SimVitD: simulation tools for powering vitamin D studies.

Rebecca Mangan   Jason Wyse   Lina Zgaga

Abstract

SimVitD provides simulation based tools to compare supplementation schemes in vitamin D studies. These tools aim to account for and characterise key sources of variability and heterogeneity in vitamin D benefit. Seasonal variation in solar radiation is pronounced, which gives a natural sinusoidal variation in vitamin D status; consequently, the relative contribution of a vitamin D supplementation to the overall vitamin D status, and it's impact, will vary seasonally. Functionality to approximate the power of a study comparing two supplementation schemes via simulation is easily accessible.

Keywords: Statistical power, sample size determination, heterogeneity of treatment effect.

1. Introduction

A large number of observational studies in humans suggest that vitamin D deficiency can have detrimental impacts on health (Theodoratou *et al.* 2014). Further, hundreds of experimental studies in animals and cell lines describe underlying (patho)physiological mechanisms associated with deficiency. Majority of an individual's vitamin D is derived from ultraviolet B (UVB) exposure (Webb and Holick 1988; O'Sullivan *et al.* 2019), and hence follows a strong seasonal pattern. The usefulness of vitamin D in providing immune protection against common maladies is affected by this seasonality. As well as a heterogeneous treatment effect (HTE) between individuals taking a vitamin D supplement, there may be a large intra-individual heterogeneity in the general effectiveness of supplementation depending on the season.

The package **SimVitD** presents simulation based tools to aid planning the comparison of supplementation schemes in vitamin D studies, which aim to account for a set of perceived sources of variability and heterogeneity in vitamin D benefit. As seasonal variation in solar radiation is pronounced, there will be a natural, sinusoidal variation in vitamin D status; consequently, the relative contribution of vitamin D supplementation to the overall vitamin D status, and it's impact, will vary seasonally: while vitamin D supplementation may contribute the majority of the vitamin D in Winter, the same dose may be relatively insignificant in the Summer. Some of the **SimVitD** schemes are shown in Figures 1a and 1b.

Calculating the study power of a randomised control trial or the sample size required for a given power is non-trivial under such circumstances. **SimVitD** uses simulation of exposures and infections at an individual level (microsimulation), to investigate the effects of various

vitamin D interventions on disease rates within a study group. Individual vitamin D status trajectories are simulated throughout the year for two or more groups, exposures and incidencesof infections are simulated based on disease risk at the time of the exposure. The power of the study (or the sample size needed to obtain a given power) can be approximated via simulation. These tools can be utilised in the planning of vitamin D studies.

The remainder of this vignette is organised as follows. Section 2 outlines our proposed models for vitamin D status profiles and supplementation schemes. Section 3 describes simulation of a body's response to vitamin D using exposures to a common infection as example. Section 4 describes the scheme for approximation of the power when comparing two supplementation schemes for vitmain D. Section 5 contains an example on usage of the package.
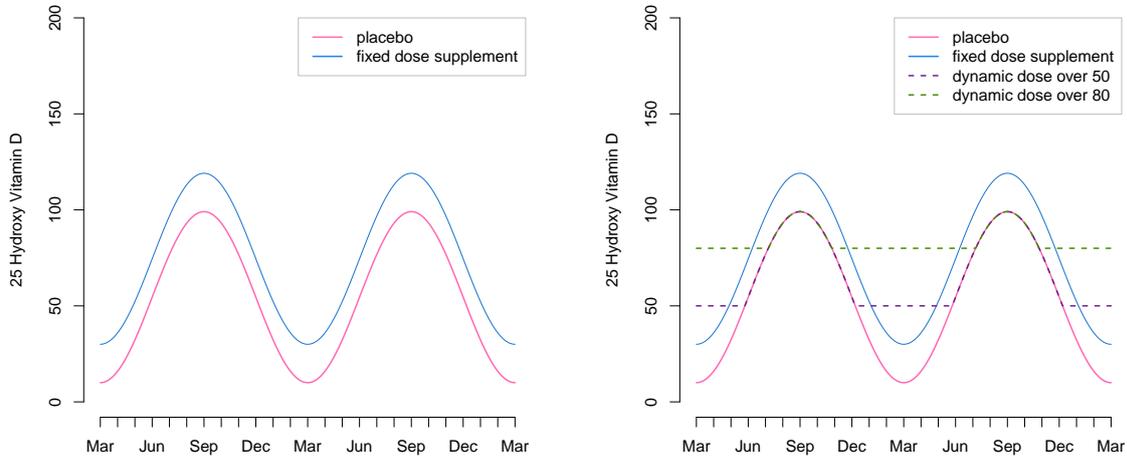
(a) Placebo and fixed dose supplementation shemes.

(b) Dynamic dose supplementation with two threshold levels, and a fixed dose supplementaion.

2. Model for vitamin D status

This section provides an overview of the main components of the proposed simulation based approach to estimating power and sample size determination in **SimVitD**.

Power calculations in **SimVitD** proceed by simulating many realisations of a study. Within each study, individuals' vitamin D status trajectories and potential exposures and protections from infections are simulated *separately*. This is akin to a microsimualation. The core steps of the simulation approach being proposed are:

  (i) simulation of an individual's vitamin D status trajectories

 (ii) simulation of an individual's exposures to infectious agents

(iii) determination of the probability of developing infection at each exposure time, dependent on an individual's vitamin D status at exposure

(iv) simulation of contracting an infection at conditional on step (iii).

