Supplementary Files

The separate sections in this supplementary material are referred to in the paper by SF. Thus, SF1 refers to section 1 (on multilevel data models).

Contents

1 Multilevel data models

The two-level model (Hussey and Hughes, Controlled clinical Trials, 2009) describes the outcome Y_{itm} of subject $m = 1, \dots, n_1$ in cluster $i = 1, \dots, I$ in measurement period $t = 1, \dots, T$ as

$$\left.\begin{array}{c} Y_{itm} = \mu + u_{0i} + \beta_t + \delta \cdot X_{it} + e_{itm}\,, \\ u_{0i} \sim \mathrm{N}(0,\sigma_2^2), \quad e_{itm} \sim \mathrm{N}(0,\sigma_1^2)\,; \quad \{\, u_{0i}\,, e_{itm}\} \text{ mutually independent,} \end{array}\right\}$$

where '{....} mutually independent' means that all effects between {...} are independent, so u_{0i} are independent, e_{itm} are independent and each pair $u_{0i},\ e_{i\prime tmm}$ is independent. Also, $\sim N(0,\sigma^2)$ means that the corresponding random effect is normally distributed with mean 0 and variance σ^2. The highest level (level 2) is measured as cohort and the lowest level (level 1) cross-sectionally.

For 3 levels, the outcome Y_{itjm} of measurement $m = 1, \dots, n_1$ of subject $j = 1, \dots, n_2$ in cluster $i = 1, \dots, I$ in measurement period $t = 1, \dots, T$ is

$$\left.\begin{array}{c} Y_{itjm} = \mu + u_{00i} + u_{0i(t)j} + +\beta_t + \delta \cdot X_{it} + e_{itjm}\,, \\ u_{00i} \sim N(0,\sigma_3^2), \qquad u_{0i(t)j} \sim N(0,\sigma_2^2), \qquad e_{itjm} \sim N(0,\sigma_1^2)\,, \\ \{u_{00i},\ u_{0i(t)j},\ e_{itjm}\} \text{ independent.} \\ u_{0i(t)j}\,, u_{0i(t')j} \text{ equal/unequal for t}\neq\text{t' if level 2 measured as cohort/crossectional} \end{array}\right\}$$

Note that there is one intermediate level (level 2) that can either be measured as cohort or cross-sectionally

Similarly, for 4 levels, the outcome Y_{itjlm} of measurement $m = 1, \dots, n_1$ of subject $k = 1, \dots, n_2$ in sub-cluster $j = 1, \dots, n_3$ in cluster $i = 1, \dots, I$ in measurement period $t = 1, \dots, T$ is

$$\left.\begin{array}{c} Y_{itjkm} = \mu + u_{000i} + u_{00i(t)j} + u_{0i(t)jk} + \beta_t + \delta \cdot X_{it} + e_{itjkm}\,, \\ u_{000i} \sim N(0,\sigma_4^2), u_{00i(t)j} \sim N(0,\sigma_3^2), u_{0i(t)jk} \sim N(0,\sigma_2^2),\ e_{itjkm} \sim N(0,\sigma_1^2)\,, \\ \{u_{000i},\ u_{00i(t)j},\ u_{0i(t)jk},\ e_{itjkm}\} \text{ mutually independent;} \\ u_{00i(t)j}, u_{00i(t')j} \text{ equal/unequal for t}\neq\text{t' if level 3 measured as cohort/crossectional;} \\ u_{0i(t)jk}, u_{00i(t')jk} \text{ equal/unequal for t}\neq\text{t' if level 2 measured as cohort/crossectional.} \end{array}\right\}$$

And for p levels we have independent random effects u for level $p, p-1, \dots, 3, 2$ with variances $\sigma_p^2, \dots, \sigma_2^2$.

In terms of the cluster averages $Y_{it\cdot}$ (at each time point/period), we have a repeated measurement design and the above models lead to equal correlations between averages at different time points in the same cluster and to the same variance at each time point.

For example, for $p = 3$, and time/period t and s, and dropping the index (t) in $u_{0i(t)j}$ for readibility

$$covar(Y_{it\cdot}, Y_{is\cdot}) =$$

$$cov\left(\sum_{j=1}^{n_2}\sum_{m=1}^{n_1}\frac{\left[\mu + u_{00i} + u_{0ij} + \beta_t + X_{it} + e_{itjm}\right]}{n_2 n_1}\,, \sum_{j\prime=1}^{n_2}\sum_{m\prime=1}^{n_1}\frac{\left[\mu + u_{00i} + u_{0ij'} + \beta_s + X_{is} + e_{isj'm'}\right]}{n_2 n_1}\right.,$$

$$= \frac{1}{n_2^2 n_1^2} \sum_{j,j'=1}^{n_2} \sum_{m,m'=1}^{n_1} cov\left[\mu + u_{00i} + u_{0ij} + \beta_t + X_{it} + e_{itjm}\right], \left[\mu + u_{00i} + u_{0ij'} + \beta_s + X_{is} + e_{isj'm'}\right]$$

and because constants have covariance 0, this reduces to

$$= \frac{1}{n_2^2 n_1^2} \sum_{j,j'=1}^{n_2} \sum_{m,m'=1}^{n_1} cov\left[u_{00i} + u_{0ij} + e_{itjm}\right], \left[u_{00i} + u_{0ij'} + e_{isj'm'}\right]$$

And because $\{u_{00i}, u_{0ij}, e_{itjm}\}$ are independent

$$= \frac{1}{n_2^2 n_1^2} \left[\sum_{j,j'=1}^{n_2} \sum_{m,m'=1}^{n_1} \underbrace{cov(u_{00i}, u_{00i})}_{=\sigma_3^2} + \sum_{j,j'=1}^{n_2} \sum_{m,m'=1}^{n_1} \underbrace{cov(u_{0ij}, u_{0ij'})}_{=\begin{cases} level\ 2\ cohort: \sigma_2^2\ if\ j=j' \\ level\ 2\ crosssection: \sigma_2^2\ if\ j=j' and\ t=s \end{cases}} \right.$$
$$\left. + \sum_{j,j'=1}^{n_2} \sum_{m,m'=1}^{n_1} \underbrace{cov(e_{itjm}, e_{isj'm'})}_{=\sigma_1^2\ if\ t=s\ \&j=j'\&m=m'} \right]$$

$$= \begin{cases} \text{level 2 as cohort:} & \begin{cases} if\ t = s: \frac{1}{n_2^2 n_1^2}[n_2^2 n_1^2 \sigma_3^2 + n_2 n_1^2 \sigma_2^2 + n_2 n_1 \sigma_1^2] = \sigma_3^2 + \frac{\sigma_2^2}{n_2} + \frac{\sigma_1^2}{n_2 n_1} \\ if\ t \neq s: \frac{1}{n_2^2 n_1^2}[n_2^2 n_1^2 \sigma_3^2 + n_2 n_1^2 \sigma_2^2] = \sigma_3^2 + \frac{\sigma_2^2}{n_2} \end{cases} \\ \text{level 2 as crosssection:} & \begin{cases} if\ t = s: \frac{1}{n_2^2 n_1^2}[n_2^2 n_1^2 \sigma_3^2 + n_1^2 \sigma_2^2 + n_2 n_1 \sigma_1^2] = \sigma_3^2 + \frac{\sigma_2^2}{n_2} + \frac{\sigma_1^2}{n_2 n_1} \\ if\ t \neq s: \frac{1}{n_2^2 n_1^2}[n_2^2 n_1^2 \sigma_3^2] = \sigma_3^2 \end{cases} \end{cases}$$

In general, $var(Y_{it\cdot}) = \sigma_p^2 + \sigma_{p-1}^2/n_{p-1} + \sigma_{p-2}^2/(n_{p-1}n_{p-2}) + \cdots + \sigma_1^2/(n_{p-1}n_{p-2}\cdots n_2 n_1)$. Also, if the levels $p, p-1, \ldots, k$ are measured as cohort and the levels $k-1, \ldots, 2, 1$ as cross-section, then the variances that are in common in cluster i at time t and s are those of level $p, p-1, \ldots, k$, so

$$covar(Y_{it\cdot}, Y_{is\cdot}) = \sigma_p^2 + \frac{\sigma_{p-1}^2}{n_{p-1}} + \frac{\sigma_{p-2}^2}{n_{p-1}n_{p-2}} + \cdots + \frac{\sigma_k^2}{n_{p-1}\, n_{p-2} \cdots n_2 n_1}.$$

1.1 Two types of intracluster correlations

For $p \geq 3$, we could define at least two types of intracluster correlations in the above multilevel model. One approach is to consider the true correlation of level 1 unit within their level 2 unit (i.e, $(\sigma_p^2 + \sigma_{p-1}^2 + \cdots + \sigma_2^2)/(\sigma_p^2 + \sigma_{p-1}^2 + \cdots + \sigma_2^2 + \sigma_1^2) = \rho_{12}$), of the level 2 units within their level 3 unit (i.e., $(\sigma_p^2 + \sigma_{p-1}^2 + \cdots + \sigma_3^2)/(\sigma_p^2 + \sigma_{p-1}^2 + \cdots + \sigma_2^2) = \rho_{23}$), of the level 3 units within their level 4 unit (i.e., $(\sigma_p^2 + \sigma_{p-1}^2 + \cdots + \sigma_4^2)/(\sigma_p^2 + \sigma_{p-1}^2 + \cdots + \sigma_3^2) = \rho_{12}$), and so on. The other approach is to look at the correlation of level 1 units within their level 2 unit (i.e, $(\sigma_p^2 + \sigma_{p-1}^2 + \cdots + \sigma_2^2)/(\sigma_p^2 + \sigma_{p-1}^2 + \cdots + \sigma_2^2 + \sigma_1^2) = \rho_{12} = \rho^{12}$), the correlation of level 1 units within their level 3 unit ((i.e,

$(\sigma_p^2 + \sigma_{p-1}^2 + \cdots + \sigma_3^2)/(\sigma_p^2 + \sigma_{p-1}^2 + \cdots + \sigma_2^2 + \sigma_1^2) = \rho^{13}$), the correlation of level 1 units within their level 4 unit (i.e, $(\sigma_p^2 + \sigma_{p-1}^2 + \cdots + \sigma_4^2)/(\sigma_p^2 + \sigma_{p-1}^2 + \cdots + \sigma_2^2 + \sigma_1^2) = \rho^{14}$), and so on.

These types of correlations can be calculated from each other. For example, $\rho_{23} = \rho^{13}/\rho^{12}$ and $\rho^{14} = \rho_{12} \cdot \rho_{23} \cdot \rho_{34}$. In general: $\rho_{k,k+1} = \rho^{1,k+1}/\rho^{1,k}$ and $\rho^{1,k} = \rho_{12} \cdot \rho_{23} \cdots \rho_{k-1,k}$.

1.2 Variance inflation factors (VIFs) for p levels of clustering.

In the above multilevel data models,
the variance of a cluster mean $var(Y_{it\bullet})$ and the variances at each level σ_i^2 can be expressed in terms of the total variance and the intraclass correlations via variance inflation factors, and vice versa.

2 level:

$$var(Y_{it\bullet}) = \sigma_2^2 + \frac{\sigma_1^2}{n_1} = \frac{\sigma_{tot}^2}{n_1} \cdot [1 + (n_1 - 1)\rho_{12}]$$

With $\sigma_{tot}^2 = \sigma_1^2 + \sigma_2^2$, and

$$\rho_{12} = \frac{\sigma_2^2}{\sigma_2^2 + \sigma_1^2}$$

The equality holds because

$$\sigma_2^2 = \rho_{12} \cdot \sigma_{tot}^2, \qquad \sigma_1^2 = (1 - \rho_{12}) \cdot \sigma_{tot}^2$$

3 level:
We have

$$\sigma_3^2 + \frac{\sigma_2^2}{n_2} + \frac{\sigma_1^2}{n_1 n_2} = \frac{\sigma_{tot}^2}{n_1 n_2} \cdot [1 + (n_1 - 1)\rho_{12}] \cdot [1 + (n_2 - 1)\tilde{\rho}_{23}] \quad (*)$$

With

$$\rho_{12} = \frac{\sigma_3^2 + \sigma_2^2}{\sigma_3^2 + \sigma_2^2 + \sigma_1^2}$$

(the correlation of level 1 units in level 2 units)
and

$$\tilde{\rho}_{23} = \tilde{\rho}_{23}(n_1) = \rho_{23} \cdot \frac{n_1 \rho_{12}}{[1 + (n_1 - 1)\rho_{12}]}$$

where

$$\rho_{23} = \frac{\sigma_3^2}{\sigma_3^2 + \sigma_2^2},$$

(the correlation of level 2 units within level 3 units).

Proof. We use repeatedly that $A + \frac{B}{n} = (A + B) \cdot \left[\frac{1 + (n-1)\rho}{n}\right]$ with $\rho = \frac{A}{A+B}$.
First to see that, with $\rho_2 = \sigma_3^2 / \left(\sigma_3^2 + \sigma_2^2 + \frac{\sigma_1^2}{n_1}\right)$,

$$\sigma_3^2 + \frac{\sigma_2^2}{n_2} + \frac{\sigma_1^2}{n_1 n_2} = \sigma_3^2 + \frac{\sigma_2^2 + \frac{\sigma_1^2}{n_1}}{n_2} = \left[\frac{1 + (n_2 - 1)\rho_2}{n_2}\right] \cdot \left(\sigma_3^2 + \sigma_2^2 + \frac{\sigma_1^2}{n_1}\right)$$

And then to see that the last factor is

$\left(\sigma_3^2+\sigma_2^2+\frac{\sigma_1^2}{n_1}\right)=\left[\frac{1+(n_1-1)\rho_1}{n_1}\right]\cdot(\sigma_3^2+\sigma_2^2+\sigma_1^2)$ (#)

with $\rho_1=(\sigma_3^2+\sigma_2^2)/(\sigma_3^2+\sigma_2^2+\sigma_1^2)$.

Now we backward look at the ρ_i's.
First $\rho_1=\rho_{12}$ as defined above. Using that we can express
$\rho_2=\frac{\sigma_3^2}{\sigma_3^2+\sigma_2^2+\frac{\sigma_1^2}{n_1}}=\underbrace{\left[\frac{\sigma_3^2}{\sigma_3^2+\sigma_2^2}\right]}_{\rho_{23}}\cdot\underbrace{\frac{\sigma_3^2+\sigma_2^2}{\left[\sigma_3^2+\sigma_2^2+\frac{\sigma_1^2}{n_1}\right]}}_{(\#)}=\rho_{23}\cdot\frac{\sigma_3^2+\sigma_2^2}{\left[\frac{1+(n_1-1)\rho_1}{n_1}\right]\cdot(\sigma_3^2+\sigma_2^2+\sigma_1^2)}=\rho_{23}\cdot\rho_{12}\cdot\left[\frac{n_1}{1+(n_1-1)\rho_1}\right]=\tilde{\rho}_{23}$.

The variances at each level can be expressed in the total variance and the intracluster correlations:

$$\sigma_3^2=\rho_{23}\cdot\rho_{12}\cdot\sigma_{tot}^2$$
$$\sigma_2^2=(1-\rho_{23})\cdot\rho_{12}\cdot\sigma_{tot}^2$$
$$\sigma_1^2=(1-\rho_{12})\cdot\sigma_{tot}^2$$

4 level design.

$\sigma_4^2+\frac{\sigma_3^2}{n_3}+\frac{\sigma_2^2}{n_2n_3}+\frac{\sigma_1^2}{n_2n_3n_1}=\frac{\sigma_{tot}^2}{n_1n_2n_3}\cdot\left[1+(n_1-1)\rho_{12}\right]\cdot\left[1+(n_2-1)\tilde{\rho}_{23}\right]\cdot\left[1+(n_3-1)\tilde{\rho}_{34}\right]$ (**)

where
$\sigma_{tot}^2=\sigma_1^2+\sigma_2^2+\sigma_3^2+\sigma_4^2,$

$$\rho_{12}=\frac{\sigma_4^2+\sigma_3^2+\sigma_2^2}{\sigma_4^2+\sigma_3^2+\sigma_2^2+\sigma_1^2}$$

(the correlation of level 1 units within their level 2 unit),

$$\tilde{\rho}_{23}=\tilde{\rho}_{23}(n_1)=\frac{\sigma_4^2+\sigma_3^2}{\sigma_4^2+\sigma_3^2+\sigma_2^2+\sigma_1^2/n_1}=\rho_{23}\cdot\frac{n_1\rho_{12}}{[1+(n_1-1)\rho_{12}]}$$

with
$\rho_{23}=\frac{\sigma_4^2+\sigma_3^2}{\sigma_4^2+\sigma_3^2+\sigma_2^2}$

(the correlation of level 2 units within their level 3 unit)
and

$$\tilde{\rho}_{34}=\tilde{\rho}_{34}(n_2,n_1)=\frac{\sigma_4^2}{\left[\sigma_4^2+\sigma_3^2+\frac{\sigma_2^2}{n_2}+\frac{\sigma_1^2}{n_1n_2}\right]}=\rho_{34}\cdot\frac{n_2\tilde{\rho}_{23}}{[1+(n_2-1)\tilde{\rho}_{23}]}$$

with

$$\rho_{34}=\frac{\sigma_4^2}{\sigma_4^2+\sigma_3^2}$$

(the correlation of level 3 units within their level 4 unit).

