SimVitD: simulation tools for exploring vitamin D trials.

Jason Wyse **Rebecca Mangan** **Lina Zgaga**

Abstract

SimVitD provides simulation based tools to investigate supplementation schemes in vitamin D trials and estimate trial power in comparisons of supplementation arms. These tools aim to account for and characterize 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.

Keywords: Statistical power, sample size determination, heterogeneous treatment effect.

1. Introduction

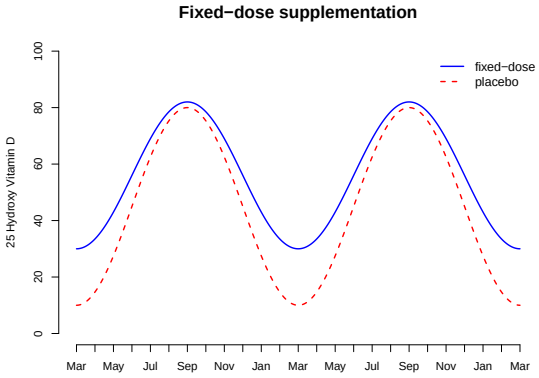
A large number of observational studies have linked vitamin D deficiency with cancer, cognition, cardio-vascular, metabolic, autoimmune, infectious diseases, mortality, and many other illnesses (Theodoratou *et al.* 2014); most recently vitamin D has been implicated in COVID-19 infection and severity (Grant *et al.* 2020; Entrenas Castillo *et al.* 2020). Vitamin D deficiency is very common world-wide: it is estimated that over 1 billion people are vitamin D deficient (Holick 2007). If disease associations are real, tackling the deficiency could have an enormous impact on public health globally. Therefore it is not surprising that there has been a considerable interest in vitamin D in the last two decades. However, randomized controlled trials (RCT) often fail to show benefit of vitamin D supplementation.

The package **SimVitD** provides simulation based tools to aid planning the comparison of supplementation arms in vitamin D trials. These tools 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.

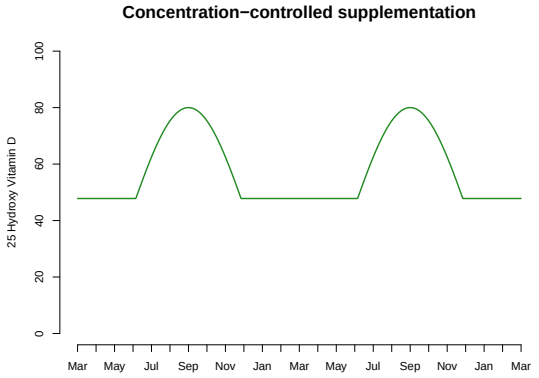
Calculating the study power of a randomised control trial or the sample size required for a given power is non-trivial under such circumstances. **SimVitD** uses simulation of exposures and infections at an individual level to investigate the effects of various vitamin D interventions on disease rates within a study group. Individual vitamin D status curves are simulated throughout the year for two or more groups, exposures and incidences of infections are simu-

lated 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 utilized in the planning of vitamin D studies.

The remainder of this vignette is organized 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 vitamin D. Section 5 contains an example on usage of the package.



(a) Placebo (no supplementation) and fixed dose equivalent supplementation shemes.



(b) Dynamic dose supplementation with threshold at 50 nmol/l equivalent.

2. Modelling individual vitamin D status

The tools in **SimVitD** simulate many instances of a study. Within each study, individuals' vitamin D status trajectories and potential exposures and protections from infections are simulated *separately*. The core steps of the simulation approach being proposed are:

(i) simulation of individual vitamin D status curves

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

(iii) determination of the probability of developing infection at each exposure time, conditional on vitamin D status at exposure

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

The package contains options to approximate power to compare supplementation arms in a randomized control trial. This can be used as an aid for planning the experimental approach and determining the required sample sizes for a desired power. Power approximations are simulation based, compiled by aggregating many independent replications of the chosen trial design.

