**Box 1: Examples of ceiling or floor effects**

***Quality and Outcomes Framework performance indicator DM22 (2006-07 to 2012-13)***

Measured the percentage of patients with diabetes who have a record of estimated glomerular filtration rate or serum creatinine testing in the previous 15 months. In 2006-07 and with 8365 general practices reporting the indicator, one practice had a score of 0% and 700 a score of 100%. The mean score was 96.4%.

***GP patient satisfaction survey, from 8307 general practices in 2008***

Survey on how easy it was for patients to get through on the phone at own doctor's surgery (no *v* yes): 81 practices scored 100%, when the mean score was 87.4%. Survey also asked about the ability for patients to get appointment within two days (no *v* yes): 93 practices scored 100%, when the mean score was 85.7%

***Investigation of prescribing safety using the Clinical Practice Research Datalink, in 2013[5]***

Investigation looked at the proportion of women with a breast cancer diagnosis who were prescribed oral or transdermal oestrogens: 319 (61%) of 523 practices had a prevalence of 0%, when the mean prevalence was 1.1%. For patients prescribed repeated amiodarone without a thyroid function test within the recommended time period, 43 (9%) of 505 practices had a prevalence of 100%, when the mean prevalence was 42%.

relative risks. Although such an approach could be acceptable in a scenario where few providers are low scoring (where the rare event approximation stands), more generally odd ratios overestimate the relative risk and their interpretation tends to be flawed.[6] In addition, such a simplification discards a considerable amount of information and reduces statistical power, while the choice of the threshold value used to dichotomise performance may be arbitrary.

A more suitable approach is based on a transformation that uses the logit function to linearise a performance, satisfaction, or safety indicator. This transformation effectively takes a scale that ranges from 0 to 1 (or 0% to 100%), and expands the scale so that it ranges from minus (–) infinity to plus (+) infinity. The indicator can then be analysed by standard linear models or similar methods, in a frequentist or Bayesian framework. Given the particular challenges associated with this approach, in this paper we provide guidance for researchers in conducting and interpreting analyses of logit transformed performance scores of quality indicators.

For practical examples, we draw on our experience of analysing family practice performance under the United Kingdom's Quality and Outcomes Framework (QOF).[7-9] The framework is a financial incentive scheme introduced by the UK government in 2004, which rewards practices on the basis of their performance on more than 100 quality indicators related to the clinical management of chronic disease, practice organisation and, patient experience.[10] For the clinical indicators, which are regularly reviewed and could be withdrawn,[11] practices are assessed on the basis of the percentage of eligible patients for whom each target is met (eg, the percentage of patients with coronary heart disease who give a blood pressure recording of ≤150/90 mm Hg). The main research questions in relation to performance on the QOF indicators relate to how practice performance varies between practices with different characteristics and patient profiles, and how performance changes over time.

We have focused on performance indicators in primary care to provide practical examples and place the methods into context. However, the methods are relevant for the analysis of any percentage score that aggregates binary or continuous information from a lower level unit (eg, patient) to a higher level unit (eg, general practice population) and where non-linearity is present. Examples of satisfaction and safety include the percentage of people who are happy with access to their preferred general practitioner or who are prescribed a drug that puts them at risk.

## Approach

### Logit transformation

The first step is to assemble practice scores (in the case of the QOF, the proportion of eligible patients for whom a given target has been achieved) and model these scores on the logit (log-odds) scale. There are two main options to do this and to achieve an expansion in the scale from [0, 1] to [± infinity]: simple and empirical transformation.

In the simple logit transformation, the score $p$ ($0 \leq P \leq 1$) is transformed into a log odds: $\text{logit}(p)=\ln(p/(1-p))$ (fig 1). For example, a difference in the untransformed score ($p$) from 0.97 to 0.98 (that is, 97% to 98%) represents the same difference in transformed score ($\text{logit}(p)=0.41$) as a difference from 0.55 to 0.65 (that is, 55% to 65%). The transformed score can then be modelled with standard linear models and the analysis of transformed scores also ensures that predicted achievement scores lie between 0% and 100%. The main drawback of the simple logit transformation is that achievement scores of 0% and 100% become minus and plus infinity, respectively, following transformation. As a consequence, these observations will be interpreted as missing values by statistical software packages and removed from analyses. If there are a large number of scores at the ceiling or floor values, this effectively renders simple logit transformation ineffective.

The empirical logit transformation offers an improvement over the simple logit transformation at the ceiling and floor points by making a separate transformation at these values.[12 13] For scores where $p$ is strictly greater than 0 and less than 1, the simple logit transformation is applied as above. For scores where $p$ is equal to 0 or 1, the empirical logit transformation is given by formula below, where $n$ is the number of observations over which $p$ is calculated:[14 15]

$$\text{Logit}(p) = \ln\left[\left(p+\frac{0.5}{n}\right)/\left(1-p+\frac{0.5}{n}\right)\right]$$

Details of equation

In the QOF setting, $n$ would be the number of patients for which an indicator is evaluated (the denominator for the indicator), for example, the number of patients diagnosed with diabetes. Not only does this transformation overcome the problems described above, but it also has an additional benefit; scoring 100% on an indicator evaluated on a large number of people ($n$) is rewarded with a higher transformed score