Appendix A

Brief description of the model

The model used a number of assumptions to combine three sources of data: 1) epidemiological evidence on the relationship between the RR of all-cause mortality and quintile of exercise capacity; 2) individual patient data reporting the distribution of physical activity levels in adults in England; 3) life tables for males and females, showing for each age the probability that an individual will die before their next birthday. From this data, the model produced two natural history scenarios. In these natural history scenarios the numbers of deaths before the age of 60 of a cohort of 100,000 50 year old males, and 100,000 50 year old females, were simulated. Individuals were assigned physical activity levels drawn from a distribution which reflected typical levels observed in English adults. Annual risks of mortality were adjusted upwards for more sedentary people and downwards for more active people. This was done by assigning an activity category to each person, ranging between 1 and 100. Category 1, means that the individual’s physical activity level is in the bottom 1% of the range of exercise levels observed, category 2 means the individual’s physical activity level is within the 1-2% range, up to category 100, which means the individual’s physical activity is in the top 1% of the distribution. In the natural history scenarios, an equal number of people can be expected to be in each of the 100 physical activity categories by definition; in practice the categories were only approximately equal because truncation error meant they were not fully continuous.

Each of the 100 relative physical activity categories was assigned a RR of all-cause mortality. People who were in the lowest categories have a RR significantly above 1, meaning that each year it is supposed that their annual mortality risk is greater than the national average; people in the highest categories have RRs significantly below 1, meaning they are assumed to have a reduced annual mortality rate relative to their peers.
In the intervention scenarios, where it is assumed a physical activity intervention has been successful in increasing mean physical activity levels, each individual’s physical activity in the baseline scenario was increased by a given amount. In one of the intervention scenarios, each individual increased their physical activity levels by the same amount. In the other two scenarios, the increase in physical activity depended on the baseline levels of physical activity, and the distribution of increases was either left-skewed, meaning more sedentary people gained most, or right-skewed, meaning that the most active people gained most. The differences in the numbers of lives saved assuming the skewless, left-skewed, or right-skewed scenarios provides an indication of how the direction of skew – whether it is the least active or the most active to respond most to an intervention – modifies the public health benefits of physical activity interventions. An additional analysis explores this further by estimating the relationship between the relative effectiveness of an intervention in saving lives and the degree of skew.

**Data used to estimate baseline physical activity levels**

We use data recorded in the Health Survey for England, 2008, to produce estimates of baseline physical activity levels for adults of working age in England. [1] The 2008 edition of the Health Survey for England focused on physical activity and fitness, and included a sub-sample of adults who wore an accelerometer for a week following the survey. [1] This provided an objective measure of the amount of physical activity each participant performed. The sub-sample of patients with valid accelerometry data were used, and assumed to be representative of the physical activity levels of the general population. Data from participants aged between 25 years and 60 years were used. This is because reduced rates of car ownership and stable employment were assumed to affect patterns of physical activity in the under 25s, and retirement was assumed to reduce typical levels of occupational physical activity performed by people aged 60 and over. An additional series of analyses, where data from adults aged 18 to 60 were used, were conducted and produced substantially similar results.
Data used to estimate the relationship between relative physical activity
and relative risk (RR) of mortality

A range of epidemiological studies exist which show there is a nonlinear relationship between
measures of physical activity levels relative to others in one’s peer group, and relative risk of
mortality. Such studies show that the less active people are, the higher their relative risk of
mortality. This inverse relationship appears to be consistent regardless of the measure of physical
activity level used, the population, and the number of distinct categories into which physical activity
levels are divided. [2–12]

For this model, we used results reported in a paper published in 2002, which estimated RR of all-
cause mortality by quintile of exercise capacity. [12] The paper reported a US-based study of 6213
male participants, with a mean age of 59 years, who were referred for exercise capacity testing in
1987. Peak exercise capacity at this test was measured in metabolic equivalents (METs). Participants
were followed up using the Social Security death index until July 2000, and the relationship between
quintile of exercise capacity and RR of all-cause mortality identified. RRs were reported separately
for people with and without cardiovascular disease (CVD); we used the RR estimates for those
without CVD.

Although it should be noted that exercise capacity is not the same as physical activity level, a paper
published in 2004, involving a representative subset of the same study sample, found that the
relationship between RR and quartile of physical activity level, and the relationship between RR and
quartile of exercise capacity level, were very similar, although the exercise capacity measure was a
stronger predictor of mortality rates. [11] Results reported in the earlier 2002 paper was therefore
used because this paper used a larger sample size, and reported the more predictive measure
(exercise capacity) by quintile rather than quartile.