The package also contains options to approximate power values for different study designs. This can be used as an aid for planning the experimental approach, most notably determining the required sample sizes for a desired power. The power approximations are based upon many independent replications of the chosen trial design.

2.1. Vitamin D status trajectories

Majority of an individual's vitamin D is derived from UVB exposure (Webb and Holick 1988; O'Sullivan *et al.* 2019). Their vitamin D status will naturally vary throughout the year with peaks during periods with more exposure to sunlight and troughs when this is not the case (Kelly *et al.* 2016; O'Sullivan *et al.* 2017). With this in mind, a squared sine wave curve is used as an approximation of an individual's vitamin D status profile over time. Considering a study with n individuals,

$$V_i^{\mathrm{pl}}(t) = H_i + A_i \sin^2(t), \quad t \geq 0 \tag{1}$$

gives the profile of individual i and t is time. The parameters $H_i > 0$ and $A_i > 0$ control the overall height of the curve and the level of variability between periods with and without alot of exposure to sunlight. This guarantees strictly positive Vitamin D levels matching physical reality. In **SimVitD** an adjustment is made to account for a 1-2 month lag effect from UVB exposure to expressed vitamin D level.

Any reasonable model of vitamin D profiles in a group of individuals would allow for random variation between individuals; we thus assume that

$$A_i \sim \mathrm{Gamma}\left(\kappa_A^2, \frac{\kappa_A^2}{a_0}\right) \qquad H_i \sim \mathrm{Gamma}\left(\kappa_H^2, \frac{\kappa_H^2}{h_0}\right).$$

where these are independently drawn. The parameters a_0 and h_0 give the expected heights and profile amplitudes for individuals. The parameters κ_A and κ_H control the deviation of individual vitamin D profiles around these expected values. More convenient parameterisations of values are discussed in Section 5.

In **SimVitD** these curves can be generated using the `vitd.curve()` function. Multiple curves can be simulated at once, to generate a group of individuals, as shown for example in Figure 2. This function also allows generation of curves for some different kinds of vitamin D supplementation schemes, where an individual takes addtional vitamin D. Two types of supplmentation schemes are avaialable beyond the curve in (1). The curve in (1) corresponds to no supplementation and is called `placebo` in the function through the `type` argument.

Fixed dosing scheme

A fixed dosing scheme corresponds to an individual taking a daily vitamin D supplement of a fixed amount. In this case, the no supplement curve is modified by shifting it up by δ

$$V_i^{\mathrm{fix}}(t) = \delta + V_i^{\mathrm{pl}}(t).$$

This assumes that there is a constant daily effect of the supplement. This scheme corresponds to `fixed-dose` in the `vitd.curve()` function and is shown in Figure 1a. The value of δ can be passed to `vitd.curve()`. Note however, that δ represents the change in detected vitamin D as seen in the blood, and not the dosage of the supplement.

Dynamic dosing scheme

A dynamic dosing scheme allows an individual to be monitored regularly and their vitamin D status kept above a threshold τ. That is

$$V_i^{\mathrm{dyn}}(t) = \max\left\{\tau, V_i^{\mathrm{pl}}(t)\right\}.$$

This scheme corresponds to `dynamic-dose` in the `vitd.curve()` function. A comparison of the placebo and dosing schemes is shown in Figure 1b.

Crossover schemes from placebo to fixed or dynamic dosing

For a crossover study, an individual switches from placebo to either one of a fixed dosing or dynamic dosing scheme. A crossover time ξ is given, and then

$$V_i^{\mathrm{cross}}(t) = V_i^{\mathrm{pl}}(t)\,\mathrm{I}(t \leq \xi) + V_i^{\mathrm{treat}}(t)\,\mathrm{I}(t > \xi),$$

with $\mathrm{I}(\cdot)$ the indicator function and "treat" representing the treatment scheme used i.e. either fixed dosing or dynamic dosing.

Fluctuations and seasonal schedule

The default of the package is to work on the assumption of a northern hemisphere seasonal schedule with Summer months being June to August. Cyclic vitamin D profiles follow the assumed yearly periodic curve with troughs in March and peaks in September (Kelly *et al.* 2016; O'Sullivan *et al.* 2017). The September peak is an adjustment for the 1-2 month time lag it takes absorbed UVB to express as serum measurable vitamin D.

3. Response to vitamin D

The **SimVitD** package allows one to investigate, via simulation, potential benefits of vitamin D supplementation. It presents a toolbox, to investigate potential outcomes of studies for particular states of nature. From this perspective the package is designed to examine the immune boosting properties of having sufficient vitamin D levels. We take an unnamed infection which vitamin D protects against. The assumption is that there is less likelihood of getting the infection if one has high vitamin D levels when susceptible and exposed to it.

