortality, probabilities of single or double ovulation, maternal mortality both during child birth and independent of childbirth. It allows for up to 20 parities, makes an object that is a list of seven objects that can be further analysed and saves that object in the working directory as an .Rda file for future use. A set of input variables needed by the function, based on data from Gambia is provided by the function "prepGambiaBasic".

1

Usage

```
paritiesinputs(Title, inputs)
```

Arguments

Title	a term in quotation marks that identifies the simulation; it becomes a column in the objects "sim" and "LONG" returned by the function.
inputs	a named vector of 44 input variables produced by "prepGambiaBasic". See the help file for "prepGambiaBasic" for i) a definition of each input variable, ii) the value assigned to each, and iii) how to alter the default values.

Details

This function does the simulation, producing an object "sim" with a row for each mother for 20 parities, and another version of "sim", called "LONG" with a row for each mother for each parity. It also returns The Title, the input values used for the simulation, a summary of each parity (tallies), details of surviving menopause, and the postnatal interval. See the "Value" section below for a description of these objects.

The function repeats a series of steps that produce 26 variables (all num except for ID as int, and Title as factor) for each of 20 parities, based on information supplied as input variables explained in the help page for "prepGambiaBasic". The function prints to the R console Parity, mean q and mean u as an indication of its progress, and as a check that the simulation is behaving as expected.

The function saves an .Rda file to the working directory. See the "Value" section below.

The following list provides the details of the 26 variables it produces for each parity.

INITIALIZED VARIABLES (4)

ID - int: values from 1 to "ind"; identifies each of the "ind" individuals

Parity - num: values from 1 to 20

BMI - num: a random number from a normal distribution with mean of "BMIm"" and a standard deviaton of "BMISD"; remains constant over all parities; BMIe is an effect size (0.05).

agecyc - num: initiated for Parity 1 as a random number from a normal distribution with mean of "agefmm" and a standard deviation of "agefmSD"; for subsequent parities, it is calculated from "agelb" and "postyr" (see definitions of these below)

PRENATAL VARIABLES (17)

q - num: proportion of double ovulations usually varying with age, switchpoint mean and switch-point standard deviation (as "agecyc", "spm", and "spSD" - see "prepGambiaBasic"- according to R code "q <- pnorm(agecyc, spm, spmSD)") but when constantQ = 1, it is set to this default value. Its value can depend on other input variables: if "constantQ" is 1, it will have the input variable value; if "constantQ" = 0 and "BMIoff" = 1 - increase q by age only; if "constantQ" = 0 and "BMIoff" = 0 - increase q by age and BMI, dependent on size of BMIe, and reduce asr when BMI is less than or greater than 23.75, according to Prospective Studies Collaboration. 2009. "Body-mass index and cause-specific mortality in 900 000 adults: collaborative analyses of 57 prospective studies." Lancet 373, 1083-1096 Figure 2, and , and adjust stills and stillt according to P.W.G. Tennant, J. Rankin, and R. Bell. 2011. "Maternal body mass index and the risk of fetal and infant death: a cohort study from the North of England." Human Reproduction 26, 1501-1511, Figure 1. Adjustments to asr and stills and stillt have small effects.

p - num: proportion of single ovulations p = (1-q)

u - num: probability of live birth calculated as 0.55 * 0.89^(agecyc - 18) where 0.55 comes from O'Connor, K. A., Holman, D. J. & Wood, J. W. Declining fecundity and ovarian ageing in natural

fertility populations. Maturitas 30, 127-136, doi:10.1016/s0378-5122(98)00068-1 (1998), and 0.89 is an average of several data sets that used MicroSoft Excel's Solver to fit this non-linear equation (Hazel et al. in preparation). u delcines with agecyc from 0.55 at agecyc = 18. In Parity 1, agecyc is set to 14 even if the combination of agefmm and agefmSD predict lower agecyc.

r - num: probability of early loss calculated as 1 - (u + s + t); see previous variable and next two variables for u, s, and t; r increases with age from about 0.38 at agecyc = 18, and approaches 1.0.

s - num: probability of abortion calculated as $0.185559 - 0.010243 * agecyc + 0.000153 * agecyc^2$ based on u and fits to Danish data in Andersen, A.-M. N., Wohlfahrt, J., Christens, P., Olsen, J. & Melbye, M. Maternal age and fetal loss: population based register linkage study. BMJ 320, 1708-1712, doi:10.1136/bmj.320.7251.1708 (2000). This is a slightly U-shaped function that starts at about 0.05 at agecyc = 18.

t - num: probability of late loss calculated as $11.849 * agecyc^{-2.916}$ based on u and fits to Danish data (Anderson et al. 2002). This function produces a value of about 0.003 at agecyc = 18, and declines by an order of magnitude at agecyc ~ 40.

