Office for Statistics Regulation
1 Drummond Gate
London SW1V 2QQ

T: 0845 604 1857
E: authority.enquiries@statistics.gov.uk
W: www.statisticsauthority.gov.uk

Director General for Regulation

Lewis Macdonald MSP
Convener, Health and Sport Committee
Scottish Parliament

4 December 2018

Dear Mr Macdonald

Concerns raised about the Scottish Intensive Care Society Audit of Critical Care

We have investigated the issues raised by an anonymous complaint about the Scottish Intensive Care Society Audit of Critical Care (SICSAG)[1]. As requested by the complainants, I am reporting our findings to you. We have focused our investigation on the key points relevant to the Code of Practice for Statistics published in February 2018 and its three pillars of trustworthiness, quality and value (see the subheadings below). We would be happy to discuss any of the issues raised in more detail with you or the authors of the letter.

Due to the importance of the issues raised, and the technical nature of some of the points, we commissioned an external review from an expert based at the Nuffield Trust to help us reach our judgement. His report (attached as Annex A) supported us in reaching our views.

Quality: sound methods

- Following the external review of methods (provided at Annex A), we are reassured that the statistical methodology applied by SICSAG in the reporting of intensive care mortality is appropriate, and that the approach taken for handling possible outliers follows international good practice in this area. The external review addressed the key methodological issues raised by the complainant. We also asked the reviewer to comment on the methods used to avoid large numbers of units being inappropriately classified as abnormal (over-dispersion) following advice from an international expert in this field.

Quality: assured quality

- Data 'gaming' is a very serious breach that would undermine the audit process and invalidate related statistics. The data validation process that Information Services Division Scotland (ISD) apply should identify data entry inconsistencies, whether accidental or deliberate, but we recommend that additional checks should be carried out to provide further reassurances. This is also supported by the external review.

Trustworthiness: orderly release

- Publication timeliness and orderly release dates are important markers of trustworthiness. We are confident that ISD's processes in this area are compliant with the Code of Practice for Statistics. It was noted that in 2016 and 2017 the release date of the SICSAG report coincided with reporting of School Higher Examination Results for Scotland. ISD have

[1] http://www.sicsag.scot.nhs.uk/publications/main.htm

released the SICSAG report on a consistent date since 2012 (2nd Tuesday in August). It is only in 2016 and 2017, due to a change in the week Examination Results were published (from 1st Tuesday in August to 2nd Tuesday in August), that the two coincided. We would not expect ISD to change its release date in the event of the release clashing with the release of school exam results.

Value: clarity and insight

- We consider that some aspects of the interpretation of the results could have been set out more clearly in the report. For example, as the external review highlights, more details of the relationship between outliers and the review process that results from their identification would have provided valuable context and additional insight to these statistics. It could also have provided reassurance that the statistics were being used appropriately.
- We welcome the changes made to the report to improve the clarity of the interpretation of outliers as a result of ISD's own review following the complaints raised. We also encourage ISD to reflect on the external review's additional feedback about areas to improve when they next review the report's commentary.

Value: accessibility

- Between the 2015 and 2016 reports the volume of additional information available to download as tables and figures reduced. These changes should be informed by user needs and reasons for changes should be explained clearly. We recommend that ISD review the information provided alongside the published reports to ensure that user needs are being met. ISD should also explain the reasoning for decisions taken - including where information ceases to be published - more clearly to users.

We understand that the complainants wish to remain anonymous. ISD are keen to have a meeting between all interested parties to discuss the concerns raised. We support that and would welcome your assistance to facilitate such a meeting. If this is not possible, we would still welcome a meeting with yourself, ISD and SICSAG.

Please thank the complainants for bringing these matters to my attention.

Your sincerely

Ed Humpherson
Director General for Regulation

Annex A: External Review of complaints raised and ISD's response

Background

In July 2018, the Office for Statistics Regulation (OSR) received complaints about the Scottish Intensive Care Society Audit of Critical Care's reporting of intensive care mortality in Scotland. To support the judgement of the issues raised, OSR commissioned an external review of the complaints and ISD's response. The review was conducted by Chris Sherlaw-Johnson of the Nuffield Trust. The report of the review is presented below.

Nuffield Trust Report for the Office for Statistics Regulation in relation to concerns raised about the Scottish Intensive Care Society Audit of Critical Care

Summary

1. The statistical methodology is appropriate for the purposes of representing variation in ICU mortality across the different hospitals.
2. Some of the accompanying text relating to outliers could be altered to clarify messages to the public and reassure them that outliers are taken seriously as potential concerns relating to quality of care. This requires more detail of the process of responding to outliers.
3. More could be made of the importance of the governance reviews conducted by Health Boards of hospitals that are statistical outliers and describing their scope. If it is possible to do so, then the findings could be made public and reiterated in the report in the following year.
4. If a statistical outlier in one year is no longer an outlier the next, then it could be helpful to speculate why it might be (e.g. better coding, improved care following review, random variation etc.).
5. If there are concerns about inconsistent coding used for the APACHE score then these should be addressed as a separate line of enquiry, if that is not being done already.

Method employed by SICSAG

The interpretation and use mortality outliers have been well discussed by, for example, health care regulators, the Mid Staffs Inquiry and at various times across the media. One of the clear messages that comes from these and other discussions is that on its own a mortality outlier is not a judgement about the quality of care provided by the hospital, because what causes the outlier could be outside the control of the hospital being measured. For example, it may be that any risk adjustment being applied does not capture all the relevant risk factors, there may be measurement error or differences in the way certain measurements are made, or patients may be included in the count of deaths that are generally excluded elsewhere. Therefore, the most appropriate way to handle outliers is to treat them as triggers for asking questions about the care that is being delivered, and this appears to be the approach used by SICSAG.