Exposures to infection

An individual's exposures to infection over the period of a prospective vitamin D study are simulated from a Possion process. In the case of seasonally concentrated infections (e.g. flu), a non-homegeneous Poisson process (NHPP) with rate function $\lambda(t), t \geq 0$ is used. Simulations from an NHPP in **SimVitD** are done through the R package **poisson** (Brock and Slade 2015).
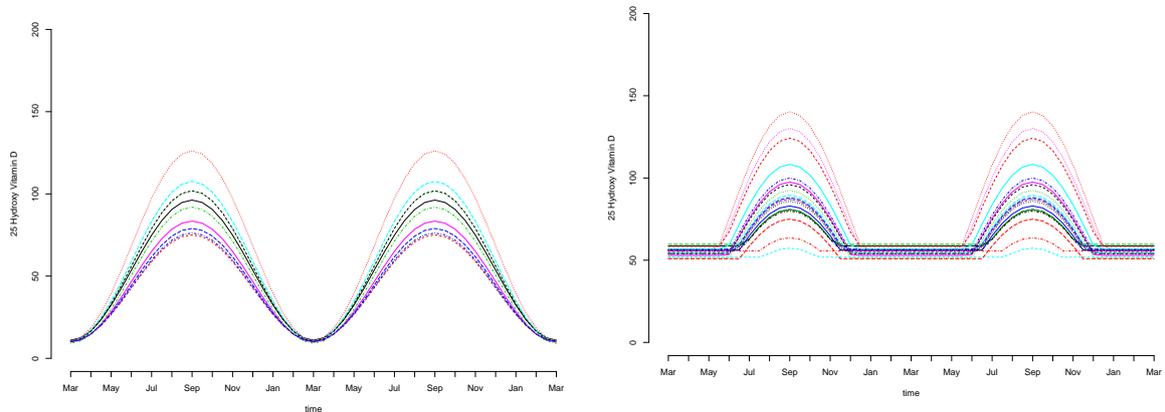
(a) Vitamin D profiles from placebo group

(b) Vitamin D profiles from treatment group with a threshold level of 50

Figure 2: Examples of vitamin D profiles simulated using the `vid.curve()` function. (a) Typical curves displaying random variation for a group of participant's without any supplementation (placebo). (b) Curves for a group where vitamin D is monitored and supplemented so as to be consistently be above 50 (units).

The function $\lambda(t)$ is defined by rescaling an overall rate λ_0 which gives the rate of exposures when most intense

$$\lambda(t) = \lambda_0 \, v(t), \qquad 0 \le v(t) \le 1, \qquad t \ge 0.$$

The function $v(t)$ may be passed by the user. A convenience function `intensity.function()` gives a step function for simple Summer/Winter rates, as shown in Figure 3.

Likelihood of infection

The likelihood an individual gets infection after an exposure depends on their vitamin D status at exposure. This is modulated by a baseline (healthy) prevalence p_0 and a relative risk curve. The probability p_0 gives the probability of a vitamin D replete individual contracting an infection after exposure. The relative risk curve is a member of generalised logistic family

$$g(x) = \ell + \frac{u - \ell}{1 + \mathrm{e}^{a + b\,x}},$$

where x is the vitamin D status. The parameters ℓ, u give the lowest and highest relative risk values. The value of u states how much more likely one is to get infection when completely depleted in vitamin D compared with when one is fully replete. The values of a and b are determined by providing points of inflection of the relative risk curve. These points default to 10 and 70 nmol/l in the package; these are values where we see the greatest change in immunity due to vitamin D supplementation. See Figures 6a and 6b for examples.

Summary of simulation steps

We now summarise the simulation of exposures and infections. Consider individual i and let $T_1, \ldots, T_N$ denote the times at which they are exposed. Exposure times only within the frame

of the study are used: $T_{\text{start}} < T_k \leq T_{\text{end}}$, $k = 1, \ldots, N$.

$$
\begin{aligned}
T_1, \ldots, T_N &\sim \text{NHPP}(\lambda(t)) && \text{simulate the exposure time} \\
L_k &= V_i^{\text{pl}}(T_k) && \text{find the individual's vitamin D status } k = 1, \ldots, N \\
P_k &= p_0\, g(L_k) && \text{get the probability of infection after exposure } k = 1, \ldots, N \\
I_k &\sim \text{Bernoulli}(P_k) && \text{simulate developing infection at exposures } k = 1, \ldots, N.
\end{aligned}
$$

In the case of infections ($I_k = 1$), **SimVitD** includes an option to impose a "holding time"; an exponentially distributed amount of time where the infected individual is not susceptible to a new infection.
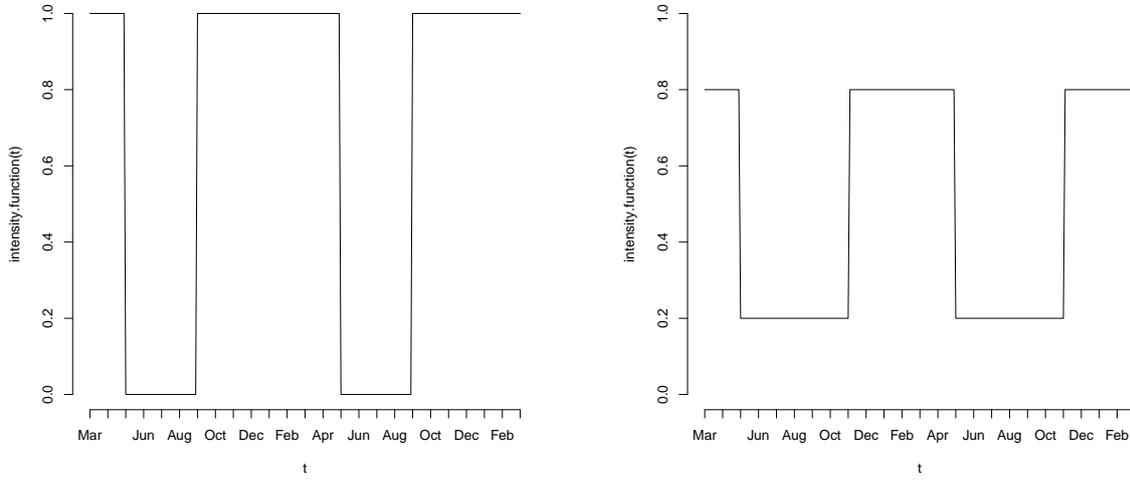
(a) Intensity function for flu season with summer rate = 0 and winter rate = 1