The RRs reported by exercise capacity quintiles were used to produce a continuous function
mapping RR of all-cause mortality onto relative physical activity. Relative physical activity was
defined over a range between 0 (least active) to 1 (most active). The RRs by quintile reported in the
2002 paper were scaled by dividing them by the third quintile, so that people who did more physical
activity than average had RRs less than 1, and people who did less physical activity than average had
RRs greater than 1. These rescaled RRs were then converted into a series of five points, where the
horizontal coordinates were defined as the midpoints of the quintile (0.1 for the first quintile, 0.3 for
the second quintile, 0.5 for the third quintile, and so on); and the vertical coordinates were the RRs
reported for each quintile. A curvilinear line was fit through these points by assuming a power law
relationship between RR and relative physical activity; the parameter values of this power law were
identified by fitting a linear regression of the logarithm of RR against the logarithm of relative
physical activity level. The resulting curve had a very close fit to the points ($R^2 = 0.98$), which
compared favourably to the fit produced by assuming a linear relationship ($R^2 = 0.86$).

Power laws tend towards infinity as the predictor variable (relative exercise level) tends towards
zero, predicting that the most inactive people have very highly elevated RRs. This may not be
clinically plausible given that that the simulated population are not assumed to be drawn from a
terminally ill patient population, and so the maximum predicted RR was therefore capped at 10, a
high but plausible value. No similar lower cap was applied to the amount by which people doing
more than the median level of physical activity could reduce their RR, and so estimates from the
simulation model are likely to be relatively conservative. The capped power law was used to produce
an RR value for each of the 100 physical activity categories using a numerical integration approach.

**Use of life table data**

The probability of a person dying in each year between the ages of 50 and 60 years inclusive was
estimated using office for National Statistics life table data [13]. Without any adjustment, the
probability of a fifty year old reaching the age of sixty is the product of the probabilities of surviving
to each age from 51 to 60 conditional on surviving the previous year. The conditional probabilities of
surviving each year are the complements of the probability of dying within the next year. In the
simulations, the proportions of people surviving in each of the 100 physical activity categories were estimated by multiplying the conditional probabilities of dying at each age from 51 to 60 years by the RR for that particular physical activity category. The health effects of different hypothetical interventions by changing the proportion of the population cohort in each of the 100 physical activity categories described above.

Method of varying concentration of additional physical activity by baseline activity

In order to simulate variation in the distribution of additional physical activity as a function of baseline levels of physical activity, a probability distribution known as a Beta distribution was used. For the model, the predictor (input) variable for the distribution is the amount of baseline physical activity, over the range 0 (least active) to 1 (most active), and the response (output) variable is the amount of additional physical activity conditional on the baseline activity level. The Beta distribution has two parameters, $\alpha$ and $\beta$. By setting both parameters to 1, the Beta distribution, Beta $(1, 1)$, becomes a Uniform distribution, which is equivalent to supposing that all participants increase their physical activity levels by the same amount. By adding a small amount, say 0.05, to the $\alpha$ parameter and subtracting this same amount to the $\beta$ parameter, a slightly left-skewed distribution, Beta$(1.05, 0.95)$, is produced, which is equivalent to supposing that the less physically active people do slightly more additional physical activity than the more physically active people. Subtracting this amount from the alpha parameter and adding it to the beta parameter produces Beta $(0.95, 1.05)$, a slightly right-skewed distribution in which the people who do the most additional physical activity were more physically active to begin with. The two distributions Beta $(1.05, 0.95)$ and Beta $(0.95, 1.05)$ are mirror images of each other, with Beta $(1.05, 0.95)$ skewed to the left by the same amount that Beta $(0.95, 1.05)$ is skewed to the right. The further the two parameters are from 1, the more skewed the distributions become. Because of this, the effect of both the magnitude (slightly skewed or very skewed) can be assessed independently of the direction (left skewed or right skewed) of the skew.
In the results below, a left-skewed scenario is described which uses Beta (1.15, 0.85), and a right-skewed scenario described which uses Beta (0.85, 1.15). The effect of both direction and magnitude of skew is also explored for a wide range of distributions by varying a ‘skew’ parameter $\tau$ over a range of values from -0.20 to 0.20, and defining the resulting Beta distributions as Beta ($1 - \tau, 1 + \tau$).

