Main figures for article on 20 years of the npde adventure, for AAPS J

Emmanuelle Comets

19/02/2021

Introduction

This notebook is an R Markdown document containing R code for key figures in the paper **Developing tools to evaluate non-linear mixed effect models: 20 years on the npde adventure**, by Emmanuelle Comets and France Mentré, submitted to AAPS Journal for the special edition *Women in Science*. It is bundled in a Zenodo archive along with all the files necessary to reproduce the code and figures herein.

This document can be used to reproduce the main figures in the paper. It includes the citation data used in the article, and shows the data management applied to extract relevant keywords. A number of libraries are required for this code to run (see first code chunk).

At the end, we also show how to compute the ndpe for the viral load example used in the paper, producing the default graphs from the new npde 3.0 package. The npde package submitted to CRAN on the same day is included in the Zenodo archive (Linux version).

Loading libraries

```
knitr::opts_chunk$set(echo = TRUE)
# Libraries
library(gridExtra)
library(ggplot2)
library(grid)
library(RColorBrewer)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following object is masked from 'package:gridExtra':
##
##     combine
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(devtools)
```

```
## Loading required package: usethis
```

```
library(mclust)
```

```
## Package 'mclust' version 5.4.6
## Type 'citation("mclust")' for citing this R package in publications.
```

```
# Reading Excel files (?)
library("readxl")

# Word count libraries
library("tm")
```

```
## Loading required package: NLP
```

```
##
## Attaching package: 'NLP'
```

```
## The following object is masked from 'package:ggplot2':
##
##     annotate
```

```
library("SnowballC")
library("wordcloud")
```

Most cited articles

Database of citations for 5 most cited articles

The citation data was downloaded from Web of Science on October 1st, 2020, for the 5 foundation articles mentioned in the submitted article.

```
cite.list<-dir(pattern="citations")
cite.data<-vector(length=length(cite.list),mode="list")
for(ipub in 1:length(cite.list)) {
  cite.data[[ipub]]<- read_excel(cite.list[ipub])
}
```

```
## New names:
## * `` -> ...68
## New names:
## * `` -> ...68
## New names:
## * `` -> ...68
```

```
nampub<-unlist(strsplit(cite.list,"citations_"))
nampub<-unlist(strsplit(nampub,".xls"))
nampub<-nampub[nampub!=""]
```

Checking autocitations for each article

```
for(ipub in 1:length(cite.list)) {
  datcite<-cite.data[[ipub]]
  id1<-grep("Comets",datcite$Authors)
  id2<-grep("Mentr",datcite$Authors)
  id3<-grep("Brendel",datcite$Authors)
  id<-unique(c(id1,id2,id3))
  cat("Publication:",nampub[ipub],"\n")
```

```
  cat("Number of citations:",dim(datcite)[1],"including",length(id),"auto-citations \n")
  cat("        number of articles:",dim(datcite[datcite$`Document Type`=="Article",])[1],"\n")
  print(table(datcite$`Publication Type`))
  print(table(datcite$`Document Type`))
}
```

```
## Publication: cometsPackagenpde
## Number of citations: 264 including 13 auto-citations
##         number of articles: 246
##
##   B   J   S
##   2 261   1
##
##                           Article            Article; Book Chapter
##                               246                                2
##            Article; Early Access               Editorial Material
##                                 8                                1
## Editorial Material; Book Chapter                           Letter
##                                 1                                3
##                 Meeting Abstract                           Review
##                                 1                                2
## Publication: escolano06
## Number of citations: 63 including 10 auto-citations
##         number of articles: 53
##
##  B  J  S
##  2 60  1
##
##                           Article            Article; Book Chapter
##                                53                                2
##            Article; Early Access Editorial Material; Book Chapter
##                                 1                                1
##                            Letter                 Meeting Abstract
##                                 2                                1
##                            Review
##                                 3
## Publication: karlJPKPD10
## Number of citations: 49 including 5 auto-citations
##         number of articles: 39
##
##  B  J
##  1 48
##
##               Article Article; Book Chapter Article; Early Access
##                    39                    1                     3
##                Letter      Meeting Abstract                Review
##                     3                    2                     1
## Publication: karlPharmRes06
## Number of citations: 217 including 14 auto-citations
##         number of articles: 199
##
##   B   J
##   2 215
##
```

```
##               Article Article; Book Chapter Article; Early Access
##                   199                     2                      2
##    Editorial Material                Letter      Meeting Abstract
##                     1                     2                      1
##             News Item                Review
##                     1                     9
## Publication: thamTuto
## Number of citations: 91 including 5 auto-citations
##         number of articles: 77
##
##  J
## 91
##
##               Article Article; Early Access    Editorial Material
##                    77                     6                      2
##                Review
##                     6
```

Data selection

- Tabulating data type and citations
- create dataset with all articles and letters

```
# Create dataset with all articles and letters, for all articles for the moment
fulldat<-books<-NULL
for(ipub in 1:length(cite.list)) {
  datcite<-cite.data[[ipub]]
  id1<-grep("Comets",datcite$Authors)
  id2<-grep("Mentr",datcite$Authors)
#  id3<-grep("Brendel",datcite$Authors)
  id<-unique(c(id1,id2))
  cat("Publication:",nampub[ipub],"\n")
  cat("Number of citations:",dim(datcite)[1],"including",length(id),"auto-citations \n")
  cat("        number of articles:",dim(datcite[datcite$`Document Type`=="Article",])[1],"\n")
  print(table(datcite$`Publication Type`))
  print(table(datcite$`Document Type`))
  dat1<-datcite[datcite$`Document Type`=="Article" | datcite$`Document Type`=="Article; Early Access" |
  fulldat<-rbind(fulldat,dat1)
  dat1<-datcite[datcite$`Document Type`=="Article; Book Chapter" ,1:67]
  books<-rbind(books,dat1)
}
```

```
## Publication: cometsPackagenpde
## Number of citations: 264 including 13 auto-citations
##         number of articles: 246
##
##   B   J   S
##   2 261   1
##
##                           Article            Article; Book Chapter
##                               246                                2
##            Article; Early Access               Editorial Material
##                                 8                                1
## Editorial Material; Book Chapter                           Letter
```

```
##                                1                                3
##                Meeting Abstract                           Review
##                                1                                2
## Publication: escolano06
## Number of citations: 63 including 10 auto-citations
##         number of articles: 53
##
##  B  J  S
##  2 60  1
##
##                           Article            Article; Book Chapter
##                                53                                2
##            Article; Early Access Editorial Material; Book Chapter
##                                1                                1
##                            Letter                 Meeting Abstract
##                                2                                1
##                            Review
##                                3
## Publication: karlJPKPD10
## Number of citations: 49 including 5 auto-citations
##         number of articles: 39
##
##  B  J
##  1 48
##
##               Article Article; Book Chapter Article; Early Access
##                    39                    1                     3
##                Letter      Meeting Abstract                Review
##                     3                    2                     1
## Publication: karlPharmRes06
## Number of citations: 217 including 14 auto-citations
##         number of articles: 199
##
##   B   J
##   2 215
##
##               Article Article; Book Chapter Article; Early Access
##                   199                    2                     2
##    Editorial Material                Letter      Meeting Abstract
##                     1                    2                     1
##             News Item                Review
##                     1                    9
## Publication: thamTuto
## Number of citations: 91 including 5 auto-citations
##         number of articles: 77
##
##  J
## 91
##
##               Article Article; Early Access    Editorial Material
##                    77                    6                     2
##                Review
##                     6
```

```
cat("Total of ",dim(fulldat)[1],"citations\n")
```

```
## Total of  644 citations
```

```
books<-books[!duplicated(books$`UT (Unique WOS ID)`),]
books$`Source Title`
```

```
## [1] "PHARMACOKINETIC-PHARMACODYNAMIC MODELING AND SIMULATION, SECOND EDITION"
## [2] "INTRODUCTION TO POPULATION PHARMACOKINETIC/PHARMACODYNAMIC ANALYSIS WITH NONLINEAR MIXED EFFECTS
```

```
# Removing duplicates
fulldat<-fulldat[!duplicated(fulldat$`UT (Unique WOS ID)`),]
  id1<-grep("Comets",fulldat$Authors)
  id2<-grep("Mentr",fulldat$Authors)
  id<-unique(c(id1,id2))
cat("Number of unique articles citing any of the",length(nampub),"articles:",dim(fulldat)[1],"including
```

```
## Number of unique articles citing any of the 5 articles: 514 including 25 auto-citations
```

```
removecol<-c(3:5,7,11:12,15:19,30,48:51,56:57)

write.table(fulldat[,-c(removecol)], file.path("uniquepublications.csv"), quote=TRUE, row.names=F)

# Annotated file

workdat<-read.table(file.path("uniquepublications_annotated.csv"),stringsAsFactors = FALSE, header=TRUE)

# Removing our 5 articles
cite.pubmedid<-c(18215437,0,20033477,16906454,27884052) #pubmedID for the 5 articles, one (Escolano) mi
cite.pubmedid<-cite.pubmedid[cite.pubmedid>0]
workdat1<-workdat[!(workdat$Pubmed.Id %in% cite.pubmedid),] #510 publications
```