Proof. Again we repeatedly apply $A+\frac{B}{n}=(A+B)\cdot\left[\frac{1+(n-1)\rho}{n}\right]$ with $\rho=\frac{A}{A+B}$.
First, with $\rho_3=\sigma_4^2/\left(\sigma_4^2+\sigma_3^2+\frac{\sigma_2^2}{n_2}+\frac{\sigma_1^2}{n_2n_1}\right)$,

$$\sigma_4^2+\frac{\sigma_3^2}{n_3}+\frac{\sigma_2^2}{n_2n_3}+\frac{\sigma_1^2}{n_2n_3n_1}=\sigma_4^2+\frac{\sigma_3^2+\frac{\sigma_2^2}{n_2}+\frac{\sigma_1^2}{n_2n_1}}{n_3}=\left[\frac{1+(n_3-1)\rho_3}{n_3}\right]\cdot\left(\sigma_4^2+\sigma_3^2\ +\frac{\sigma_2^2}{n_2}+\frac{\sigma_1^2}{n_2n_1}\right)$$

Then the last factor is, with $\rho_2=(\sigma_4^2+\sigma_3^2)/\left(\sigma_4^2+\sigma_3^2+\sigma_2^2+\frac{\sigma_1^2}{n_1}\right)$

$$\sigma_4^2+\sigma_3^2\ +\frac{\sigma_2^2}{n_2}+\frac{\sigma_1^2}{n_2n_1}=\sigma_4^2+\sigma_3^2+\frac{\sigma_2^2+\frac{\sigma_1^2}{n_1}}{n_2}=\left[\frac{1+(n_2-1)\rho_2}{n_2}\right]\cdot\left(\sigma_4^2+\sigma_3^2\ +\sigma_2^2+\frac{\sigma_1^2}{n_1}\right)\quad (\#\#)$$

And this last factor is

$$\sigma_4^2+\sigma_3^2\ +\sigma_2^2+\frac{\sigma_1^2}{n_1}=\left[\frac{1+(n_1-1)\rho_1}{n_1}\right]\cdot(\sigma_4^2+\sigma_3^2\ +\sigma_2^2+\sigma_1^2)\quad (\#)$$

With $\rho_1=\frac{\sigma_4^2+\sigma_3^2\ +\sigma_2^2}{\sigma_4^2+\sigma_3^2\ +\sigma_2^2+\sigma_1^2}$

Now we work backwards to define the equalities for the ρ_{ij}'s . First $\rho_1=\rho_{12}$ by definition. Further by using (#) of two lines above:

$$\rho_2=\frac{(\sigma_4^2+\sigma_3^2)}{\left(\sigma_4^2+\sigma_3^2+\sigma_2^2+\frac{\sigma_1^2}{n_1}\right)}=\underbrace{\left[\frac{\sigma_4^2+\sigma_3^2}{\sigma_4^2+\sigma_3^2+\sigma_2^2}\right]}_{\rho_{23}}\cdot\underbrace{\frac{\sigma_4^2+\sigma_3^2+\sigma_2^2}{\left(\sigma_4^2+\sigma_3^2+\sigma_2^2+\frac{\sigma_1^2}{n_1}\right)}}_{=(\#)}$$

$$=\rho_{23}\cdot\rho_{12}\cdot\left[\frac{1+(n_1-1)\rho_1}{n_1}\right]^{-1}=\widetilde{\rho_{23}}\qquad (\#\#\#)$$

Actually this result we can use (in the last equality) to reformulate

$$\rho_3=\frac{\sigma_4^2}{\sigma_4^2+\sigma_3^2+\frac{\sigma_2^2}{n_2}+\frac{\sigma_1^2}{n_2n_1}}=\underbrace{\left[\frac{\sigma_4^2}{\sigma_4^2+\sigma_3^2}\right]}_{\rho_{34}}\cdot\underbrace{\frac{\sigma_4^2+\sigma_3^2}{\left[\sigma_4^2+\sigma_3^2+\frac{\sigma_2^2}{n_2}+\frac{\sigma_1^2}{n_2n_1}\right]}}_{=(\#\#)}$$

$$=\rho_{34}\cdot\frac{\sigma_4^2+\sigma_3^2}{\left[\frac{1+(n_2-1)\rho_2}{n_2}\right]\cdot\left(\sigma_4^2+\sigma_3^2\ +\sigma_2^2+\frac{\sigma_1^2}{n_1}\right)}=\rho_{34}\cdot\widetilde{\rho_{23}}\cdot\left[\frac{1+(n_2-1)\rho_2}{n_2}\right]^{-1}$$

$$=\rho_{34}\cdot\frac{n_2\widetilde{\rho_{23}}}{1+(n_2-1)\widetilde{\rho_{23}}}=\widetilde{\rho_{34}}$$

.

The variances at each level can be expressed in the total variance and the intracluster correlations:

$$\sigma_4^2=\rho_{34}\cdot\rho_{23}\cdot\rho_{12}\cdot\sigma_{tot}^2$$
$$\sigma_3^2=(1-\rho_{34})\cdot\rho_{23}\cdot\rho_{12}\cdot\sigma_{tot}^2$$
$$\sigma_2^2=(1-\rho_{23)}\cdot\rho_{12}\cdot\sigma_{tot}^2$$
$$\sigma_1^2=(1-\rho_{12})\cdot\sigma_{tot}^2$$

p level design

We prove with induction that for all $p=3,4,\ldots$ the following holds:

$$var(Y_{it\bullet})=\sigma_p^2+\frac{\sigma_{p-1}^2}{n_{p-1}}+\frac{\sigma_{p-2}^2}{n_{p-2}n_{p-1}}+\cdots+\frac{\sigma_k^2}{n_kn_{k+1}\cdot\ldots\cdot n_{p-1}}+\cdots+\frac{\sigma_1^2}{n_1n_2\cdot\ldots\cdot n_{p-1}}$$

$$= \frac{\sigma_{tot}^2}{n_1 n_2 \cdot \ldots \cdot n_{p-1}} \left[1 + (n_1 - 1)\rho_{12}\right] \cdot \left[1 + (n_2 - 1)\tilde{\rho}_{23}\right] \cdot \ldots$$
$$\ldots \cdot \left[1 + \left(n_{p-1} - 1\right)\tilde{\rho}_{p-1,p}\right] \;,$$

where
$$\sigma_{tot}^2 = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_p^2$$

intracluster correlations are defined by

$$\rho_{k,k+1} = \frac{\sigma_p^2 + \sigma_{p-1}^2 + \cdots + \sigma_{k+1}^2}{\sigma_p^2 + \sigma_{p-1}^2 + \cdots + \sigma_2^2 + \sigma_k^2}$$

and their sample estimates defined for $k = 1,2,\cdots,p-1$ recursively by
$$\tilde{\rho}_{k,k+1} = \rho_{k,k+1} \cdot \frac{n_{k-1}\tilde{\rho}_{\,k-1,k}}{\left[1 + (n_{k-1} - 1)\tilde{\rho}_{\,k-1,k}\right]}, \qquad \tilde{\rho}_{12} = \rho_{12}$$
together with the equality
$$\frac{\sigma_p^2}{\sigma_p^2 + \sigma_{p-1}^2 + + \frac{\sigma_{p-2}^2}{n_{p-2}} + \frac{\sigma_{p-3}^2}{n_{p-2} \cdot n_{p-3}} + \cdots + \frac{\sigma_1^2}{n_{p-2} \cdot n_{p-3} \cdots n_1}} = \tilde{\rho}_{p-1,p}$$

Proof. The above holds for $p = 3,4$ (noting that the last equality is the proof of $\rho_2 = \tilde{\rho}_{23}$ and $\rho_3 = \tilde{\rho}_{34}$ in the proof for the 3 level and 4 level case respectively).

Now assume the above holds for p. We prove it holds for $p + 1$.

First: using again $A + \frac{B}{n} = (A + B) \cdot \left[\frac{1+(n-1)\rho}{n}\right]$ with $\rho = \frac{A}{A+B}$, we see
$$var(Y_{it\bullet}) = \sigma_{p+1}^2 + \frac{\sigma_p^2}{n_p} + \frac{\sigma_{p-1}^2}{n_{p-1}n_p} + \frac{\sigma_{p-2}^2}{n_{p-2}n_{p-1}n_p} + \cdots + \frac{\sigma_k^2}{n_k n_{k+1} \cdot \ldots \cdot n_{p-1} \cdot n_p} + \cdots$$
$$+ \frac{\sigma_1^2}{n_1 n_2 \cdot \ldots \cdot n_{p-1} \cdot n_p}$$

$$= \sigma_{p+1}^2 + \frac{\sigma_p^2 + \frac{\sigma_{p-1}^2}{n_{p-1}} + \frac{\sigma_{p-2}^2}{n_{p-2}n_{p-1}} + \cdots + \frac{\sigma_k^2}{n_k n_{k+1} \cdot \ldots \cdot n_{p-1}} + \cdots + \frac{\sigma_1^2}{n_1 n_2 \cdot \ldots \cdot n_{p-1}}}{n_p}$$

$$= \frac{\left[1 + \left(n_p - 1\right)\rho\right]}{n_p} \cdot \left[\underbrace{\breve{\sigma}_p^2}_{=\sigma_{p+1}^2 + \sigma_p^2} + \frac{\sigma_{p-1}^2}{n_{p-1}} + \frac{\sigma_{p-2}^2}{n_{p-2}n_{p-1}} + \cdots + \frac{\sigma_1^2}{n_1 n_2 \cdot \ldots \cdot n_{p-1}}\right]$$
which is, by assumption that the factorization result holds for p:

$$= \frac{\left[1 + \left(n_p - 1\right)\rho\right]}{n_p} \cdot \left[1 + (n_1 - 1)\rho_{12}\right] \cdot \left[1 + (n_2 - 1)\tilde{\rho}_{23}\right] \cdot \ldots \cdot \left[1 + \left(n_{p-1} - 1\right)\tilde{\rho}_{p-1,p}\right]$$
$$\cdot \frac{\sigma_{tot}^2}{n_1 n_2 \cdot \ldots \cdot n_{p-1}}$$
With
$$\sigma_{tot}^2 = \sigma_1^2 + \sigma_2^2 + \cdots + \breve{\sigma}_p^2 = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_p^2 + \sigma_{p+1}^2$$

$$\rho_{k,k+1} = \frac{\breve{\sigma}_p^2 + \sigma_{p-1}^2 + \cdots + \sigma_{k+1}^2}{\breve{\sigma}_p^2 + \sigma_{p-1}^2 + \cdots + \sigma_2^2 + \sigma_k^2} = \frac{\sigma_{p+1}^2 + \sigma_p^2 + \sigma_{p-1}^2 + \cdots + \sigma_{k+1}^2}{\sigma_{p+1}^2 + \sigma_p^2 + \sigma_{p-1}^2 + \cdots + \sigma_2^2 + \sigma_k^2}$$

and $\tilde{\rho}_{k,k+1}$'s recursively defined for $k = 1,2,\cdots,p-1$. Thus, also in the $\tilde{\rho}_{k,k+1}$, $\sigma_p^2 \to \sigma_{p+1}^2 + \sigma_p^2 =: \breve{\sigma}_p^2$.
So we are done if we can show that $\rho = \tilde{\rho}_{p,p+1}$ defined by the recursive formula.
From the way we defined ρ:

$$\rho = \frac{\sigma_{p+1}^2}{\sigma_{p+1}^2 + \sigma_p^2 + \frac{\sigma_{p-1}^2}{n_{p-1}} + \frac{\sigma_{p-2}^2}{n_{p-2}n_{p-1}} + \cdots + \frac{\sigma_k^2}{n_k n_{k+1} \cdot \ldots \cdot n_{p-1}} + \cdots + \frac{\sigma_1^2}{n_1 n_2 \cdot \ldots \cdot n_{p-1}}}$$

$$= \frac{\sigma_{p+1}^2}{\sigma_{p+1}^2 + \sigma_p^2} \cdot \frac{\overbrace{\breve{\sigma}_p^2}^{=\sigma_{p+1}^2 + \sigma_p^2}}{\underbrace{\breve{\sigma}_p^2}_{=\sigma_{p+1}^2 + \sigma_p^2} + \frac{\sigma_{p-1}^2}{n_{p-1}} + \frac{\sigma_{p-2}^2}{n_{p-2}n_{p-1}} + \cdots + \frac{\sigma_k^2}{n_k n_{k+1} \cdot \ldots \cdot n_{p-1}} + \cdots + \frac{\sigma_1^2}{n_1 n_2 \cdot \ldots \cdot n_{p-1}}}$$

$$= \rho_{p,p+1} \cdot \frac{\breve{\sigma}_p^2}{\breve{\sigma}_p^2 + \frac{\sigma_{p-1}^2}{n_{p-1}} + \frac{\sigma_{p-2}^2}{n_{p-2}n_{p-1}} + \cdots + \frac{\sigma_k^2}{n_k n_{k+1} \cdot \ldots \cdot n_{p-1}} + \cdots + \frac{\sigma_1^2}{n_1 n_2 \cdot \ldots \cdot n_{p-1}}}$$

(now we rewrite the denominator using $A + \frac{B}{n} = (A+B) \cdot \left[\frac{1+(n-1)\breve{\rho}}{n}\right]$ with $\breve{\rho} = \frac{A}{A+B}$, and $n = n_{p-1}$)

$$= \rho_{p,p+1} \cdot \frac{\breve{\sigma}_p^2}{\left[\frac{1 + (n_{p-1} - 1)\breve{\rho}}{n_{p-1}}\right] \cdot \left(\breve{\sigma}_p^2 + \sigma_{p-1}^2 + \frac{\sigma_{p-2}^2}{n_{p-2}} + \cdots + \frac{\sigma_1^2}{n_1 n_2 \cdot \ldots \cdot n_{p-2}}\right)}$$

(and this becomes using the result on $\tilde{\rho}_{p-1,p}$ that holds for p with $\sigma_p^2 \to \breve{\sigma}_p^2 = \sigma_{p+1}^2 + \sigma_p^2$)

$$= \rho_{p,p+1} \cdot \frac{\tilde{\rho}_{p-1,p}}{\left[\frac{1 + (n_{p-1} - 1)\breve{\rho}}{n_{p-1}}\right]} = \rho_{p,p+1} \cdot \frac{n_{p-1} \cdot \tilde{\rho}_{p-1,p}}{1 + (n_{p-1} - 1)\tilde{\rho}_{p-1,p}}$$

because $\breve{\rho} = \frac{A}{A+B} = \frac{\breve{\sigma}_p^2}{\breve{\sigma}_p^2 + \sigma_{p-1}^2 + \frac{\sigma_{p-2}^2}{n_{p-2}} + \cdots + \frac{\sigma_1^2}{n_1 n_2 \cdot \ldots \cdot n_{p-2}}} = \tilde{\rho}_{p-1,p}$,
again by using the result for $\tilde{\rho}_{p-1,p}$ that holds for p with $\sigma_p^2 \to \breve{\sigma}_p^2 = \sigma_{p+1}^2 + \sigma_p^2$.
End proof.