2.1. Generative model for individual status

As most of an individual's vitamin D is derived from synthesis in skin following UVB exposure (Webb and Holick 1988; O'Sullivan *et al.* 2019), vitamin D status will naturally vary throughout the year. 25OHD concentration (marker of vitamin D status) tends to peak late in the summer, following the period with the strongest UVB radiation; the status trough will follow the period of lowest exposure (Kelly *et al.* 2016; O'Sullivan *et al.* 2017). It is useful to note that the peak and trough in an individual's status will depend on geo-location; there will be variability within, say, the northern hemisphere (O'Sullivan *et al.* 2017). This paper works on the assumption of a northern hemisphere seasonal schedule with summer months being June to August. Cyclic status profiles follow the assumed yearly periodic curve with a trough in February or March and peak in August or September (Kelly *et al.* 2016; O'Sullivan *et al.* 2017). The peak 25OHD occurs with 1-2 month time lag following the peak UVB radiation (here we assume 2 months); this reflects the period of pronounced vitamin D accumulation arising from abundant production of the nutrient in the skin.

Consider a group of trial participants, indexed by i. A phase shifted cosine curve with a lower threshold is used to model a participant's vitamin D status. This borrows from Degerud *et al.* (2016) who used cosinor models to describe longitudinal vitamin D readings in a cohort of individuals. This is a natural model to capture the variation affected by UVB, which is by far the most dominant source of vitamin D, as food sources are scarce. The curve used is

$$V_i^{\text{pl}}(t) = \max\left\{\mu_i + H_i + A_i \cos(2\pi t - \pi)\,,\, \tau\right\}, \quad t \geq 0. \tag{1}$$

Here, μ_i is a mean level that could be described by participant specific characteristics i.e. $\mu_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}$ with $\mathbf{x}_i$ a vector of covariates. The lower threshold of τ nmol/l of circulating 25OHD is a detectability threshold. In (1) t is time measured in years i.e. the interval $[0, 1]$ corresponds to one year. The parameter H_i gives a perturbation of the mean around μ_i similar to a random effect, and $A_i \geq 0$ controls the change in status between periods with and without significant UVB exposure. The phase adjustment ν accounts for a lag effect from UVB exposure to expressed circulating vitamin D level. This can be used to make adjustments for geo-location effects e.g. northern/southern hemisphere. Here, $\nu = \pi$ is assumed, meaning that the study starts at the beginning of March ($t = 0$), and at that point vitamin D is lowest.

Variation in individual 25OHD concentration is accounted for by generating the amplitude and a height perturbation in (1) via

$$A_i \sim \text{Gamma}\left(\alpha_A, \beta_A\right) \qquad H_i \sim \text{N}\left(0, \sigma_\mu^2\right),$$

independently drawn for each individual. The shape and rate parameters α_A, β_A are chosen to have a specified expected value and standard deviation μ_A, σ_A. The height perturbation standard deviation is σ_μ (note μ) simulates a random effect for overall height. These specifications should be made to sufficiently represent typical population variation in trial participants.

2.2. Intervention schemes

A number of possible approaches may be under consideration when planning a prospective trial. The curve in (1) corresponds to no supplementation and is referred to as *placebo* (option `type="placebo"` in function `vitd.curves()`). Two further supplementation/intervention schemes are available in **SimVitD**. The schemes are flexible and can be used to characterize many potential assumptions.

Fixed-dose scheme

A fixed-dose scheme (`type="fixed-dose"`) corresponds to an individual taking a daily supplement of a fixed amount, and this is the prevailing approach in RCTs. Supplementation will provide more of a boost when vitamin D levels are low. To mimic this, the no supplement curve is modified by adding a flexible function $F_i(t)$

$$V_i^{\text{fix}}(t) = V_i^{\text{pl}}(t) + F_i(t).$$

where

$$F_i(t) = \delta_i \left[\omega_i + \frac{1}{2}(1-\omega_i)\ (1+\sin(2\pi t - 3\pi/2))\right] \qquad \omega_i \in [0,1].$$

The parameter ω_i gives the proportion of the fixed-dose which is always utilized. As supplementation will have more impact in periods of deficiency, we allow the uptake from the remaining $1-\omega_i$ proportion of the fixed dose to vary according to a complementary cosine function. The package simulates ω_i from a Beta distribution where the mean and standard deviation $(\mu_\omega, \sigma_\omega)$ are specified. For example, passing $\mu_\omega = 1, \sigma_\omega = 0$ will result in $F_i(t) = \delta_i$.

The parameter δ_i represents an individual's overall derived benefit from that dosage, accounting for variation in overall assimilation. This can be simulated at an individual level using a truncated distribution capped at the administered equivalent dose level. If an equivalent of δ nmol/l is administered, δ_i is sampled from the density

$$f(t) = \frac{\gamma\, \mathrm{e}^{-\gamma(\delta - t)}}{1 - \mathrm{e}^{-\gamma\delta}}\, \mathbb{I}(0 < t < \delta).$$

Simulation can be carried out straightforwardly by inversion. As $\gamma \to \infty$, sampled values concentrate on an atom at δ, corresponding to the same overall equivalent dose for every participant.

