Supplementary materials

A Details of mean score variance

The sandwich variance (6) is computed from $\mathbf{B}$ and $\mathbf{C}$, where $\mathbf{B}$ has components

$$\begin{aligned}
\mathbf{B}_{SS} &= -d\mathbf{U}_S/d\boldsymbol{\beta}_S &&= \sum_i h'(\hat{\boldsymbol{\beta}}_S \mathbf{x}_{Si})\mathbf{x}_{Si}\mathbf{x}_{Si}^T \\
\mathbf{B}_{SP} &= -d\mathbf{U}_S/d\boldsymbol{\beta}_P &&= -\sum_i (1-r_i)h'(\hat{\boldsymbol{\beta}}_P \mathbf{x}_{Pi} + \Delta_i)\mathbf{x}_{Si}\mathbf{x}_{Pi}^T \\
\mathbf{B}_{PS} &= -d\mathbf{U}_P/d\boldsymbol{\beta}_S &&= \mathbf{0} \\
\mathbf{B}_{PP} &= -d\mathbf{U}_P/d\boldsymbol{\beta}_P &&= \sum_i r_i h'(\hat{\boldsymbol{\beta}}_P \mathbf{x}_{Pi})\mathbf{x}_{Pi}\mathbf{x}_{Pi}^T
\end{aligned}$$

and $\mathbf{C}$ has components

$$\begin{aligned}
\mathbf{C}_{SS} &= \sum_i e_{Si}^2 \mathbf{x}_{Si}\mathbf{x}_{Si}^T \\
\mathbf{C}_{SP} = \mathbf{C}_{PS}^T &= \sum_i e_{Si}e_{Pi}\mathbf{x}_{Si}\mathbf{x}_{Pi}^T \\
\mathbf{C}_{PP} &= \sum_i e_{Pi}^2 \mathbf{x}_{Pi}\mathbf{x}_{Pi}^T
\end{aligned}$$

where $e_{Si} = y_i^*(\hat{\boldsymbol{\beta}}_P) - h(\hat{\boldsymbol{\beta}}_S^T \mathbf{x}_{Si})$ and $e_{Pi} = r_i\{y_i - h(\hat{\boldsymbol{\beta}}_P^T \mathbf{x}_{Pi})\}$.

B Modifications for clustered data

If data are clustered, as in a cluster-randomised trial, we need to modify the variance calculations in Sections 2.1 and 2.2 and the small-sample corrections in Sections 3.1 and 3.2. Let m be the total number of clusters, m_{obs} be the number of clusters with at least one observed outcome, and m_{mis} =

$m - m_{obs}$ be the number of clusters with no observed outcome. Let the data be subscripted by cluster membership $c = 1, \ldots, m$ as well as individual i.

For the full sandwich variance method of Section 2.1, we only need to redefine the matrix $\mathbf{C} = \sum_c \mathbf{U}_c(\hat{\boldsymbol{\beta}})\mathbf{U}_c(\hat{\boldsymbol{\beta}})^T$ where $\mathbf{U}_c(\hat{\boldsymbol{\beta}}) = \sum_i \mathbf{U}_{ci}(\hat{\boldsymbol{\beta}})$ (Rogers, 1993).

For the two linear regressions method of Section 2.2, we similarly take $\mathrm{var}\left(\hat{\boldsymbol{\beta}}_P\right)$ and $\mathrm{var}\left(\hat{\boldsymbol{\beta}}_S - \hat{\boldsymbol{\beta}}_P\right)$ as clustered sandwich variances.

For the small-sample methods of Section 3, we assume the standard methods use a small-sample correction factor $f = \frac{n-1}{n-p^*}\frac{m}{m-1}$, and use $m-1$ degrees of freedom for linear regression (StataCorp, 2011). We replace n and m by n_{eff} and m_{eff}, calculated by the two methods explained below.

For the full sandwich variance method, we compute n_{eff} as in Section 3.1, and compute $m_{eff} = m_{obs} + (I_{mis}/I_{mis^*})m_{mis}$.

For the two linear regressions method, the variance with small-sample correction is (as before) $\widehat{\mathrm{var}}\left(\hat{\boldsymbol{\beta}}_P\right) + \widehat{\mathrm{var}}\left(\hat{\boldsymbol{\beta}}_S - \hat{\boldsymbol{\beta}}_P\right) = V_{small}$. The corresponding variance without small-sample correction is $\frac{n_{obs}-p}{n_{obs}-1}\frac{m_{obs}-1}{m_{obs}}\widehat{\mathrm{var}}\left(\hat{\boldsymbol{\beta}}_P\right) + \frac{n-p}{n-1}\frac{m-1}{m}\widehat{\mathrm{var}}\left(\hat{\boldsymbol{\beta}}_S - \hat{\boldsymbol{\beta}}_P\right) = V_{large}$. The heuristic $V_{small} \approx \frac{n_{eff}-1}{n_{eff}-p}\frac{m_{eff}}{m_{eff}-1}V_{large}$ leads to the equation $|V_{small}| = \left(\frac{n_{eff}}{n_{eff}-p}\right)^p |V_{large}|$. However, we have two unknowns n_{eff} and m_{eff}, so we take a second equation representing the variance with small-sample correction only for the number of clusters: $\frac{m_{obs}-1}{m_{obs}}\widehat{\mathrm{var}}\left(\hat{\boldsymbol{\beta}}_P\right) +$

$\frac{m-1}{m}\widehat{\mathrm{var}}\left(\hat{\boldsymbol{\beta}}_S - \hat{\boldsymbol{\beta}}_P\right) = V_{largen}$ say, with the heuristic $V_{small} \approx \frac{m_{eff}}{m_{eff}-1}V_{largen}$ and the second equation $|V_{small}| = \left(\frac{m_{eff}}{m_{eff}-1}\right)^p |V_{largen}|$. We solve the second equation for m_{eff} and then the first equation for n_{eff}.