Appendix 1: The National Reporting and Learning System

The NRLS, housed at the National Patient Safety Agency (NPSA), is a voluntary, national reporting system set up in 2003 for the NHS in England and Wales. To date, it is one of the largest patient safety reporting systems in the world and contains over eight million records of patient safety incidents.[34] Incidents are reported by staff at a local level and corrective measures taken where appropriate. Subsequently, these reports are anonymised for personal identifiers and uploaded to the NRLS. An alternative route by which information is uploaded to the NRLS is through an online reporting form available on the NPSA website, which is also open to members of the public. Each NRLS report refers to an unintended or unexpected incident that could have or did lead to harm for one or more patients receiving NHS-funded care. It includes the reporting of those incidents which did not lead to harm despite an error taking place, and those which did not lead to harm because the incident was prevented from reaching the patient. These incidents are further stratified into different levels of harm.[35] The database has 75 data fields, including patient demographics, specialty, location of incident, category of incident and a free-text description of the incident.[36] Each incident reported as leading to death or serious harm is reviewed individually by trained clinical staff and a range of outputs is produced to provide solutions to patient safety problems. These include one-page reports called *Rapid Response Reports*, quarterly data summaries and topic-specific information such as preventing inpatient falls in hospitals. There is constant consultation with subject-matter experts including professional organisations. NHS organisations also have deadlines imposed on them by which time they should have implemented any findings from these reports.[34]

Appendix 2: Creation of the OEI

The OEI is the sum of the number of errors (propensity, P) and the degree of harm (severity, S). This should enable us to identify hospitals with large numbers of errors and similarly those units with the greatest degree of harm. It is reasonable to assume as more procedures are carried out, a larger number of errors will be reported, although we are also cognizant that there is potentially a high risk of errors in units undertaking relatively fewer procedures.

Calculating the error propensity

For each hospital, P was calculated as:

$$P = 100\frac{n}{N},$$

where n was the number of procedures where any error had occurred and N was the total number of procedures; P had a range of 100, with 0 representing the lowest error propensity and 100 representing the largest error propensity.

The standard error of P (P_{SE}) was calculated as:

$$P_{SE} = \sqrt{\frac{P(100 - P)}{N}}$$

The consequences of a medical error can vary from negligible to fatal. The error propensity index treats all reported errors equally. However, there is a qualitative difference between hospitals which have the same E_p, but in one the proportion of death harm is nearly double the other. Therefore, it is important to capture the severity of the error. Each error report

in the database contains a NPSA code for severity which is ordinal in character. We propose

a severity index based on proportion of each harm category weighed for its severity.

For each hospital, the severity (S) was calculated as a weighted sum: $S = \sum w_i n_i / n$, where

w_i is the weight for the ith error severity category; n_i is the number of procedures where i^{th}

error severity category occurred; and n is the number of procedures where any error

occurred.

Method of determining weights

We can give greater weight to less common events by using the inverse probability weight.

The relative frequency of each harm category was calculated using the inverse probability

weights (IPW = 1/ relative frequency) and IPW relative to the no harm category. There are

two drawbacks to this way of assigning weights: one, it is data specific so that another

dataset with a different distribution will yield different weights; and two, it gives, perhaps

correctly but inconveniently, high values to severe harm and death, in which case error

severity may be measured just by counting these events. Although this proposition is

attractive in its simplicity, it is not useful in terms of error monitoring and developing

policies. Our finding that greater harm categories are less frequent is a confirmation of the

famous Heinrich ratio, which states that for every major injury, there are 29 minor injuries

and 300 near misses.[37] Referring to the ratio, an expert group on learning from adverse

events in the NHS argued for the importance of reporting near misses: "Not all unsafe

systems produce bad outcomes all the time. The potential for disasters may exist, but for

any number of reasons those disasters might not occur at all, or occur very rarely – what

has been termed 'a dynamic non-event'. If there are no bad outcomes to monitor, safety

information systems need to collect, analyse, and disseminate information from incidents and near misses, as well as from regular proactive checks on the system's 'vital signs'."[38]

We therefore chose a weighing system computed as 2^i where i is the ordinal number of error severity category, from 0 for no harm to 4 for death. The error severity was also rescaled to a range of 0 to 100 as done with S.

For this purpose, the harm categories were assigned numerical values from 0 to 4 (no harm = 0, low harm = 1, moderate harm = 2, high harm = 3, and death = 4) to reflect their natural order of severity. The weight assigned to the ith harm category was 2^i.

$$S = \frac{100}{16}\left(\left(\sum_{i=0}^{4} 2^i \frac{n_i}{n}\right)\right)$$

where, n_i is number of procedures where i^{th} harm category occurred and n is the number of procedures where any error occurred. The constant term 100/16, 16 or 2^4, being the maximum value possible (when all reported errors were deaths) for the variable part of the formula, was used to adjust the scale of the index so that 100 was the maximum value, representing a situation where all errors reported resulted in deaths. The minimum value would be 6.67, representing the case where all reported errors produced no harm. We intentionally avoided rescaling S, 0 to 100, to differentiate between the situation where no errors were reported and some errors were reported but they were all in the no harm category.

The standard error of S was computed as:

$$S_{SE} = \frac{100}{16}\left(\sqrt{\sum_{i=0}^{4} 2^{2i}\left(\frac{p_i - p_i^2}{n}\right)}\right)$$

Orthopaedic Error Index, OEI

We defined the OEI, *E*, as the weighted sum of error propensity and error severity.

$$E = 0.5P + 0.5S$$

This index gives equal weights for propensity, which captures the overall number of errors and severity of errors, because both aspects are considered important in dealing with errors.**Error! Bookmark not defined.** The weights were chosen so that *E* has a range of 0 to 100. The standard error of *E* was computed as:

$$E_{SE} = \sqrt{\frac{P_{SE}^2 + S_{SE}^2}{4}}$$

To identify reporting bias, we used the relationship between number of procedures and OEI. For this purpose we first meta-regressed OEI on number of procedures and saved the predicted values of OEI.