Extracting addresses of authors

```
# Extracting country of corresponding author
address<-workdat1$Reprint.Addresses
address<-strsplit(address,",",fixed=TRUE)
corcountry<-c()
for(i in address) {
  x<-i[length(i)]
  x<-unlist(strsplit(x,".",fixed=TRUE))
  corcountry<-c(corcountry,x[x!=""])
}
# Removing postal code from the US addresses
idx<-grep("USA",corcountry)
region<-corcountry
for(i in idx) {
  x<-unlist(strsplit(corcountry[i]," ",fixed=TRUE))
  x<-x[x!=""]
  if(length(x)==3 & !is.na(as.double(x[2]))) x<-x[-c(2)]
  if(length(x)!=2) cat("Problem with",corcountry[i])
  corcountry[i]<-paste(x[1],x[2])
  region[i]<-state.name[state.abb==x[1]]
}
```

```
## Warning: NAs introduits lors de la conversion automatique

## Warning: NAs introduits lors de la conversion automatique

## Warning: NAs introduits lors de la conversion automatique

## Warning: NAs introduits lors de la conversion automatique

## Warning: NAs introduits lors de la conversion automatique

## Warning: NAs introduits lors de la conversion automatique
```

```
table(corcountry)
```

```
## corcountry
##        Australia          Belgium  Bosnia & Herceg           Canada
##               16               20                1                8
##         Colombia   Czech Republic          Denmark          England
##                1                2                3               17
##          Finland           France          Germany            Italy
##                2              141                9                3
##            Japan            Kenya      Netherlands      New Zealand
##                5                1               83                5
##  Peoples R China           Poland           Russia         Scotland
##               50                3                1                1
##      South Korea            Spain           Sweden      Switzerland
##               11               22               16                6
##         Thailand          Uruguay           CA USA           CO USA
##                3                1               18                1
##           FL USA           IA USA           IL USA           IN USA
##                2                3                1                2
##           KY USA           MA USA           MD USA           NC USA
##                1                5                8                6
##           NJ USA           NY USA           PA USA           RI USA
##                3                6                5                1
##           TN USA           TX USA           UT USA           VA USA
##                5                2                6                1
##           WA USA           WI USA
##                1                2
```

```
corcountry2<-corcountry
corcountry2[grep("USA",corcountry)]<-"USA"
x<-table(corcountry2)
addr<-as.data.frame(x)
colnames(addr)<-c("country","nb")
```

Plots for paper

World map of the papers citing the foundation papers

Creating a world map and colouring the countries according to the number of publications

```
# World map
worldmap<-map_data("world")
worldmap2<-worldmap[worldmap$lat>(-51),]
```

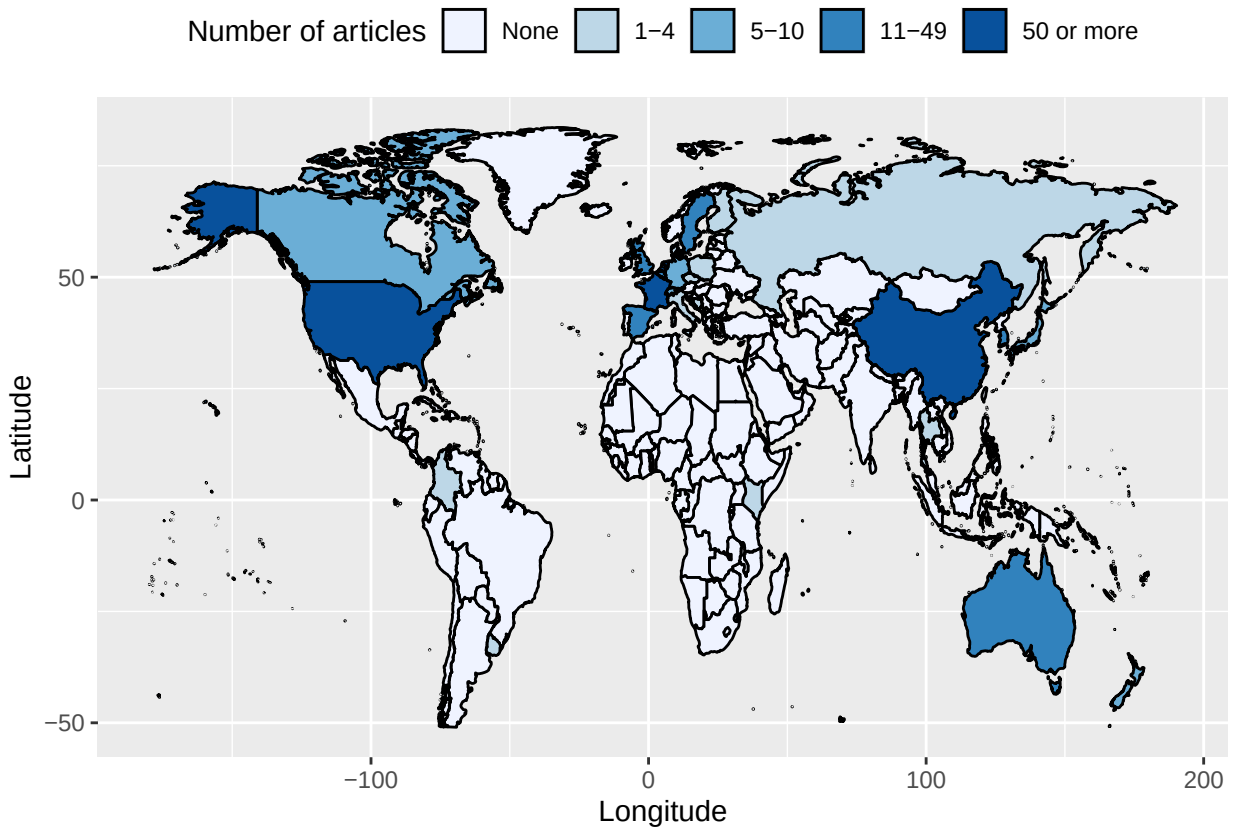
```
trim.leading <- function (x)  sub("^\\s+", "", x)
addr$country<-trim.leading(addr$country)

# Checking we have all regions
pmatch(addr$country,worldmap$region)
```

```
##  [1]  6983  9542    NA 14515 32530 33859 34727    NA   878 38500 34099 53067
## [13] 53828 56086 65901 68330    NA 72786 74493    NA 57152 36406 85239 26088
## [25] 86509 90079 90222
```

```
addr$country[addr$country=="Bosnia & Herceg"]<-"Bosnia and Herzegovina"
addr$country[addr$country=="Peoples R China"]<-"China"
addr$country[addr$country=="England"]<-"UK"
# Regroup Scotland and the UK
addr$nb[addr$country=="UK"]<-addr$nb[addr$country=="UK"]+addr$nb[addr$country=="Scotland"]
addr<-addr[addr$country!="Scotland",]
pmatch(addr$country,worldmap$region)
```

```
##  [1]  6983  9542 10951 14515 32530 33859 34727 39488   878 38500 34099 53067
## [13] 53828 56086 65901 68330 28281 72786 74493 57152 36406 85239 26088 86509
## [25] 90079 90222
```

```
# Adding colours to the dataset... Continuous scale
npaper<-addr$nb[match(worldmap2$region,addr$country)]
npaper[is.na(npaper)]<-0
worldmap3<-cbind(worldmap2,npaper=npaper)

#ggplot( data=worldmap3, aes(x=long, y=lat, fill=npaper, group=group)) + geom_polygon(color="black", si

# Adding colours to the paper, regrouping in categories
npaper[npaper<5 & npaper>1]<-1
npaper[npaper<11 & npaper>=5]<-2
npaper[npaper<50 & npaper>=11]<-3
npaper[npaper>=50]<-4
npaper<-npaper+1
worldmap3<-cbind(worldmap3,ntab=npaper)

#cols<-c(colorRampPalette(c("white", "royalblue"))(6))
#cols<-cols[-c(1)]
mycolours<-brewer.pal(5,"Blues")

mymap<-ggplot( data=worldmap3, aes(x=long, y=lat, fill=as.factor(ntab), group=group)) + geom_polygon(co

print(mymap)
```