A numerical integration approach was used to produce estimates increases in physical activity levels for each of the 100 physical activity categories. In all hypothetical physical activity interventions, the mean increase in physical activity was assumed to be 10. (i.e. ten additional minutes of MVPA per day)
References


Appendix B: The R code used to implement the model

# This R script file contains the complete model described in
# Minton, Dimairo, Everson-Hock, Scott, Goyder (2013), "Exploring the relationship between baseline
# physical activity levels and mortality reduction associated with increases in physical activity:
# a modelling study"

# ADDITIONAL FILES REQUIRED
# =============
# Some additional sources of data are required:

# "HSE2008/hse08ai.dta" - A file from the Health Survey for England 2008, in Stata (.dta) format,
# which provides estimates of the baseline distribution of physical activity
# - available to download from data-archive.ac.uk

# "LifeExpectancies_male.csv" and "LifeExpectancies_female.csv" - UK lifetables, from the Office
# for National Statistics, converted into separate .csv files for males and females
# - available to download from: http://www.ons.gov.uk/ons/taxonomy/index.html?nscl=Interim+Life+Tables

# As I do not own the copyright to these files, they are not included, but both are free to download
# from the locations indicated above

# STRUCTURE OF R CODE

# ====================

# The R code below is structured as follows:

# 1) 'Housekeeping' - removes existing data from the R workspace and sets the working directory
# 2) 'Constants' - defines various constants for use in the model
these may be modified to assess the dependence of assumptions on the model results

3) 'Functions' - A series of user-defined functions which are made use of within the model and later analyses

4) 'Main Script' - Code which makes use of the code run previously in order to perform the main analyses

# including produce graphs

# Within the 'Main Script', sections of code which reproduce the figures and tables in the manuscript are indicated

# At the end of the main body of code, additional comments are provided suggesting how to develop the model in order to test and develop the current model

# Housekeeping

rm(list=ls()) # Remove workspace contents
setwd("X:/Booster/Booster Model Manuscript/rCode") # Set working director - change to a local directory
# g for 'global'

g.lower.age <- 25  # lower age of interest in years - can be changed
#g.lower.age <- 18  # sensitivity analysis run previously - starting at 18 rather than 25 years
g.upper.age <- 60  # upper age of interest in years - can be changed

g.additional.minutes.mvpa <- 10  # assumed additional number of extra minutes on average

g.num.digits <- 100  # number of discrete categories (a large number used in order to make 'semi-
                           continuous)
g.max.break.val <- 200  # maximum value in breaks

g.filename <- "HSE2008/hse08ai.dta"  # location of HSE file for extracting distribution of minutes

#g.max.rr <- 5  # cap the RR implied by the power law
g.max.rr <- 10 # cap the RR implied by the power law

g.target <- 30 # Recommended target level of MVPA - can be changed

g.sim.from.age <- 50 # Starting age of simulated cohort

g.sim.to.age <- 60 # End age of simulated cohort

# FUNCTIONS

# Proportion of sample who met a target
PropMetTarget <- function(X, target=g.target){
  return(length(which(X > target))/length(X))
}

# Report cumulative distribution
MakeCumulative <- function(X){
    y <- max(X) * cumsum(sort(X))/sum(X)
    x <- (1:length(X))/length(X)

    Dta <- data.frame(x=x, y=y)

    return(Dta)
}

# Get distribution from HSE
# baseline physical activity level distribution
# - from Health Survey for England, 2008
GetDist <- function(filename=g.filename, lower.age=g.lower.age, upper.age=g.upper.age){
    require(foreign)

    DataAI <- read.dta("HSE2008/hse08ai.dta", convert.factors=F)

    mins <- DataAI$averagemvpaminutespervalidday2 # from accelerometry data
    ages <- DataAI$age
# I want to include only those people who gave valid accelerometry readings and are
# aged between 25 and 60 years

mvpa.baseline <- mins[mins>=0 & ages > lower.age & ages < upper.age]

