ICU
MANAGEMENT & PRACTICE

ICU Volume 15 - Issue 3 - 2015 - Management

Benchmarking: Lessons Learnt

Dr. Matti Reinikainen
******@***pkssk.fi

Chief Physician - Department of Intensive Care Medicine, North Karelia Central Hospital

Prof. Hans Kristian Flaatten
ICU Management & Practice
Editorial Board Member
******@***uib.no

Professor - Faculty of Medicine, University of Bergen, Norway
ICU Management & Practice
Editorial Board Member

LinkedIn

Benchmarking —comparing your own results with those of others—has the potential to reveal areas in which your unit could improve. However, there are pitfalls you should be aware of.

When he was the CEO of Xerox Corporation, David T. Kearns stated, *"Quality improvement can't be measured in a meaningful way against standards of your own internal devising"* (Kearns 1990). This sentence captures the basic concept of benchmarking: unless you are willing to compare your results with those of others, your impressions of quality may be too optimistic. Many aspects of intensive care unit (ICU) performance can be benchmarked. These include safety of care, measures of economic performance and patient and family satisfaction. This article concentrates on the challenges of comparing severity of illness-adjusted mortality figures.

Beginning of Benchmarking in Intensive Care

In the early 1980s, Knaus and co-workers presented a severity-of-illness scoring system for ICU patients, called the Acute Physiology and Chronic Health Evaluation (APACHE), and demonstrated its use in comparing patient populations from different ICUs and their outcomes (Knaus et al. 1981; 1982). Although it was developed in America, APACHE was soon implemented in other countries. Hospitals in the USA, France, Spain and Finland participated in the first multinational study using this scoring system (Wagner et al. 1984).

APACHE was soon followed by the Simplified Acute Physiology Score (SAPS), developed in France (Le Gall et al. 1983). Both systems have been developed further, resulting in several updated versions. The basic principle is similar in all versions: the patient's age, conditions at admission, comorbidity and abnormal current physiological measurements are given points to produce a score that is converted into a predicted probability of the patient dying during the actual hospitalisation. For an individual patient, the score reflects the severity of illness at the beginning of the intensive care period, but the associated probability of death is never the same as the final outcome: the probability is between 0 and 1 (but never exactly 0 or 1), whereas the outcome is

either survival or death. Probabilities of death are useful as an aggregate measure of risk in a large group of patients, as the sum of individual probabilities provides the expected number of deaths (Le Gall 2005). Dividing the number of deaths that occur during a certain period by the expected number of deaths gives the standardised mortality ratio (SMR).

After these severity-of-illness scoring systems made it possible to adjust for differences in patient case mix, comparing mortality outcomes of different ICUs has become popular. Large benchmarking programmes are running in many countries. The authors of this article represent the nationwide ICU quality and benchmarking programmes in Finland (the Finnish Intensive Care Consortium, started in 1994) and Norway (the Norwegian Intensive Care Registry, since 1999).

Benchmarking: A Suitable Concept for Industrial Plants, but What About Intensive Care?

Benchmarking involves comparing the performance of one organisation to that of similar organisations, with the aim to identify best practices and best performers and learn from them. The idea originated in the manufacturing industry and is not easily applied to intensive care medicine, as ICUs work with more complex processes than simply making multiple copies of standard wares out of standard raw materials.

Even so, we believe that we can apply certain ***essential lessons*** learnt by benchmarking pioneers in the manufacturing industry to intensive care:

1. **It is dangerous to become complacent** (Kearns 1990). No matter how good your ICU is today, if you stop learning and improving, you are going to regress. Beware of self-satisfaction!

2. **If you do not document your results, it is easy to imagine** that you and your team are doing a good job. This impression may be false.

3. **Comparing your current results with your previous results** is better than not documenting them at all, but it **may not be enough**. Comparing your results with those of others may open your eyes to areas in which you could improve.

4. **Intensive care medicine is team work.** The expertise of one individual doctor or the excellence of one nursing team has little impact on the overall performance of the healthcare system. Benchmarking provides an opportunity to detect weaknesses in the treatment chain of some patient groups.

Some may argue that there is little evidence that benchmarking improves quality of care. It is true that no randomised controlled trial has demonstrated the superiority of an intensive care programme with benchmarking over one without. However, the benefits of benchmarking have been proven in other fields of medicine (Kiefe et al. 2001; Hermans et al. 2013). Generally, poor quality means there is much room for improvement, whereas it may be hard to improve high-quality processes further. It is not surprising that the impact of feedback from benchmarking seems to be larger when the baseline level of performance is low (Jamtvedt et al. 2006).

Measuring and even defining quality in intensive care is difficult. In a comparison of quality indicators (QI) within intensive care in eight countries, no single indicator was used in all countries. The most common QIs were the standardised mortality rate (in six of eight countries) and patient/family satisfaction (five of eight) (Flaatten 2012). Generally, documenting severity of illness and mortality figures is considered essential in an ICU benchmarking programme (Moreno et al. 2010).

Pitfalls of SMR and How to Avoid Them

Potential confounders influencing SMR calculations include properties of the model used to adjust for differences in baseline risk, factors affecting the measurement of severity of illness and the choice of mortality endpoint.