The next eight variables are calculated from an interative process for each individual that:

- checks that "agecyc" is not NA (caused by previous death)

- checks if "agecyc" is less than "meno"

- samples an array of 14 outcomes of each ovulation cycle (1.1 to 2.99 - see "preRes" below)

- continues to sample the ovulation outcomes until a successful pregnancy occurs (outcome for "preRes" one of 1.4, 2.4, 2.7, 2.9 or 2.99), all the while keeping track of several things:

i) accumulated time (based on input variables "vm", "vSD", "wm", "wSD", "xm", and "xSD" - see "prepGambiaBasic" for definitions of these variables, and preRes below for how they are used),

ii) tallies of single and double ovulations,

iii) whether a single or double ovulation resulted in a successful pregnancy, and

iv) which of the several successful outcomes of the cycle which lead to successful pregnancy it was (i.e., one of "preRes" = one of 1.4, 2.4, 2.7, 2.9 or 2.99).

This iterative process is inefficient and probably makes up most of the time that a simulation takes.

nova - num: number of ova expended from start of recycling to successful pregnancy (including the ovum that led to successful pregnancy)

sovul - num: number of singleton ova expended from start of recycling to successful pregnancy (including the ovum that led to successful pregnancy)

douvl - num: number of doubleton ova expended from start of recycling to successful pregnancy (including the ova that led to successful pregnancy)

SorT - num: the number of ova that led to a successful pregnancy; 1 = singleton, 2 = double

preyr - num: the number of years taken from start of recycling to successful pregnancy; this generally increases with age because of the decline of u

preRes - num: a numerical code representing the outcome of a singe ovulation cycle; v, w, and x times referred to below come from vectors of values derived from sampling normal distributions with means and standard deviations specified by vm and vSD, and so on.

["preRes"code] [Product involving p, q, r, s, t, and u] [= description of outcome and times to add]

1.1 pr = probability single ovulation and early loss, add v time

1.2 ps = probability single ovulation and abortion, add w time

1.3 pt = probability single ovulation and late loss, add x time

1.4 pu = probability single ovulation live birth, time added depends on post embryonic mortality

2.1 qr^2 = probability double ovulation and both suffer early loss, add v time

2.2 qs^2 = probability double ovulation and both suffer abortion, add w time

2.3 qt^2 = probability double ovulation and both suffer late loss, add x time. preRes 2.3 can be modified by the input variable vt (virtual twin syndrome = proportion of singletons from double ovulations (preRes = 2.99) that disappear according to the "vanishing twin" syndrome; they get added to preRes = 2.3 (both twins die before birth). The value must be between 0.0 and 1.0. The default value is 0.

2.4 qu^2 = probability double ovulation and both live born, time added depends on postnatal mortality

2.5 q2rs = probability double ovulation and one suffers early loss and suffers one abortion, add w time

2.6 q2rt = probability double ovulation and one early loss and one late loss, add x time

2.7 q2ru = probability double ovulation and one early loss and one live born, time added depends on postnatal mortality

2.8 q2st = probability double ovulation and one abortion and one late loss, add x time

2.9 q2su = probability double ovulation and one abortion and one live born, time added depends on postnatal mortality

2.99 q2tu = probability double ovulation and one late loss and other live born, time added depends on postnatal mortality. preRes 2.3 can be modified by the input variable vt (virtual twin syndrome = proportion of singletons from double ovulations (preRes = 2.99) that disappear according to the "vanishing twin" syndrome; they get added to preRes = 2.3 (both twins die before birth). The value must be between 0.0 and 1.0. The default value is 0.

nsurv - num: maternal survival from start of cycling to birth calculated for each individual as a sample with replacement of 1 or 0 with probabilities of 1 - "asr"* and "asr"* where "asr"* is "asr" raised to power of ("preyr" + 0.75) in years. That is individuals get killed-off during the time they are cycling before getting pregnant ("preyr") and while they are pregnant (0.75 years), at the adult survival rate (input variable "asr") which is per year. If a maternal death is indicated, both the mother and potential offspring die, and "agelb" and "live" variables (see below) are assigned NA; the mother does not contribute further to the simulation.

cbsurv - num: childbirth survival calculated for each individual as a sample with replacement of 1 or 0 with probabilities of 1 - "scbt" and "scbt" for twins when "preRes" is = 2.4 (twins born), and "twinsoff" = 0, and with probabilities of 1 - "scbs" and "scbs" for twins when "twinsoff" = 1 and for singletons of whatever source (preRes = 1.4, 2.7, 2.9, and 2.99). If a maternal death is indicated, both the mother and potential offspring die, and "agelb" and "live" variables (see below) are assigned NA; the mother does not contribute further to the simulation.