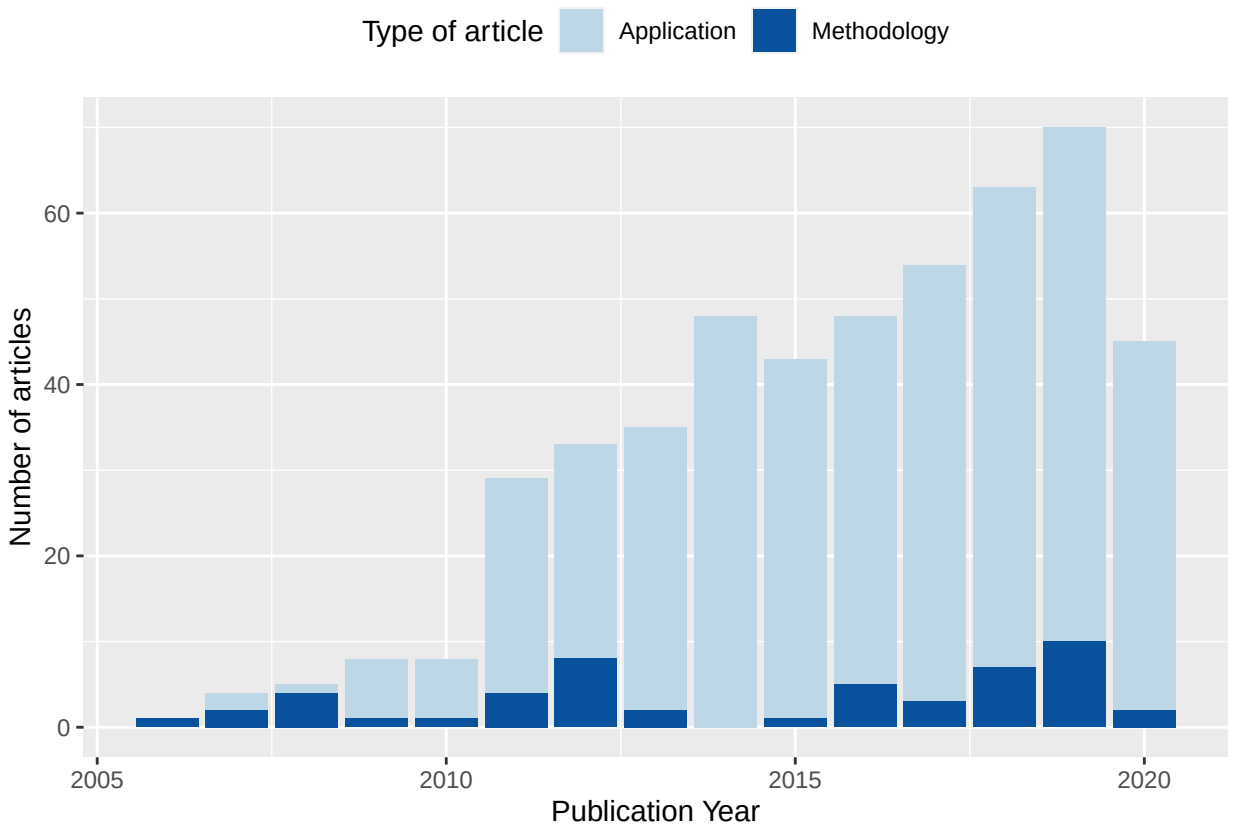


Histogram of number of publications per year

Histograms of number of (unique) publications citing at least one of the foundation articles (including autocitations, using workdat1 below will exclude the 5 autocitations):

```
histTime<-ggplot(data=workdat, aes(x=Publication.Year, fill=as.factor(Methodology.Application))) + geom_
    theme(legend.position = "top")  + scale_fill_manual(values=mycolours[c(2,5)],name = "Type of articl

print(histTime)
```

```
## Warning: Removed 20 rows containing non-finite values (stat_count).
```




```
cat("Articles citing npde including the foundation articles, methodology versus applications:\n")
```

```
## Articles citing npde including the foundation articles, methodology versus applications:
```

```
print(table(workdat$Methodology.Application))
```

```
##
##   A   M
## 463  51
```

```
cat("Articles citing npde (excluding the foundation articles), methodology versus applications:\n")
```

```
## Articles citing npde (excluding the foundation articles), methodology versus applications:
```

```
print(table(workdat1$Methodology.Application))
```

```
##
##   A   M
## 463  47
```

Word cloud of keywords

Only on the applied articles. Word cloud to view :

- types of models PK or PD
- types of studies
- age groups: children/elderly
- ...