(b) Intensity function for non flu season with summer rate = 0.2 and winter rate = 0.8

Figure 3: Examples of intensity functions simulated using `intensity.function()`

4. Study power

SimVitD has functionality to approximate the power of study comparing two supplementation schemes for three kinds of tests. It is perceivable that the three tests used might be the kind used on data arising from such a study. Denote the supplementation schemes being compared by A (placebo) and B (treatment). Scheme B is hypothesised to boost the immune system and reduce infection incidence. Assume an equal number of participants n assigned to both schemes.

Types of tests

Define

$$
\begin{aligned}
p_A &= \text{Pr\{individual gets} \geq 1 \text{ infection receiving scheme } A\} \\
m_A &= \text{ median number of infections for individual receiving scheme } A
\end{aligned}
$$

and define p_B and m_B similarly. **SimVitD** can approximate power for testing the sets of hypotheses

$$H_0 : p_A \leq p_B \qquad H_A : p_A > p_B \qquad (2)$$
$$H_0 : m_A \leq m_B \qquad H_A : m_A > m_B. \qquad (3)$$

A test of (2) is carried out using a proportions test (`prop.test` in package **stats**), and (3) is tested using a Wilcoxon rank sum test (`wilcox.test` in package **stats**) for comparing two populations. There is also a crossover study design considered where individuals switch supplementation at a specified stage in the study. The median difference Δ, of the number of infections on scheme A less scheme B in the crossover is tested,

$$H_0 : \Delta \leq 0 \qquad H_A : \Delta > 0 \qquad (4)$$

using paired samples, through a Binomial test (`binom.test` in package **stats**).

Approximating the power

The power of a test is

$$\text{Power} = \Pr\{\text{reject } H_0 \,|\, H_A \text{ is true}\}.$$

The rejection decision depends on a significance level α, giving the probability of a Type I error,

$$\alpha = \Pr\{\text{reject } H_0 \,|\, H_0 \text{ is true}\}.$$

SimVitD approximates the power for n participants in each of schemes A and B (or n total in a crossover design) by simulating a large number, N, of studies, and carrying out the test in each of these. The scheme is outlined in Figure 4.
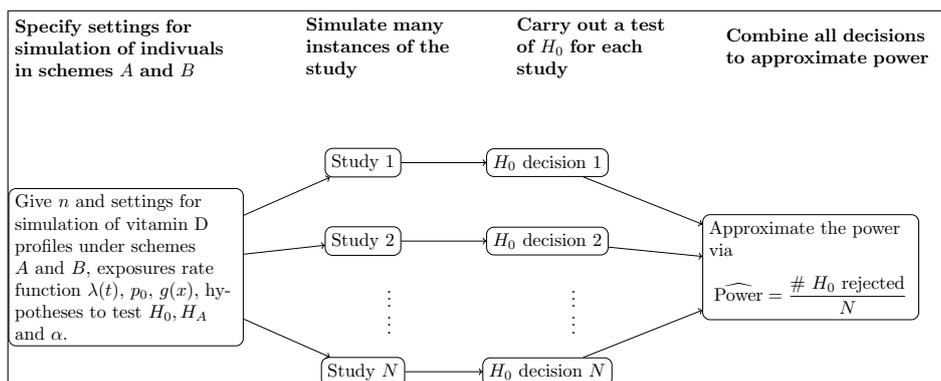

Figure 4: Flowchart showing the simulation process for estimating the power of a study.

Each of the N instances of the study are simulated under H_A being true. For each instance, the test of hypothesis for H_0 is applied to the study's data. The decision (reject or not) is recorded. A sample estimate of the power is then given by

$$\widehat{\text{Power}} = \frac{\# \ H_0 \text{ rejected}}{N},$$

the proportion of times H_0 was rejected when H_A was the actual state of nature. By the law of large numbers, we have

$$\widehat{\text{Power}} \rightarrow \text{Power}$$

as $N \rightarrow \infty$ and we also have

$$\text{error}\left\{\widehat{\text{Power}}\right\} \propto \frac{1}{\sqrt{N}}.$$

Functionality for approximation of the power in **SimVitD** is through the `power.calc()` function. Investigation of the error in approximation of the power is also possible through the `mc.error` argument.

5. Using the package

This section aims to outline an example use case for the package. In this example one group is given a dynamic dosing scheme, keeping vitamin D levels above 50 nmol/l. This is compared to a group given no supplementation (i.e. placebo). The primary endpoint is the number of respiratory tract infections seen in the two groups. The study takes place in Ireland where mean maximum vitamin D levels are approximately 80 nmol/l and mean minimum levels are approximately 10 nmol/l. The researcher aims to enrol a diverse cohort of participants, who will be randomly assigned to one of the two groups. Thus, a large scatter around these mean levels may be expected. The study will last one year, indicated by the arguments `start` and `end`.