# browser()
return(mvpa.baseline)

# Get Power Law coefficients

GetPowerCoefs <- function(Data){
  # power law
  # y = alpha * x^beta
  # log(y) = log(alpha) + beta * log(x)

  logMod <- lm(log(y) ~ log(x), data=Data)
  alpha <- exp(logMod$coef[1])
  beta <- logMod$coef[2]

  coefs <- list(alpha=alpha, beta=beta)
  return(coefs)
# Digitise & Integrate

```r
DigIntegrate <- function(digits=g.num.digits, mean.effect=.gadditional.minutes.mvpa, tau.skew=0.15, pow.coef, max.rr=g.max.rr){
  alpha <- pow.coef$alpha
  beta <- pow.coef$beta
  left.beta.mult <- rep(NA, digits)
  right.beta.mult <- rep(NA, digits)
  rr.quantile <- rep(NA, digits)

  digit.vector <- seq(0, 1, by=1/digits)
  for (i in 1:digits){
    left.beta.mult[i]   <- integrate( function (x) dbeta(x, 1 - tau.skew, 1 + tau.skew), lower=digit.vector[i], upper=digit.vector[i+1])[["value"]]
    right.beta.mult[i]  <- integrate( function (x) dbeta(x, 1 + tau.skew, 1 - tau.skew), lower=digit.vector[i], upper=digit.vector[i+1])[["value"]]
    rr.quantile[i]      <- integrate( function (x) (alpha * x ^ beta), lower=digit.vector[i], upper=digit.vector[i+1])[["value"]]
  }
  to.add.left <- mean.effect * left.beta.mult / mean(left.beta.mult)
```

to.add.right <- mean.effect * right.beta.mult / mean(right.beta.mult)

# I want to divide by the median rather than the mean, as I want half of the population to
# have mortality rates above the central value, and half to have values below the central value
rr.quantile <- rr.quantile / median(rr.quantile)

# I also want to set the maximum relative risk predicted to 5, to make this a relatively conservative
# estimate

rr.quantile[rr.quantile > max.rr] <- max.rr

output <- list(to.add.left=to.add.left, to.add.right=to.add.right, rr.quantile=rr.quantile)
return(output)

# Calculate additional amount of physical activity for each baseline exercise digit, given mean effect
# and amount of skew in effect
DigIntegrate.skew <- function(digits=g.num.digits, mean.effect=g.additional.minutes.mvpa, tau.skew=0.15){
  skew.beta.mult <- rep(NA, digits)
  digit.vector <- seq(0, 1, by=1/digits)
for (i in 1:digits){
  skew.beta.mult[i] <- integrate( function (x) dbeta(x, 1 + tau.skew, 1 - tau.skew),
  lower=digit.vector[i], upper=digit.vector[i+1])[['value']]}

to.add.skew <- mean.effect * skew.beta.mult / mean(skew.beta.mult)
return(to.add.skew)
}

# Calculate RR to apply to each exercise digit
DigIntegrate.rr <- function(digits=g.num.digits, pow.coef, max.rr=g.max.rr){
  alpha <- pow.coef$alpha
  beta <- pow.coef$beta
  rr.digits <- rep(NA, digits)

  digit.vector <- seq(0, 1, by=1/digits)
  for (i in 1:digits){
    rr.digits[i] <- integrate( function (x) (alpha * x ^ beta), lower=digit.vector[i],
    upper=digit.vector[i+1])[['value']]}
}
rr.digits <- rr.digits / median(rr.digits)
rr.digits[rr.digits > max.rr] <- max.rr
return(rr.digits)
}

# Extract mortality risks from lifetables
GetMorts <- function(sex="male", lower.age=g.sim.from.age, upper.age=g.sim.to.age){
  if(sex="male"){
    LT.Data <- read.csv("Lifetables/LifeExpectancies_male.csv")
  }
  if(sex="female"){
    LT.Data <- read.csv("Lifetables/LifeExpectancies_female.csv")
  }
  # q_x := the mortality rate between ages x and x+1
  unadjusted.morts <- LT.Data$q_x[LT.Data$x > lower.age & LT.Data$x <=upper.age]
  return(unadjusted.morts)
}

# q_x := the mortality rate between ages x and x+1
unadjusted.morts <- LT.Data$q_x[LT.Data$x > lower.age & LT.Data$x <=upper.age]
return(unadjusted.morts)
# Calculate exercise digit-specific 'lifetables'

```r
CalcAlive <- function(unadjusted.morts, digits, rr.digit) {
  prop.alive.by.digit <- rep(1, digits)
  years.run <- length(unadjusted.morts)
  for (j in 1:digits) {
    for (i in 1:years.run) {
    }
  }
  return(prop.alive.by.digit)
}
```