Harmonising keywords

Regrouping keywords in broad(er) categories, harmonising.

```
# Focus on applications using npde
workdat2<-workdat[workdat$Methodology.Application=="A",]
keyw<-c()
for(i in 1:dim(workdat2)[1]) {
  vec<-c(workdat2$Keywords.Plus[i],workdat2$Author.Keywords[i])
  vec<-vec[!is.na(vec)]
  if(length(vec)>1) vec<-paste(vec,collapse=";")
  vec<-unlist(strsplit(vec,";",fixed=TRUE))
  vec<-trimws(tolower(vec))
  keyw<-c(keyw,unique(vec))
}
# sort(unique(keyw[grep("pharmacodynamic",keyw)]))

# Table of keywords
keyw[grep("pk",keyw)]
```

```
##  [1] "pk/pd modeling"
##  [2] "pkpd modeling"
##  [3] "pk"
##  [4] "pk-pd"
##  [5] "pkpd modelling"
##  [6] "pkpd model"
##  [7] "pk-pd, pharmacokinetic-pharmacodynamic"
##  [8] "pkpd modelling"
##  [9] "population pk modelling"
## [10] "pk/pd indexes"
## [11] "population pkpd modeling"
## [12] "probability of pk-pd target attainment"
## [13] "adpkd"
## [14] "pk/pd"
## [15] "population pk"
## [16] "population pk/pd"
## [17] "pbpk model"
## [18] "pk"
## [19] "pharmacokinetic/pharmacodynamic (pk/pd) modeling"
## [20] "pk"
## [21] "pk/pd modeling"
## [22] "population pkpd"
## [23] "pk/pd"
## [24] "pk/pd modeling"
## [25] "population pk"
## [26] "population pk/pd modeling"
```

```
keyw<-gsub("pharmacokinetics", "pk", keyw)
keyw<-gsub("pharmacokinetic", "pk", keyw)
keyw<-gsub("pharmacodynamics", "pd", keyw)
keyw<-gsub("pharmacodynamic", "pd", keyw)
keyw<-gsub("pkpd", "pk/pd", keyw)
keyw<-gsub("pk pd", "pk/pd", keyw)
keyw<-gsub("pk-pd", "pk/pd", keyw)
keyw<-gsub("modeling", "modelling", keyw)
keyw<-gsub("pharmacokinetick-pharmacodynamic", "pk/pd", keyw)
```

```
keyw<-gsub("pk/pd, pharmacokinetic-pharmacodynamic", "pk/pd", keyw)
keyw<-gsub("modeling", "model", keyw)
keyw<-gsub("modelling", "model", keyw)
keyw<-gsub("models", "model", keyw)
keyw<-gsub("healthy-volunteers", "healthy volunteers", keyw)
# Regrouping children, neonates, infants and preterm
keyw<-gsub("children", "paediatrics", keyw)
keyw<-gsub("neonates", "paediatrics", keyw)
keyw<-gsub("neonate", "paediatrics", keyw)
keyw<-gsub("pediatric-patients", "paediatrics", keyw)
keyw<-gsub("pediatrics", "paediatrics", keyw)
keyw<-gsub("pediatric", "paediatrics", keyw)
keyw<-gsub("preterm", "paediatrics", keyw)
keyw<-gsub("infants", "paediatrics", keyw)
keyw<-gsub("paediatrics", "paediatric", keyw)
keyw<-gsub("model based", "model-based", keyw)
keyw<-gsub("hiv-1", "hiv", keyw)
keyw<-gsub("infections", "infection", keyw)

# adults/children
keyw[keyw=="premenopausal women" | keyw== "african women" | keyw=="postmenopausal women"]<-"women"
keyw<-gsub("young-adults", "adult", keyw)
keyw<-gsub("adults", "adult", keyw)
keyw<-gsub("adult patients", "adult", keyw)
keyw<-gsub("adult chinese patients", "adult", keyw)
keyw<-gsub("chinese adult patients", "adult", keyw)
keyw[keyw== "paediatric pk population model"]<- "paediatric population pk model"
keyw[keyw=="chinese adult"]<-"adult"
keyw[keyw=="young paediatric" | keyw=="young-paediatric" | keyw=="zambian paediatric" | keyw=="paediatr
keyw[keyw=="hiv-infected paediatric" | keyw=="hiv-infected children"]<-"hiv-infected paediatric"
keyw[keyw=="paediatric and young paediatric" | keyw=="paediatric younger" | keyw=="premature paediatric
keyw[keyw=="critically ill paediatric"]<-"critically-ill paediatric"
keyw[keyw=="newborn-paediatric" | keyw=="paediatric and adolescents" | keyw=="uninfected paediatric"]<-
keyw[keyw=="elderly hospitalized-patients" | keyw=="elderly-patients" | keyw=="older patients" | keyw==
keyw[keyw=="volunteers" | keyw=="healthy-subjects"]<-"healthy volunteers"

# ICU
keyw<-gsub("intensive care unit", "intensive care", keyw)
keyw<-gsub("intensive-care-unit", "intensive care", keyw)
keyw<-gsub("intensive-care patients", "intensive care", keyw)
keyw<-gsub("intensive care unit patients", "intensive care", keyw)
keyw<-gsub("intensive care patients", "intensive care", keyw)
keyw<-gsub("intensive-care", "intensive care", keyw)
keyw<-gsub("critically-ill patients", "intensive care", keyw)

keyw[keyw=="drug monitoring"]<-"therapeutic drug monitoring"
keyw[keyw=="drug-monitoring service"]<-"therapeutic drug monitoring"
keyw[keyw=="drug metabolism"]<-"drug-metabolism"
keyw[keyw=="drug development"]<-"drug-development"
keyw[keyw=="drug disposition"]<-"drug-disposition"
keyw[keyw=="drug interaction"| keyw=="drug interactions" | keyw=="drug-interactions" | keyw=="drug-drug
keyw[keyw=="drug"]<-"drugs"
keyw[keyw=="mediated drug disposition" | keyw=="target-mediated drug disposition" | keyw=="tmdd, target
```

```
# PK/PD/model
keyw[keyw=="pk and pd"]<-"pk/pd"
keyw[keyw=="pd pd"]<-"pd"
keyw[keyw=="pk/pd (pk/pd) model"]<-"pk/pd model"
keyw[keyw=="nlme model"]<-"nonlinear mixed-effect model"
keyw[keyw=="pharmacodynamic model" | keyw=="pd model"| keyw=="pharmacodynamic model"]<-"pd model"
keyw[keyw=="mixed effects model" | keyw=="mixed-effects model" | keyw=="mixed effect model" | keyw=="mi
keyw[keyw=="nonlinear mixed effects model" | keyw=="nonlinear mixed-effect model" | keyw=="nonlinear mi
keyw[keyw=="pd model"]<-"pd model"
keyw[keyw=="pk/pd model"]<-"pk/pd model"
keyw[keyw=="pk/pd, pk/pd"]<-"pk/pd"
keyw[keyw=="population pk and pd"]<-"population pk/pd"
keyw[keyw=="kpd model" | keyw=="k-pd model"]<-"k-pd model"
keyw[keyw=="pk/pd analysis"]<-"pk/pd"
keyw[keyw=="population pk analysis"]<-"population pk"
keyw[keyw=="population pk model analysis" ]<-"population pk model"
keyw[keyw=="pk/pd relationship"]<-"pk/pd model"
keyw[keyw=="probability of pk/pd target attainment" | keyw=="pk/pd target attainment rate"]<-"pk/pd tar

# Transplantation
keyw[keyw=="renal transplantation" | keyw=="renal-transplantation" | keyw=="renal-transplant recipients
keyw[keyw=="stem cell transplantations"]<-"stem-cell transplantation"
keyw[keyw=="liver-transplantation" | keyw=="liver transplant"]<-"liver transplantation"
keyw[keyw=="marrow-transplantation"]<-"marrow transplantation"
keyw[keyw=="paediatric liver-transplantation" | keyw=="paediatric liver-transplantation" | keyw=="paedia
keyw[keyw=="marrow-transplantation"]<-"marrow transplantation"
keyw[keyw=="kidney-transplant recipients" | keyw=="kidney-transplantation" | keyw=="kidney transplant pa
keyw[keyw=="hematopoeitic stem cell transplantation" | keyw=="hematopoietic-cell transplantation" | keyw
keyw[keyw=="transplant recipients" | keyw=="transplant patients"]<-"transplantation"

# Cancer
keyw[keyw=="breast-cancer patients" | keyw=="breast-cancer"| keyw== "her2-positive breast-cancer" | keyw
keyw[keyw=="cancer patients" | keyw=="cancer-patients" | keyw=="cancer clinical-trials"]<-"cancer"
keyw[keyw=="lung-cancer patients" | keyw=="lung-cancer" | keyw=="cell lung-cancer"]<-"lung cancer"
keyw[keyw=="colorectal-cancer patients" | keyw== "colorectal-cancer"|keyw=="metastatic colorectal-cancer
keyw[keyw=="anti-cancer agents" | keyw=="anticancer agents" ]<-"anticancer drugs"
keyw[keyw=="ovarian epithelial cancer" | keyw=="ovarian-cancer patients" |keyw=="ovarian-cancer"]<-"ova

# Animals
keyw[keyw=="murine model" | keyw=="juvenile mice"]<-"mice"
keyw[keyw=="animal-model" | keyw=="animals"]<-"animal"

keyw[keyw=="medication" | keyw=="medications"]<-"drugs"

# Parameters
keyw[keyw=="drug clearance" | keyw=="plasma-clearance" | keyw=="renal drug clearance" | keyw=="renal cl
keyw[keyw=="drug-metabolism"| keyw=="first-pass metabolism"]<-"metabolism"
keyw[keyw=="drug-disposition" | keyw=="disposition"]<-"elimination"
keyw[keyw=="drug absorption"|keyw=="oral-drug absorption"|keyw=="systemic absorption"]<- "absorption"
keyw[keyw=="glomerular filtration rate (gfr)" | keyw=="glomerular-filtration-rate" | keyw=="glomerular
keyw[keyw=="severe sepsis" | keyw=="late-onset sepsis"]<-"sepsis"
keyw[keyw=="human-plasma" | keyw=="human plasma"]<-"plasma"
```

```
# Model purpose
keyw[keyw=="model validation" ]<-"model evaluation"
keyw[keyw=="mathematical-model"]<-"mathematical model"
keyw[keyw=="mechanism-based model" ]<-"mechanistic model"
keyw[keyw=="single-dose pk"]<-"single dose pk"
keyw[keyw=="vitro" | keyw=="in vitro experiments" | keyw=="in-vitro activity" | keyw=="in-vitro" | keyw=
keyw[keyw=="pharmacogenetic" | keyw=="pk-pharmacogenetics"]<-"pharmacogenetics"
keyw[keyw=="genetic-polymorphism"]<-"genetic polymorphisms"

# Study
keyw[keyw=="drug dose-response relationship" | keyw=="dose-response relationship" ] <- "dose-response"
keyw[keyw=="clinical-pharmacology"]<-"clinical pharmacology"
keyw[keyw=="clinical trials" | keyw=="clinical-trials"]<-"clinical trial"
keyw[keyw=="phase i clinical trial" | keyw=="phase-i" | keyw=="phase-ii" | keyw=="phase i" | keyw=="phas
keyw[keyw=="randomized-controlled-trial" | keyw=="randomized controlled-trial" | keyw=="randomized-trial
keyw=="controlled-trial"]<-"randomized controlled trial"
keyw[keyw=="trials"]<-"trial"
keyw[keyw=="clinical-efficacy"]<-"efficacy"
keyw[keyw=="clinical-outcomes" | keyw=="clinical-response"]<-"clinical response"
keyw[keyw=="drug-development" | keyw=="model-based drug development"]<-"drug development"
# Dose individualisation
keyw[keyw %in% c("dose individualisation","dose adjustment","adaptive dosing","dosage optimization", "do
# Dose selection (population)
keyw[keyw %in% c("drug dosage calculations","dosage","dosing recommendation","child appropriate dosage f

keyw[keyw=="safety pharmacology"|keyw=="long term safety"]<-"safety"
keyw[grep("toxicity",keyw)]<-"toxicity"

# Conditions
keyw[keyw=="migraine disorders" | keyw=="migraine prevention"]<-"migraine"
keyw[keyw=="hiv-infected adults patients" | keyw=="hiv-infected individuals" | keyw=="immunodeficiency-v
keyw[keyw=="infectious diseases" | keyw=="infected patients" ]<-"infection"
keyw[keyw=="virus-infected patients" | keyw=="virus-infected subjects" | keyw=="virus-infection"  ]<-"vi
keyw[keyw=="impaired renal-function" | keyw=="renal-function"]<-"renal function"
keyw[keyw=="renal dysfunction" | keyw== "impaired renal-function" | keyw=="chronic-renal-failure" | keyw
