Appendix

P-values, Z-scores and funnel plots

We assume an *indicator* Y with a *target* θ_0 which specifies the desired expectation, so that Exp$(Y \mid \theta_0) = \theta_0$. The target is assumed known and measured without error.

For each observation y we calculate a standard P-value

$$p_i = P(Y > y_i \mid \theta_0, \rho_i),$$

where ρ_i is a measure of measurement precision such as the sample size. For discrete Binomial and Poisson distributions p_i could be calculated more precisely as

$$p_i = P(Y > y_i \mid \theta_0, \rho_i) + \frac{1}{2} P(Y = y_i \mid \theta_0, \rho_i).$$

These P-values are then used to test the hypothesis that the trust is 'on-target', *i.e.* Exp$(Y) = \theta_0$.

For indicators with suspected over-dispersion, sample sizes will generally be large enough so that the indicator can be reported as y_i and s_i = estimated standard error of y_i, although a log-transformation may be adopted for small samples. The general definition of a Z-statistic is

$$z_i = \frac{y_i - \theta_0}{s_{0i}}, \qquad (1)$$

where s_{0i} = standard error of y_i given the trust is on target: hence $s_{0i} = \sqrt{\mathrm{Var}(\mathrm{Y} \mid \theta_0, \rho_\mathrm{i})}$. It is important to note that s_{0i} may not necessarily be the same as the reported s_i, and hence some care is required in calculating the Z-scores. For example, if y_i is an observed proportion between 0 and 1, then

$$s_i = \sqrt{\frac{y_i(1 - y_i)}{n_i}},$$

where n_i is the 'effective' sample size. s_i^0 can then be estimated to be

$$s_{0i} = \sqrt{\frac{\theta_0(1-\theta_0)}{n_i}}.$$

The difference between these two standard errors explains why the caterpillar plots and the funnel plots can give slightly different classifications - the correct hypothesis tests are based on the funnel plots.

Funnel plots simply indicate values of Y that would reject the null hypothesis of θ_0 using the appropriate P-value. For normal approximations, funnel plot limits $\theta_0 \pm z_p s_0$ are defined as a function of standard errors s_0, where z_p is the appropriate standard normal deviate.

Interval target

Suppose a target interval (θ_0^L, θ_0^H) is assumed as the Lowest and Highest acceptable rate: these might not necessarily be symmetric about the average, and thought should be given as to what is meant by both 'poor' and 'good' performance. Then for observations within this target, no P-value can be calculated, while for observations above the target interval, θ_0^H is used as the target, and similarly for below the interval. The funnel is therefore drawn at $\theta_0^L - z_p s_0, \theta_0^H + z_p s_0$.

Over-dispersion model

Following the standard approach of generalised linear modelling [13] we shall introduce an over-dispersion factor ϕ that will inflate the null variance, so that

$$\mathrm{Var}(Y \mid \theta_0, \rho, \phi) = \phi\ \mathrm{Var}_0(Y \mid \theta_0, \rho).$$

Suppose we have a sample of I units that we shall assume (for the present) all to be on-target. ϕ may be estimated as follows:

$$\hat{\phi} = \frac{1}{I} \sum_i z_i^2, \qquad (2)$$

where z_i is the standardised Pearson residual defined in (1). $I\hat{\phi}$ is a standard test of heterogeneity, and is distributed as a χ^2_{I-1} distribution under the null hypothesis that all are hitting an (estimated) target. Over-dispersion might only be assumed if $\hat{\phi}$ is 'significantly' greater than 1, although it would be more appropriate to avoid such a preliminary significance test to avoid a discontinuity in its use. The current control limits in the funnel plot can then be inflated by a factor $\sqrt{\hat{\phi}}$ around θ_0. For example, based on the approximate normal control limits, over-dispersed control limits can then be plotted as

$$\theta_0 \pm z_p \sqrt{\hat{\phi}} s_0, \qquad (3)$$

equivalent to creating a 'modified' Z-score $z_i / \sqrt{\hat{\phi}}$ and comparing to standard normal deviates.

Robust 'winsorised' Z-scores can be used in estimating ϕ - see below.

Random-effects model

This assumes that Exp$(Y_i) = \theta_i$, and that for 'on-target' trusts θ_i is distributed with mean θ_0 and standard deviation τ. τ can be estimated using a standard 'method of moments' [14]

$$\hat{\tau}^2 = \frac{I\hat{\phi} - (I-1)}{\sum_i w_i - \sum_k w_i^2 / \sum_i w_i} \qquad (4)$$

where $w_i = 1/s_i^2$, and $\hat{\phi}$ is the estimate of heterogeneity: if $\hat{\phi} < (I-1)/I$, then $\hat{\tau}^2$ is set to 0 and complete homogeneity is assumed. Otherwise the adjusted Z-scores are given by

$$z_i^D = \frac{y_i - \theta_0}{\sqrt{s_0^2 + \tau^2}},$$

and the funnel plot boundaries are given by

$$\theta_0 \pm z_p \sqrt{s_0^2 + \tau^2} : \quad (5)$$

a more refined procedure would base s_0^2 on the estimated rate under the random-effects model.

Winsorising Z-scores

Winsorising consists of shrinking in the extreme Z-scores to some selected percentile.

1. Rank cases according to their naive Z-scores.
2. Identify Z_q and Z_{1-q}, the $100q$% most extreme top and bottom naive Z-scores, where q might, for example, be 0.1.
3. Set the lowest $100q$% of Z-scores to Z_q, and the highest $100q$% of Z-scores to Z_{1-q}. These are the Winsorised statistics.

This retains the same number of Z-scores but discounts the influence of outliers. An alternative would be 'trimming', in which the extremes are discarded entirely. In theory an adjustment can be made to allow for this Winsorisation [9], but this is not carried out here.