The variances at each level can be expressed in the total variance and the intracluster correlations:

$$\sigma_p^2 = \rho_{p-1,p} \cdot \rho_{p-2,p-1} \cdot \cdots \cdot \rho_{12} \cdot \sigma_{tot}^2$$
$$\sigma_{p-1}^2 = (1 - \rho_{p-1,p}) \cdot \rho_{p-2,p-1} \cdot \cdots \cdot \rho_{12} \cdot \sigma_{tot}^2$$
$$\cdots$$

$$\sigma_3^2 = (1-\rho_{34}) \cdot \rho_{23} \cdot \rho_{12} \cdot \sigma_{tot}^2$$
$$\sigma_2^2 = (1-\rho_{23)} \cdot \rho_{12} \cdot \sigma_{tot}^2$$
$$\sigma_1^2 = (1-\rho_{12}) \cdot \sigma_{tot}^2$$

1.3 The correlation $\boldsymbol{\rho}$ between means at different times of the same cluster

Also in the above multilevel data models,
the variance of a cluster mean $\boldsymbol{var(Y_{it\bullet})}$ and the correlation between means at different times of the same cluster $\rho = corr(Y_{it\bullet}, Y_{is\bullet})$ can be expressed as follows

2 levels (Hussey and Hughes):

$$\tau^2 = {\sigma_2}^2, \quad \sigma^2 = \frac{{\sigma_1}^2}{n_1} \quad \xrightarrow{yields} \quad \rho = \frac{{\sigma_2}^2}{\left[{\sigma_2}^2 + \frac{{\sigma_1}^2}{n_1}\right]} = \frac{n_1\rho_{12}}{[1+(n_1-1)\rho_{12}]}$$

$$var(Y_{it\bullet}) = {\sigma_2}^2 + \frac{{\sigma_1}^2}{n_1}$$

Where n_1 is number of subjects in each cluster at each time point.

3 level:
If only the clusters are repeatedly measured (subjects and measurements cross-sectional):

$$\tau^2 = {\sigma_3}^2, \quad \sigma^2 = \frac{{\sigma_2}^2}{n_2} + \frac{{\sigma_1}^2}{n_1 n_2}, \qquad \xrightarrow{yields} \quad \rho = \frac{{\sigma_3}^2}{\left[{\sigma_3}^2 + \frac{{\sigma_2}^2}{n_2} + \frac{{\sigma_1}^2}{n_1 n_2}\right]}$$

So

$$\rho = \frac{\rho_{12}\rho_{23}\sigma_{tot}^2}{\frac{\sigma_{tot}^2}{n_1 n_2} \cdot [1+(n_1-1)\rho_{12}] \cdot [1+(n_2-1)\tilde{\rho}_{23}]} = \frac{\rho_{12}\rho_{23} \cdot n_1 n_2}{[1+(n_1-1)\rho_{12}] \cdot [1+(n_2-1)\tilde{\rho}_{23}]}$$

If subjects within the clusters are repeatedly measured (measurements cross-sectional):

$$\tau^2 = {\sigma_3}^2 + \frac{{\sigma_2}^2}{n_2}, \quad \sigma^2 = \frac{{\sigma_1}^2}{n_1 n_2}, \qquad \xrightarrow{yields} \quad \rho = \frac{{\sigma_3}^2 + \frac{{\sigma_2}^2}{n_2}}{\left[{\sigma_3}^2 + \frac{{\sigma_2}^2}{n_2} + \frac{{\sigma_1}^2}{n_1 n_2}\right]}$$

So:

$$\rho = \frac{\rho_{12} \frac{[1+(n_2-1)\rho_{23}]}{n_2} \sigma^2}{\frac{\sigma^2}{n_1 n_2} \cdot [1+(n_1-1)\rho_{12}] \cdot [1+(n_2-1)\tilde{\rho}_{23}]} = \frac{\rho_{12} n_1 [1+(n_2-1)\rho_{23}]}{[1+(n_1-1)\rho_{12}] \cdot [1+(n_2-1)\tilde{\rho}_{23}]}$$

In all cases:

$$var(Y_{it\bullet}) = {\sigma_3}^2 + \frac{{\sigma_2}^2}{n_2} + \frac{{\sigma_1}^2}{n_1 n_2}.$$

4 level:

If only the clusters are repeatedly measured (sub-clusters and subjects and measurements cross-sectional):

$$\tau^2 = {\sigma_4}^2, \quad \sigma^2 = \frac{{\sigma_3}^2}{n_3} + \frac{{\sigma_2}^2}{n_2 n_3} + \frac{{\sigma_1}^2}{n_1 n_2 n_3}, \qquad \xrightarrow{yields} \quad \rho = \frac{{\sigma_4}^2}{\left[{\sigma_4}^2 + \frac{{\sigma_3}^2}{n_3} + \frac{{\sigma_2}^2}{n_2 n_3} + \frac{{\sigma_1}^2}{n_1 n_2 n_3}\right]}$$

so

$$\rho = \frac{\rho_{12}\rho_{23}\rho_{34}\sigma_{tot}^2}{\frac{\sigma_{tot}^2}{n_1 n_2 n_3} \cdot \left[1 + (n_1 - 1)\rho_{12}\right] \cdot \left[1 + (n_2 - 1)\tilde{\rho}_{23}\right]\left[1 + (n_3 - 1)\tilde{\rho}_{34}\right]}$$

i.e.

$$\rho = \frac{\rho_{12}\rho_{23}\rho_{34} \cdot n_1 n_2 n_3}{\left[1 + (n_1 - 1)\rho_{12}\right] \cdot \left[1 + (n_2 - 1)\tilde{\rho}_{23}\right]\left[1 + (n_3 - 1)\tilde{\rho}_{34}\right]}$$

If sub-clusters within the clusters are repeatedly measured (subjects and measurements cross-sectional):

$$\tau^2 = {\sigma_4}^2 + \frac{{\sigma_3}^2}{n_3}, \quad \sigma^2 = \frac{{\sigma_2}^2}{n_2 n_3} + \frac{{\sigma_1}^2}{n_1 n_2 n_3}, \qquad \xrightarrow{yields} \quad \rho = \frac{{\sigma_4}^2 + \frac{{\sigma_3}^2}{n_3}}{\left[{\sigma_4}^2 + \frac{{\sigma_3}^2}{n_3} + \frac{{\sigma_2}^2}{n_2 n_3} + \frac{{\sigma_1}^2}{n_1 n_2 n_3}\right]}$$

So

$$\rho = \frac{\rho_{12}\rho_{23}\frac{[1+(n_3-1)\rho_{34}]}{n_3}\sigma_{tot}^2}{\frac{\sigma_{tot}^2}{n_1 n_2 n_3} \cdot [1+(n_1-1)\rho_{12}] \cdot [1+(n_2-1)\tilde{\rho}_{23}][1+(n_3-1)\tilde{\rho}_{34}]} = \frac{\rho_{12}\rho_{23} n_1 n_2 [1+(n_3-1)\rho_{34}]}{[1+(n_1-1)\rho_{12}] \cdot [1+(n_2-1)\tilde{\rho}_{23}][1+(n_3-1)\tilde{\rho}_{34}]}.$$

If subjects within sub-clusters within the clusters are repeatedly measured (measurements cross-sectional):

$$\tau^2 = {\sigma_4}^2 + \frac{{\sigma_3}^2}{n_3} + \frac{{\sigma_2}^2}{n_2 n_3}, \quad \sigma^2 = \frac{{\sigma_1}^2}{n_1 n_2 n_3}, \qquad \xrightarrow{yields} \quad \rho = \frac{{\sigma_4}^2 + \frac{{\sigma_3}^2}{n_3} + \frac{{\sigma_2}^2}{n_2 n_3}}{\left[{\sigma_4}^2 + \frac{{\sigma_3}^2}{n_3} + \frac{{\sigma_2}^2}{n_2 n_3} + \frac{{\sigma_1}^2}{n_1 n_2 n_3}\right]}$$

so

$$\rho = 1 - \frac{\frac{{\sigma_1}^2}{n_1 n_2 n_3}}{\left[{\sigma_4}^2 + \frac{{\sigma_3}^2}{n_3} + \frac{{\sigma_2}^2}{n_2 n_3} + \frac{{\sigma_1}^2}{n_1 n_2 n_3}\right]} = 1 - \frac{\frac{(1-\rho_{12})\sigma_{tot}^2}{n_1 n_2 n_3}}{\frac{\sigma_{tot}^2}{n_1 n_2 n_3} \cdot [1+(n_1-1)\rho_{12}] \cdot [1+(n_2-1)\tilde{\rho}_{23}][1+(n_3-1)\tilde{\rho}_{34}]} =$$

$$= 1 - \frac{1 - \rho_{12}}{\left[1 + (n_1 - 1)\rho_{12}\right] \cdot \left[1 + (n_2 - 1)\tilde{\rho}_{23}\right]\left[1 + (n_3 - 1)\tilde{\rho}_{34}\right]}$$

(since $1 - \rho_{12} = \frac{\sigma_1^2}{\sigma_4^2 + \sigma_3^2 + \sigma_2^2 + \sigma_1^2}$).

In all cases:

$$var(Y_{it\bullet}) = {\sigma_4}^2 + \frac{{\sigma_3}^2}{n_3} + \frac{{\sigma_2}^2}{n_2 n_3} + \frac{{\sigma_1}^2}{n_1 n_2 n_3}$$

p-levels

2 Formulas for the variance of the treatment effect estimator $var(\hat{\delta})$

Using that $\rho = corr(Y_{it\bullet}, Y_{it'\bullet}) = \tau^2/(\sigma^2 + \tau^2)$ and $\tau^2 + \sigma^2 = Var(Y_{it\bullet})$, it follows that $\sigma^2 = (1-\rho) \cdot var(Y_{it\bullet})$ and $\tau^2 = \rho \cdot Var(Y_{it\bullet})$, so that

$$var(\hat{\delta}) = \frac{I\sigma^2(\sigma^2 + T\tau^2)}{f(X)\sigma^2 + g(X)\tau^2} = \frac{I \cdot (1-\rho) \cdot [1 + (T-1)\rho]}{f(X) \cdot (1-\rho) + g(X) \cdot \rho} \cdot var(Y_{it\bullet}) \ .$$

Inversion of formula (3) in Girling and Hemming reads, in their definition of σ^2, ρ, and K:

$$var(\hat{\delta}) = \frac{\sigma^2 \cdot (1-\rho)}{K \cdot T} \cdot \frac{1}{(a_D - b_D \cdot R)}$$

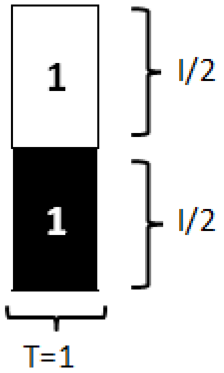
but their σ^2 is shorthand $var(Y_{it\bullet})$, their $\rho = Corr(Y_{it\bullet}, Y_{is\bullet})$ is the same as ours, and their K is the total number of clusters, so our I. Thus their formula reads in our notation:

$$var(\hat{\delta}) = \frac{(1-\rho)}{I \cdot T \cdot (a_D(X) - b_D(X) \cdot R)} \cdot var(Y_{it\bullet}) \ .$$

2.1 Variance inflation factors compared to the parallel group post-test design (cPG1)

In a parallel group post-test (i.e. 1 measurement) cluster randomized design, there are$I$ cluster randomized 1:1 to intervention and control:

Parallel group design with
1 measurement: PG1



and the treatment estimator is

$$\hat{\delta}_{cPG1} = \frac{\sum_{i=\frac{I}{2}+1}^{I} Y_{it\bullet}}{I/2} - \frac{\sum_{i=1}^{I/2} Y_{it\bullet}}{I/2}$$

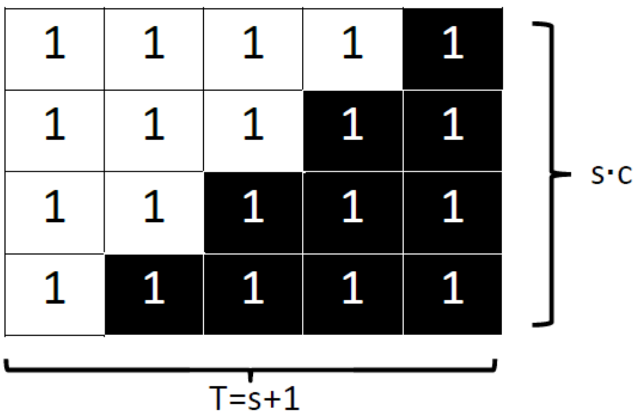
which has variance

$$var(\hat{\delta}_{cPG1}) = 2var\left(\frac{\sum_{i=1}^{I/2} Y_{it\bullet}}{I/2}\right) = \frac{4var(Y_{it\bullet})}{I} \quad ,$$

So

$$var(Y_{it\bullet}) = \frac{I}{4} \cdot var(\hat{\delta}_{cPG1}) \quad .$$

3 The standard stepped wedge design



The above desing illustrates a standard stepped wedge design with $s = 4$ sequences.
If c is the number of clusters per arm, s the number of steps, then the total number of clusters is $I = cs$ and $T = s + 1$,

$$f(X) = f(c,s) = c^2 \cdot \frac{(s^3 - s)}{6}, \qquad g(X) = g(c,s) = c^2 \cdot \frac{(s^4 - s^2)}{12} + f(c,s) = \left(1 + \frac{s}{2}\right) f,$$

A derivation of this is given in the appendix of Woertman, De Hoop, Moerbeek *et al.*, Journal of Clinical Epidemiology, 2013,using that $t = b = 1$ for the standard stepped wedge, and realizing that their notation uses k instead of s for the number of sequences and i instead of c for the number of clusters per sequence.

Then the variance in terms of σ^2 and τ^2 is:

$$var(\hat{\delta}) = \frac{I\sigma^2(\sigma^2 + T\tau^2)}{f(X)\sigma^2 + g(X)\tau^2}$$

$$= \frac{\sigma^2 \cdot [\sigma^2 + (s+1)\tau^2]}{c \cdot \left[\frac{s^2}{6} - \frac{1}{6}\right] \cdot \sigma^2 + \ c \cdot \left[\frac{s^3}{12} - \frac{s}{12} + \frac{s^2}{6} - \frac{1}{6}\right] \cdot \tau^2}$$

$$= \frac{6 \cdot \sigma^2 \cdot [\sigma^2 + (s+1)\tau^2]}{cs \cdot (s - 1/s) \cdot [\sigma^2 + (s/2 + 1)\tau^2]}$$

$$= \frac{6 \cdot}{I \cdot (s - 1/s)} \cdot \sigma^2 \cdot \frac{[\sigma^2 + (1+s)\tau^2]}{[\sigma^2 + (1 + s/2)\tau^2]}$$

$$= \frac{6 \cdot}{I \cdot (s - 1/s)} \cdot \sigma^2 \cdot \left[1 + \frac{s/2 \cdot \tau^2}{\sigma^2 + (1 + s/2)\tau^2}\right]$$