Vitamin D status trajectories for the supplemetation schemes described in Section 2.1 are generated using the `vitd.curve()` function. The values of κ_A, a_0, κ_H and h_0 giving the amplitudes and height of simulated vitamin D status trajectories are determined through the prior expectations about the groups under study. The value of a_0 is determined from the difference of the arguments `Min.Height` and `Max.Height` in the package, while the value of h_0 is identical to the value passed for `Min.Height`. The values a_0 and h_0 give the expected amplitude and mean level of the Vitamin D profile curve. The values of κ_A and κ_H quantify the spread about the mean level in multiples of the mean. Regarding the inputs into our package, the value of κ_A, for example, is derived from the `Spread.Max` argument via 100/`Spread.Max`; this translates as larger values of `Spread.Max` corresponding to more variance in the generated Vitamin D profiles. The value of κ_H is derived analogously using the `Spread.Min` argument.

Simulating and plotting vitamin D status profiles

The two groups are generated and plotted using the below code.

```
R> control <- vitd.curve( 20, type = "placebo", start = 0, end = 1, Min.Height = 10,
...                                  Max.Height = 80, Spread.Min = 10, Spread.Max = 10 )
R> plot(control)
R> treatment <- vitd.curve( 20, type = "dynamic-dose", start = 0, end = 1,
...                                  Min.Height = 10, Max.Height = 80, Flat.Height = 50,
...                                  Spread.Min = 10, Spread.Max = 10, Spread.FH = 10 )
R> plot(treatment)
```
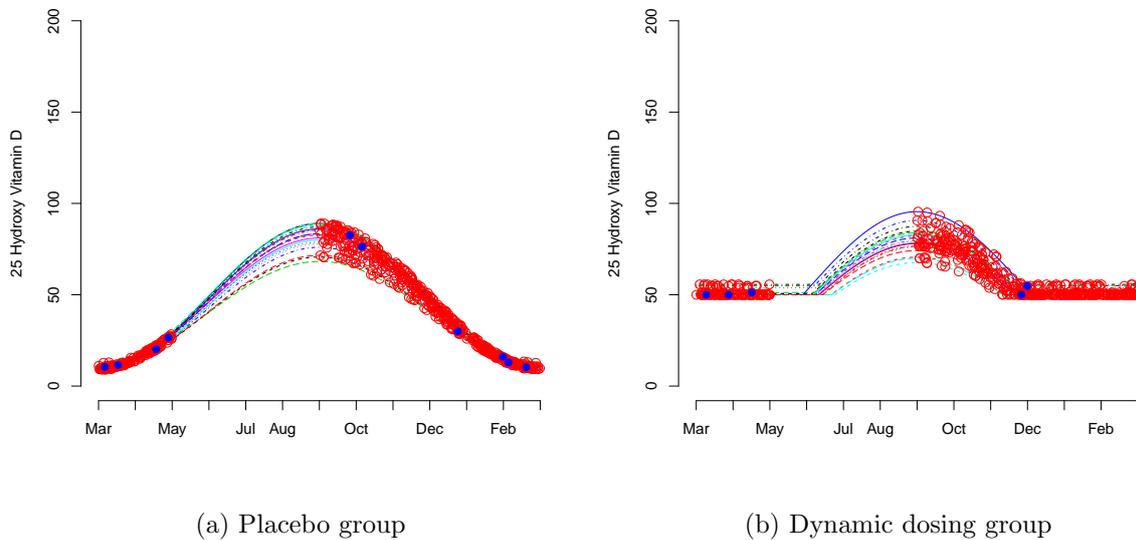
(a) Placebo group

(b) Dynamic dosing group

Figure 5: Vitamin D profiles for the placebo and dynamic dosing groups with exposures in red and infections in blue

Exposures to infection and resulting cases

Next exposure times to respiratory tract infections are simulated. The same intensity function is used for both groups, with a mean of one exposure per week and with no exposures occurring outside of flu season.

```
R> intensfun <- intensity.function( summer.rate = 0, winter.rate = 1, flu = TRUE )
R> control_expos <- exposure.levels( control, rate = 1, intensfun, end = 1 )
R> treatment_expos <- exposure.levels( treatment, rate = 1, intensfun, end = 1 )
R> control_inf <- infection.count( control_expos, baseline = 0.03,
...                                          RR = 3, holding.time = 2 )
R> treatment_inf <- infection.count( treatment_expos, baseline = 0.03,
...                                          RR = 3, holding.time = 2 )
```

The exposures and infections may be plotted over the vitamin D status profiles.

```
R> plot(control)
R> plot(control_expos)
R> infection.count.plot(control_expos, control_inf)
R> plot(treatment)
R> plot(treatment_expos)
R> infection.count.plot(treatment_expos, treatment_inf)
```