keyw[keyw=="human liver"]<-"liver"
keyw[keyw=="chronic hepatitis-b" | keyw=="chronic hepatitis-c" | keyw=="hepatitis c infection" | keyw==
keyw[keyw=="morbid obesity" | keyw=="obese-patients" | keyw=="morbidly obese-patients"]<-"obesity"
keyw[keyw=="overweight"]<-"under/overweight"
keyw[keyw=="crohn's disease" | keyw=="crohns-disease"]<-"crohn disease"
keyw[keyw=="inflammatory-bowel-disease"]<-"inflammatory bowel disease"
keyw[keyw=="parkinson's disease" | keyw=="parkinsons-disease"]<-"Parkinson's disease"

# Treatments
keyw[keyw=="clinical-evaluation"]<-"clinical evaluation"
keyw[keyw=="multidrug-resistant" | keyw=="multidrug-resistance" | keyw=="mrp2 (multi-drug resistance pro
keyw[keyw=="antituberculosis drugs"]<-"antitubercular drugs"
keyw[keyw=="antibody-drug conjugate"]<-"antibody drug conjugate"
keyw[keyw=="plasma clozapine"]<-"clozapine"
keyw[keyw=="lopinavir plasma-concentrations"]<-"lopinavir"
keyw[keyw=="combinations" | keyw=="combination therapy" | keyw=="fixed-dose combination"]<-"combination
keyw[keyw=="recombinant-human-erythropoietin"]<-"recombinant erythropoietin"
keyw[keyw=="antimicrobial agents" | keyw=="combination antimicrobial therapy" | keyw=="nebulized antimi
```

```r
keyw[keyw=="microbial response"]<-"antimicrobial activity"

# Other
keyw[grep("polymorphism",keyw)]<-"genetic polymorphism"
keyw[grep("pain",keyw)]<-"pain"
keyw[keyw=="p-glycoprotein inhibition"]<-"p-glycoprotein inhibitor"
keyw[keyw %in% c("developmental changes","developmental expression","developmental pharmacology")]<-"dev
keyw[keyw %in% c("monte carlo simulation","monte-carlo simulation","monte-carlo-simulation","monte carlo
keyw[keyw=="simulations" | keyw=="computer simulation" ]<-"simulation"
keyw[keyw %in% c("serum concentrations","serum concentration","plasma-concentrations", "blood-concentra
keyw[keyw=="trough concentrations"]<-"trough concentration"

if(FALSE) {
keyw[keyw==""]<-""
}

keyw[keyw=="pk"]<-"pharmacokinetics"
keyw[keyw=="pd"]<-"pharmacodynamics"


# Removing duplicate keyword
keyw<-keyw[keyw!="1-infected paediatric"]
```

```r
# Looking at some specific keywords
sort(unique(keyw[grep("pd",keyw)]))
```

```
##  [1] "antituberculosis pd"     "k-pd model"
##  [3] "pd breakpoints"          "pd model"
##  [5] "pk/pd"                   "pk/pd adequacy"
##  [7] "pk/pd indexes"           "pk/pd model"
##  [9] "pk/pd target attainment" "population pd"
## [11] "population pk/pd"        "population pk/pd model"
## [13] "update"                  "updated guidance"
```

```r
sort(unique(keyw[grep("pk",keyw)]))
```

```
##  [1] "adpkd"                                "altered aminoglycoside pk"
##  [3] "aminoglycoside pk"                    "clinical pk"
##  [5] "developmental pk"                     "dose pk"
##  [7] "gentamicin pk"                        "integrated pk"
##  [9] "nonlinear pk"                         "paediatric population pk model"
## [11] "pbpk model"                           "piperacillin population pk"
## [13] "pk bridging"                          "pk interaction"
## [15] "pk model"                             "pk parameters"
## [17] "pk properties"                        "pk/pd"
## [19] "pk/pd adequacy"                       "pk/pd indexes"
## [21] "pk/pd model"                          "pk/pd target attainment"
## [23] "population pk"                        "population pk model"
## [25] "population pk/pd"                     "population pk/pd model"
## [27] "propacetamol pk"                      "single dose pk"
## [29] "steady-state pk"                      "tissue pk"
## [31] "whole-body physiologically based pk"
```

```r
#sort(unique(keyw[grep("drug monitoring",keyw)]))
sort(unique(keyw[grep("adult",keyw)])) # regrouped and split
```

```
## [1] "adult"                              "adult renal transplant recipients"
## [3] "adult t-cell lymphoma"              "hiv-infected adult"
```

```
sort(unique(keyw[grep("paediatric",keyw)])) # regrouped and split
```

```
##  [1] "birth-weight paediatric"          "breast-fed paediatric"
##  [3] "critically-ill paediatric"        "febrile paediatric"
##  [5] "hiv-infected paediatric"          "malnourished paediatric"
##  [7] "paediatric"                       "paediatric age range"
##  [9] "paediatric anesthesia"            "paediatric brain tumors"
## [11] "paediatric burns"                 "paediatric clinical pharmacology"
## [13] "paediatric covariate model"       "paediatric drug development"
## [15] "paediatric drug prescription"     "paediatric intensive care"
## [17] "paediatric liver transplantation" "paediatric migraineurs"
## [19] "paediatric oncology patients"     "paediatric pharmacology"
## [21] "paediatric population pk model"   "postoperative paediatric"
```

```
sort(unique(keyw[grep("neonate",keyw)]))
```

```
## character(0)
```

```
sort(unique(keyw[grep("transplant",keyw)]))
```

```
## [1] "adult renal transplant recipients"
## [2] "heart-transplantation"
## [3] "hematopoietic stem cell transplantation"
## [4] "kidney transplantation"
## [5] "liver transplantation"
## [6] "paediatric liver transplantation"
## [7] "renal transplant"
## [8] "solid-organ transplantation"
## [9] "transplantation"
```

```
sort(unique(keyw[grep("cancer",keyw)]))
```

```
##  [1] "anticancer agent"           "anticancer drug resistance"
##  [3] "anticancer drugs"           "breast cancer"
##  [5] "cancer"                     "cancer resistance protein"
##  [7] "cancer-cells"               "colorectal cancer"
##  [9] "gastric-cancer"             "life-threatening cancer"
## [11] "lung cancer"                "ovarian cancer"
```

```
sort(unique(keyw[grep("tumo",keyw)]))
```

```
##  [1] "advanced solid tumors"      "antitumor-activity"
##  [3] "brain-tumors"               "neuroendocrine tumors"
##  [5] "paediatric brain tumors"    "solid tumors"
##  [7] "tumor burden"               "tumor grade"
##  [9] "tumor growth model"         "tumor response"
## [11] "tumor static concentration" "tumor-size"
## [13] "tumors"                     "tumour size prediction"
```

```
sort(unique(keyw[grep("development",keyw)]))
```

```
## [1] "developmental pk"           "drug development"
## [3] "model development"          "paediatric drug development"
```

```
sort(unique(keyw[grep("dog",keyw)])) # no cat
```

```
## [1] "clopidogrel" "dogs"
```

```
sort(unique(keyw[grep("mice",keyw)])) # no mouse
```

```
## [1] "mice"
```

```
sort(unique(keyw[grep("monkey",keyw)])) # no primate, monkey
```

```
## character(0)
```

```
sort(unique(keyw[grep("monitoring",keyw)]))
```

```
## [1] "monitoring-data"             "therapeutic drug monitoring"
```

```
sort(unique(keyw[grep("evaluation",keyw)]))
```

```
## [1] "clinical evaluation" "evaluation"          "external evaluation"
## [4] "model evaluation"
```

```
sort(unique(keyw[grep("internal",keyw)]))
```

```
## [1] "internalization"
```

```
sort(unique(keyw[grep("intensive",keyw)]))
```

```
## [1] "intensive care"            "paediatric intensive care"
```

```
sort(unique(keyw[grep("drug",keyw)]))
```

```
##  [1] "antibody drug conjugate"             "anticancer drug resistance"
##  [3] "anticancer drugs"                    "antiepileptic drugs"
##  [5] "antiretroviral drugs"                "antitubercular drugs"
##  [7] "direct-acting antiviral (daa) drugs" "drug articaine"
##  [9] "drug development"                    "drug exposure"
## [11] "drug resistance"                     "drug selection"
## [13] "drug susceptibility"                 "drug transporters"
## [15] "drug-eluting stents"                 "drug-interaction"
## [17] "drug-metabolizing-enzymes"           "drug-treatment"
## [19] "drugs"                               "drugs of abuse"
## [21] "immunosuppressive drugs"             "multidrug resistance"
## [23] "nonsteroidal antiinflammatory drugs" "paediatric drug development"
## [25] "paediatric drug prescription"        "prodrug oseltamivir"
## [27] "pulmonary drug delivery"             "therapeutic drug monitoring"
## [29] "veterinary drugs"
```

```
sort(unique(keyw[grep("model",keyw)]))
```

```
##  [1] "2-part model"
##  [2] "antagonist interaction-model"
##  [3] "cell life span model"
##  [4] "cell transit model"
##  [5] "circulation model"
##  [6] "compartment model"
##  [7] "composite null model"
##  [8] "disease model"
##  [9] "generalized linear mixed-effect model"
## [10] "greenlab model"
## [11] "hierarchical model"
## [12] "hollow-fiber infection model"
## [13] "indescribable model"
```

```
## [14] "k-pd model"
## [15] "markov model"
## [16] "mathematical model"
## [17] "mechanistic model"
## [18] "mixed-effects model"
## [19] "model"
## [20] "model and simulation"
## [21] "model development"
## [22] "model evaluation"
## [23] "model framework"
## [24] "model-based meta-analysis"
## [25] "model-based research"
## [26] "nonlinear mixed-effects model"
## [27] "paediatric covariate model"
## [28] "paediatric population pk model"
## [29] "pbpk model"
## [30] "pd model"
## [31] "pharmacometric model"
## [32] "physiological model"
## [33] "pk model"
## [34] "pk/pd model"
## [35] "plant growth model"
## [36] "population model"
## [37] "population pk model"
## [38] "population pk/pd model"
## [39] "response surface model"
## [40] "semimechanistic model"
## [41] "system model"
## [42] "systematic model comparison"
## [43] "translational model"
## [44] "tuberculosis pharmacometric model"
## [45] "tumor growth model"
## [46] "turnover model"
```

```
sort(unique(keyw[grep("distribution",keyw)]))
```

```
## [1] "asymptotic-distribution" "biodistribution"
## [3] "body-fat distribution"   "distribution terms"
## [5] "distributions"           "tissue distribution"
## [7] "vegf distribution"
```

```
sort(unique(keyw[grep("volunteer",keyw)]))
```

```
## [1] "healthy volunteers"
```

```
sort(unique(keyw[grep("recipient",keyw)]))
```

```
## [1] "adult renal transplant recipients" "recipients"
```