The value for "twinsoff" influences the next three variables.

First, if "twinsoff" = 1, a successful pregnancy from a double ovulation (code 2.4 in "preRes" above) results in a live birth ("live" see below) of only 1, not 2 as expected from the code 2.4. "twinsoff" = 1 allows double ovulation and all associated prenatal outcomes but alows the birth of only a single offspring. That offspirng is treated as a singleton for childbirth survival of the mother and all postnatal survival probabilities.

Second, the rate of stillbirths depends on the value of "preRes": for 2.4 and "twinsoff" = 0, the rate is "stillt", otherwise it is "stills" because "twinsoff" = 1 prevents twins. The values for "Searstill" come from vectors with length "ind", produced from sampling with replacement 0, 1, or 2 with probabilities of (1-stillt)^2, 2*(1-stillt)*stillt, stillt^2) for twin births or with probabilities of 1 - "stills" and "stills" for singleton births from code 1.4 or 2.4 with "twinsoff" = 1

Third, "nsurv" and "cbsurv" kill both mother and any potential live birth. "Searstill" kills only the potential live birth, and the mother survives. If either of "nsurv" or "cbsurv" for an individual mother are 0, the values for "agelb" and "live" are set to NA, because that pregnancy killed mother

and offspring. If "Searstill" is 0, and both "nsurv" and "cbsurv" are 1, "agelb" gets a numeric value because the mother survives the stillbirth, but "live" is assigned NA because of the death by stillbirth.

Searstill - num: calculated from samples of 0 or 1 and "stills" for singletons (preRes = 1.4, 2.7, 2.9, 2.99), or samples of 0, 1 or 2 and "stillt" for twins (preRes = 2.4). If the mother has not been killed off by "nsurv" or "cbsurv" already, "Searstill" can impose the death of 0, 1 or 2 offspring if it is a twin pregnancy (preRes = 2.4), with probabilities associated with "stillt", or 0 or 1 offspring if it is a singleton pregnancy (preRes = 1.4, 2.7, 2.9, 2.99) with probabilities associated with "stills". See previous section for details and caveats. See "prepGambiaBasic" for "stills" and "stillt".

agelb - num: age at giving birth, calculated by summing "agecyc" and "preyr"; NA if mother died according to "nusurv" and "cbsurv" or suffered .

live - num: number of live offspring, either 1 or 2; NA if mother died according to "nusurv" and "cbsurv" and see paragraph above "Searstill" for details and caveats.

POSTNATAL VARIABLES (6)

The next four variables describe, for each parity, the survival of the offspring. They depend on whether the birth was a singleton (from a single ovulation or a double ovulation that produced a single birth) or a pair of twins, and the 8 input variables, "sb1" to "t215" described in "prepGambiaBasic" and calculated from R. Sear,* D. Shanley, I. A. MCGregor & R. Mace. The fitness of twin mothers: evidence from rural Gambia. Evol. Biol. 14, 443- (2001) Figure 2.

The process is similar for each variable. Vectors "ind" long are made by sampling with replacement from 0 or 1 or 2 with probabilities (1 - survival rate)^2, 2 * (1 - survival rate) * survival rate, and survival rate^2 for twins, and from 0 or 1 with probabilities of (1- survival rate) and survival rate for singletons.

month1 - num: the number of offspring surviving from birth to 1 month

year1 - num: the number of offspring surviving from 1 month to 12 months or 1 year

year2 - num: the number of offspring surviving from 1 year to 2 years, considered to be the age of weaning

year15 - num: the number of offspring surviving from 2 years to 15 years

postyr - num: the number of years from having a live birth until the mother starts cycling again for the next parity; this requires calculations involving *Mothers' times to start cycling again* (i.e., input variables "jm" to "mSD" in "prepGambiaBasic") which are used to produce vectors "ind" long of samples from random normal distributions with the appropriate means and standard deviations, dependant on the fate of the mothers' offspring as described by input variables "sb1" to "t215" in "prepGambiaBasic".

Title - character: argument "Title" used in the call to the function.