Exploring seasonal variation in risk

rr.curve.plot() visualises where exposures occur on the relative risk curve. Figures 6a and 6b show the output of
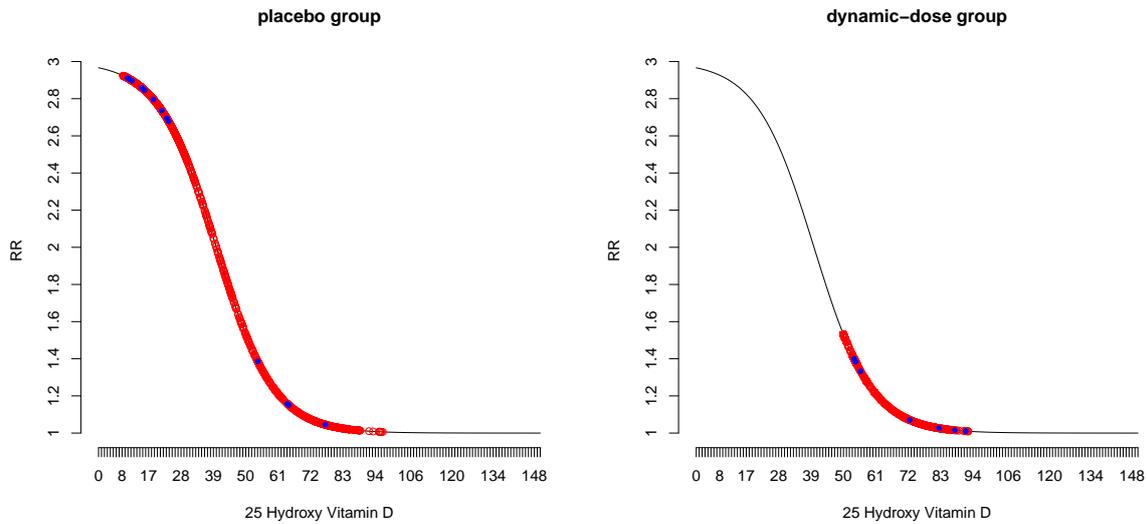
(a) Relative risk at exposures with infections indi-
cated for placebo group.

(b) Relative risk at exposures with infections indi-
cated for dynamic dosing group.

Figure 6: Relative risk curve with exposures and infections overlain. Output from
`rr.curve.plot()`.

```
R> rr.curve.plot(control_expos, control_inf )
R> rr.curve.plot(treatment_expos, treatment_inf )
```

elucidating the discrepancy between the placebo and dynamic dosing schemes by indicating
what the risk level is for infection in each group.

The `rr.profile.plot()` function gives a visualisation tool to explore the seasonal variation
in risk for an individual. It indicates where exposures occurred in the vitamin D status profile
and the corresponding relative risk side-by-side. The exposures resulting in infection are also
indicated. An example is shown in Figures 7a and 7b which are obtained from

```
R> rr.profile.plot( control, control_expos, control_inf )
R> rr.profile.plot( treatment, treatment_expos, treatment_inf )
```

Approximating power of detecting difference in treatment

The summaries from the `infection.count` objects show a difference in the mean between
the placebo and dynamic dosing groups. The mean of the dynamically dosed group being
lower than the mean of the control group. The `power.calc` function is used to determine the
sample size needed to obtain at least 80% power.

```
R> pow <- power.calc( num.participants = c(20,40,60), num.sims = 500,
...          test.type = "count", sig.level = 0.05,
...          vitdcurves.placebo = control, vitdcurves.treatment  = treatment,
```
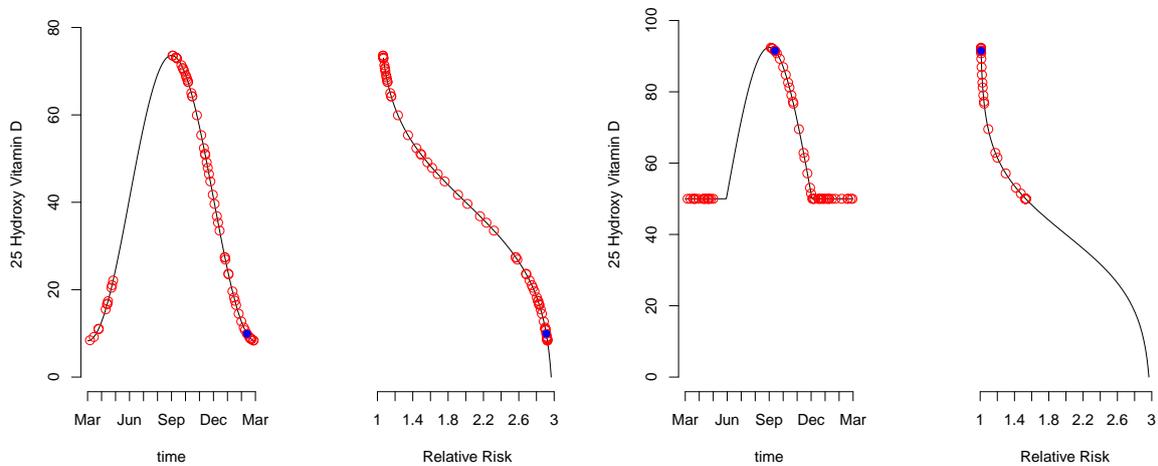
(a) Single participant from placebo group   (b) Single participant from dynamic dose group

Figure 7: Profiles and relative risk for single participants with exposures in red and infections in blue. Output from `rr.profile.plot()`.

```
...          baseline = 0.03, RR = c(2,3,4), rate = 1, intensity.func = intensfun, holding.t
R> plot( pow, x.legend = 20, y.legend = 1,
...          main.legend = "Relative Risk", legend.size = 0.8 )
R> abline( h = 0.8, lty = 2 )
```

A large value of `num.sims` should be used. A value of at least 500 is recommended.

The power calculation shows that in order to get a power of 80% the study would need to enrol at least 60 participants in each group, when a relative risk of 3 is assumed.