In terms of ρ and $var(Y_{it\bullet})$ it is:

$$var(\hat{\delta}) = \frac{I \cdot (1-\rho) \cdot [1 + (T-1)\rho]}{f(X) \cdot (1-\rho) + g(X) \cdot \rho} \cdot var(Y_{it\bullet})$$

$$= \frac{(1-\rho) \cdot [1 + s\rho]}{c \cdot \left[\frac{s^2}{6} - \frac{1}{6}\right] + \ c \cdot \left[\frac{s^3}{12} - \frac{s}{12}\right] \cdot \rho} \cdot var(Y_{it})$$

$$= \frac{6 \cdot (1-\rho) \cdot [1 + s\rho]}{cs \cdot (s - 1/s) \cdot \left[1 + \frac{s}{2}\rho\right]} \cdot var(Y_{it}) = \frac{6 \cdot (1-\rho) \cdot [1 + s\rho]}{I \cdot (s - 1/s) \cdot \left[1 + \frac{s}{2}\rho\right]} \cdot var(Y_{it})$$

and for two levels, we can see this replicates variance formula as derived in the appendix of (Woertman, De Hoop, Moerbeek *et al.*, Journal of Clinical Epidemiology, 2013) as follows.
Proof: in the 2-level case, $\rho = corr(Y_{it}, Y_{is}) = \frac{\sigma_2^2}{\sigma_2^2 + \frac{\sigma_1^2}{n_1}} = \frac{n_1\rho_{12}}{d}$, and $\rho_{12} = \sigma_2^2/(\sigma_1^2 + \sigma_2^2)$ is the ICC (ratio of level 2 variance to total variance of level 1 and 2), where use $d := 1 + (n_1 - 1)\rho_{12}$ as short-cut.
Then as in the Woertman, De Hoop paper using ***k* as the number of steps (instead of s) and *i* as the number of clusters per sequence (instead of *c*)**:
$1 - \rho = \frac{1-\rho_{12}}{d}$, $\quad 1 + k\rho = [1 + (kn_1 + n_1 - 1)\rho_{12}]/d$,

$$1+\frac{k}{2}\rho=\left[1+\left(\frac{k}{2}n_1+n_1-1\right)\rho_{12}\right]/d,$$

so

$$\frac{6\cdot(1-\rho)\cdot[1+k\rho]}{ik\cdot(k-1/k)\cdot\left[1+\frac{k}{2}\rho\right]}\cdot[1+(n_1-1)\rho_{12}]\cdot\frac{\sigma_{tot}^2}{n_1}=$$

$$=\frac{6\cdot\frac{1-\rho_{12}}{d}\cdot[1+(kn_1+n_1-1)\rho_{12}]/d}{ik\cdot(k-1/k)\cdot\left[1+\left(\frac{k}{2}n_1+n_1-1\right)\rho_{12}\right]/d}\cdot d\cdot\frac{\sigma_{tot}^2}{n_1}$$

$$=\frac{6\cdot(1-\rho_{12})\cdot[1+(kn_1+n_1-1)\rho_{12}]}{(k-1/k)\cdot\left[1+\left(\frac{k}{2}n_1+n_1-1\right)\rho_{12}\right]}\cdot\frac{\sigma_{tot}^2}{n_1 ik}$$

which is the Woertman-De Hoop formula, i.e.,

$$var(\hat{\delta})=\frac{1+\rho(Nkt+Nb-1)}{1+\rho(\frac{1}{2}Nkt+Nb-1)}\cdot\frac{6(1-\rho)}{t\left(k-\frac{1}{k}\right)}\cdot\frac{\sigma_t^2}{Nik},$$

when we realize that $b=1, t=1$, so one measurement at baseline and after each step) after identification of $N=n_1,\ \sigma_t^2=\sigma_{tot}^2,\ ik=I, k=T-1$. *End proof.*

3.1 variance inflation factor compared to cPG1

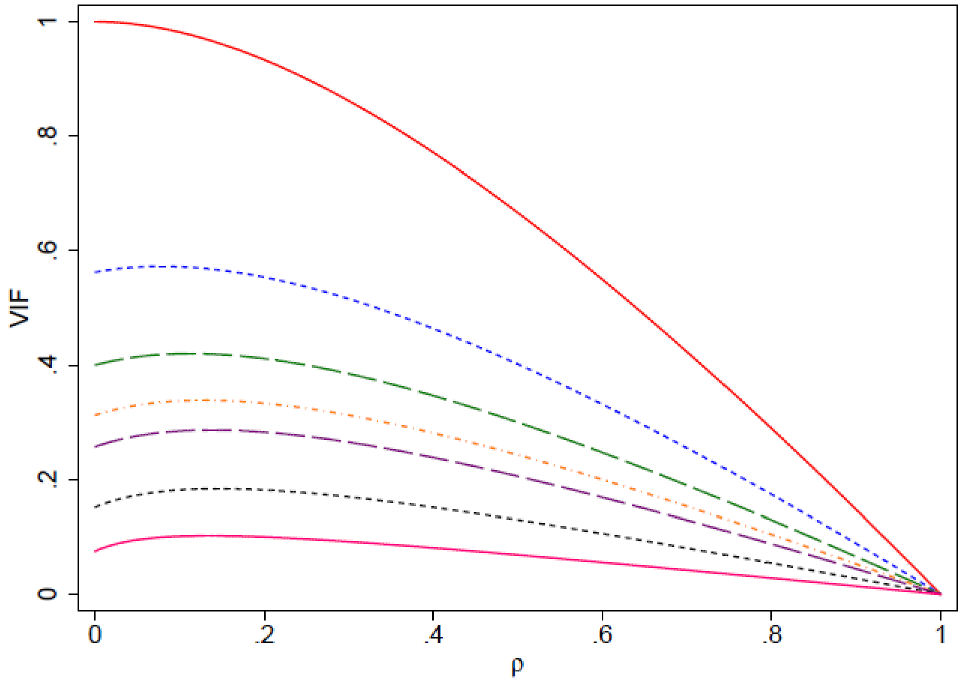
From 2.1 we have

$$var(\hat{\delta})=\frac{6\cdot(1-\rho)\cdot[1+s\rho]}{I\cdot\left(s-\frac{1}{s}\right)\cdot\left[1+\frac{s}{2}\rho\right]}\cdot var(Y_{it})=\frac{3}{2}\cdot\frac{(1-\rho)\cdot[1+s\rho]}{\left(s-\frac{1}{s}\right)\cdot\left[1+\frac{s}{2}\rho\right]}var(\hat{\delta}_{cPG1}) \quad .$$

3.1 Behavior of the VIF_{SW} as a function of ρ

As a function of ρ, $var(\hat{\delta})=\frac{6\cdot(1-\rho)\cdot[1+s\rho]}{I\cdot(s-1/s)\cdot\left[1+\frac{s}{2}\rho\right]}$ is proportional to $\frac{(1-\rho)\cdot[1+s\rho]}{\left[1+\frac{s}{2}\rho\right]}$, which has a derivative with respect to ρ that is zero (for positive ρ) at

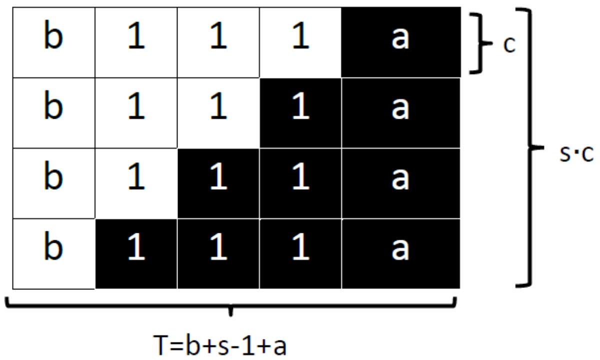
$$\rho_{VIFmax}=\frac{-2s+\sqrt{2s^2+s^3}}{s^2}.$$



The Figure above pictures $VIF_{SW:cPG1}$ as a function of ρ for number of sequences s increasing from the top to the lowest line as $s = 2,3,4,5,6,10,20$.

Hence the value of ρ for which $VIF_{SW:cPG1}$ will reach it maximum and hence $var(\hat{\delta})$ will be the largest will approach 0. Also, the Figure illustrates that maximum over all ρ of $VIF_{SW:cPG1}$ will approach to 0 when the number of sequences s increases.

4 Stepped wedge with more/fewer/no data collection at baseline or at final step



As an example, the above Figure shows a stepped wedge with $s = 4$ sequences an different data collection at baseline and end of the trial. At baseline b times as many data is collected as between switches, and at the final step, a times as many. To be precise, if we denote the the first time point by 1 and the last by t_e, and the sample size in a cluster at time point $t = 2, \dots, t_e - 1$ by N, then $N_1 = b \cdot N$ and $N_{t_e} = a \cdot N$. Formulawise, this comes down to using $T = b + s - 1 + a$ (Thompson *et al.*), so as if there are b measurements before and a measurement after and $s - 1$ measurements in between.

In the appendix 1 to Thompson et al., we find expressions for $IU-W$ and $U^2+ITU-TW-IV$ where U,W,V are the notations of Hussey and Hughes for S,C,R as in our paper.
These lead to,

$$f=IU-W=\frac{c^2}{6}\cdot s\cdot(s^2-1)\ ,$$

and

$$g=U^2+ITU-TW-IV=\frac{c^2}{6}\cdot s\cdot(s^2-1)\cdot\left(b+\frac{s}{2}-1+a\right)=f\cdot\left(b+\frac{s}{2}-1+a\right)\ ,$$

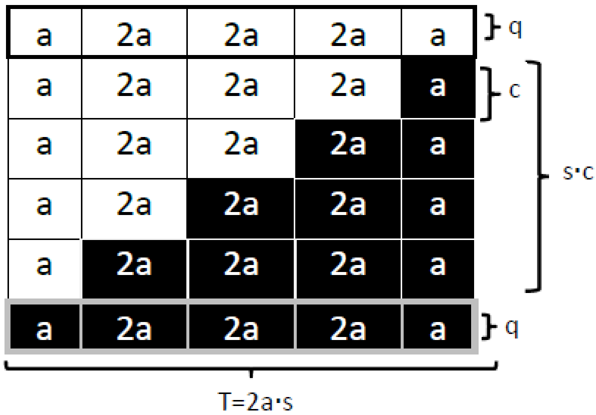
where we substituted our notation s for the number of sequences instead of Thompson's notation k and we substituted c for the number of clusters per sequence instead Thompson's notation i.

4. 1 variance inflation factor compared to cPG1

From 2.1 we have

$$var(\hat{\delta})=\frac{6\cdot(1-\rho)\cdot[1+[a+b-2+s]\cdot\rho]}{I\cdot\left(s-\frac{1}{s}\right)\cdot[1+[a+b-2+s]\cdot\rho]}\cdot var(Y_{it})$$
$$=\frac{3}{2}\cdot\frac{(1-\rho)\cdot[1+[a+b-2+s]\cdot\rho]}{\left(s-\frac{1}{s}\right)\cdot[1+[a+b-2+s]\cdot\rho]}\cdot var(\hat{\delta}_{cPG1})\ .$$

5 Hybrid stepped wedge design



The Figure above illustrates a hybrid design with $s=4$ sequences.
In a hybrid design, a fraction β of the in total I clusters (so $\beta\cdot I$ clusters) is allocated to a *modified* stepped wedge design with s sequences, i.e. with a measurements at $t=1,T$, and $2a$ measurements at $t=2,\ldots,T-1$. In each sequence of the stepped wedge part, there are c clusters, so $s\cdot c=\beta\cdot I$.
The remaining clusters, i.e. $(1-\beta)\cdot I=:2q$ clusters, are allocated to a parallel group design with q clusters per arm measured at the same measurement times.

Formula (3) in Girling & Hemming gives

$$var(\hat{\delta}) = (1-\rho) \cdot \frac{1}{a_D - b_D \cdot R} \cdot \frac{Var(Y_{it\bullet})}{I \cdot T},$$

where we have used that in the notation of their formula (3): their $\sigma^2 = Var(Y_{it\bullet})$, their $\rho = corr(Y_{it\bullet}, Y_{is\bullet})$ is the same as ours, and their K is the total number of clusters, so our I.

Since $= \frac{T \cdot \rho}{1+(T-1) \cdot \rho}$,

$$\frac{1}{a_D - b_D \cdot R} = \frac{1 + (T-1) \cdot \rho}{a_D + [(a_D - b_D) \cdot T - a_D] \cdot \rho} \ ,$$

Girling and Hemming's formula comes down to

$$var(\hat{\delta}) = \frac{1}{I \cdot T} \cdot \frac{(1-\rho) \cdot [1 + (T-1) \cdot \rho]}{a_D + [(a_D - b_D) \cdot T - a_D] \cdot \rho} \cdot var(Y_{it\bullet})$$

Comparing this with our formulation with abbreviating $f(X)$ to f and $g(X)$ to g, we have

$$var(\hat{\delta}\,) = \frac{I \cdot (1-\rho) \cdot [1 + (T-1)\rho]}{f \cdot (1-\rho) + g \cdot \rho} \cdot var(Y_{it\bullet}) = I \cdot \frac{(1-\rho) \cdot [1 + (T-1)\rho]}{f + (g-f) \cdot \rho} \cdot var(Y_{it\bullet}) \ \ ,$$

so that we can identify:

$$\begin{array}{ccc} f & \leftrightarrow & I^2 \cdot T \cdot a_D \ , \\ g & \leftrightarrow & I^2 \cdot T^2 \cdot (a_D - b_D) \ . \end{array}$$

In terms of the number of sequences (s) in the stepped wedge part (with c clusters per sequence) and the number of clusters (q) that are completely in the control condition and completely in the intervention condition, we can calculate f.
According to Table 1 in Girling and Hemming, $4 \cdot a_D = 1 - \frac{\beta^2}{3} \cdot \left(1 + \frac{2}{g^2}\right)$, where g is the number of sequences in the modified stepped wedge part, so in our notation, $g = s$.
Realizing that $I = (s \cdot c + 2q)$ and $(I \cdot \beta = s \cdot c$, we have

$$I^2 \cdot 4a_D = I^2 - \frac{I^2\beta^2}{3} \cdot \left(1 + \frac{2}{s^2}\right) = I^2 - \frac{I^2\beta^2}{3} - \frac{2I^2\beta^2}{3s^2} = (s \cdot c + 2q)^2 - \frac{(s \cdot c)^2}{3} - \frac{2(s \cdot c)^2}{3s^2} =$$

$$= \frac{2s^2c^2}{3} - \frac{2}{3}c^2 + 4q^2 + 4s \cdot c \cdot q \ ,$$

so because $T = 2as$

$$4f = T \cdot (I^2 \cdot 4a_D) = 2as \cdot \left(\frac{2s^2c^2}{3} - \frac{2}{3}c^2 + 4q^2 + 4s \cdot c \cdot q\right) =$$
$$= a \cdot c^2 \cdot \left(\frac{4}{3}s^3 - \frac{4}{3}s\right) + a \cdot c \cdot (8s^2q) + 8asq^2 \ .$$

Thus,

$$f = a \cdot c^2 \cdot \left(\frac{1}{3}s^3 - \frac{1}{3}s\right) + a \cdot c \cdot (2s^2q) + a \cdot (2sq^2) \ .$$

For calculation of g, we use from Table 1 in Girling and Hemming that

$$(4a_D - 4b_D) = \tfrac{\beta}{3} \cdot \left[2 + \tfrac{1}{g^2} - \beta \cdot \left(1 + \tfrac{2}{g^2}\right)\right] = \tfrac{\beta}{3} \cdot \left[2 + \tfrac{1}{s^2} - \beta \cdot \left(1 + \tfrac{2}{s^2}\right)\right],$$

so that

$$I^2(4a_D - 4b_D) = I^2\left\{\frac{\beta}{3} \cdot \left[2 + \frac{1}{s^2} - \beta \cdot \left(1 + \frac{2}{s^2}\right)\right]\right\} = \frac{2}{3} \cdot (I\beta) \cdot I + \frac{1}{3} \cdot \frac{I\beta \cdot I}{s^2} - \frac{1}{3} \cdot (I\beta)^2 - \frac{2}{3} \cdot \frac{(I\beta)^2}{s^2} =$$

$$= \frac{2}{3} \cdot (sc) \cdot (sc + 2q) + \frac{1}{3} \cdot \frac{sc \cdot (sc + 2q)}{s^2} - \frac{1}{3} \cdot (sc)^2 - \frac{2}{3} \cdot \frac{(sc)^2}{s^2} =$$

$$= c^2 \cdot \left(\frac{1}{3}s^2 - \frac{1}{3}\right) + c \cdot \left(\frac{4}{3}sq + \frac{2}{3} \cdot \frac{q}{s}\right) .$$

Then

$$4g = T^2 \cdot [I^2(4a_D - 4b_D)] = (2as)^2 \cdot \left[c^2 \cdot \left(\frac{1}{3}s^2 - \frac{1}{3}\right) + c \cdot \left(\frac{4}{3}sq + \frac{2}{3} \cdot \frac{q}{s}\right)\right] =$$

$$= a^2 \cdot c^2 \cdot \left(\frac{4}{3}s^4 - \frac{4}{3}s^2\right) + a^2 \cdot c \cdot \left(\frac{16}{3}s^3 q + \frac{8}{3} \cdot sq\right) ,$$

so

$$g = a^2 \cdot c^2 \cdot \left(\frac{1}{3}s^4 - \frac{1}{3}s^2\right) + a^2 \cdot c \cdot \left(\frac{4}{3}s^3 q + \frac{2}{3} \cdot sq\right) .$$

5. 1 variance inflation factor compared to cPG1

From 2.1 we have

$$var(\hat{\delta}) = (1 - \rho) \cdot \frac{1}{a_D - b_D \cdot R} \cdot \frac{Var(Y_{it\bullet})}{I \cdot T} = \frac{(1 - \rho)}{T} \cdot \frac{1}{4a_D - 4b_D \cdot R} \cdot var(\hat{\delta}_{cPG1})$$

where

$$4a_D = 1 - \frac{\beta^2}{3}\left(1 + \frac{2}{s^2}\right) \quad , \quad 4b_D = \left(1 - \frac{\beta}{3}\left[2 + \frac{1}{s^2}\right]\right) .$$

6 General properties of the functions $f(X)$ and $g(X)$

We have that $K_1 I^2 \leq f^2 \leq K_2 I^2, f > 0$ for some constants $K_1, K_2 > 0$.