Keywords covering several notions

Some keywords cover different notions. The usual text parsing approach would be to cut strings into single words, but here trying to keep expressions => split the keywords into 2

```
# Splitting keywords
splitkw<-function(vec, pattern, replist) {
```

```
  idx<-which(vec==pattern)
  vec[idx]<-replist[1]
  for(i in 2:length(replist))
    vec<-c(vec,rep(replist[i],length(idx)))
  return(vec)
}
# splitkw(1:5,2,c(10,11,12)) # Testing the function
# splitkw(rep(1:5, each=3),2,c(10,11,12))

#
keyw<-gsub("models", "model", keyw)
keyw<-splitkw(keyw,"population pk/pd model",c("population pk/pd","model"))
keyw<-splitkw(keyw,"population pk model",c("population pk","model"))
keyw<-splitkw(keyw,"pk/pd models",c("pk/pd","model"))
keyw<-splitkw(keyw,"pk models",c("pk","model"))
keyw<-splitkw(keyw,"pk model",c("pk","model"))
keyw<-splitkw(keyw,"pbpk model",c("pbpk","model"))
keyw<-splitkw(keyw,"pd models",c("pd","model"))
keyw<-splitkw(keyw,"whole-body physiologically based pk",c("pbpk","model"))

# Adult + specific disease
keyw<-splitkw(keyw,"hiv-infected adult",c("hiv","adult","infectious disease"))
keyw<-splitkw(keyw,"adult t-cell lymphoma",c("adult","cancer"))
keyw<-splitkw(keyw,"infected women",c("women","infection"))
# Children + specific condition
keyw<-splitkw(keyw,"paediatric intensive care",c("paediatric","intensive care"))
keyw<-splitkw(keyw,"paediatric population pk model",c("paediatric","population pk","model"))
keyw<-splitkw(keyw,"hiv-infected paediatric",c("hiv","paediatric"))
keyw<-splitkw(keyw,"paediatric intensive care",c("paediatric","intensive care"))
keyw<-splitkw(keyw,"critically ill paediatric",c("paediatric","intensive care"))
keyw<-splitkw(keyw,"critically-ill paediatric",c("paediatric","intensive care"))
keyw<-splitkw(keyw,"paediatric liver transplantation",c("paediatric","liver transplant","transplantatio
keyw<-splitkw(keyw,"paediatric oncology patients" ,c("paediatric","cancer"))
keyw<-splitkw(keyw, "paediatric brain tumors",c( "paediatric","cancer"))
keyw<-splitkw(keyw,"paediatric pharmacology",c("paediatric","pharmacology"))
keyw<-splitkw(keyw,"paediatric clinical pharmacology",c("paediatric","pharmacology"))
keyw<-splitkw(keyw,"birth-weight paediatric",c("paediatric","neonates"))
keyw<-splitkw(keyw,"breast-fed paediatric",c("paediatric","neonates"))
keyw<-splitkw(keyw,"paediatric pharmacology",c("paediatric","pharmacology"))
keyw<-splitkw(keyw,"paediatric clinical pharmacology",c("paediatric","pharmacology"))
keyw<-splitkw(keyw,"paediatric anesthesia",c("paediatric","anesthesia"))
keyw<-splitkw(keyw,"paediatric drug development",c("paediatric","drug development"))
keyw<-splitkw(keyw,"paediatric population pk model",c("paediatric","population pk","model"))
keyw<-splitkw(keyw,"paediatric liver transplantation",c("paediatric","liver transplantation"))
keyw<-splitkw(keyw,"paediatric migraineurs",c("paediatric","migraine"))
keyw<-splitkw(keyw,"childhood migraine",c("paediatric","migraine"))
keyw<-splitkw(keyw,"neonatal sepsis",c("paediatric","neonates","sepsis"))

# Animals
keyw<-splitkw(keyw,"mice",c("mice","animal"))
keyw<-splitkw(keyw,"veterinary drugs",c("animal","drugs"))

# Cancer
```

```
# Transplantation (any)
keyw<-splitkw(keyw,"renal transplant",c("renal transplant","transplantation"))
keyw<-splitkw(keyw,"adult renal transplant recipients",c("renal transplant","transplantation","adult"))
keyw<-splitkw(keyw,"heart-transplantation",c("heart-transplantation","transplantation"))
keyw<-splitkw(keyw,"liver transplantation",c("liver transplantation","transplantation"))
keyw<-splitkw(keyw,"hematopoietic stem cell transplantation",c("hematopoietic stem cell transplantation
keyw<-splitkw(keyw,"solid-organ transplantation",c("solid-organ transplantation","transplantation"))

# Drugs
keyw<-splitkw(keyw,"piperacillin population pk",c("piperacillin","population pk"))
keyw<-splitkw(keyw,"gentamicin pk",c("gentamicin","pk"))
keyw<-splitkw(keyw,"anticancer drug resistance",c("cancer","drug resistance"))

#keyw<-splitkw(keyw,"",c())

if(FALSE) {
keyw<-splitkw(keyw,"",c())
keyw<-splitkw(keyw,"",c())
keyw<-splitkw(keyw,"",c())
}
```