Exploring the Monte Carlo error in power approximation

Monte Carlo error in the power approximation may be explored by using the `mc.error` argument to `pow.calc()`.

```
R> pow_mc <- power.calc( num.participants = c(20,40,60), num.sims = 500,
...          test.type = "count", sig.level = 0.05,
...          vitdcurves.placebo = control, vitdcurves.treatment = treatment,
...          baseline = 0.03, RR = c(2,3,4), rate = 1,
...          intensity.func = intensfun, holding.time = 2, mc.error=10 )
R> plot(pow_mc)
```

Crossover design

For the crossover design there are two options. Participants swap from placebo to either one of fixed or dynamic dosing. In the code below, a two year trial is considered, with a crossover from placebo to dynamic dosing after the first year.
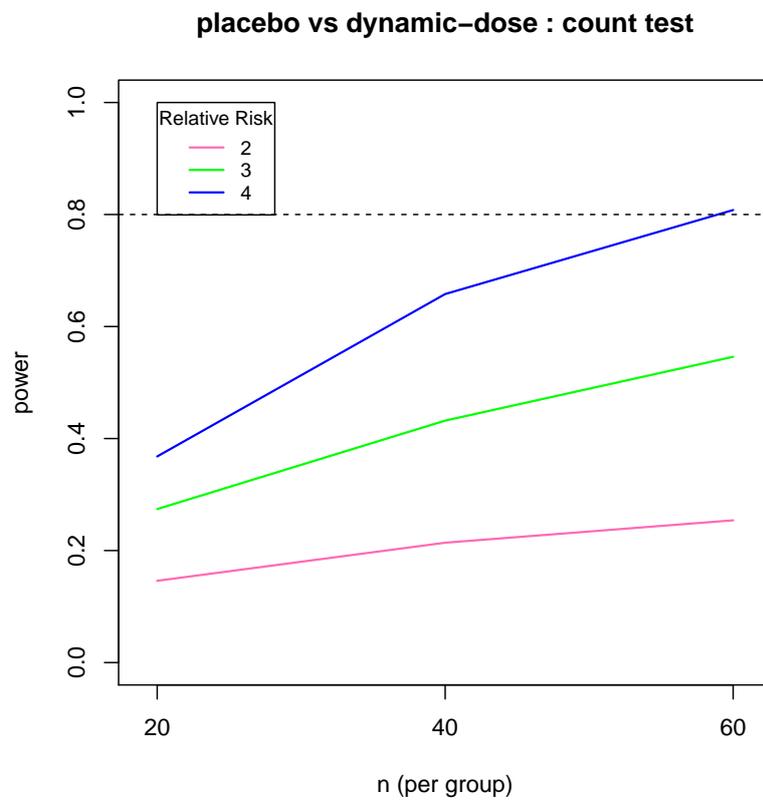
Figure 8: Output from `plot(pow)` showing esitmate of the power at each relative risk level and specified value of n.
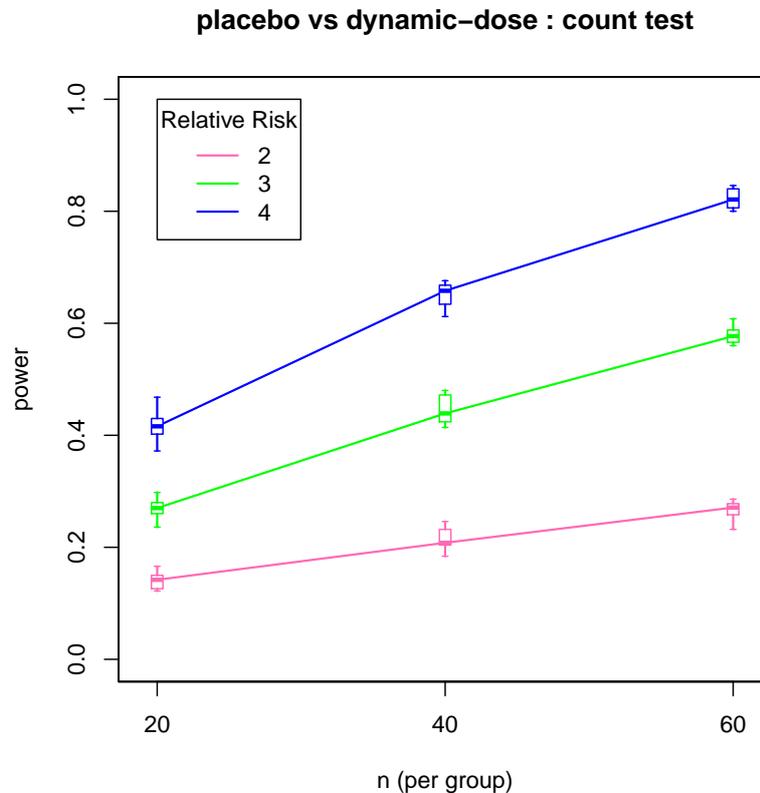
placebo vs dynamic–dose : count test

Figure 9: Output from `plot(pow_mc)` showing the result of the estimation of power over ten runs at each relative risk level and value of n.

```
R> crossover <- vitd.curve( 20, type = "cross-placebo-dynamic-dose",
...                                start = 0, end = 2, cross = 1,
...                                Max.Height = 80, Min.Height = 10, Flat.Height = 50,
...                                Spread.Min = 10, Spread.Max = 10, Spread.FH = 20 )
R> plot(crossover)
```

The plot output is shown in Figure 10 where the cross from placebo to supplementation is clear.