Proof. Recall that $f(X) = S \cdot I - C$, where $S = \Sigma_{it} X_{it}$, $C = \Sigma_t(\Sigma_i X_{it})^2$, $R = \Sigma_i(\Sigma_t X_{it})^2$ are the total sum of 1's in the matrix X, and the sum of the squares of the column sums, and the sum of the squares of the row sums.

As we dealing with I clusters that have T measurements, we have that the design matrix X has I rows and T columns. We consider the reasonably general situation that some of the clusters have at some of the measurements the intervention condition and the control condition other measurements. Say pI rows (clusters) will have in at least one column (measurement) a 1 (intervention condition) and 0's (control condition) in the rest of the columns (measurements). Then:

- $S = \beta pI$ for some $\beta \geq 1$ (as the total number of 1's in X will be at pI). Consequently,
- $S \cdot I = \beta p I^2$.
- The design can be divided in $k \leq pI$ arms where all clusters in an arm have the same sequence of intervention condition and control condition over the T measurements and at

least once the intervention (we call this a non-null arm). There may be an arm consisting of clusters that remain in the control condition throughout the study; we call this the null arm. Let I_i be the number of clusters in non-null arm i. Then $\sum_1^k I_i = pI$. Then there are at least k columns having I_i 1's (say the first time that the I_i clusters in non-null arm $i = 1,2, \ldots, k$ have the intervention. Then $C \geq \sum_1^k {I_i}^2$. Note that this also holds if some (or all) of the arms have their first 1 in the same column. If the smallest $I_i = \alpha \cdot I$, with $\alpha \geq 1$. Consequently,

- $C \geq k\alpha^2 I^2$.
- Then: $f = (\beta p - k\alpha^2)I^2$
- Now f > 0 always as per following argument. In the matrix X, which has I rows (clusters) and T columns (measurements), there are n_0 columns with only 0's, n_1 columns with only one 1, n_2 columns with two 1's, …, n_I columns with I 1's, so only 1's. Then $n_0 + n_1 + n_2 + n_3 + \cdots + n_I = T$, but more importantly: $S = 0n_0 + 1n_1 + 2n_2 + 3n_3 + \cdots + In_I$ and hence $S \cdot I = n_1 I + 2In_2 + 3In_3 + \cdots + I^2 n_I$. Now $C = n_1 + n_2 \cdot 2 \cdot 2 + n_3 \cdot 3 \cdot 3 + \cdots + n_I \cdot I \cdot I$. Comparing $S \cdot I$ and C term by term we see that for n_i that are non-zero: $n_1 I > n_1, n_2(2 \cdot I) > n_2(2 \cdot 2), n_3(3 \cdot I) > n_3(3 \cdot 3), \ldots, n_{I-1}((I-1) \cdot I) > n_{I-1}(I-1) \cdot (I-1),\ n_I(I \cdot I) = n_I(I \cdot I)$, so $S \cdot I > C$, because not all $n_i = 0$.

End proof.

Alternative proof. As we dealing with I clusters that have T measurements, we have that the design matrix X has I rows and T columns and

$$S = \sum_{t=1}^{T}\sum_{i=1}^{I} X_{it}\,,\quad I = \sum_{j=1}^{I} 1\,,\quad S \cdot I = \left(\sum_{t=1}^{T}\sum_{i=1}^{I} X_{it}\right)\cdot\left(\sum_{j=1}^{I} 1\right) = \sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{j=1}^{I} X_{it} \cdot 1\ .$$

Also,

$$C = \sum_{t=1}^{T}\left(\sum_{i=1}^{I} X_{it}\right)^2 = \sum_{t=1}^{T}\left(\sum_{i=1}^{I} X_{it}\right)\cdot\left(\sum_{j=1}^{I} X_{jt}\right) = \sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{j=1}^{I} X_{it} \cdot X_{jt}\ .$$

Substracting gives:

$$f = S \cdot I - C = \sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{j=1}^{I} X_{it} \cdot (1 - X_{jt}) = \sum_{t=1}^{T}\underbrace{\left(\sum_{i=1}^{I} X_{it}\right)}_{E_t}\cdot\underbrace{\left(\sum_{j=1}^{I}(1 - X_{jt})\right)}_{Z_t}\ ,$$

where Z_t is the number of 0's in column t and E_t is the number of 1's in column t. Thus, the number of clusters in the control and intervention condition at time t.

Thus, $f \geq 0$. Since $E_t, Z_t \geq 0$, we can have $f = 0$ only if we have a design with for each time t either $E_t = 0$, i.e. all clusters in the control condition, or $Z_t = 0$, i.e. all clusters in the intervention condition. That is, the design only allows before-after comparisons, possibly multiple of those comparisons and possibly none (if $X_{it} = 1$ for all i, t or $X_{it} = 0$ for all i, t). This makes only sense for estimating a treatment effect if there is at one before-after comparison *and* one can assume that there is no time trend.

End proof.

We have that $g \geq 0$

Proof. As the variance of the design is given by

$$\mathrm{var}(\hat{\delta}\,) = \frac{I\sigma^2(\sigma^2 + T\tau^2)}{f(X)\sigma^2 + g(X)\tau^2},$$

with $\sigma^2, \tau^2, T, I > 0$ we have that for any design (maxtrix X),

$$f(X)\sigma^2 + g(X)\tau^2 > 0.$$

Now, in a multilevel design, we can get σ^2 as small as we want if the sample size at all levels below the clusters level becomes extremely large, while τ^2 will then converge to the variance component of the highest level i.e. a non-zero positive value (see e.g. Table 1). Then $g(X)$ must be ≥ 0. Note that $g(X)$ can be zero, e.g. for a parallel group design. *End proof.*

7 Factors influencing power

7.1 Increasing the total number of clusters I or the number of sequences s

From the section 2.1 and the sections on the specific designs (3,4,5) we have

$$var(\hat{\delta}) = Var(\hat{\delta}_{cPG1}) = VIF_{rm} \cdot \frac{4var(Y_{it\bullet})}{I} \quad ,$$

and from section 1.2, we have

$$var(Y_{it\bullet}) = [1 + (n_1 - 1)\rho_{12}] \cdot [1 + (n_2 - 1)\tilde{\rho}_{23}] \cdot \ldots \cdot \left[1 + (n_{p-1} - 1)\tilde{\rho}_{p-1,p}\right] \cdot \frac{\sigma_{tot}^2}{n_1 n_2 \cdot \ldots \cdot n_{p-1}}$$

so

$$var(\hat{\delta}) = VIF_{rm:cPG1} \cdot VIF_p \cdot \frac{4\sigma_{tot}^2}{I \cdot n_1 \cdot n_2 \cdots n_{p-1}}$$

where

$$VIF_{rm:cPG1} = \left\{ \begin{array}{c} VIF_{SW_s:cPG1} = \frac{3}{2} \cdot \frac{(1-\rho) \cdot (1 + s \cdot \rho)}{\left(s - \frac{1}{s}\right) \cdot \left(1 + \frac{s}{2} \cdot \rho\right)} \quad , \\ \\ VIF_{SW_s(a,b):cPG1} = \frac{3}{2} \cdot \frac{(1-\rho) \cdot (1 + [a + b - 2 + s] \cdot \rho)}{\left(s - \frac{1}{s}\right) \cdot \left(1 + \left[a + b - 2 + \frac{s}{2}\right] \cdot \rho\right)} \, , \\ \\ VIF_{H(\beta,s):cPG1} = \frac{(1-\rho)}{T} \cdot \frac{1}{1 - \frac{\beta^2}{3}\left(1 + \frac{2}{s^2}\right) + R \cdot \left(1 - \frac{\beta}{3}\left[2 + \frac{1}{s^2}\right]\right)} \quad , \end{array} \right\}$$

and

$$VIF_p = [1 + (n_1 - 1)\rho_{12}] \cdot [1 + (n_2 - 1)\tilde{\rho}_{23}] \cdots \left[1 + (n_{p-1} - 1)\tilde{\rho}_{p-1,p}\right] .$$

Everything else being equal, we see that increasing the total number of clusters I decreases $var(\hat{\delta})$ and that increasing the number of sequences s decreases $var(\hat{\delta})$ except for the hybrid design and except for the situation described now:

For the standard stepped wedge and the stepped wedge with more/less/no data collection at baseline and end, the following holds. As shown by Thompson et al., when the total cluster size is over all measurements $m = n_1 \cdots n_{p-1} \cdot T$ is kept constant, then this total cluster size is inversely proportional to the number of sequences ($m \sim 1/s$, because $T = a + b + s - 1$). Because of this, the factor

$$VIF_{rm:cPG1} \cdot \frac{1}{n_1 \cdot n_2 \cdots n_{p-1}} \sim \frac{1}{s} \cdot \frac{1}{m}$$

is rather constant when s increases and thus

$$var(\hat{\delta}) = VIF_{rm;cPG1} \cdot VIF_p \cdot \frac{4\sigma_{tot}^2}{I \cdot n_1 \cdot n_2 \cdots n_{p-1}}$$

will not decrease to 0 as s increases.

7.2 Regardless of the design, increasing sample size at any level increases the power

Regardless of the design (i.e., f, g), we can reformulate

$$var(\hat{\delta}) = I \cdot \sigma^2 \cdot \frac{(\sigma^2 + T\tau^2)}{f(X)\sigma^2 + g(X)\tau^2}, \quad X \text{ design matrix}$$

For simplicity we can rewrite the last factor as

$$F(x, y) = \frac{x + a \cdot y}{cx + b \cdot y}$$

with $x = \sigma^2 > 0, y = \tau^2 > 0, a = T \geq 1, b = g \geq 0, c = f > 0$.

Note that

$$\tau^2 = \sigma_p^2 + \frac{\sigma_{p-1}^2}{n_{p-1}} + \frac{\sigma_{p-2}^2}{n_{p-2}n_{p-1}} + \cdots + \frac{\sigma_k^2}{n_{p-1} \cdot \ldots \cdot n_k}$$

And

$$\sigma^2 = \frac{\sigma_{k-1}^2}{n_{p-1} \cdot \ldots \cdot n_k n_{k-1}} + \cdots + \frac{\sigma_1^2}{n_{p-1} \cdot \ldots \cdot n_2 n_1}$$

for some k between 2 and p. Now let the sample size at a particular level increase from $n_{u,1}$ to $n_{u,2}$. We show that $x_1 F(x_1, y_1) \geq x_2 F(x_2, y_2)$, where $x_1 = \sigma^2$ for $n_u = n_{u,1}$ etc. Then $(\hat{\delta}_1) \geq var(\hat{\delta}_2)$.

Proof. We have to show that

$$x_1 \cdot \frac{x_1 + a \cdot y_1}{cx_1 + b \cdot y_1} \geq x_2 \cdot \frac{x_2 + a \cdot y_2}{cx_2 + b \cdot y_2}$$

Equivalently:

$$cx_1^2 x_2 + x_1^2 y_2 b + x_1 x_2 y_1 ac + aby_1 y_2 x_1 \geq cx_2^2 x_1 + x_1 x_2 y_2 ac + x_2^2 y_1 b + aby_1 y_2 x_2$$

Equivalenty:

$$(cx_1 x_2 + aby_1 y_2) \cdot (x_1 - x_2) + (x_1^2 y_2 - x_2^2 y_1) b + x_1 x_2 ac(y_1 - y_2) \geq 0 \ (*)$$

Note that if $u \geq k$, then increasing $n_{u,1}$ to $n_{u,2}$ will decrease τ^2 and σ^2 will remain the same, while if $u < k$, the opposite occurs. Thus, in every case case, $x_1 \geq x_2$ and $y_1 \geq y_2$ and the therefore the first and last term of (*) are ≥ 0.

We are done if we can show that $x_1^2 y_2 - x_2^2 y_1 \geq 0$.
To this end, we distinguish again the cases $u \geq k$ and $u < k$.
If $u \geq k$, we can write $y_1 = \tau^2 | [n_u = n_{u,1}] = A + B/n_{u,1}$
where $A = \sigma_p^2 + \frac{\sigma_{p-1}^2}{n_{p-1}} + \frac{\sigma_{p-2}^2}{n_{p-2} n_{p-1}} + \cdots + \frac{\sigma_{u+1}^2}{n_{u+1} \cdot \ldots \cdot n_{p-1}}$ and

$$B = \frac{\sigma_u^2}{\cancel{n_u} \cdot n_{u+1} \cdot \ldots \cdot n_{p-1}} + \frac{\sigma_{u-1}^2}{\cancel{n_u} \cdot n_{u-1} \cdot \ldots \cdot n_{p-1}} + \cdots + \frac{\sigma_{k+1}^2}{n_{k+1} \cdot \ldots \cdot \cancel{n_u} \cdot \ldots \cdot n_{p-1}}.$$

Furthermore, $x_1 = \sigma^2 | [n_u = n_{u,1}] = C/n_{u,1}$, where $C = \frac{\sigma_k^2}{n_k n_{k+1} \cdot \ldots \cdot \cancel{n_u} \cdot \ldots \cdot n_{p-1}} + \cdots + \frac{\sigma_1^2}{n_1 n_2 \cdot \ldots \cdot \cancel{n_u} \cdot \ldots \cdot n_{p-1}}$.

Similarly, $y_2 = A + B/n_{u,2}$ and $x_2 = C/n_{u,2}$.

Then

$$x_1^2 y_2 - x_2^2 y_1 = \left[\frac{C}{n_{u,1}}\right] A + B/n_{u,2}] - \left[\frac{C}{n_{u,2}}\right]^2 A + B/n_{u,1}]^2 = C^2 A \cdot \left[\frac{1}{n_{u,1}^2} - \frac{1}{n_{u,2}^2}\right] + \frac{C^2 B}{n_{u,1} n_{u,2}} \left[\frac{1}{n_{u,1}} - \frac{1}{n_{u,2}}\right]$$

Which is ≥ 0, because $n_{u,1} \leq n_{u,2}$.

If $u < k$, then $y_2 = y_1$, and $x_1^2 y_2 - x_2^2 y_1 = (x_1^2 - x_2^2) y_1 \geq 0$, because $x_1 \geq x_2$.
End proof.

7.3 How to achieve arbitrary high power levels

If the following condition holds:

$$\text{there is a constant } K > 0, \text{such that } \frac{(\sigma^2 + T\tau^2)}{\sigma^2 + g(X)/f(X)\tau^2} \leq K, \quad (\tau^2, \sigma^2 > 0) \qquad (@)$$

then we can get any power we want by increasing the sample size at any levels that is repeated or at the first cross-sectional level below the repeated levels.
In particular: this holds for the standard stepped wedge design, the stepped wedge with more/less/no data collection at baseline/end and the hybrid, but not for the parallel group design.

Proof. Because $f > 0$, the expression in (@) is defined. Moreover, $f \sim I^2$, say $f \geq K_2 \cdot I^2$ by the result earlier.
Then

$$var(\hat{\delta}) = \frac{I}{f} \cdot \sigma^2 \cdot \frac{(\sigma^2 + T\tau^2)}{\sigma^2 + \left(\frac{g}{f}\right)\tau^2} \leq \frac{K_2}{I} \cdot K \cdot \sigma^2 .$$

As

$$\sigma^2 = \frac{\sigma_{k-1}^2}{n_{p-1} \cdot \ldots \cdot n_{k-1}} + \frac{\sigma_{k-2}^2}{n_{p-1} \cdot \ldots \cdot n_{k-1} n_{k-2}} + \cdots + \frac{\sigma_2^2}{n_{p-1} \cdot \ldots \cdot n_{k-1} \cdot \ldots \cdot n_2} + \frac{\sigma_1^2}{n_{p-1} \cdot \ldots \cdot n_{k-1} \cdot \ldots \cdot n_1}$$

for some $k \geq 2$, we have that σ^2 and thus $var(\hat{\delta})$ can get arbitrarily small whenever n_i sufficiently increases for $i = k-1, k, k+1, \ldots, p-1$. Of course, if $n_p = I$ sufficiently increases, then $var(\hat{\delta})$ can get arbitrarily small as well. Since $i = k, k+1, \ldots, p-1, p$ are the repeated levels and $i = k-1$ is the first level below it, increasing any of corresponding n_i will increase power to any degree desired.
For the standard stepped wedge design ($a = b = 1$) and the stepped wedge with more/less/no data collection

$g = \left(a + b + \frac{s}{2} - 1\right) f, T = a + b + s - 1$, and

$$\frac{[\sigma^2 + (a + b + s - 1)\tau^2]}{\left[\sigma^2 + \left(a + b + \frac{s}{2} - 1\right)\tau^2\right]} = 1 + \frac{\frac{s}{2}\tau^2}{\left[\sigma^2 + \left(a + b + \frac{s}{2} - 1\right)\tau^2\right]}$$

is minimally 1 and maximally $1 + \frac{s}{2a+2b+s-2}$, so $K = 2$ does the job.