Value

The function returns an object called "outputs" which is a list of seven components, described below:

Title	a character object, as specified by Title in the call to the function
inputs	a named vector of 42 input variables
tallies	This data frame gets calculated as the 20 parities are created by the simulation. This is a brief account of how the 20 rows are related to each other, using Parity1 column as an example.
	"livemothersbegin" will be "ind", and no mothers will be listed for "postnatalmeno" (but will be in later parities, depending on the value for "meno").

"postnatal_meno" is the number of mothers excluded because their age exceeds "meno" after their postnatal interval ("postyr") following the birth in their last parity ("agelb")

"naturaldeaths" are the number of deaths of mothers during cycling and pregnancy according to input variable "asr"

"childbirthdeaths" are the number of deaths of mothers (their offspring die too)

"prenatal_meno" is the number of mothers exceeding age of menopause ("meno") during the "preyr" interval while they cycle until they attain a pregnancy

"livemothersbirth" is "livemothersbegin" less the sum of "postnatal_meno", "naturaldeaths", "childbirthdeaths" and "prenatal_meno" (except when it does not; I could not find a few extra deaths in late parities)

"Searstillbirths" is the number of mothers with stillbirths, in which the offspring die, but the mothers survive.

"motherswithliveoffspring" is "livemothersbirth" less "Searstillbirths" and the sum of "singletons" and "pairs of twins"

"singletons" is the number of single births

"pairsoftwins" is the number of twin births

"twinsper1000births" is 1000 x ("pairsoftwins" / ("pairsoftwins" + "singletons")); the stillbirths are not considered as live births.

"postdeaths" is the number of mothers dying during the "postyr" interval, according to "asr".

"motherspost" is the numbers of mothers surviving the "postyr" interval (and the number starting the next parity)

The last 7 variables are means. The means are calculated from "LONG", and the sample sizes for each parity are the "livemothersbegin", and not "ind" of the initiating cohort. The means are of the variables agelb, q, u, r, preyr, postyr, ibi (inter-birth interval, in months).

sim   a data frame of "ind" rows and 521 variables named X1 to X521. See LONG below for names for these variables that repeat for the 20 parities. X521 is "Title".

LONG   a data.frame of "ind" * 20 rows and 27 variables described in "Details" above.

menosurv   a data.frame of "ind" rows and 20 variables (for Parity 1 to 20) of zero or one indicating whether the mother has reached menopause during preyr interval

pnsurv   a data frame of "ind" rows and 20 variables (for Parity 1 to 20) of zero or one indicating whether the mother suffered death at the natural mortality rate (asr) during the postnatal interval.

The function also saves a binary R file to the working directory, named "Title_.rda" where Title is the "Title" supplied in the call to the function. Loading "Title_.rda" puts an object called "outputs" in R's Global Environment, overwriting any prexisting object called "outputs" (so be careful). "outputs" can be used in a call to "analyses" and/or "plottallies" for a series of summaries and plots of the simulation saved in "Title_.rda".

Author(s)

Bob Black, Joe Tomkins, Wade Hazel

Examples

```
# get a set of input variables
inputs <- prepGambiaBasic()

# call the function, assigning a name to what it returns
outputs <- paritiesinputs('gb', inputs)

################## end Example ############################
```

---

prepGambiaBasic   *Specify input variables for "GambiaBasic"*

Description

This specifies the default values for 44 input variables for the Gambia data.

Usage

```
prepGambiaBasic(Title = "GambiaBasic", ind = 1000, agefmm = 18.0, agefmSD = 1.8,
          BMIm = 20.7, BMISD = 2.3, BMIe = 0.01, meno = 40.0, q = 0.11,
          spm = 41.1, spSD = 11.6,
          drb = 0.89, dra = 0.55, vm = 1.5, vSD = 0.05, wm = 2.5, wSD = 0.15,
          xm = 7.5, xSD = 0.5, asr = 0.993, scbs = 0.992, scbt = 0.967,
          stillt = 0.889, stills = 0.962, vt = 0.0, sb1 = 0.93, s112 = 0.86,
          s12 = 0.878, s215 = 0.662, tb1 = 0.57, t112 = 0.698, t12 = 0.796,
          t215 = 0.537, jm = 2.3, jSD = 0.05, k1m = 6.0, k1SD = 1.0,
          k2m = 18.0, k2SD = 1.0, mm = 26.0, mSD = 2.0,
          BMIoff = 1.0, constantQ = 0.0, twinsoff = 0.0, console = 1.0, ...)
```

