Care Quality Commission

GP Insight

NHS GP Practices

Statistical Methodology

June 2017

Contents

1. Introduction

This document describes, in some detail, the statistical methods we have used to analyse the data that supports the monitoring of NHS GP practices as part of GP Insight.

Our general approach for the model is to assess variation by comparing a GP Practice's observed outcomes with the outcomes of others. Where appropriate, we account for the relative sizes of practices, for example, indicator GPHLIAP: 'Number of antibacterial prescription items prescribed per Specific Therapeutic group Age-sex Related Prescribing Unit (STAR PU)'.

We use a cross-sectional analysis, which assesses variation by comparing practice outcomes over a fixed period of time. Previous values or trends are not accounted for.

2. Analysis of cross-sectional data using z-scores

2.1 Z-scores

With cross-sectional data we measure the deviation of observed values from an expected or target value. Where we can transform the data into a standard normal distribution we generate z-scores which reflect the number of standard deviations.
Suppose the practice value for an indicator is y, and it has an expected or target value t, we can express the deviation of the indicator from the expected value as a z-score, defined as:

$$z = \frac{y - t}{s_0}$$

where s_0 is the standard deviation of y if the practice's observed outcomes were randomly distributed about t.

Here z is referred to as the unadjusted z-score. Under a null hypothesis that a practice's true level of outcomes is exactly the same as the expected value, z has mean 0 and standard deviation 1, and if we assume normality, then p-values 0.025 and 0.001 correspond to $z = \pm 1.96$ and $z = \pm 3.10$ respectively, which corresponds very closely to 2 and 3 standard deviations from the mean.

The default expected values against which a practice is compared are calculated by comparing rates observed for an individual practice against the mean rate of all practices included in GP Insight (sum of numerators for all practices in GP Insight / sum of denominators for all practices in GP Insight).
However, for some items we standardise by case mix (for example, by age and sex) in order to compare observed outcomes against what you would expect if the rate for each patient was the same as for similar patients over the whole country.
Often the raw data are not normally distributed, in which case we use one of the following appropriate transformations:

2.1.1 Z-scores from proportions

Assume an observed proportion $y = r/n$, with an expected or target proportion p. The observed proportion is transformed to render it more normally distributed by applying an arcsine transformation to the square root of the observed proportion:

$$Y = arcsin\sqrt{\frac{r}{n}}$$

The expected value can be approximated by:

$$T = arcsin\sqrt{p}$$

and the standard deviation *(s)* is approximated by:

$$s = \frac{1}{2\sqrt{n}}$$

Hence the transformed unadjusted z-score:

$$z = \frac{Y - T}{s} = 2\sqrt{n}\left(arcsin\sqrt{\frac{r}{n}} - \arcsin\sqrt{p}\right)$$

2.1.2 Z-scores from percentages

We assume an indicator that consists of an observed percentage p, where the numerator and denominator are not available.

There is no preliminary scaling for the percentage measure as there is no measure of accuracy for these items. Hence:

$Y = p$

$T = t$

Where t is either the mean percentage across all providers ($\bar{p}$) or the target percentage.

The standard deviation *(s^0)* is that of the percentages across all providers, given by:

$$s^0 = \sqrt{\frac{\sum p^2}{N} - \left(\frac{\sum p}{N}\right)^2}$$

Where N is the number of providers with indicator values.

Thus, the unadjusted Z-score for a provider is given by:

$$Z = \frac{(Y-T)}{s^0} = \frac{(p-t)}{\sqrt{\frac{\sum p^2}{N} - \left(\frac{\sum p}{N}\right)^2}} \qquad (5)$$

2.1.3 Z-scores from ratios of counts

We assume a ratio indicator of the form y = O_1/O_2, where O_1 and O_2 are both counts, and an average or target ratio *t*.

In order to deal with zero/low counts we add 0.5 to all observations, and, noting that a log transformation reduces positive skewness, the transformed indicator becomes:

$$Y = log_e\left(\frac{O_1 + 0.5}{O_2 + 0.5}\right)$$

with an expected value approximately equal to

$$\mathrm{T} = log_e(t)$$

and a standard deviation:

$$s = \sqrt{\frac{O_1}{(O_1 + 0.5)^2} + \frac{O_2}{(O_2 + 0.5)^2}}$$

Thus the transformed, unadjusted z-score becomes:

$$z = \frac{Y - T}{s} = \frac{log_e[(O_1 + 0.5)/(O_2 + 0.5)] - log_e(t)}{\sqrt{O_1/(O_1 + 0.5)^2 + O_2/(O_2 + 0.5)^2}}$$

If either O_1 or O_2 is much bigger than the other, say when one represents a population, it will have a negligible impact on the score.