For the hybrid design, $\frac{g}{f}$ and T are positive, non-zero constants (depending on design parameters, independent of σ^2 and τ^2). There is a number K such that $T \leq K \cdot \frac{g}{f}$, and $K \geq 1$. Then

$$\frac{(\sigma^2 + T\tau^2)}{\sigma^2 + \frac{g}{f} \cdot \tau^2} \leq \frac{(K\sigma^2 + T\tau^2)}{\sigma^2 + \frac{g}{f} \cdot \tau^2} \leq \frac{\left(K\sigma^2 + K \cdot \frac{g}{f}\tau^2\right)}{\sigma^2 + \frac{g}{f} \cdot \tau^2} \leq K \ .$$

Note that this holds for any design with $\frac{g}{f}$ and T are positive, non-zero constants.

However, for the parallel group design with T repeated measurements: $f = I^2/4$, $g = 0$, and thus

$$\frac{(\sigma^2 + T\tau^2)}{\sigma^2 + g(X)/f(X)\tau^2} = 1 + T\left(\frac{\tau^2}{\sigma^2}\right)$$

And this can get arbitrarily large when τ^2 becomes very large compared to σ^2, so this expression is not bounded over all $\tau^2, \sigma^2 > 0$.
End proof.

Corrollary:
We can compensate underpower due to few higher level units (e.g. clusters) by adding more units at a lower level provided this is a repeated measurements level or the first level below the levels that are repeatedly measured.
As an example, we have a four-level design such as in the CHANGE trial with organisations (level 4) and nursing homes (level 3) repeatedly measured but nurses (level 2) and of course observations (level 1) measured cross-sectionally, then we can compensate a lack of power by too few organizations by increasing the number of nursing homes per organization or the number of nurses per nursing home.
Practically, the power formula states:
$\frac{\delta}{\sqrt{var(\hat{\delta})}} - z_{1-\frac{\alpha}{2}} = z_{1-\beta}$, which implies that if power is 50% : $\frac{\delta}{\sqrt{var(\hat{\delta})}} - 1.96 = 0$ and if power is 80%: $\frac{\delta}{\sqrt{var(\hat{\delta})}} - 1.96 = 0.84$, so $\sqrt{var(\hat{\delta})}$ has to 1.96/(0.84+1.96) =0.7 smaller to get from 50% power to 80% power i.e. $var(\hat{\delta})$ has to be 0.49 smaller i.e. the variance should be halved.
Similarly, to get from a power of 30% ($z_{0.3} = 0.53$) power to 80% power, $var(\hat{\delta})$ has to be 1/3.8 times smaller, so approximately 4 times smaller.
Restricting to a standard stepped wedge design, we have from earlier results

$$var(\hat{\delta}) = \frac{6 \cdot}{I \cdot (s - 1/s)} \cdot \sigma^2 \cdot \left[1 + \frac{s/2 \cdot \tau^2}{\sigma^2 + (1 + s/2)\tau^2}\right]$$

So

$$\frac{6\cdot}{I\cdot(s-1/s)}\cdot\sigma^2\cdot 1 \le var(\hat{\delta}) \le \frac{6\cdot}{I\cdot\left(s-\frac{1}{s}\right)}\cdot\sigma^2\cdot 2.$$

Thus, it is sufficient to make adjust the sample sizes at level 3 or 2 such that σ^2 4 times smaller to get from 50% to 80% power and 8 times smaller to get from 30% to 80% power.

Proof. If $\tilde{\sigma}^2 = \sigma^2/4$, then $var\left(\underbrace{\hat{\delta}}_{\tilde{\sigma}^2}\right) \le \frac{6\cdot}{I\cdot\left(s-\frac{1}{s}\right)}\cdot\frac{\sigma^2}{4}\cdot 2 = \frac{1}{2}\cdot\left(\frac{6\cdot}{I\cdot\left(s-\frac{1}{s}\right)}\cdot\sigma^2\right) \le \frac{1}{2}\cdot\ var\left(\underbrace{\hat{\delta}}_{\sigma^2}\right)$. And similarly for $\tilde{\sigma}^2 = \sigma^2/4$. *End proof.*

As $\sigma^2 = \frac{{\sigma_2}^2}{n_2 n_3} + \frac{{\sigma_1}^2}{n_1 n_2 n_3}$ in this case (see Table 1), this means that n_3 or n_2 has to be 4 or 8 times larger.

7.4 Power as a function of intracluster correlations $\rho_{u,u+1}$

The power of as a function of an intracluster correlation can be obtained by combining the function $\rho \to VIF(\rho)$ that depends on the design and the function $\rho_{u,u+1} \to \rho$ that depends on the multilevel data generating model. For the latter we prove:

ρ increases if any intracluster correlation increases, i.e.,
For any p-level design and any $1 \le u \le p-1$, the function $\rho_{u,u+1} \to \rho = \tau^2/(\tau^2+\sigma^2)$ is increasing.

To get the idea of the proof, let us first consider the 4-level design. Then
$\sigma_4^2 = \rho_{34}\cdot\rho_{23}\cdot\rho_{12}\cdot\sigma_{tot}^2$
$\sigma_3^2 = (1-\rho_{34})\cdot\rho_{23}\cdot\rho_{12}\cdot\sigma_{tot}^2$
$\sigma_2^2 = (1-\rho_{23)}\cdot\rho_{12}\cdot\sigma_{tot}^2$
$\sigma_1^2 = (1-\rho_{12})\cdot\sigma_{tot}^2$
Because σ_{tot}^2 occurs in every σ_i^2, it factors out in every formula and therefore we may as well omit it. From the above it is clear that the all σ_i^2 will have a non-zero derivative with respect to ρ_{12}; only $\sigma_2^2, \sigma_3^2,\ \sigma_4^2$ will have a non-zero derivative with respect to ρ_{23}; and only $\sigma_3^2,\ \sigma_4^2$ will have a non-zero derivative with respect to ρ_{34}.
By way of an example, we show that $\rho_{23} \to \rho$ has a positive (or at least non-negative) derivative if $\tau^2 = \sigma_4^2 + \frac{\sigma_3^2}{n_3} + \frac{\sigma_2^2}{n_3 n_2}$ and $\sigma^2 = \frac{\sigma_1^2}{n_1 n_2 n_3}$. We then have that $\frac{\partial\tau^2}{\partial\rho_{23}} = \rho_{34}\cdot\rho_{12} + \frac{(1-\rho_{34})\cdot\rho_{12}}{n_3} - \frac{\rho_{12}}{n_3 n_2}$ and $\frac{\partial\sigma^2}{\partial\rho_{23}} = 0$.
Now,

$$\frac{\partial\rho}{\partial\rho_{23}} = \frac{(\tau^2+\sigma^2)\cdot\frac{\partial\tau^2}{\partial\rho_{23}} - \tau^2\cdot\frac{\partial(\tau^2+\sigma^2)}{\partial\rho_{23}}}{(\tau^2+\sigma^2)^2} = \frac{\frac{\partial\tau^2}{\partial\rho_{23}}\cdot\sigma^2 - \frac{\partial\sigma^2}{\partial\rho_{23}}\cdot\tau^2}{(\tau^2+\sigma^2)^2}$$

So this is positive if $\frac{\partial\tau^2}{\partial\rho_{23}}\cdot\sigma^2 - \frac{\partial\sigma^2}{\partial\rho_{u,u+1}}\cdot\tau^2 \ge 0$, and this is

$$\left(\rho_{34}\cdot\rho_{12} + \frac{(1-\rho_{34})\cdot\rho_{12}}{n_3} - \frac{\rho_{12}}{n_3 n_2}\right)\cdot\frac{\sigma_1^2}{n_1 n_2 n_3} - 0\cdot\left[\sigma_4^2 + \frac{\sigma_3^2}{n_3} + \frac{\sigma_2^2}{n_3 n_2}\right] =$$
$$\ge \left[\frac{\rho_{34}\cdot\rho_{12} + (1-\rho_{34})\cdot\rho_{12} - \rho_{12}}{n_3 n_2}\right]\cdot\frac{\sigma_1^2}{n_3 n_2 n_1} = 0$$

Which concludes what we wanted to show. Note that the $\ge$ is a strict $>$ when $\rho_{12} > 0$ combined with either $\rho_{34} > 0$ and $n_3 n_2 > 1$ or $\rho_{34} < 1$ and $n_2 > 1$.

The general situation is similar and goes like this:

Proof. Consider a p-level design, so $p \geq 2$. Then

$$\tau^2 = \sigma_p^2 + \frac{\sigma_{p-1}^2}{n_{p-1}} + \frac{\sigma_{p-2}^2}{n_{p-1} \cdot n_{p-2}} + \cdots + \frac{\sigma_k^2}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_{k+1} n_k}$$

and

$$\sigma^2 = \frac{\sigma_{k-1}^2}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_k n_{k-1}} + \cdots + \frac{\sigma_2^2}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_3 \cdot n_2} + \frac{\sigma_1^2}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_2 \cdot n_1}$$

for a $2 \leq k \leq p-1$ and

$$\rho = \tau^2 / (\tau^2 + \sigma^2)$$

Let $1 \leq u \leq p-1$.
We prove that $\rho_{u,u+1} \to \rho$ is increasing by showing that its derivative

$$\frac{\partial \rho}{\partial \rho_{u,u+1}} \geq 0$$

in several steps.

Step 1. It suffices to show that

$$\frac{\partial \tau^2}{\partial \rho_{u,u+1}} \cdot \sigma^2 \geq \frac{\partial \sigma^2}{\partial \rho_{u,u+1}} \cdot \tau^2.$$

This follows from

$$\frac{\partial \rho}{\partial \rho_{u,u+1}} = \frac{(\tau^2 + \sigma^2) \cdot \dfrac{\partial \tau^2}{\partial \rho_{u,u+1}} - \tau^2 \cdot \dfrac{\partial(\tau^2 + \sigma^2)}{\partial \rho_{u,u+1}}}{(\tau^2 + \sigma^2)^2}$$

When we note that the denominator is always positive.

Step 2.
Because

$$\rho_{u,u+1} = \frac{\sigma_p^2 + \sigma_{p-2}^2 + \cdots + \sigma_{u+1}^2}{\sigma_p^2 + \sigma_{p-2}^2 + \cdots + \sigma_{u+1}^2 + \sigma_u^2}$$

we have, **if we set $\boldsymbol{\sigma_{tot}^2 = 1}$** (note that this can be done without loss of generality because the factor σ_{tot}^2 appears in every right hand side) that the terms involved in making τ^2 and σ^2 can be expressed as follows:

$\sigma_p^2 = \rho_{p-1,p} \cdot \rho_{p-2,p-1} \cdot \ldots \cdot \rho_{w,w+1} \cdot \ldots \cdot \rho_{v,v+1} \cdot \ldots \cdot \rho_{23} \cdot \rho_{12}$
$\dfrac{\sigma_{p-1}^2}{n_{p-1}} = \dfrac{\left(1 - \rho_{p-1,p}\right) \cdot \rho_{p-2,p-1} \cdot \ldots \cdot \rho_{w,w+1} \cdot \ldots \cdot \rho_{v,v+1} \cdot \ldots \cdot \rho_{23} \cdot \rho_{12}}{n_{p-1}}$
$\vdots$ $\vdots$

$\frac{\sigma_{w+1}^2}{n_{p-1}\cdot n_{p-2}\cdot \ldots \cdot n_{w+1}} = \frac{(1-\rho_{w+1,w+2})\cdot \rho_{w,w+1}\cdot \rho_{w-1,w}\cdot \ldots \cdot \rho_{v,v+1}\cdot \ldots \cdot \rho_{23}\cdot \rho_{12}}{n_{p-1}\cdot n_{p-2}\cdot \ldots \cdot n_{w+1}}$
$\frac{\sigma_{w}^2}{n_{p-1}\cdot n_{p-2}\cdot \ldots \cdot n_{w}} = \frac{(1-\rho_{w,w+1})\cdot \rho_{w-1,w}\cdot \rho_{w-2,w-1}\cdot \ldots \cdot \rho_{v,v+1}\cdot \ldots \cdot \rho_{23}\cdot \rho_{12}}{n_{p-1}\cdot n_{p-2}\cdot \ldots \cdot n_{w}}$
$\frac{\sigma_{w-1}^2}{n_{p-1}\cdot n_{p-2}\cdot \ldots \cdot n_{w}\cdot n_{w-1}} = \frac{(1-\rho_{w-1,w})\cdot \rho_{w-2,w-1}\cdot \rho_{w-3,w-2}\cdot \ldots \cdot \rho_{v,v+1}\cdot \ldots \cdot \rho_{23}\cdot \rho_{12}}{n_{p-1}\cdot n_{p-2}\cdot \ldots \cdot n_{w}\cdot n_{w-1}}$
$\vdots$ $\vdots$
$\frac{\sigma_{k}^2}{n_{p-1}\cdot n_{p-2}\cdot \ldots \cdot n_{k}} = \frac{(1-\rho_{k,k+1})\cdot \rho_{k-1,k}\cdot \rho_{k-2,k-1}\cdot \ldots \cdot \rho_{v,v+1}\cdot \ldots \cdot \rho_{23}\cdot \rho_{12}}{n_{p-1}\cdot n_{p-2}\cdot \ldots \cdot n_{k}}$
$\vdots$ $\vdots$
$\frac{\sigma_{v+1}^2}{n_{p-1}\cdot n_{p-2}\cdot \ldots \cdot n_{v+1}} = \frac{(1-\rho_{v+1,v+2})\cdot \rho_{v,v+1}\cdot \rho_{v-1,v}\cdot \ldots \cdot \rho_{23}\cdot \rho_{12}}{n_{p-1}\cdot n_{p-2}\cdot \ldots \cdot n_{v}\cdot n_{v+1}}$
$\frac{\sigma_{v}^2}{n_{p-1}\cdot n_{p-2}\cdot \ldots \cdot n_{v}} = \frac{(1-\rho_{v,v+1})\cdot \rho_{v-1,v}\cdot \rho_{v-2,v-1}\cdot \ldots \cdot \rho_{23}\cdot \rho_{12}}{n_{p-1}\cdot n_{p-2}\cdot \ldots \cdot n_{v}}$
$\vdots$ $\vdots$
$\frac{\sigma_{2}^2}{n_{p-1}\cdot n_{p-2}\cdot \ldots \cdot n_{2}} = \frac{(1-\rho_{23})\cdot \rho_{12}}{n_{p-1}\cdot n_{p-2}\cdot \ldots \cdot n_{2}}$
$\frac{\sigma_{1}^2}{n_{p-1}\cdot n_{p-2}\cdot \ldots \cdot n_{2}\cdot n_{1}} = \frac{(1-\rho_{12})}{n_{p-1}\cdot n_{p-2}\cdot \ldots \cdot n_{2}\cdot n_{1}}$

The idea of the proof is that only the terms containing σ_i^2 with $i \geq u$ have a non-zero derivative with respect to $\frac{\partial \rho}{\partial \rho_{u,u+1}}$. However, these terms could occur only in τ^2, or in both τ^2 and σ^2.