Concentration-controlled scheme

A concentration-controlled scheme allows an individual to be monitored regularly and their status kept above a 25(OH)D threshold ρ_i, for example in a randomized concentration-controlled trial (RCCT) design Kraiczi *et al.* (2003). That is

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

A comparison of the placebo, fixed-dose equivalent supplementation and concentration-controlled schemes is shown in Figure 1b. The individual target level is simulated via $\rho_i \sim \text{Gamma}\,(\alpha_\rho, \beta_\rho)$.

with α_ρ, β_ρ determined to give a specified expectation μ_ρ and standard deviation σ_ρ. This is option `type="dynamic-dose"` in `vitd.curve()`.

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). This is the northern hemisphere option.

Glossary of parameters for `vitd.curve()`

Table 1 gives a summary of the parameters used to generate individual vitamin D curves and the corresponding argument names in `vitd.curve()`.

Parameter	Argument name
μ	`mu`
μ_A	`amplitude`
μ_ρ	`dyn.dose.thresh`
σ_μ	`sd.mu`
σ_A	`sd.amplitude`
σ_ρ	`sd.dyn.dose.thresh`
δ	`supp.dose`
γ	`supp.dose.rate`
μ_ω	`weight`
σ_ω	`sd.weight`
τ	`min.thresh`
North	`north.hemi`

Table 1: Summary of parameters and corresponding argument names in **SimVitD**.

3. Benefit of vitamin D supplementation

The **SimVitD** models benefits of vitamin D supplementation through modelling of random exposures to a specific event, and occurrences of that event upon exposure. We use infection, but this can be thought of in more abstract terms. One could, for example, be examining protection against allergic reaction following exposure to allergens, asthma attack, or relapse of autoimmune disease, or many others. Vitamin D protects against this unnamed infection such that higher levels of vitamin D imply a lower probability of contracting infection upon exposure.

Exposures to infection

An individual's exposures over the period of a trial are simulated from a non-homogeneous Poisson process (Lewis and Shedler 1979). In the case of seasonally concentrated infections (e.g. flu, see Figure 2 (a)), rate function $\lambda(t), t \geq 0$ reflects different rates of exposure at

different times. Simulations from an NHPP can be carried out conveniently in `R` using the `poisson` (Brock and Slade 2015) package. The rate function can also, of course, represent a constant rate of exposure. Instituting this function is an opportunity to incorporate domain expertise into the planning stage of the trial. The expected number of exposures over a time window $(0, \tau]$ is given by $\int_0^\tau \lambda(t)\,\mathrm{d}t$. The function $\lambda(t)$ is defined by rescaling an overall rate λ_0 which gives the rate of exposures when most intense

$$\lambda(t) = \lambda_0\, \upsilon(t), \qquad 0 \leq \upsilon(t) \leq 1, \qquad t \geq 0.$$

The function $\upsilon(t)$ may be passed by the user. A convenience function `intensity.function()` gives a step function for simple Summer/Winter rates.

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 completely sufficient vitamin D individual contracting an infection after exposure. The relative risk curve is a member of generalized logistic family

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

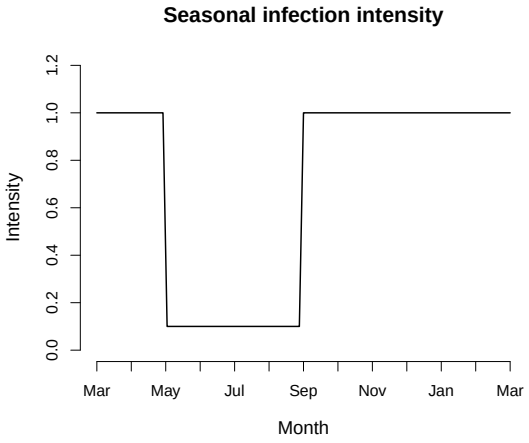
where x is the vitamin D status. The parameters ℓ, u give the lowest and highest relative risk values. The value of u states how much more likely one is to get infection when completely depleted in vitamin D compared with when one is fully replete. Figure 2 shows examples. The values of a and b are determined by providing points of inflection of the relative risk curve. These points default to 10 and 70 nmol/L in the package; there is much debate around the reference values chosen and this is another opportunity to explicitly represent domain experience in trial planning. See Figures 5a and 5b for examples.

