Health-care quality registers

OUTCOME-ORIENTATED RANKING OF HOSPITALS IS UNRELIABLE

J. Ranstam, P. Wagner, O. Robertsson, L. Lidgren

From Lund University Hospital, Lund, Sweden

An outcome-orientated ranking of hospitals is becoming increasingly popular and is routinely used by Swedish health-care authorities. Whereas uncertainty about an outcome is usually presented with 95% confidence intervals, ranking’s based on the same outcome are typically presented without any concern for bias or statistical precision. In order to study the effect of incomplete registration of re-operation on hospital ranking we performed a simulation study using published data on the two-year risk of re-operation after total hip replacement.

This showed that whereas minor registration incompleteness has little effect on the observed risk of revision, it can lead to major errors in the ranking of hospitals. We doubt whether a level of data entry sufficient to generate a correct ranking can be achieved, and recommend that when ranking hospitals, the uncertainties about data quality and random events should be clearly described as an integral part of the results.
We then performed further simulations using incompleteness rates ranging from 0% to 15%. We also calculated the adjusted risk estimate taking account of the incompleteness of the register. A re-operation risk estimate calculated using register data is related to the register’s degree of completeness. With incomplete re-operation registration the observed re-operation risk will underestimate the true re-operation risk. If the level of completeness is known, the risk estimate can, however, be adjusted for incompleteness. For example, with a level of completeness of 0.9 and a re-operation risk of 0.2, the adjusted risk estimate is calculated as 0.2/0.9 = 0.22.

The statistical package (R Foundation for Statistical Computing, Vienna, Austria) was used for the calculations.4

### Results

In the case of risk of re-operation, adjustment for register incompleteness is fairly straightforward. With randomly distributed and reasonably low incompleteness rates, and low risks for re-operation, register incompleteness has a negligible effect. For example, the average observed re-operation rate within two years is 0.015.2 An estimated re-operation incompleteness of 5% results in an adjusted re-operation risk of 0.016.

By contrast, a revision incompleteness of 5% has major consequences for hospital ranking. The simulation study showed that 63 of 77 hospitals (95% CI 54 to 70) were incorrectly ranked (Fig. 1). Furthermore, the maximum rank error was 14 (95% CI 8 to 29), which cannot be considered a minor consequence (Fig. 2).

When the simulation program was rerun with varying completeness rates, the incidence of incorrect ranking was substantial even with incompleteness rates as low as 1% to 2% (Fig. 3). The result was similar for maximum rank error (Fig. 4). In fact, if the incompleteness rate is 4% or more, it is of little importance whether it is 5%, 10% or 15%.

To investigate whether incorrect ranking was less of a problem with higher revision rates, i.e. with greater variance in the distribution of hospital-specific re-operation rates, the simulation program was rerun once again with a fictitiously increased re-operation rate of ten times. This indicated that incorrect ranking is related to re-operation rate, but even when the real re-operation rate is increased ten times, the error remains substantial (Figs 5 and 6).
Discussion

Performance measurement using ranking has previously been criticised on the grounds of statistical imprecision,\(^5\) and this weakness has been demonstrated in empirical studies which have found ranks to be ‘not reliable indicators of performance or best practice’,\(^6\) and even ‘extremely unreliable statistical summaries of performance’.\(^7\) Consequently, serious concerns have been exposed about using registry data for ranking purposes.\(^8\)

We have previously described the effects on ranking of issues related to statistical precision and suggested a statistical method that minimised the problem.\(^9\) To the best of our knowledge, the current report is the only one that quantifies ranking errors caused by incomplete registration.

It is almost impossible to avoid some incompleteness in large registers. The Swedish Knee Arthroplasty Register is a good example.\(^10\) Despite regular annual enquiries, corrections by participating hospitals, a postal survey to all living patients and cross-referencing with the official reimbursing (ICD10) databases, only 95% of all revisions can be identified.\(^11\)

The data completeness of other Swedish national healthcare quality registers varies greatly and in several instances is unknown. However, not even those registers that claim a very high completeness\(^12\) present higher completeness rates than 96% to 97%.

Our results clearly show that even such high rates of completeness are insufficient for valid ranking, at least when Swedish hospitals are ranked according to the two-year risk of reoperation after hip replacement. For most of the registers we doubt whether it is practically possible to achieve the level of data quality required to make a correct ranking.

Risk estimates related to implants, surgical technique and other similar clinical factors remain relatively robust measures in incomplete registers as long as the incompleteness is randomly distributed. Furthermore, when such estimates are presented as a basis for clinical decision, they usually appear with confidence intervals. This makes such data a sound basis for rational decisions.

On the other hand, as we show in our example, observed rankings without any indication of variability due to the
method of sampling cannot be used as a basis for rational decision making.

Statistical methods for computing confidence intervals for ranks have been described.\(^\text{6}\) It is possible to develop these so as to account for the uncertainties caused by register incompleteness.

If, in spite of their unreliability, hospital rankings are presented as a basis for clinical improvement, the actual margin of error should be assessed and clearly described for each individual hospital so that unnecessary mistakes in ranking are avoided.

No benefits in any form have been received or will be received from a commercial party related directly or indirectly to the subject of this article.

References