Creating word cloud figure

```
dm<-data.frame(keyw=unique(keyw),count=NA,stringsAsFactors = FALSE)
for(i in unique(keyw)) {
  dm$count[dm$keyw==i]<-length(keyw[keyw==i])
}
dm<-dm[order(dm$count, decreasing = TRUE),]
d <- data.frame(word = dm$keyw, freq=dm$count)
d[d$freq>=10,]
```

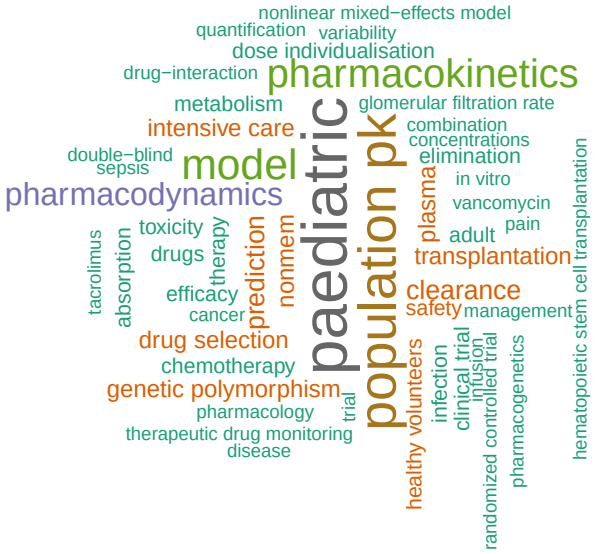
```
##                                    word freq
## 1                            paediatric  258
## 2                          population pk  194
## 3                                 model  149
## 4                       pharmacokinetics  136
## 5                       pharmacodynamics   87
## 6                             prediction   57
## 7                              clearance   52
## 8                         transplantation   48
## 9                                 nonmem   43
## 10                         intensive care   41
## 11                         drug selection   40
## 12                                 plasma   37
## 13                  genetic polymorphism   36
## 14                                safety   35
## 15                  healthy volunteers   33
## 16                               therapy   32
## 17                             metabolism   31
## 18                               efficacy   31
## 19                            elimination   30
## 20                              infection   30
## 21                               toxicity   29
```

```
## 22                                  adult   29
## 23                              absorption   26
## 24                                   drugs   25
## 25                           clinical trial   22
## 26                  dose individualisation   22
## 27                             chemotherapy   22
## 28                                  cancer   21
## 29              glomerular filtration rate   20
## 30                               vancomycin   20
## 31                            concentrations   19
## 32                              pharmacology   19
## 33                                 infusion   18
## 34                                 in vitro   18
## 35                              combination   18
## 36                                    trial   18
## 37                               variability   17
## 38 hematopoietic stem cell transplantation   17
## 39                             quantification   17
## 40                           drug-interaction   17
## 41             therapeutic drug monitoring   16
## 42                               management   16
## 43                             double-blind   16
## 44                                     pain   16
## 45             randomized controlled trial   15
## 46                               tacrolimus   15
## 47                           pharmacogenetics   15
## 48                                   sepsis   15
## 49                                 disease   14
## 50           nonlinear mixed-effects model   14
## 51                               parameters   14
## 52                               simulation   14
## 53                             breast cancer   13
## 54                            renal function   13
## 55                                    liver   13
## 56                              clinical pk   13
## 57                                 kinetics   13
## 58                                      hiv   12
## 59                              pk/pd model   12
## 60                              antibiotics   12
## 61                           renal transplant   12
## 62                                resistance   12
## 63                              performance   11
## 64                                      age   11
## 65                      mixed-effects model   11
## 66                                gentamicin   11
## 67                              metabolites   10
## 68                           serum creatinine   10
## 69                            glucuronidation   10
## 70                           drug development   10
## 71                                  obesity   10
## 72                               recipients   10
## 73                                   impact   10
```

```
cat(dim(d[d$freq>=10,])[1],"keywords occurring 10 times or more, and", dim(d[d$freq<10 & d$freq>=5,])[1]
```

```
## 73 keywords occurring 10 times or more, and 91 keywords occurring between 5 and 10 times
```

```
#d[d$freq<=10 & d$freq>5,]

set.seed(1234567)

wordcloud(words = d$word, freq = d$freq, min.freq = 5,scale=c(2,.5),
    max.words=50, random.order=FALSE, rot.per=0.35, colors=brewer.pal(8, "Dark2"))
```



Diagnostic plots for viral load example

Install package in development mode

This is a snapshot of the version of npde released at the time of the revision of the paper submitted to AAPS Journal. The npde package version 3.0 has been submitted to CRAN and should be available shortly on CRAN. We use development mode to avoid messing up any existing installation of npde.

```
dev_mode() # development mode
```

```
## Dev mode: ON
```

```
install.packages(pkgs="npde_3.0.tar.gz",repos=NULL) # local installation
```

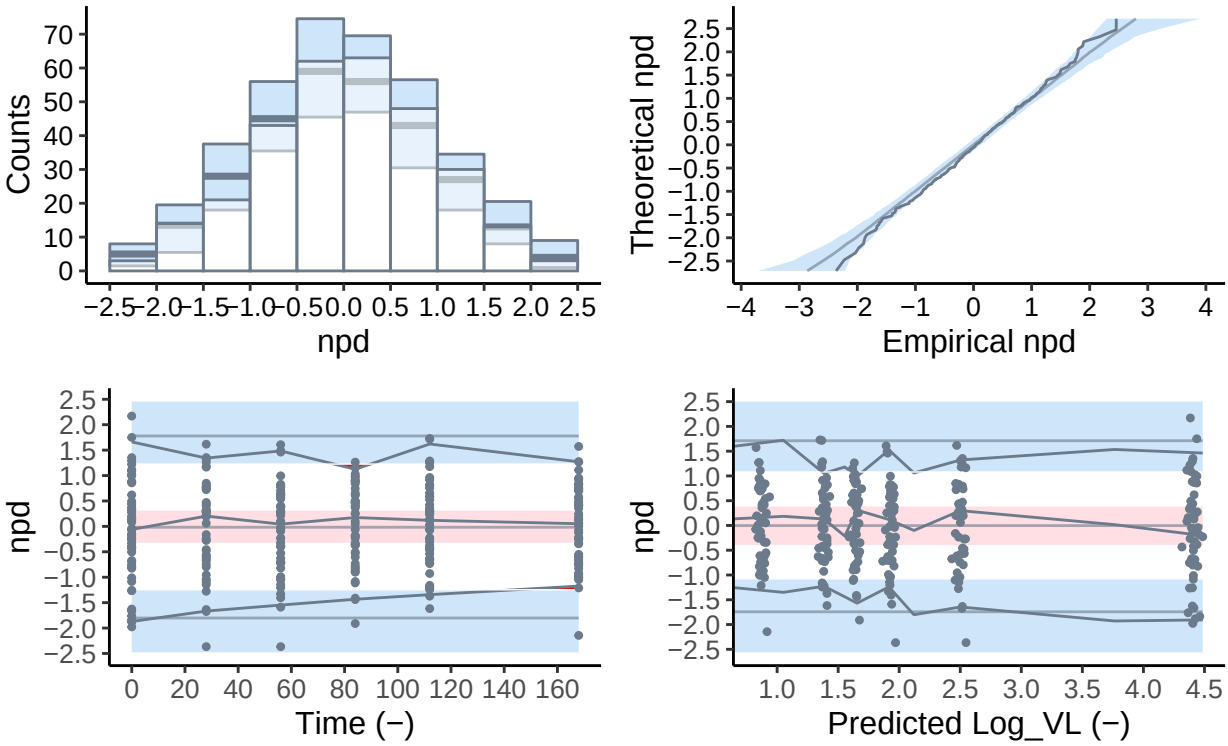
```
## Installing package into '/home/eco/R-dev'
## (as 'lib' is unspecified)
```

```
# install.package("npde) # installation from CRAN
library(npde)
```