Arguments

Title	a name in quotation marks used to identify the set of input variables. See usage, above, for default values for Title and the other arguments.
ind	number of mothers starting the simulation
agefmm	mean age at first menstruation (this is "agecyc" for Parity 1)
agefmSD	standard deviation of mean age at first menstruation
BMIm	Body Mass Index mean
BMISD	mean BMI standard deviation
BMIe	Effect size of Body Mass Index used to adjust q (proportion of cycles with double ovulations - see below) for age and BMI when BMIoff (see below) is 0. The default value is an arbitrary value that seemed to work. Increased from 0.01 to 0.05 in TwinSim0.8.6, 20190207 and back to 0.01 in TwinSim0.9.0
meno	age at menopause
q	proportion of double ovulations; proportion of single ovulations is p = (1-q); used only when constantQ = 1. Otherwise, q varies with age according to R code "q <- pnorm(agecyc, spm, spmSD)", where "agecyc" is age at start of cycling for a new parity, and "spm" and "spSD" are defined below). If BMIoff = 0, q is also adjusted by BMIe.

spm	Switchpoint mean
spSD	Switchpoint standard deviation
drb	slope of decline of u with age
dra	interept of decline of u with age at agecyc 18
vm	Mean months to next cycle after early loss.
vSD	Standard deviation mean of months to next cycle after early loss.
wm	Mean months to next cycle after abortion.
wSD	Standard deviation of mean of months to next cycle after abortion.
xm	Mean months to next cycle after late loss.
xSD	Standard deviation of months to next cycle after late loss.
asr	Annual survival rate from Sear et al. 2001. asr gets reduced for values of BMI greater than and less than 23.75 by a small amount if BMIoff = 0.
scbs	Probability of surviving singleton childbirth from Sear et al. 2001.
scbt	Probability of surviving twin childbirth from Sear et al. 2001.
stillt	Rate of surviving possibility of still birth: twins; from Sear et al. 2001. If BMIoff = 0, stillt gets reduced by a small amount for values greater than and less than BMI = 23.
stills	Rate of surviving possibility of still birth: singleton from Sear et al. 2001. If BMIoff = 0, stills gets reduced by a small amount for values greater than and less than BMI = 23.
vt	proportion of singletons from double ovulations (preRes = 2.99) that disappear according to the "vanishing twin" syndrome; they get added to preRes = 2.3 (both twins die before birth). The value must be between 0.0 and 1.0.
sb1	Survival of offspring calculated from Sear et al. 2001 Figure 2 - Singleton: survival rate birth to 1 month
s112	Singleton: survival rate 1 month to 12 months from Sear et al. 2001.
s12	Singleton: survival rate 1 year to to years (weaning) from Sear et al. 2001.
s215	Singleton: survival rate 2 years to 15 years from Sear et al. 2001.
tb1	Twin: survival rate birth to 1 month from Sear et al. 2001.
t112	Twin: survival rate 1 month to 12 months from Sear et al. 2001.
t12	Twin: survival rate 1 year to to years (weaning) from Sear et al. 2001.
t215	Twin: survival rate 2 years to 15 years from Sear et al. 2001.
jm	Months to add if live birth dies before 1 month
jSD	Standard deviation of months to add if live birth dies before 1 month
k1m	Months to add if live birth dies before 1 year
k1SD	Standard deviation of months to add if live birth dies before 1 year
k2m	Months to add if live birth dies before 2 years
k2SD	Standard deviation of months to add if live birth dies before 2 years
mm	Months to add if live birth reaches 2 years (weaning)
mSD	Standard deviation of months to add if live birth reaches 2 years (weaning)
BMIoff	a switch for determining the adjustment of q for BMI and of asr for BMI: 1 is no adjustment of q for BMI; 0 makes the adjustment

constantQ	0 allows q (see below) to vary by parity;
	1 makes q constant for all parities. Use in conjunction with BMIoff:
	constantQ = 0 and BMIoff = 1 - adjust q by age only
	constantQ = 0 and BMIoff = 0 - adjust q by age and BMI
twinsoff	a switch to control the production of twins from double ovulations: twins = 0, no twins = 1
console	a switch to control printing to the console. console = 0 turns off the printing.
...	used to alter the default values; see examples for an illustration of its use

Value

a named vector of length 44, used directly as input to "paritiesinputs". See the second example below for how to modify the default values listed in the "Usage"section above.

Author(s)

Bob Black

Examples

```
# prints to R console a formatted listing of the input variables
# and returns a named vector of length 44
gb <- prepGambiaBasic()

# see the named vector
gb

# alter some of the default values
gba <- prepGambiaBasic("GambiaBasic_altered",
                       ind = 100, agefmm = 15, agefmSD = 0)

############## end of Exampples #####################
```

Index