Summary of simulation steps

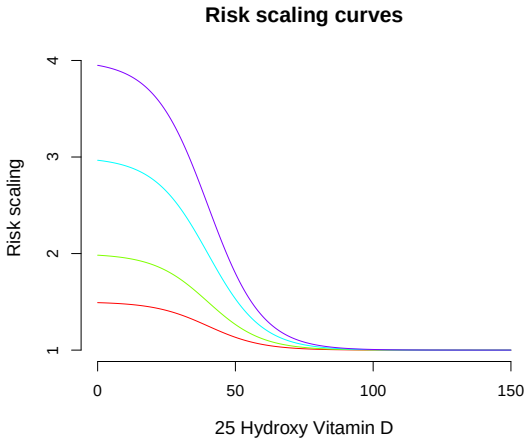
We give an overall summary of the generative model of exposures and infections. Consider individual i and let $T_1, \ldots, T_M$ denote the times at which they are exposed. Exposure times only within the time frame of the trial are used: $\tau_{\text{start}} < T_k \leq \tau_{\text{end}}$, $k = 1, \ldots, M$.

$T_1, \ldots, T_M \sim \text{NHPP}(\lambda(t))$	simulate individual's exposure times
$L_k = V_i(T_k)$	find their status $k = 1, \ldots, M$
$P_k = p_0\, g(L_k)$	get the probability of infection after exposure $k = 1, \ldots, M$
$I_k \sim \text{Bernoulli}(P_k)$	simulate whether infection is developed from exposure $k = 1, \ldots, M$.

In the case of infections ($I_k = 1$), one may also wish to impose a non-susceptible period or "holding time". For example, an exponentially distributed amount of time where the infected individual is not susceptible to a new infection. This is available through the `holding.time` argument in `infection.count()`.



(a) Intensity function for flu season with summer rate of 0.1 and winter rate of 1



(b) Logistic curves with $\ell = 1$ and $u = 1.5, 2, 3, 4$

Figure 2: Example of intensity function and risk scaling curves.

4. Study power

SimVitD approximates the power of study comparing two supplementation trial arms. Determining whether there is a benefit of supplementation will most ordinarily be carried out by investigation of the number of events (e.g. infections) that occurred in participants assigned to each arm over the trial duration, or some function thereof.

Types of comparisons

There are two tests based on counts of infections available in **SimVitD**. Define

$$\begin{aligned} \theta_s &= \Pr\{\text{individual gets} \geq 1 \text{ infection in arm } s\} \\ \mu_s &= \text{expected number of infections for individual in arm } s \end{aligned}$$

The power of the tests

$$H_0 : \theta_{\text{pl}} \leq \theta_{\text{supp}} \qquad H_A : \theta_{\text{pl}} > \theta_{\text{supp}} \tag{2}$$

$$H_0 : \mu_{\text{pl}} \leq \mu_{\text{supp}} \qquad H_A : \mu_{\text{pl}} > \mu_{\text{supp}} \tag{3}$$

will be of interest. It can be shown that for a trial where one arm does receive a supplement, then the generative model introduced above will generate data under H_A. Thus this generative model can serve as a proxy in a meaningful way.

Approximating the power

The power of a test is

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

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 a placebo arm and $rn, r \in \mathbb{N}$ participants in an intervention arm with $r = 1$ corresponding to the same number in each arm. The trial is simulated N times and a bootstrap (Efron and Tibshirani 1993) test of the hypothesis [i.e. (2) or (3)] is carried out in each of these. The proportion of the N trials which are rejected gives an empirical estimate of the power to detect differences between trial arms. The scheme is outlined in Figure 3. The tests of (2) and (3) are carried out using a non-parametric bootstrap two sample tests from the package **wBoot** (Weiss 2016).