# Calculate the distribution of people by PA digit

```r
GetDists <- function(dta, num.digits) {
  output <- rep(NA, num.digits)
  for (i in 1:num.digits) {
    # Code for calculating distribution
  }
  return(output)
}
```
output[i] <- length(which(dta==i))
}
output <- output/sum(output)
return(output)
}

# Get the distribution of values
mvpa.base <- GetDist()

# calculate proportion meeting target of 30 minutes per day
prop.above.threshold.base <- PropMetTarget(mvpa.base, 30)
# 45.8% met target in baseline condition


# plot this as a histogram

```r
#png("PhysAct_Baseline.png", 800,600) # uncomment this line to produce an image file
hist(mvpa.base, breaks=seq(0, g.max.break.val, by=2.5), xlim=c(0,200),
    col="grey",
    xlab="minutes of moderate to vigorous physical activity",
    main="Minutes of moderate or physical physical activity
n(Baseline)",
    ylim=c(0,150)) -> mvpa.base.hist
abline(v=30, lwd=2, lty="dashed")
#dev.off() # uncomment this line if also uncommenting the png line above
```

# comparator 1: equal increase condition

```r
mvpa.comp_equal <- mvpa.base + g.additional.minutes.mvpa
prop.above.threshold.comp_equal <- PropMetTarget(mvpa.comp_equal, 30)
# 63.6% met target in equal baseline condition
```

#png("PhysAct_comp_equal.png", 800,600) # uncomment this line to produce an image file
```r
hist(mvpa.comp_equal, breaks=seq(0,g.max.break.val, by=2.5), xlim=c(0,200),
    col="grey",
    xlab="minutes of moderate to vigorous physical activity",
    main="Minutes of moderate or physical physical activity
n(Baseline)",
    ylim=c(0,150)) -> mvpa.comp_equal.hist
abline(v=30, lwd=2, lty="dashed")
#dev.off() # uncomment this line if also uncommenting the png line above
```
xlab="minutes of moderate to vigorous physical activity",
main="Minutes of moderate or physical physical activity\n(Comparison 1: Equal gain)",
ylim=c(0,150)) -> mvpa.comp_equal.hist
abline(v=30, lwd=2, lty="dashed")
#dev.off() # uncomment this line if also uncommenting the png line above

mvpa.base.quants <- quantile(mvpa.base, seq(0,1, by=1/g.num.digits))
mvpa.base.quantcat <- cut(mvpa.base, mvpa.base.quants, label=F, include.lowest=T)

# get Distribution of additions

left.skew <- DigIntegrate.skew(g.num.digits, mean.effect=g.additional.minutes.mvpa, tau.skew= -0.15)
right.skew <- DigIntegrate.skew(g.num.digits, mean.effect=g.additional.minutes.mvpa, tau.skew= 0.15)

mvpa.comp_left <- mvpa.base # left-skewed comparison
mvpa.comp_right <- mvpa.base # right-skewed comparison
for (i in 1:length(mvpa.comp_left)){
    mvpa.comp_left[i] <- mvpa.comp_left[i] + left.skew[ max(1, mvpa.base.quantcat[i] ) ]
    mvpa.comp_right[i] <- mvpa.comp_right[i] + right.skew[ max(1, mvpa.base.quantcat[i] ) ]
    # uncomment the line below for a 'debug mode', in which more information is displayed
    # cat(mvpa.base[i], "\t", mvpa.comp_left[i], "\t", max(1, mvpa.base.quantcat[i]), "\t", left.skew[max(1,
    # mvpa.base.quantcat[i]]), "\n")
}

# Apply upper thresholds
mvpa.comp_left[mvpa.comp_left > max(mvpa.base.hist$breaks)] <- max(mvpa.base.hist$breaks, na.rm=T)
mvpa.comp_right[mvpa.comp_right > max(mvpa.base.hist$breaks)] <- max(mvpa.base.hist$breaks, na.rm=T)

prop.above.threshold.comp_left <- PropMetTarget(mvpa.comp_left, 30)

# 64.6% met target in this condition

prop.above.threshold.comp_right <- PropMetTarget(mvpa.comp_right, 30)