We now distinguish three cases: $u > k, u = k, u < k.$

Case $u > k$.
Then $p \geq k + 2$ (because $u \leq p - 1$ and $u > k$) and u is like w in the table above, so writing w instead of u:

$$\tau^2 = A\rho_{w,w+1} + B(1 - \rho_{w,w+1}) + C, \qquad \sigma^2 = D$$

where A, B, C, D are ≥ 0 and independent of $\rho_{w,w+1}$, because they are made of $\rho_{w',w'+1}, n_i$ with $w' \neq u$. To be specific:

$$A = \left[\sigma_p^2 + \frac{\sigma_{p-1}^2}{n_{p-1}} + .. + \frac{\sigma_{w+1}^2}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_{w+1}}\right] / \rho_{w,w+1} =$$

$$= \rho_{p-1,p} \cdot \rho_{p-2,p-1} \cdot \ldots \cdot \cancel{\rho_{w,w+1}} \cdot \ldots \cdot \rho_{v,v+1} \cdot \ldots \cdot \rho_{23} \cdot \rho_{12}$$

$$+ \frac{(1 - \rho_{p-1,p}) \cdot \rho_{p-2,p-1} \cdot \ldots \cdot \cancel{\rho_{w,w+1}} \cdot \ldots \cdot \rho_{v,v+1} \cdot \ldots \cdot \rho_{23} \cdot \rho_{12}}{n_{p-1}} + \cdots$$

$$+ \frac{(1 - \rho_{w+1,w+2}) \cdot \cancel{\rho_{w,w+1}} \cdot \rho_{w-1,w} \cdot \ldots \cdot \rho_{v,v+1} \cdot \ldots \cdot \rho_{23} \cdot \rho_{12}}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_{w+1}}$$

$$B = \frac{\sigma_w^2}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_w} / (1 - \rho_{w,w+1}) = \frac{\cancel{(1 - \rho_{w,w+1})} \cdot \rho_{w-1,w} \cdot \rho_{w-2,w-1} \cdot \ldots \cdot \rho_v \cdot \ldots \cdot \rho_{23} \cdot \rho_{12}}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_w}$$

And

$$C = \left[\frac{\sigma_{w-1}^2}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_w \cdot n_{w-1}} + \cdots + \frac{\sigma_k^2}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_k}\right]$$

$$= \frac{(1 - \rho_{w-1,w}) \cdot \rho_{w-2,w-1} \cdot \rho_{w-3,w-2} \cdot \ldots \cdot \rho_{v,v+1} \cdot \ldots \cdot \rho_{23} \cdot \rho_{12}}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_w \cdot n_{w-1}} + \cdots$$

$$+ \frac{(1 - \rho_{k,k+1}) \cdot \rho_{k-1,k} \cdot \rho_{k-2,k-1} \cdot \ldots \cdot \rho_{v,v+1} \cdot \ldots \cdot \rho_{23} \cdot \rho_{12}}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_k}$$

And

$$D = \frac{\sigma_{k-1}^2}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_k \cdot n_{k-1}} + \cdots + \frac{\sigma_2^2}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_2} + \frac{\sigma_1^2}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_1}$$

$$= \frac{(1 - \rho_{k-1,k}) \cdot \rho_{k-2,k-1} \cdot \rho_{k-3,k-2} \cdot \ldots \cdot \rho_{v,v+1} \cdot \ldots \cdot \rho_{23} \cdot \rho_{12}}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_w \cdot n_{w-1}} + \cdots +$$

$$\frac{(1 - \rho_{23}) \cdot \rho_{12}}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_2} + \frac{(1 - \rho_{12})}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_1}$$

Note that $A \geq B$, because

$$A = \left[\sigma_p^2 + \frac{\sigma_{p-1}^2}{n_{p-1}} + .. + \frac{\sigma_{w+1}^2}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_{w+1}}\right] / \rho_{w,w+1} \geq \frac{\sigma_p^2 + \sigma_{p-1}^2 + \cdots + \sigma_{w+1}^2}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_{w+1}} / \rho_{w,w+1} =$$

$$\frac{\sigma_p^2 + \sigma_{p-1}^2 + \cdots + \sigma_{w+1}^2}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_{w+1}} / \frac{\sigma_p^2 + \sigma_{p-1}^2 + \cdots + \sigma_{w+1}^2}{\sigma_p^2 + \sigma_{p-1}^2 + \cdots + \sigma_{w+1}^2 + \sigma_w^2} = \frac{\sigma_p^2 + \sigma_{p-1}^2 + \cdots + \sigma_{w+1}^2 + \sigma_w^2}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_{w+1}}$$

While

$$B = \frac{\sigma_w^2}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_w} / \left(1 - \rho_{w,w+1}\right) = \frac{\sigma_w^2}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_{w+1} \cdot n_w} / \frac{\sigma_w^2}{\sigma_p^2 + \sigma_{p-1}^2 + \cdots + \sigma_w^2}$$

$$\frac{\sigma_p^2 + \sigma_{p-1}^2 + \cdots + \sigma_w^2}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_{w+1} \cdot n_w} = \frac{1}{n_w} \cdot \frac{\sigma_p^2 + \sigma_{p-1}^2 + \cdots + \sigma_w^2}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_{w+1}}$$

So $B \leq A/n_w$.

Now

$$\frac{\partial \tau^2}{\partial \rho_{w,w+1}} = A - B, \quad \frac{\partial \sigma^2}{\partial \rho_{w,w+1}} = 0$$

So

$$\frac{\partial \tau^2}{\partial \rho_{w,w+1}} \cdot \sigma^2 = (A - B)D \geq 0 = \frac{\partial \sigma^2}{\partial \rho_{w,w+1}} \cdot \tau^2$$

<u>*Case $u = k$.*</u>

Here two subcases can arise:

If $k = p - 1$, then

$$\tau^2 = A\rho_{p-1,p} \quad , \qquad \sigma^2 = B\left(1 - \rho_{p-1,p}\right) + C$$

with

$A = \sigma_p^2 = \cancel{\rho_{p-1,p}} \cdot \rho_{p-2,p-1} \cdot \ldots \cdot \rho_{23} \cdot \rho_{12}$

$$B = \frac{\sigma_{p-1}^2}{n_{p-1}} = \frac{\cancel{\left(1 - \rho_{p-1,p}\right)} \cdot \rho_{p-2,p-1} \cdot \ldots \cdot \rho_{23} \cdot \rho_{12}}{n_{p-1}}$$

$$C = \frac{\sigma_{p-2}^2}{n_{p-1} \cdot n_{p-2}} + \cdots + \frac{\sigma_2^2}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_2} + \frac{\sigma_1^2}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_1}$$

$$= \frac{(1 - \rho_{p-2,p-1)} \cdot \rho_{k-3,k-2} \cdot \rho_{k-4,k-3} \cdot \ldots \cdot \rho_{23} \cdot \rho_{12}}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_w \cdot n_{w-1}} + \cdots +$$

$$\frac{(1 - \rho_{23}) \cdot \rho_{12}}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_2} + \frac{(1 - \rho_{12})}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_1}$$

Clearly,

$$\frac{\partial \tau^2}{\partial \rho_{p-1,p}} = A, \quad \frac{\partial \sigma^2}{\partial \rho_{w,w+1}} = -B,$$

so

$$\frac{\partial \tau^2}{\partial \rho_{p-1,p}} \cdot \sigma^2 = AB\left(1 - \rho_{p-1,p}\right) + AC \geq 0 \geq -BA\rho_{p-1,p} = \frac{\partial \sigma^2}{\partial \rho_{p-1,p}} \cdot \tau^2$$

If $k \leq p - 2$,

Then

$$\tau^2 = A\rho_{k,k+1} + B\left(1 - \rho_{k,k+1}\right), \qquad \sigma^2 = D$$

with A, B, C are ≥ 0 and independent of $\rho_{k,k+1}$. Indeed,

$$A = \left[\sigma_p^2 + \frac{\sigma_{p-1}^2}{n_{p-1}} + .. + \frac{\sigma_{k+1}^2}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_{k+1}}\right] / \rho_{k,k+1} =$$

$$= \rho_{p-1,p} \cdot \rho_{p-2,p-1} \cdot \ldots \cdot \cancel{\rho_{k,k+1}} \cdot \ldots \cdot \rho_{v,v+1} \cdot \ldots \cdot \rho_{23} \cdot \rho_{12}$$

$$+ \frac{(1 - \rho_{p-1,p}) \cdot \rho_{p-2,p-1} \cdot \ldots \cdot \cancel{\rho_{k,k+1}} \cdot \ldots \cdot \rho_{v,v+1} \cdot \ldots \cdot \rho_{23} \cdot \rho_{12}}{n_{p-1}} + \cdots$$

$$+ \frac{(1 - \rho_{k+1,k+2)} \cdot \cancel{\rho_{k,k+1}} \cdot \rho_{k-1,k} \cdot \ldots \cdot \rho_{v,v+1} \cdot \ldots \cdot \rho_{23} \cdot \rho_{12}}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_{k+1}}$$

$$B = \frac{\sigma_k^2}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_k} / (1 - \rho_{k,k+1}) = \frac{\cancel{(1 - \rho_{k,k+1})} \cdot \rho_{k-1,k} \cdot \rho_{k-2,k-1} \cdot \ldots \cdot \rho_v \cdot \ldots . \cdot \rho_{23} \cdot \rho_{12}}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_k}$$

$$D = \frac{\sigma_{k-1}^2}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_k \cdot n_{k-1}} + \cdots + \frac{\sigma_2^2}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_2} + \frac{\sigma_1^2}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_1}$$

$$= \frac{(1 - \rho_{k-1,k)} \cdot \rho_{k-2,k-1} \cdot \rho_{k-3,k-2} \cdot \ldots \cdot \rho_{v,v+1} \cdot \ldots . \cdot \rho_{23} \cdot \rho_{12}}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_w \cdot n_{w-1}} + \cdots +$$

$$\frac{(1 - \rho_{23}) \cdot \rho_{12}}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_2} + \frac{(1 - \rho_{12})}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_1}$$

And exactly as for the *Case $u > k$,* we can prove $A \geq B$ and thus

$$\frac{\partial \tau^2}{\partial \rho_{k,k+1}} \cdot \sigma^2 = (A - B)D \geq 0 = \frac{\partial \sigma^2}{\partial \rho_{k,k+1}} \cdot \tau^2$$

Case $u < k$.
We have again two subcases:

If $u = k - 1$, then

$$\tau^2 = A\rho_{k-1,k} \ , \qquad \sigma^2 = B(1 - \rho_{k-1,k}) + C$$

$$A = \left[\sigma_p^2 + \frac{\sigma_{p-1}^2}{n_{p-1}} + .. + \frac{\sigma_k^2}{n_{p-1} \cdot n_{p-2} \cdot \ldots \cdot n_k}\right] / \rho_{k-1,k} =$$

$$= \rho_{p-1,p} \cdot \rho_{p-2,p-1} \cdot \ldots \cdot \cancel{\rho_{k-1,k}} \cdot \ldots \cdot \rho_{23} \cdot \rho_{12}$$

$$+ \frac{(1 - \rho_{p-1,p}) \cdot \rho_{p-2,p-1} \cdot \ldots \cdot \cancel{\rho_{k-1,k}} \cdot \ldots \ldots \cdot \rho_{23} \cdot \rho_{12}}{n_{p-1}} + \cdots$$

$$+\frac{(1-\rho_{k,k+1}\cdot\cancel{\rho_{k-1,k}}\cdot\rho_{k-2,k-1}\cdot\ldots\cdot\rho_{23}\cdot\rho_{12}}{n_{p-1}\cdot n_{p-2}\cdot\ldots\cdot n_k}$$

$$B=\frac{\sigma_{k-1}^2}{n_{p-1}\cdot n_{p-2}\cdot\ldots\cdot n_k\cdot n_{k-1}}/(1-\rho_{k-1,k})\ =\frac{\cancel{(1-\rho_{k-1,k})}\cdot\rho_{k-2,k-1}\cdot\rho_{k-3,k-2}\cdot\ldots\cdot\rho_{23}\cdot\rho_{12}}{n_{p-1}\cdot n_{p-2}\cdot\ldots\cdot n_v\cdot n_{k-1}}$$

$$C=\frac{\sigma_{k-2}^2}{n_{p-1}\cdot n_{p-2}\cdot\ldots\cdot n_{k-1}\cdot n_{k-2}}+\cdots+\frac{\sigma_2^2}{n_{p-1}\cdot n_{p-2}\cdot\ldots\cdot n_2}+\frac{\sigma_1^2}{n_{p-1}\cdot n_{p-2}\cdot\ldots\cdot n_1}=$$

$$=\frac{(1-\rho_{k-2,k-1)}\cdot\rho_{k-3,k-2}\cdot\rho_{k-4,k-3}\cdot\ldots\cdot\rho_{23}\cdot\rho_{12}}{n_{p-1}\cdot n_{p-2}\cdot\ldots\cdot n_{k-1}\cdot n_{k-2}}+\cdots+$$

$$\frac{(1-\rho_{23})\cdot\rho_{12}}{n_{p-1}\cdot n_{p-2}\cdot\ldots\cdot n_2}+\frac{(1-\rho_{12})}{n_{p-1}\cdot n_{p-2}\cdot\ldots\cdot n_1}$$

So:

$$\frac{\partial\tau^2}{\partial\rho_{k-1,k}}\cdot\sigma^2=AB\big(1-\rho_{k-1,k}\big)+AC\geq 0\geq -BA\rho_{k-1,k}=\frac{\partial\sigma^2}{\partial\rho_{k-1,k}}\cdot\tau^2$$

If $u\leq k-2$, then u is like v in the table, so

$$\tau^2=A'\rho_{v,v+1}\,,\qquad \sigma^2=A\rho_{v,v+1}+B(1-\rho_{v,v+1})+C$$

With

$$A'=\left[\sigma_p^2+\frac{\sigma_{p-1}^2}{n_{p-1}}+\cdots+\frac{\sigma_k^2}{n_{p-1}\cdot n_{p-2}\cdot\ldots\cdot n_k}\right]/\rho_{v,v+1}=$$

$$=\rho_{p-1,p}\cdot\rho_{p-2,p-1}\cdot\ldots\cdot\cancel{\rho_{v,v+1}}\cdot\ldots\cdot\rho_{23}\cdot\rho_{12}+\frac{\big(1-\rho_{p-1,p}\big)\cdot\rho_{p-2,p-1}\cdot\ldots\cdot\cancel{\rho_{v,v+1}}\cdot\ldots\cdot\rho_{23}\cdot\rho_{12}}{n_{p-1}}$$

$$+..+\frac{(1-\rho_{k,k+1)}\cdot\rho_{k-1,k}\cdot\rho_{k-2,k-1}\cdot\ldots\cdot\cancel{\rho_{v,v+1}}\cdot\ldots\cdot\rho_{23}\cdot\rho_{12}}{n_{p-1}\cdot n_{p-2}\cdot\ldots\cdot n_k}$$

$$A=\left[\frac{\sigma_{k-1}^2}{n_{p-1}\cdot n_{p-2}\cdot\ldots\cdot n_{k-1}}+\cdots+\frac{\sigma_{v+1}^2}{n_{p-1}\cdot n_{p-2}\cdot\ldots\cdot n_{v+1}}\right]/\rho_{v,v+1}=$$

$$=\frac{(1-\rho_{k-1,k-2})\cdot\rho_{k-2,k-1}\cdot\rho_{k-3,k-2}\cdot\ldots\cdot\cancel{\rho_{v,v+1}}\cdot\ldots\cdot\rho_{23}\cdot\rho_{12}}{n_{p-1}\cdot n_{p-2}\cdot\ldots\cdot n_k\cdot n_{k-1}}+\cdots$$
$$\ldots+\frac{(1-\rho_{v+1,v+2)}\cdot\cancel{\rho_{v,v+1}}\cdot\rho_{v-1,v}\cdot\ldots\cdot\rho_{23}\cdot\rho_{12}}{n_{p-1}\cdot n_{p-2}\cdot\ldots\cdot n_v\cdot n_{v+1}}$$

$$B=\frac{\sigma_v^2}{n_{p-1}\cdot n_{p-2}\cdot\ldots\cdot n_v}/\,(1-\rho_{v,v+1})=\frac{\cancel{(1-\rho_{v,v+1})}\cdot\rho_{v-1,v}\cdot\rho_{v-2,v-1}\cdot\ldots\cdot\rho_{23}\cdot\rho_{12}}{n_{p-1}\cdot n_{p-2}\cdot\ldots\cdot n_v}$$

$$C=\frac{\sigma_{v-1}^2}{n_{p-1}\cdot n_{p-2}\cdot\ldots\cdot n_{v-1}}+\cdots+\frac{\sigma_2^2}{n_{p-1}\cdot n_{p-2}\cdot\ldots\cdot n_2}+\frac{\sigma_1^2}{n_{p-1}\cdot n_{p-2}\cdot\ldots\cdot n_2\cdot n_1}$$

$$= \frac{(1-\rho_{v-1,v})\cdot\rho_{v-2,v-1}\cdot\rho_{v-3,v-2}\cdot\ldots\cdot\rho_{23}\cdot\rho_{12}}{n_{p-1}\cdot n_{p-2}\cdot\ldots\cdot n_{v-1}} + \cdots + \frac{(1-\rho_{23})\cdot\rho_{12}}{n_{p-1}\cdot n_{p-2}\cdot\ldots\cdot n_2}$$
$$\ldots + \frac{(1-\rho_{12})}{n_{p-1}\cdot n_{p-2}\cdot\ldots\cdot n_2\cdot n_1}$$

Then

$$\frac{\partial\tau^2}{\partial\rho_{k-1,k}}\cdot\sigma^2 - \frac{\partial\sigma^2}{\partial\rho_{k-1,k}}\cdot\tau^2 = A'\left[A\rho_{v,v+1} + B(1-\rho_{v,v+1}) + C\right] - (A-B)\left(A'\rho_{v,v+1}\right) =$$
$$= A'B + A'C \geq 0$$

End proof.