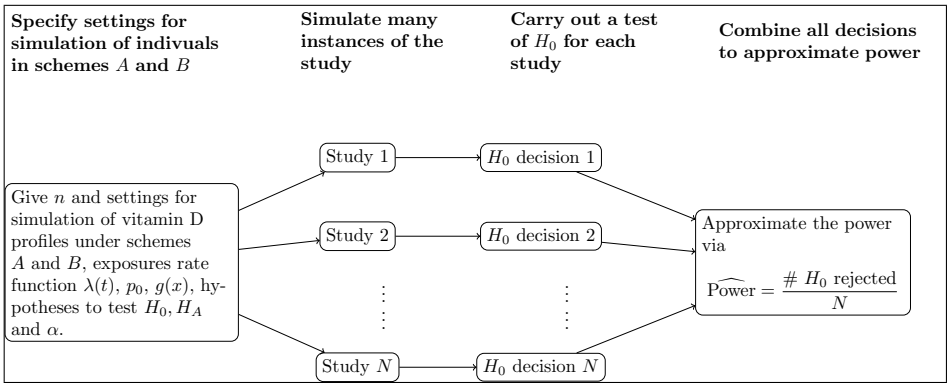


Figure 3: Flowchart showing the simulation process for estimating the power of a study.

The sample estimate of the power is

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

the proportion of times H_0 was rejected when H_A was the ground truth. The law of large numbers gives

$$\widehat{\text{Power}} \to \text{Power}$$

as $N \to \infty$ 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 Monte Carlo in approximation of the power is also possible through the `mc.error` argument. There is an option to parallelize these calculations through the arguments `parallel` and `num.cores`.

5. Using the package

This section walks thorough an example using the package. We examine a two-armed trial with a population having baseline 25OHD concentrations of 35 nmol/L with minimum detectable levels of 10 nmol/L. This translates to $\mu = 35$. The researchers aim expect scatter around the expected maximum and minimum levels to be reflected through $\mu_A = 15$ $\sigma_H = 5$, $\sigma_A = 5$. The primary endpoint is the number of infections contracted over the trial duration.

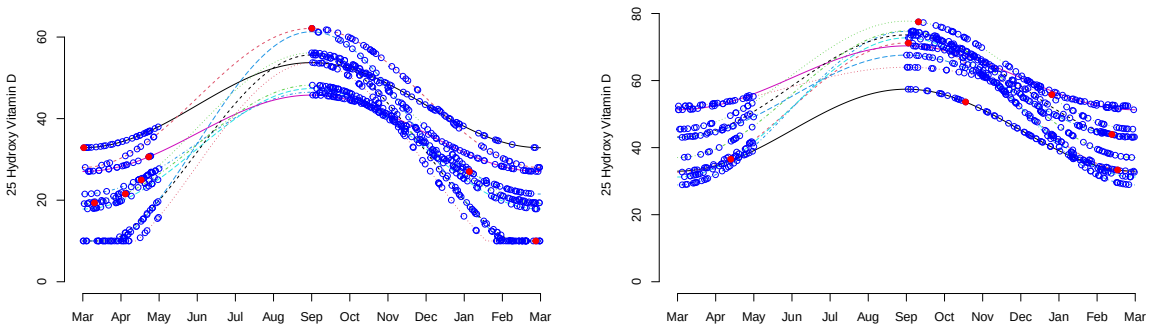
This is a one year study, indicated by the arguments `start` and `end`.

Participants will receive either a placebo or a fixed-dose vitamin D supplement equivalent to 20 nmol/L increase in 25OHD daily, with little variability in the derived uptake (i.e. $\gamma = \infty$ and $\mu_\omega = 1, \sigma_\omega = 0$).

Simulating and plotting vitamin D status profiles

Example curves for ten individuals from each arm are generated and plotted.

```
> # placebo example
> pl <- vitd.curve( n=10, type = "placebo", start = 0, end = 1, mu=35, sd.mu = 5,
...                   amplitude = 15, sd.amplitude = 5 )
> plot(pl, ylim=c(0,85))
> # supplement
> tr <- vitd.curve( n=10, type = "fixed-dose", start = 0, end = 1, mu=35, sd.mu = 5,
...                   amplitude = 15, sd.amplitude = 5, supp.dose=20,
...                   supp.dose.rate = Inf, weight = 1, sd.weight = 0 )
> plot(tr, ylim=c(0,85))
```