Approximation of the power can also be carried out using `pow.calc()`. In this case, only the `vitdcurves.treatment` argument will be passed. Figure 11 shows the approximate power.

```
R> pow_cross <- power.calc( num.participants = c(20,40,60), num.sims = 500,
...                  test.type = "crossover", sig.level = 0.05,
...                  vitdcurves.treatment  = crossover,
...                  baseline = 0.03, RR = c(2,3,4), rate = 1,
...                  intensity.func = intensfun, holding.time = 2, mc.error=10 )
R> plot(pow_cross)
```
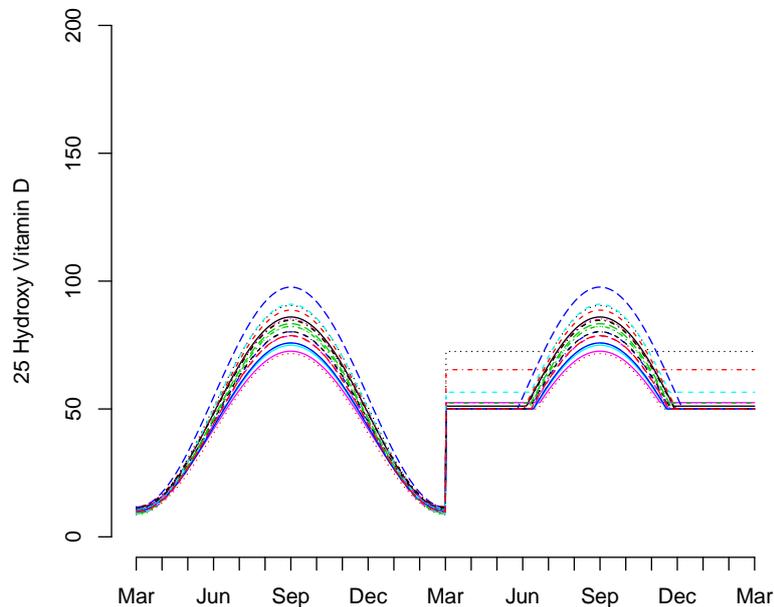
Figure 10: Status curves for two year crossover from placebo to dynamic dosing.

References

Brock K, Slade D (2015). *poisson: Simulating Homogenous and Non-Homogenous Poisson Processes*. R package version 1.0, URL https://CRAN.R-project.org/package=poisson.

Kelly D, *et al.* (2016). "The contributions of adjusted ambient ultraviolet B radiation at place of residence and other determinants to serum 25-hydroxyvitamin D concentrations." *British Journal of Dermatology*, **174**, 1068–1078.

O'Sullivan F, *et al.* (2017). "Ambient UVB Dose and Sun Enjoyment Are Important Predictors of Vitamin D Status in an Older Population." *The Journal of Nutrition*, **147**(5), 858–868.

O'Sullivan F, *et al.* (2019). "Sunshine is an Important Determinant of Vitamin D Status Even Among High-dose Supplement Users: Secondary Analysis of a Randomized Controlled Trial in Crohn's Disease Patients." *Photochemistry and Photobiology*.

Theodoratou E, *et al.* (2014). "Vitamin D and multiple health outcomes: umbrella review of systematic reviews and meta-analyses of observational studies." *British Medical Journal*, **348**, 2035.

Webb AR, Holick MF (1988). "The role of sunlight in the cutaneous production of vitamin D3." *Annual Review of Nutrition*, **8**, 375–399.
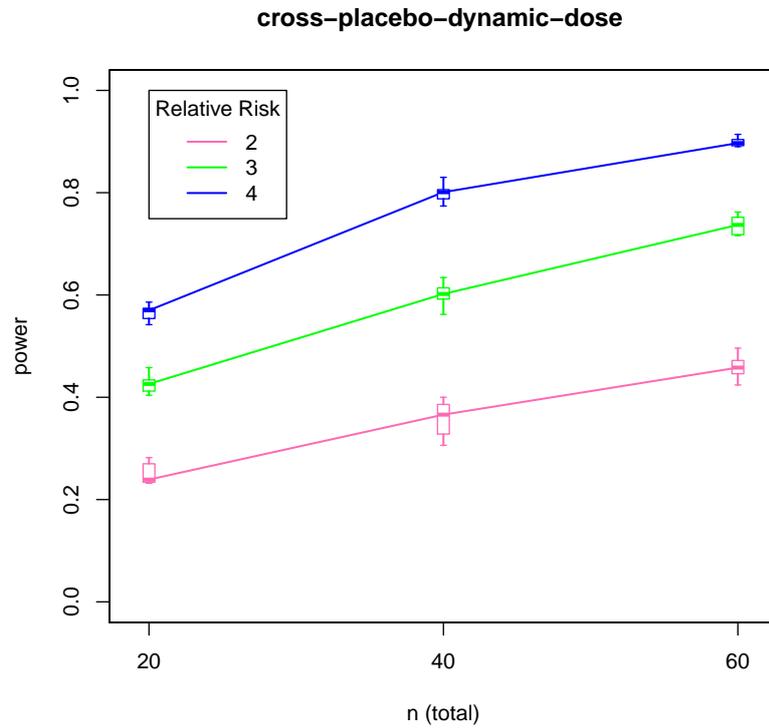
cross–placebo–dynamic–dose

Figure 11: Power curves for two year crossover from placebo to dynamic dosing.

Affiliation:

Jason Wyse & Rebecca Mangan (formerly)
Discipline of Statistics and Information Systems
School of Computer Science and Statistics
Trinity College Dublin
College Green
Dublin 2, Ireland
E-mail: wyseja@tcd.ie

Lina Zgaga
Discipline of Public Health and Primary Care
Institute of Public Health
School of Medicine
Trinity College Dublin
College Green
Dublin 2, Ireland
E-mail: zgagal@tcd.ie