8 Derivation of VIFs (variance inflation factors) compared to t-test and cluster randomized parallel group post-test (i.e. one measurement) design

Recall that in the Hussey and Hughes formulation

$$var(\hat{\delta}\,) = \frac{I\cdot(1-\rho)\cdot[1+(T-1)\rho]}{f(I,X)\cdot(1-\rho) + g(I,X)\cdot\rho}\cdot var(Y_{it\bullet})$$

regardless the type of repeated measurements design (e.g., parallel group, stepped wedge, or cross-sectional) and

$$var(Y_{it\cdot}) = [1+(n_1-1)\rho_{12}]\cdot[1+(n_2-1)\tilde{\rho}_{23}]\cdot\ \ldots\ldots\cdot\ \left[1+\left(n_{p-1}-1\right)\tilde{\rho}_{p-1,p}\right]\cdot\frac{\sigma_{tot}^2}{n_1 n_2\cdot\ldots\cdot n_{p-1}}$$
$$= VIF_p\cdot\frac{\sigma_{tot}^2}{n_1 n_2\cdot\ldots\cdot n_{p-1}}$$

Note that , $\frac{4}{I}\cdot var(Y_{it\bullet}) = VIF_p\cdot\frac{4\sigma_{tot}^2}{n_1 n_2\cdot\ldots\cdot n_{p-1}\cdot I} = VIF_p\cdot\frac{2\sigma_{tot}^2}{(N_{tot}/2)}$ and the latter factor is just the variance of a t-test (z-test) with two arms of size $N_{tot}/2$ (the total sample size is N_{tot}) and an outcome measure with variance σ_{tot}^2. Thus

$$var(Y_{it\bullet}) = \frac{I}{4}\cdot VIF_p\cdot var_{t-test}$$

In particular, for a post-test parallel arm *cluster* randomized trial: $T=1,\ \ S=\frac{I}{2},\ \ C=\frac{I^2}{4},\ \ R=\frac{I}{2},$ so that $f=\frac{I^2}{4}, g=0$, and thus

$$var\left(\hat{\delta}_{cPG1}\right) = \frac{4}{I}\cdot var(Y_{it\bullet}) = VIF_p\cdot var_{t-test}.$$

which extends the three level VIF in Teerenstra *et al.*, Clinical Trials 2008: 486-495.

More generally, we get for any repeated measures design:

$$var(\hat{\delta}) = \frac{I \cdot (1-\rho) \cdot [1+(T-1)\rho]}{f(I,X) \cdot (1-\rho) + g(I,X) \cdot \rho} \cdot \frac{I}{4} \cdot VIF_p \cdot var_{t-test} =$$

$$= \underbrace{\frac{(I^2/4) \cdot (1-\rho) \cdot [1+(T-1)\rho]}{f(I,X) \cdot (1-\rho) + g(I,X) \cdot \rho}}_{VIF(design:\, cPG1)} \cdot VIF_p \cdot var_{t-test}$$

where the first factor is the variance inflation due to the repeated measures design (with respect to a **c**luster randomized **p**arallel **g**roup design and **1** measurement) and the second the variance inflation due to the multilevel structure at each of measurements.

In terms of the Girling and Hemming formulation

$$var(\hat{\delta}) = \frac{(1-\rho)}{I \cdot T \cdot (a_D(X) - b_D(X) \cdot R)} \cdot var(Y_{it\bullet}) =$$
$$= \frac{(1-\rho)}{I \cdot T \cdot (a_D(X) - b_D(X) \cdot R)} \cdot \left(\frac{I}{4} \cdot VIF_p \cdot var_{t-test}\right)$$

So

$$var(\hat{\delta}) = \underbrace{\frac{(1-\rho)}{T \cdot (4a_D(X) - 4b_D(X) \cdot R)}}_{VIF(design:\, cPG1)} \cdot VIF_p \cdot var_{t-test}$$

8.1 VIF standard stepped wedge

In particular, for the standard stepped wedge design with the **number of steps denoted by k (instead of s) and the number of clusters per sequence denoted by i (instead of c)**, we have as derived earlier:

$$f(X) = (ik)^2 \cdot \frac{\left(k - \frac{1}{k}\right)}{6} \quad , \quad g(X) = (ik)^2 \cdot \frac{(k^2-1)/2}{6} - f(i,k)$$

So that

$$VIF_{SW:cRCT} = \frac{1}{4} \cdot \frac{I^2 \cdot (1-\rho) \cdot [1+(T-1)\rho]}{f(I,X) \cdot (1-\rho) + g(I,X) \cdot \rho}$$

$$= \frac{(ik)^2/4 \cdot (1-\rho) \cdot [1+k\rho]}{(ik)^2 \cdot \frac{(k-1/k)}{6} + (ik)^2 \cdot \frac{(k^2-1)/2}{6} \cdot \rho} = \frac{3}{2} \cdot \frac{(1-\rho) \cdot [1+k\rho]}{(k-1/k) + (k^2-1)/2 \cdot \rho}$$

$$= \frac{3}{2} \cdot \frac{(1-\rho) \cdot [1+k\rho]}{\left(k - \frac{1}{k}\right) \cdot \left[1 + \frac{k}{2}\rho\right]}$$

Note that $VIF_{SW} = VIF_{SW:cRCT} \cdot VIF_p$ corresponds to the Woertman&De Hoop design factor

$$DE_{SW} = \frac{1 + ICC \cdot (ktN + bN - 1)}{1 + ICC \cdot (\frac{ktN_1}{2} + bN - 1)} \cdot \frac{3(1 - ICC)}{2t(k - \frac{1}{k})}$$

for the standard stepped wedge, i.e. $b = 1, t = 1$, so one measurement at baseline and after each step) after identification of $N = n_1,\ \sigma_t^2 = \sigma_{tot}^2,\ ik = I, k = T - 1 = s$ the number of steps, and $ICC = \rho_{12}$.

Proof. In the 2-level case, $\rho = corr(Y_{it\bullet}, Y_{is\bullet}) = \frac{\sigma_2^2}{\sigma_2^2 + \frac{\sigma_1^2}{n_1}} = \frac{n_1 \rho_{12}}{d}$, where $\rho_{12} = \sigma_2^2/(\sigma_1^2 + \sigma_2^2)$ is the icc (ratio of level 2 variance to total variance of level 1 and 2) and use $d := 1 + (n_1 - 1)\rho_{12}$ as short-cut. Note that $d = VIF_2$.
Then with k as the number of steps:
$1 - \rho = \frac{1-\rho_{12}}{d}$, $\quad 1 + k\rho = [1 + (kn_1 + n_1 - 1)\rho_{12}]/d$,

$$1 + \frac{k}{2}\rho = \left[1 + \left(\frac{k}{2}n_1 + n_1 - 1\right)\rho_{12}\right]/d,$$

So that

$$VIF_{SW:cPG1} = VIF_{SW:cPG1} \cdot VIF_2 =$$

$$\frac{3}{2} \cdot \frac{(1-\rho) \cdot [1 + k\rho]}{\left(k - \frac{1}{k}\right) \cdot \left[1 + \frac{k}{2}\rho\right]} \cdot VIF_2 = \frac{3}{2} \cdot \frac{(1-\rho_{12})/d \cdot [1 + (kn_1 + n_1 - 1)\rho_{12}]/d}{\left(k - \frac{1}{k}\right) \cdot \left[1 + \left(\frac{k}{2}n_1 + n_1 - 1\right)\rho_{12}\right]/d} \cdot d$$

$$= \frac{3}{2} \cdot \frac{(1-\rho_{12}) \cdot [1 + (kn_1 + n_1 - 1)\rho_{12}]}{\left(k - \frac{1}{k}\right) \cdot \left[1 + \left(\frac{k}{2}n_1 + n_1 - 1\right)\rho_{12}\right]}$$

which is the Woertman-De Hoop result. *End proof.*

8.2 VIF for stepped wedge with more/fewer/no data collection at baseline and/or at final step

As

$$f = \frac{c^2}{6} \cdot s \cdot (s^2 - 1)\ , \quad g = f \cdot \left(b + \frac{s}{2} - 1 + a\right)\ , \quad T = b + s - 1 + a\ , \quad I = c \cdot s$$

$$VIF_{SW(a,b):cPG1} = \frac{(I^2/4) \cdot (1-\rho) \cdot [1 + (T-1)\rho]}{f \cdot (1-\rho) + g \cdot \rho} = \frac{(I^2/4) \cdot (1-\rho) \cdot [1 + (T-1)\rho]}{f + (g-f) \cdot \rho}$$
$$= \frac{I^2}{4f} \cdot \frac{(1-\rho) \cdot [1 + (b + s - 2 + a)\rho]}{\left(1 + \left[b + \frac{s}{2} - 2 + a\right]\rho\right)} = \frac{3}{2} \cdot \frac{(1-\rho) \cdot (1 + [a + b - 2 + s] \cdot \rho)}{\left(s - \frac{1}{s}\right) \cdot \left(1 + \left[a + b - 2 + \frac{s}{2}\right] \cdot \rho\right)}\ .$$

8.3 VIF for the hybrid design

From Table 1 in Girling and Hemming we have that in their notation g for the number of sequences (so our s):

$$4 \cdot a_D = 1 - \frac{\beta^2}{3} \cdot \left(1 + \frac{2}{g^2}\right) \quad , \quad 4 \cdot b_D = 1 - \frac{\beta}{3} \cdot \left(2 + \frac{1}{g^2}\right)$$

So

$$VIF_{H(\beta,s):cPG1} = \frac{(1-\rho)}{T \cdot (4a_D(X) - 4b_D(X) \cdot R\,)} =$$

$$= \frac{(1-\rho)}{T} \cdot \frac{1}{1 - \frac{\beta^2}{3} \cdot \left(1 + \frac{2}{g^2}\right) + R \cdot \left(1 - \frac{\beta}{3} \cdot \left(2 + \frac{1}{g^2}\right)\right)} \;.$$

9 Binary and rate outcomes

If we take a two-level design and a binary outcome as an example, we can model the trial hierarchically as follows. Each subject j in cluster i has an binary outcome B_{ij} that is 1 with probability p_i, when cluster i is in the control condition, and with probability $p_i + \delta$, when cluster i is in the invention condition. The probabilities p_i vary over the clusters according to some distribution with mean μ and variance s_c^2. Then the average proportion in the control condition can be calculated by conditioning on the cluster (propability p_i):

$$E(B_{ij}) = E(E(B_{ij}|p_i) = E(p_i|p_i) = \mu$$

and the same can be done to calculate the variance of Y_{ij}:

$$var(B_{ij}) = E\left(var(B_{ij}|p_i)\right) + var\left(E(B_{ij}|p_i)\right)$$

where the first term can be calculated using the equality $var(Y) = E(Y^2) - [E(Y)]^2$:

$$E(p_i(1-p_i)|p_i) = E(p_i - p_i^2|p_i) = E(p_i|p_i) - E(p_i^2|p_i) = \mu - \{E(var(p_i|p_i) + [E(p_i|p_i)]^2\}$$
$$= \mu - \{s_c^2 + \mu^2\} = \mu(1-\mu) - s_c^2$$

and can be seen as the within-cluster variance.
The second term

$$var(p_i) = s_c^2 = \sigma_2^2$$

can be seen as the between-cluster variance.
For three levels, the level 1 units B_{ijk} (in the control condition) are binomially distributed with probabilities p_{ij} that dependent on the level 2 unit j and level 3 unit i that the B_{ijk} comes from. When we think hierarchically, the true proportion p_i of level 3 unit i comes from a distribution with mean μ and variance σ_3^2 and the true proportion p_{ij} of the level 2 unit j (from that level 3 unit i) comes from a distribution with mean p_i and variance σ_2^2. Thus, the p_{ij} have mean μ and variance $s_c^2 = \sigma_3^2 + \sigma_2^2$. So we can apply the argument of above on the clustering of level 1 units within their level 2 units with variance $s_c^2 = \sigma_3^2 + \sigma_2^2$. This can be generalized to $p > 3$.
For a rate outcome, we have similarly in a two-level design that the rate outcome R_{ij} of subject j in cluster i is from a Poisson distribution with mean λ_i that varies over clusters with mean λ and variance s_c^2. Then

$$E(R_{ij}) = E(R_{ij}|\lambda_i) = \lambda$$

and

$$var(R_{ij}) = E\left(var(R_{ij}|\lambda_i)\right) + var(E(R_{ij}|\lambda_i)$$

where the first term can be interpreted as the within-cluster variance

$$E(var(R_{ij}|\lambda_i) = E(\lambda_i|\lambda_i) = \lambda$$

and the second term is the between-cluster variance

$$var(E(R_{ij}|\lambda_i) = var(\lambda_i) = s_c^2$$

.

10 Programming of the sample size and power formulas in SAS® and Excel® for the standard stepped wedge design

The SAS® program uses the multiplicative formulas for the standard error and the sample size, i.e., ρ and VIF_p are expressed in terms of $\rho_{12}, \rho_{23}, \rho_{34}, \tilde{\rho}_{23}, \tilde{\rho}_{34}$ and n_1, n_2, n_3, n_4 as in the Table of the Appendix.
The Excel program calculates ρ from τ^2 and σ^2 after having calculated these from $\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2$ which in turn are calculated from the conversion formulas in the Table of the Appendix. For the 3-level case for example, we have

$$\rho = \frac{\tau^2}{\tau^2 + \sigma^2} = \frac{{\sigma_3}^2 + \frac{{\sigma_2}^2}{n_2}}{{\sigma_3}^2 + \frac{{\sigma_2}^2}{n_2} + \frac{{\sigma_1}^2}{n_1 n_2}},$$

In which we have, apart from a common factor σ_{tot}^2:

$$\sigma_3^2 = \rho_{23}\rho_{12}$$

$$\sigma_2^2 = (1 - \rho_{23})\rho_{12}$$

$$\sigma_1^2 = (1 - \rho_{12})$$

Furthermore, if we realize that that the multilevel variance inflation factor is the ratio of the variance of a cluster mean accounting for clustering compared to the variance of a mean of a cluster considered of totally independent units, we have

$$VIF_3 = \frac{{\sigma_3}^2 + \frac{{\sigma_2}^2}{n_2} + \frac{{\sigma_1}^2}{n_1 n_2}}{\left[\frac{\sigma_3^2 + \sigma_2^2 + \sigma_1^2}{n_2 n_1}\right]} \quad .$$

11 References

Thompson JA, Fielding K, Copas A. The optimal design of stepped wedge trials with equal allocation to sequences and a comparison to other trial designs. Clinical Trials 2016; 14: 639-647.

Hussey M, Hughes J. Design and analysis of stepped wedge cluster randomized trials. Contemp Clin Trials 2007; 28(182-191).Hussey and Hughes

Girling AJ, Hemming K. Statisitcal efficiency and optimal designs for stepped wedge cluster studies under linear mixed effects models. Statistics in Medicine 2016; 35: 2149-2166.