Relationships between high mortality and quality of care

Again this has been much debated. It is reasonable to assume some sort of relationship may exist, but, given the arguments made above, it would not be a one-to-one correspondence, i.e. high risk adjusted mortality does not mean an ICU or hospital is providing poor care and, vice versa, a hospital providing poor care will not necessarily appear as a mortality outlier. There are several hospitals in London where there are significant concerns with quality of care, some in special measures, but standardised mortality rates across London are comparatively low. Mid Staffs has often been cited as a trust with high mortality rates that had significant quality of care issues and therefore mistakenly used to prove that high mortality implies poor quality. But, some trusts were more significant outliers than Mid Staffs, and, as the Keogh Review later found, some of the most significant mortality outliers did not reveal much cause for concern on investigation.

Standardising outcomes

The SICSAG work standardises mortality according to patient risk on entry to the ICU as measured by the APACHE score. This is widely used for predicting risk and appropriate in this context. Of course, like any risk score, it will not feature every possible risk factor, and the relationships between factors and outcome may vary from one population to the next. Also, an issue has been raised about the consistency of coding the presence of pre-existing conditions that are used by the score. If this is true then it is a line of enquiry when following up an outlier, and should also prompt national advice on better coding.

The use of recalibrated risk scores rebased to a national baseline is standard practice where the purpose is to focus on variation within a country, and follows the practice of other audits as well as official statistics, such as the Standardised Hospital Mortality Indicator (SHMI) for English hospitals. This, naturally, accepts that the country may have rates that deviate from other countries. But there are two questions here: one is about exploring why a particular hospital has worse mortality rates than the rest of the country and the other is a wider policy question about the overall standards of care within a country. The focus of the SICSAG report appears to be the former.

Visual presentation

The funnel plot approach for the visual reporting of variation between hospitals or units is fairly standard and is similar to recommended approaches across UK clinical audits, by regulators and internationally. Even if funnel plots are not being used, the underlying messages are the same: the statistics suggest where there may be problems, not where there are problems. And often, the funnel plot is the public face of the data where, in practice, within the service, more continuous monitoring methods are used that can pick up potential concerns sooner.

Explanation of the ARI outlier

Presenting statistical outliers to the public is always a difficult balancing act between promoting caution about interpretation and assurance that potential concerns around quality of care are being looked into in the appropriate way. In the 2016 SICSAG report the governance process is explained (to a point) and there is a cautionary sentence to suggest that it could be a false positive:

"Unit W is a tertiary referral centre, and as such severity of illness can be underscored in this group as they are referred from other hospitals already sedated, and then cannot be assigned APACHE points for depressed levels of consciousness."

I cannot speak for all members of the public, but I am not sure the balance here is right. A review was ongoing, whose findings were yet to be available, but at the same time there is an explanation that appears to be trying to tell people that there shouldn't be anything to worry about. Moreover, is it all patients who are already sedated, or just a few, and what is the likelihood that this is a reasonable explanation? It may have been better to say that this might have been the issue (i.e. of unrecorded severity) but a review is ongoing at the hospital which will report on the quality of care, and once the review is concluded SICSAG will issue a statement about what has been found and any recommendations. I would also have included an outline of the scope of the review and stated that a report will be made public soon after completion.

This could also have been picked up in the following year's report. In that later report it might also have been appropriate to discuss whether ARI is no longer an outlier because the tertiary centre issue has been corrected for, or whether it is not an outlier despite being a tertiary centre, or whether it could be due to improvements in care following the previous year's review.

Presenting "effect sizes" and measures of differences from the mean

A suggestion was made that the SICSAG report should present sizes of effect. However, presenting differences between observed and expected mortality as "effect sizes" can be misleading. This is a debate that became prominent in the media a few years ago when unhelpful

terms like “excess mortality” were being used to describe outcomes for outlying NHS trusts in England. The problem was that these differences were being interpreted as the number of deaths for which a hospital was culpable. After much guidance from statisticians and epidemiologists, this prompted the Department of Health at the time to recommend hospitals measure “avoidable” deaths using case note review, because, only through careful case-by-case analysis can you begin to understand whether a death was truly due to poor quality care.

The use of two and three standard deviation limits is also conventional. And it is correct to point out in a public document that, if you are testing multiple hospitals, false positives can occur by random and that, for example, one in 20 hospitals would breach the two standard deviation limit. The complainant raised concerns with the way this is presented in the report, suggesting that the accompanying text is trying to explain the outlier as random deviation. Because of the paragraph that follows, I don’t think it does come across this way, but I would recommend the way this is worded is looked into, perhaps with advice from members of the public.

Another issue is about the way the outlier is described as at 2SD when it is closer to the 3SD limit. I think describing the outlier as “to 2SD” may be misleading, and perhaps saying something like “an outlier that breaches the 2SD limit” would be better, and to make it clear that it is not how close the outlier is to the limit that counts but whether it crosses it.

A question has been raised about the different way of reporting the infection data later in the report. It is appropriate for these to differ as they are not similar analyses. The infection analysis looks at unadjusted rates of different types of infection with confidence limits reflecting random variation each year. It is not an analysis looking at the variation among several units against a baseline.

Over-dispersion

The funnel plots account for over-dispersion, following recommendations made in other studies. The analysis applies multiplicative rather than additive over-dispersion, but I think the finer point as to the suitability of one over the other is relatively minor in this context given the way the outputs are being used.

Chris Sherlaw-Johnson
Nuffield Trust
October 2018