(a) Placebo arm

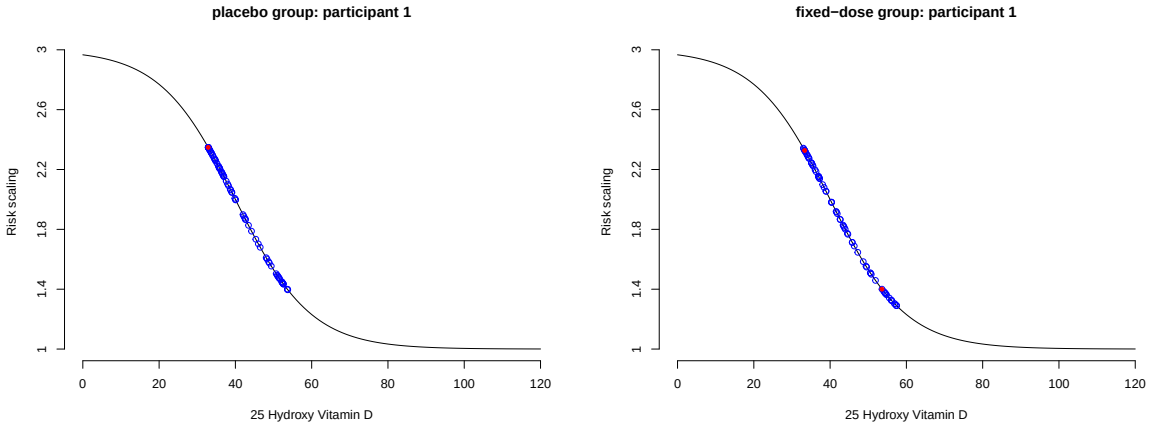
(b) Fixed-dose equivalent arm

Figure 4: Vitamin D individual curves along with exposures (blue) and infections (red) for the placebo and fixed-dose arms.

Exposures to infection and resulting cases

Next exposure times are simulated. The same intensity function is used for both groups, with a mean of 1.2 exposures per week and with no exposures occurring outside of flu season. It is assumed there is an exponentially distributed post-infection period with mean two weeks where infection cannot reoccur.

```
> # intensity of infections
> v <- intensity.function( summer.rate = 0, winter.rate = 1, flu = TRUE )
```



(a) Relative risk at exposures with infections for placebo arm individual.

(b) Relative risk at exposures with infections for fixed-dose individual.

Figure 5: Output from `rr.curve.plot()`.

```
> # simulate exposures from NHPP and infections conditional on exposures
> expo.pl <- exposure.levels( pl, rate=1.2, v, end = 1)
> inf.pl <- infection.count( expo.pl, baseline = 0.03, RR = 3, holding.time = 2 )
> expo.tr <- exposure.levels( tr, rate=1.2, v, end = 1)
> inf.tr <- infection.count( expo.tr, baseline = 0.03, RR = 3, holding.time = 2 )
```

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

```
> plot(pl)
> plot(expo.pl)
> infection.count.plot(expo.pl, inf.pl)
> plot(tr)
> plot(expo.tr)
> infection.count.plot(expo.tr, inf.tr)
```

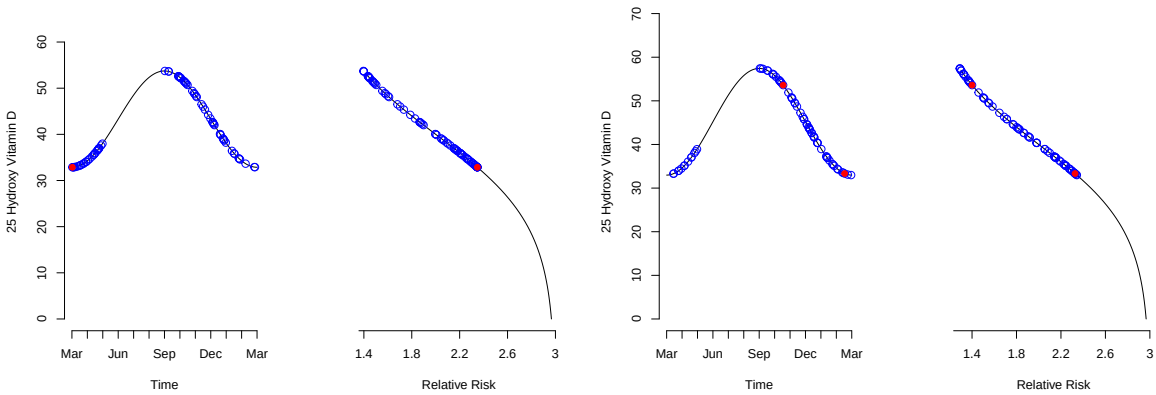
Exploring seasonal variation in risk

`rr.curve.plot()` visualizes where exposures occur on the relative risk curve. Figures 5a and 5b show the output of

```
> rr.curve.plot(expo.pl, inf.pl)
> rr.curve.plot(expo.tr, inf.tr)
```