Default diagnostic plots for npd for the viral load example used in the paper

```
## ------------------------------------------------
## Distribution of npde :
##       nb of obs: 300
```

```
##            mean= 0.03821   (SE= 0.053 )
##        variance= 0.8327   (SE= 0.068 )
##        skewness= -0.04464
##        kurtosis= -0.2207
## ------------------------------------------------
## Statistical tests (adjusted p-values):
##   t-test                 : 1
##   Fisher variance test   : 0.0959 .
##   SW test of normality   : 1
##   Global test            : 0.0959 .
## ---
## Signif. codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
## ------------------------------------------------
```



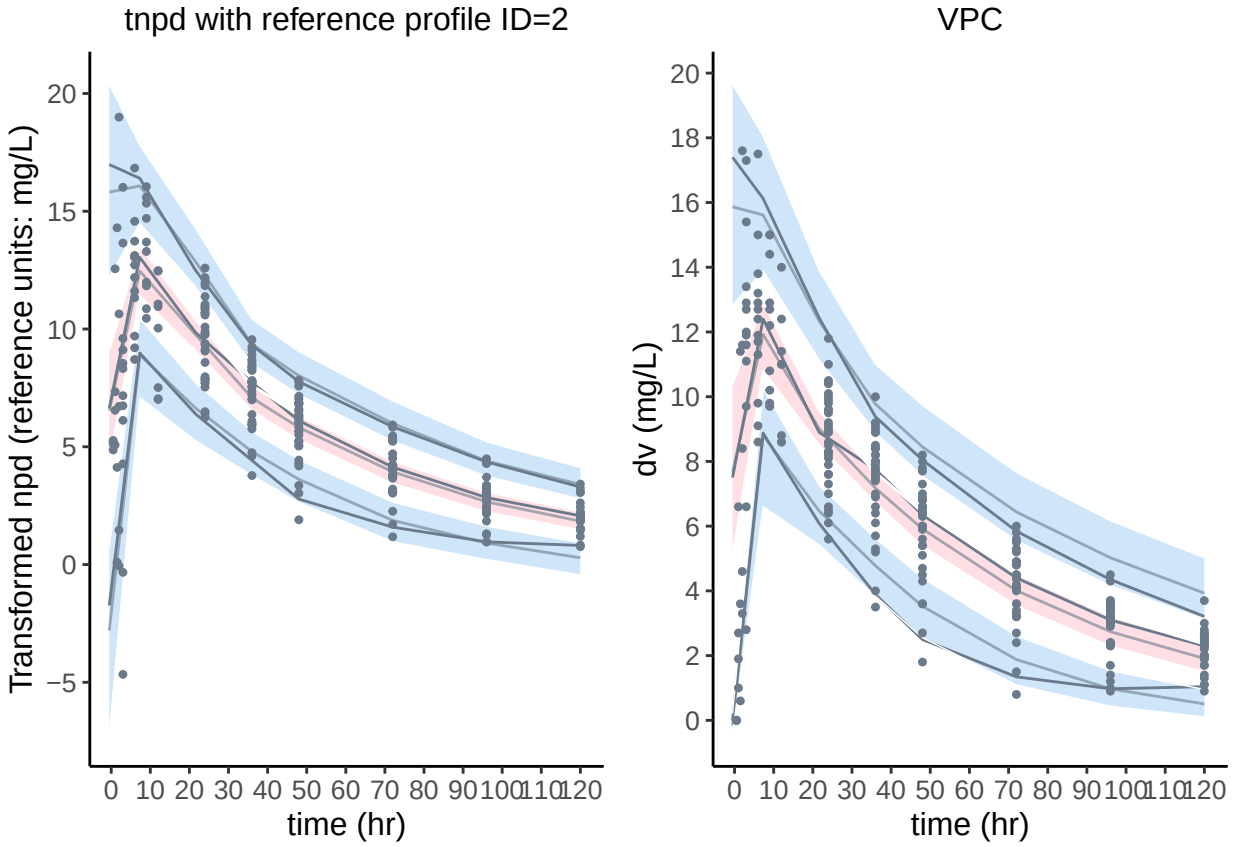
npd for the warfarin example in the documentation

Running the example:

```
## ------------------------------------------------
## Distribution of npde :
##       nb of obs: 247
##            mean= 0.02928   (SE= 0.059 )
##        variance= 0.8549   (SE= 0.077 )
##        skewness= -0.07211
##        kurtosis= -0.4172
## ------------------------------------------------
## Statistical tests (adjusted p-values):
##   t-test                 : 1
##   Fisher variance test   : 0.288
##   SW test of normality   : 1
```

```
##   Global test           : 0.288
## ---
## Signif. codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
## ----------------------------------------------
```

Reference plot for subject 2

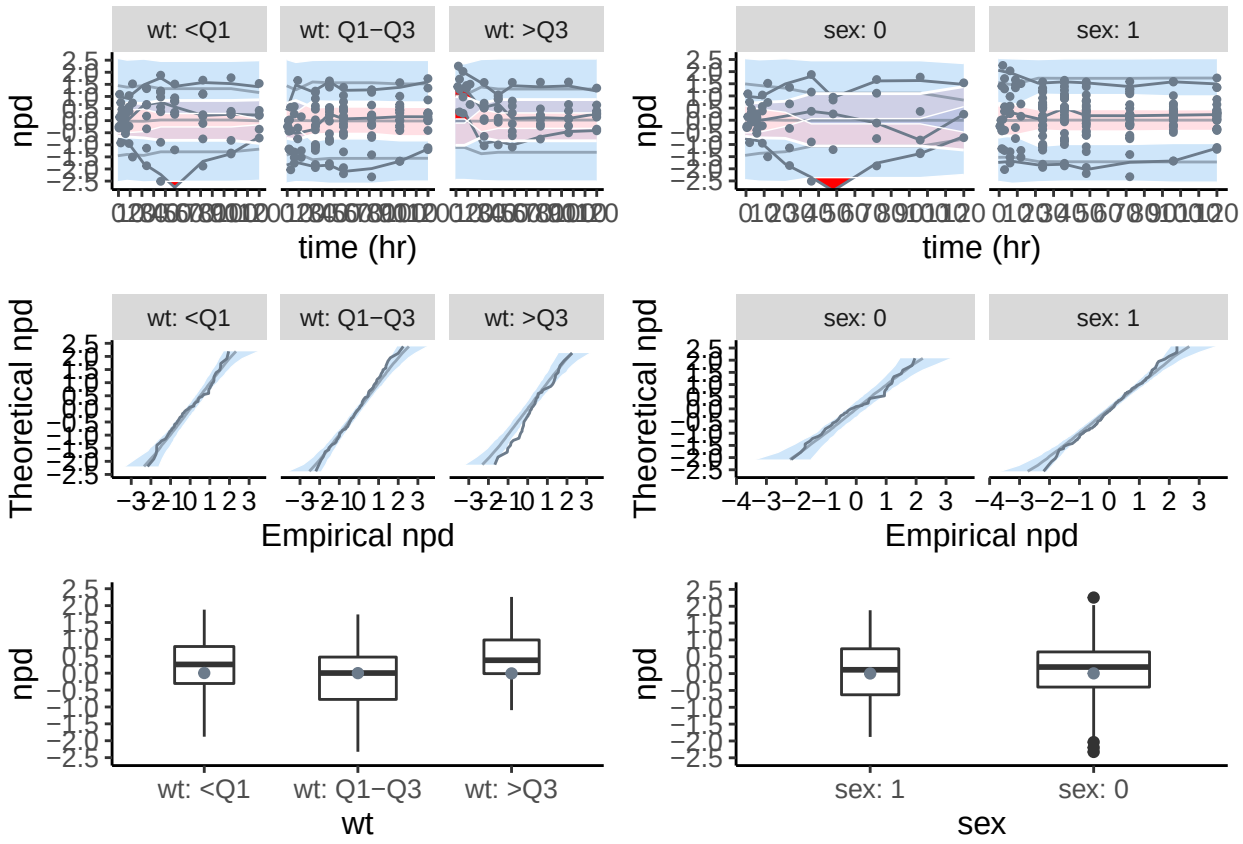


The code below shows several diagnostic plots for covariates:

- plots like scatterplots or distribution plots can be split over categories of covariates
- we can also plot the distribution of eg npd versus the categories as boxplots

```
## Warning: Removed 2 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 2 rows containing non-finite values (stat_boxplot).
```



End of code

Exit development mode.

```
dev_mode() # development mode
```

```
## Dev mode: OFF
```