# 61.1% met target in this condition
# Produce histogram

```r
hist(mvpa.comp_left, breaks=seq(0,g.max.break.val, by=2.5), xlim=c(0,200),
    col="grey",
    xlab="minutes of moderate to vigorous physical activity",
    main="Minutes of moderate or physical physical activity\n(Comparison 2: Left skew)",
    ylim=c(0,150))
abline(v=30, lwd=2, lty="dashed")
```

```r
hist(mvpa.comp_right, breaks=seq(0,g.max.break.val, by=2.5), xlim=c(0,200),
    col="grey",
    xlab="minutes of moderate to vigorous physical activity",
    main="Minutes of moderate or physical physical activity\n(Comparison 3: Right skew)",
    ylim=c(0,150))
abline(v=30, lwd=2, lty="dashed")
```

`#` produced histogram

```r
text(45, 85, "Figure 1")
```

`#` Figure 1

`#` Figure 1

```r
#`
"Figure 1"
`#`
```
The code below produces Figure 1 from the manuscript:

```r
#png("Hist_Comparisons.png", 1000, 1000) # uncomment to produce image file
split.screen(c(2,2))
screen(1)

hist(mvpa.base, breaks=seq(0, g.max.break.val, by=2.5), xlim=c(0,200),
     col="grey",
     xlab="Minutes MVPA per day",
     main="a) Baseline",
     ylim=c(0,150)) -> mvpa.base.hist
abline(v=30, lwd=2, lty="dashed")

screen(2)

hist(mvpa.comp_equal, breaks=seq(0,g.max.break.val, by=2.5), xlim=c(0,200),
     col="grey",
     xlab="Minutes MVPA per day",
     main="b) Comparison 1: Equal gain",
     ylim=c(0,150)) -> mvpa.comp_equal.hist
abline(v=30, lwd=2, lty="dashed")
```
screen(3)

hist(mvpa.comp_left, breaks=seq(0,g.max.break.val, by=2.5), xlim=c(0,200),
  col="grey",
  xlab="Minutes MVPA per day",
  main="c) Comparison 2: Left skew",
  ylim=c(0,150))

abline(v=30, lwd=2, lty="dashed")

screen(4)

hist(mvpa.comp_right, breaks=seq(0,g.max.break.val, by=2.5), xlim=c(0,200),
  col="grey",
  xlab="Minutes MVPA per day",
  main="d) Comparison 3: Right skew",
  ylim=c(0,150))

abline(v=30, lwd=2, lty="dashed")

#dev.off() # uncomment if also uncommenting png line above
# The following code produces the numbers for table 1 of the manuscript
# (NOTE: Due to simulation uncertainty the values will not be slightly different each time)

# produce a histogram object of baseline mvpa categories for manipulating later
#mvpa.baseline.hist <- hist(mvpa.baseline, breaks=50, plot=F)

Redists <- data.frame(
  base = cut(mvpa.base, mvpa.base.quants, label=F, include.lowest=T),
  comp.equal = cut(mvpa.comp_equal, mvpa.base.quants, label=F, include.lowest=T),
  comp.left = cut(mvpa.comp_left, mvpa.base.quants, label=F, include.lowest=T),
  comp.right = cut(mvpa.comp_right, mvpa.base.quants, label=F, include.lowest=T)
)

for (i in 2:4){
Redists[,i][Redists[,1]==g.num.digits] <- g.num.digits

# Mapping relationship
MapData <- data.frame(y = c(4.5, 2.5, 1.8, 1.4, 1),
                       x = c(0.1, 0.3, 0.5, 0.7, 0.9))
MapData$y <- MapData$y/MapData$y[3]

powerCoefs <- GetPowerCoefs(MapData)

# Relative risks
rr.mort.by.digit <- DigIntegrate.rr(digits=g.num.digits, pow.coef=powerCoefs)

# distribution of people by digit
dist.base <- GetDists(Redists$base, g.num.digits)
dist.comp.equal <- GetDists(Redists$comp.equal, g.num.digits)
dist.comp.left <- GetDists(Redists$comp.left, g.num.digits)
```r
dist.comp.right <- GetDistS(Redists$comp.right, g.num.digits)

# Code for calculating numbers in table 1 of manuscript for males
morts.unadjusted.by.year <- GetMorts()

prop.alive.by.digit <- CalcAlive(morts.unadjusted.by.year, g.num.digits, rr.mort.by.digit)

prop.alive.base <- sum(prop.alive.by.digit * dist.base)
prop.alive.comp.equal <- sum(prop.alive.by.digit * dist.comp.equal)
prop.alive.comp.left <- sum(prop.alive.by.digit * dist.comp.left)
prop.alive.comp.right <- sum(prop.alive.by.digit * dist.comp.right)

lives.saved.left <- 100000 * (prop.alive.comp.left - prop.alive.base)
lives.saved.equal <- 100000 * (prop.alive.comp.equal - prop.alive.base)
```
lives.saved.right <- 100000 * (prop.alive.comp.right - prop.alive.base)