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 visualization tool to explore the seasonal variation in risk for an individual. It indicates where exposures occurred in the vitamin D status profile



(a) Single participant from placebo arm

(b) Single participant from fixed-dose arm

Figure 6: Profiles and relative risk for single participants with exposures in red and infections in blue. Output from `rr.profile.plot()`.

and the corresponding relative risk side-by-side. The exposures resulting in infection are also indicated. An example is shown in Figures 6a and 6b which are obtained from

```
> rr.profile.plot( pl, expo.pl, inf.pl )
> rr.profile.plot( tr, expo.tr, inf.tr )
```

Approximating power of detecting difference between trial arms

To function `power.calc` returns the Monte Carlo based power estimate with Monte Carlo error if requested. If Monte Carlo error is requested, it is possible to parallelize runs of the trial. The function takes the size of each group along with the type of test `"proportion"` or `"count"` [i.e. (2) or (3)]. A range of $u - \ell$ values can be passed. In the code below, three $(2, 3, 4)$ are passed through the `RR` (relative risk) argument. The function is passed the example curves to define the populations in each arm of the trial.

```
> # get the Monte Carlo power estimates
> pow <- power.calc( n = c(20,40,60), ratio=1, N = 500,
...         test.type = "count", sig.level = 0.05,
...         vitdcurves.placebo = pl, vitdcurves.treatment  = tr,
...         baseline = 0.03, RR = c(2,3,4), rate = 1, intensity.func = v, holding.time = 2,
...         verbose = TRUE )
> # plot the power curves
> plot( pow, x.legend = 20, y.legend = 1,
...         main.legend = "Relative Risk", legend.size = 0.8 )
> # plot of effect sizes
> plot( pow, x.legend = 20, y.legend = 1,
...         main.legend = "Relative Risk", legend.size = 0.8, which=2 )
```

A large value of N should be used. A value of at least 500 is recommended. Figure 7 shows the output of plotting the object returned by `pow.calc()`.

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()`.

```
> # quantify the Monte Carlo error in the power- split computations over 2 cores
> pow.mcerr <- power.calc( n = c(20,40,60), ratio=1, N = 500,
...         test.type = "count", sig.level = 0.05,
...         vitdcurves.placebo = pl, vitdcurves.treatment  = tr,
...         baseline = 0.03, RR = c(2,3,4), rate = 1, intensity.func = v, holding.time = 2,
...         mc.error = 5, parallel = TRUE, num.cores = 2 )
> plot(pow.mcerr)
```

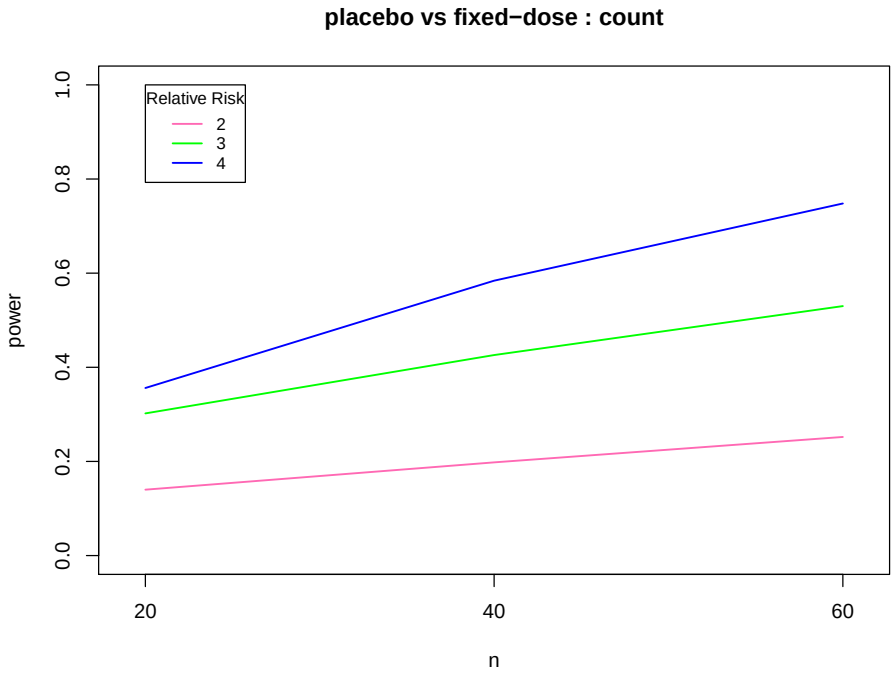


Figure 7: Output from `plot(pow)` showing estimate of the power at each relative risk level and specified value of n.