2.2 Over-dispersion

Many z-scores are likely to be over-dispersed, that is their true variances are greater than one, which may be because of insufficient benchmarking or the presence of common-cause factors that render the Poisson model inadequate. The consequence is that analyses may pick up statistically significant differences that are not of practical importance. When considering an outcome based on an 'average' or 'expected' level, it may then be reasonable to accept as inevitable a degree of between-trust variability and we therefore seek to identify practices that deviate from this distribution, rather than deviating from a single measure. In order to do this we must estimate the degree of over-dispersion (see Section 2.2.2). When estimating over-dispersion it may be better to do so using techniques that avoid undue influence of outlying trusts, such as winsorisation (see Section 2.2.1).

The significance of observed deviations then takes into account both the precision with which the indicator is measured within each practice (i.e. the sample size), and the estimated between practice variability.

2.2.1 Winsorisation

Winsorisation is the process of transforming outliers in statistical data. In this context it involves shrinking in extreme unadjusted Z-scores to the value of a selected percentile. This is done by:

1. Ranking trusts according to their unadjusted Z-scores.
2. Identifying Z_q and Z_{1-q}, the 100*q*% most extreme high and low unadjusted Z-scores, where *q* may be, for example, 0.1.
3. Setting the lowest 100*q*% of unadjusted Z-scores to Z_q and the highest 100*q*% of Z-scores to Z_{1-q}. These are the winsorised statistics.

This process retains the same number of Z-scores, but protects our estimation of over-dispersion from the influence of actual outliers.

2.2.2 Estimating over-dispersion

In calculating an adjusted Z-score for an indicator, we estimate the over-dispersion factor phi *(ϕ)* as follows:

$$\hat{\phi} = \frac{1}{n}\sum_{i=1}^{n} \hat{z}_i^2$$

Where n is the number of provider for a data item and $\hat{z}_i$ is the winsorised z-score about the mean for the *i*th provider.

Under a null hypothesis that all units only exhibit random variability around the expected value, which is derived from the data, $n\hat{\phi}$ has an approximate χ^2_{n-1}distribution.This can therefore be used as a standard test of heterogeneity.

2.2.3 Calculating adjusted Z-scores

We then use the resulting over-dispersion factor to calculate an adjusted Z-score for each observation.

The over-dispersion model we use is an additive random effects model. This model assumes that each provider has its own true underlying level t_i, and that for 'on-target' provider t_i is distributed with mean t_0 and standard deviation, τ. In other words the null hypothesis is represented by a distribution rather than a single point A standard method of moments estimate for τ^2 is:

$$\hat{\tau}^2 = \frac{n\hat{\phi} - (n-1)}{\sum_{i=1}^{n} w_i - \left(\sum_{i=1}^{n} w_i^2 / \sum_{j=1}^{n} w_j\right)}$$

Where $w_i = 1/s_i^2$ and $n\hat{\phi}$ is the test for heterogeneity. (s_i is as calculated in section 2.1 with the appropriate transformation.)

If $n\hat{\phi} \geq (n-1)$ then the adjusted Z-scores are given by:

$$z_i^* = \frac{y_i - t_0}{\sqrt{s_i^2 + \hat{\tau}^2}}$$

Where y_i is equal to the transformed observed value.

Otherwise, if $n\hat{\phi} < (n-1)$, τ^2 is set to zero, complete homogeneity is assumed and no adjustments are necessary.

3. Further reading

CQC z-scoring

Spiegelhalter D J, Sherlaw-Johnson C, Bardsley M, Blunt I, Wood C, Grigg O. Statistical methods for healthcare regulation: rating, screening and surveillance. *J Roy Statist Soc* A 2012; **175**: 1-47.

Cross-sectional analyses using z-scores and funnel plots

Spiegelhalter D J. Funnel plots for comparing institutional performance. Stat Med 2005;24:1185-1202.