# Code for calculating numbers in table 1 of manuscript for females

f.morts.unadjusted.by.year <- GetMorts("female")

f.prop.alive.by.digit <- CalcAlive(f.morts.unadjusted.by.year, g.num.digits, rr.mort.by.digit)

f.prop.alive.base <- sum(f.prop.alive.by.digit * dist.base)

f.prop.alive.comp.equal <- sum(f.prop.alive.by.digit * dist.comp.equal)

f.prop.alive.comp.left <- sum(f.prop.alive.by.digit * dist.comp.left)

f.prop.alive.comp.right <- sum(f.prop.alive.by.digit * dist.comp.right)

f.lives.saved.left <- 100000 * (f.prop.alive.comp.left - prop.alive.base)

f.lives.saved.equal <- 100000 * (f.prop.alive.comp.equal - prop.alive.base)

f.lives.saved.right <- 100000 * (f.prop.alive.comp.right - prop.alive.base)

########################################################################

30
taus <- seq(-0.20, 0.20, by=0.01)
lives.saved.m <- rep(NA, length(taus))
lives.saved.f <- rep(NA, length(taus))

for (i in 1:length(taus)){
  this.tau=taus[i]
  # get Distribution of additions
  this.skew <- DigIntegrate.skew(g.num.digits, mean.effect=g.additional.minutes.mvpa, tau.skew=
                                  this.tau)

  mvpa.this <- mvpa.base

  for (j in 1:length(mvpa.this)){
    mvpa.this[j] <- mvpa.this[j] + this.skew[ max(1, mvpa.base.quantcat[j] ) ]
  }
}
prop.above.threshold.this <- PropMetTarget(mvpa.this, 30)

mvpa.dists.base <- cut(mvpa.base, mvpa.base.quants, label=F, include.lowest=T)
mvpa.dists.this <- cut(mvpa.this, mvpa.base.quants, label=F, include.lowest=T)

mvpa.dists.this[mvpa.dists.base==g.num.digits] <- g.num.digits

# distribution of people by digit
dist.base <- GetDists(mvpa.dists.base, g.num.digits)
dist.this <- GetDists(mvpa.dists.this, g.num.digits)

prop.alive.base.m <- sum(prop.alive.by.digit * dist.base)
prop.alive.this.m <- sum(prop.alive.by.digit * dist.this)
prop.alive.base.f <- sum(f.prop.alive.by.digit * dist.base)
prop.alive.this.f <- sum(f.prop.alive.by.digit * dist.this)

lives.saved.this.m <- 100000 * (prop.alive.this.m - prop.alive.base.m)
lives.saved.this.f <- 100000 * (prop.alive.this.f - prop.alive.base.f)
# Uncomment the line below for a 'debug mode'
# cat(i, "\t", lives.saved.this.m, "\t", lives.saved.this.f, "\n")
lives.saved.m[i] <- lives.saved.this.m
lives.saved.f[i] <- lives.saved.this.f

#png("RelativeEffectiveness_Skew.png", 400, 400)
plot(lives.saved.m/lives.saved.m[taus==0] ~ taus, lwd=2, type="l", ylab="Relative Effectiveness", xlab="Skew")
lines(lives.saved.f/lives.saved.f[taus==0] ~ taus, lwd=2, lty="dashed")
abline(v=0)
abline(h=1)
legend("topright", legend=c("males", "females"), lwd=c(2,2), lty=c("solid", "dashed"))
#dev.off()

# Indicative Further Developments

# 1) To test the dependence of the results on the structural assumptions made:
# Change the values in the 'Constants' section above (prefaced with 'g') and re-run the main script

# 2) To assess the level of variation in values due to stochastic variation
# Run the models many times, saving outputs in a vector object

# 3a) To assess the dependence of the estimates on the sources of data used
# Use different sources of data for lifetables and baseline physical activity levels

# 3b) To assess the dependence of the estimates on the assumed relationship between
# physical activity level and mortality RR, and use the Woodcock estimates instead
Amend the function `Digintegrate.rr` accordingly.

For help and queries please contact Jon Minton, email: nate.minton@gmail.com.