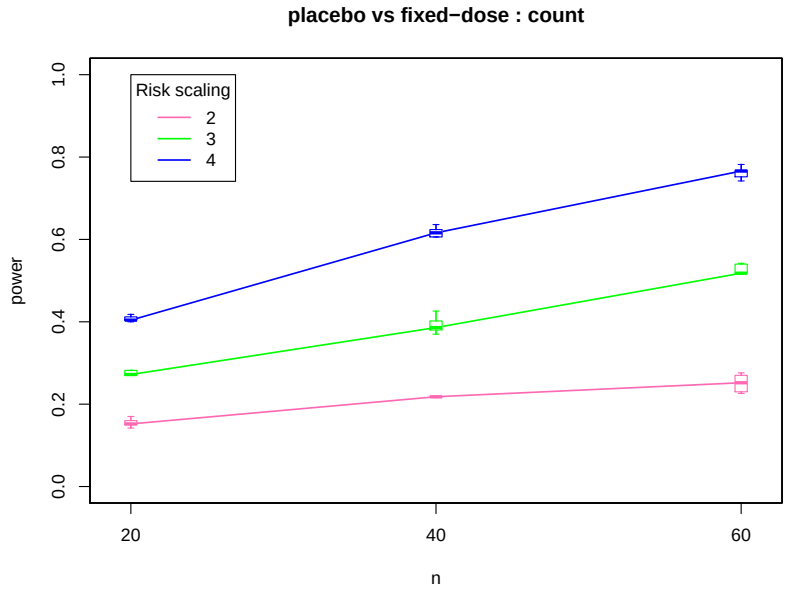


Figure 8: Output from `plot(pow_mc)` showing the result of the estimation of power over ten runs at each relative risk level and value of n.

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.

Degerud E, Hoff R, Nygard O, *et al.* (2016). "Cosinor modelling of seasonal variation in 25-hydroxyvitamin D concentrations in cardiovascular patients in Norway." *European Journal of Clinical Nutrition*, **70**, 517–522.

Efron B, Tibshirani RJ (1993). *An Introduction to the Bootstrap*. Number 57 in Monographs on Statistics and Applied Probability. Chapman & Hall/CRC, Boca Raton, Florida, USA.

Entrenas Castillo M, Entrenas Costa LM, Vaquero Barrios JM, Alcalá Díaz JF, López Miranda J, Bouillon R, Quesada Gomez JM (2020). "Effect of calcifediol treatment and best available therapy versus best available therapy on intensive care unit admission and mortality among patients hospitalized for COVID-19: A pilot randomized clinical study." *The Journal of Steroid Biochemistry and Molecular Biology*, **203**, 105751.

Grant WB, Lahore H, McDonnell SL, Baggerly CA, c B French, Aliano JL, Bhattoa HP (2020). "Evidence that Vitamin D Supplementation Could Reduce Risk of Influenza and COVID-19 Infections and Deaths." *Nutrients*, **12**, 34.

Holick MF (2007). "Vitamin D Deficiency." *New England Journal of Medicine*, **357**(3), 266–281.

Kelly D, Theodoratou E, Farrington S, Fraser R, Campbell H, Dunlop M, Zgaga L (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.

Kraiczi H, Jang T, Ludden T, Peck CC (2003). "Randomized concentration-controlled trials: Motivations, use, and limitations." *Clinical Pharmacology and Therapeutics*, **74**(3), 203–214.

Lewis PAW, Shedler GS (1979). "Simulation of nonhomogeneous Poisson processes by thinning." *Naval Research Logistics Quarterly*, **26**(3), 403–413.

O'Sullivan F, Laird E, Kelly D, van Geffen J, van Weele M, McNulty H, Hoey L, Healy M, McCarroll K, Cunningham C, Casey M, Ward M, Strain J, Molloy AM, Zgaga L (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, Raftery T, van Weele M, van Geffen J, McNamara D, O'Morain C, Mahmud N, Kelly D, Healy M, O'Sullivan M, Zgaga L (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, Tzoulaki I, Zgaga L, Ioannidis JPA (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.

Weiss NA (2016). *wBoot: Bootstrap Methods*. R package version 1.0.3, URL https://CRAN.R-project.org/package=wBoot.

Affiliation:

Jason Wyse & Rebecca Mangan
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