1. Performance of the Risk-Prediction Model

A risk-adjustment model often fits the population used for model development well. However, when one applies it to another patient population, its prognostic performance may be worse (Livingston et al. 2000). The model may systematically overestimate or underestimate the risk of death. The adequacy of risk estimation may also differ across different levels of risk: for example, the model may overestimate mortality in low-risk patient groups but underestimate mortality in high-risk patients. This is called poor calibration or poor fit of the model (Angus 2000). If the calibration is poor and there are major differences between ICUs in patient case mix, comparing SMRs is questionable. If ICUs are ranked according to SMRs, the choice of prognostic model may

heavily affect the rank of a unit (Bakshi- Raiez et al. 2007; Kramer et al. 2015). Nonetheless, risk-adjustment models developed for intensive care perform better than models based solely on administrative data, which have also been used in ranking ICUs (Brinkman et al. 2012).

Over time, outcomes tend to improve and risk-adjustment models become outdated. If benchmarking programmes use old models, it is probable that SMRs will be low for most, if not all, ICUs. This must not be interpreted as evidence of perfection. To solve the problem of worsening prognostic performance of ageing risk-adjustment models, new models have been developed. However, even if a new model fits perfectly well, its performance will deteriorate as time passes (Moreno and Afonso 2008).

An alternative approach to creating a totally new model is customising an existing model to better fit a regional patient population. A common strategy is first-level customisation, which means that the variables in the model and their relative weights are unchanged but the equation converting the severity score to probability of death is updated. Very good prognostic performance can be achieved with a customised model (Haaland et al. 2014). Whether a benchmarking programme should use an original risk-adjustment model or a locally customised or even locally created model depends on the choice of the reference population with which one wants to compare ICUs. With the original model, one can describe a patient population with a well-known severity score and compare the outcomes with those of an international reference population. However, if one wishes to compare ICUs within the benchmarking programme, then a well-fitting customised model is a better choice (Angus 2000; Moreno and Afonso 2008; Metnitz et al. 2009).

2. Factors Affecting the Measurement of Severity of Illness

Points are added to the severity score according to values of physiological parameters: the more abnormal a value, the higher the score. When data are missing, the values of the parameters in question are commonly presumed to be within the normal range, and no severity points are added. Thus patient groups with a lot of missing data may appear less severely ill than they actually are, and the calculated SMR may become erroneously high. Correspondingly, improving data completeness leads to a decrease in SMRs (Reinikainen et al. 2012).

Changing the frequency of measurement may affect the severity scoring: taking more samples increases the likelihood of obtaining abnormal values. This results in higher severity scores and lower SMRs because the scoring systems take into account the most extreme value from the observation period (Suistomaa et al. 2000). Automation of data collection with a clinical information system (CIS) increases the sampling rate of physiological data. In Finnish ICUs, the severity of illness-adjusted odds of death were 24% lower in 2005–2008 than in 2001–2004, but one-fifth of this computational improvement in outcomes could be explained by improvements in data completeness and automated data collection through the use of a CIS (Reinikainen et al. 2012). This phenomenon should be noted in benchmarking programmes if some ICUs use CIS technology and others do not.

The importance of data accuracy cannot be overstated (Angus 2000). Education and data quality monitoring are continuously needed to achieve and maintain correct and harmonised documentation practices.

Severity scores should reflect the severity of illness. However, the scoring systems are not able to distinguish a patient affected by a disease from a patient whose condition may be partly caused by substandard care prior to ICU admission or in the beginning of the ICU period.

3. Mortality Endpoints

Traditionally, vital status at hospital discharge has been used as an outcome measure. However, comparing hospital mortalities is problematic. Patients transferred to other hospitals are calculated as hospital survivors, yet some of these patients will die in the next hospital. Thus differences in hospital discharge practices can cause bias (Kahn et al. 2007).

In a recent study from Sweden, Rydenfelt et al. (2015) explored the effects of using 30-day mortality instead of hospital mortality as the outcome measure. Not surprisingly, 30-day mortality was higher than hospital mortality in almost all ICUs. What is newsworthy is that the magnitude of the difference between in-hospital and 30-day mortalities is not constant, and hospital discharge practices and patient case mix affect it: the difference increases with increasing age and severity of illness and also varies across diagnostic categories. The calculated SMR of an ICU may also be markedly influenced by the choice of mortality endpoint. Comparable findings have been published from the Netherlands: the choice of mortality endpoint (vital status at hospital discharge or at a fixed time point) affects the SMRs and SMR rank positions of ICUs (Brinkman et al. 2013). We recommend that, whenever possible, quality programmes stop using hospital mortality as the primary

endpoint and start using fixed-time mortality. Preferably, the follow-up should be longer than 30 days (e.g., 6 or 12 months).

Because of the potential impact of these confounders, one needs caution when evaluating SMRs. However, this does not mean that SMRs are without value. Constant differences in SMRs can be interpreted as an indication that one should look more deeply into the situations in different units and try to identify the factors explaining the differences (Le Gall 2005). The explanatory factors may be unrelated to quality of care, but it is also possible that true quality differences exist.

Conclusion

Comparing results of intensive care is not without problems. An ICU leader should be aware of the potential confounders affecting SMR calculations. Nevertheless, benchmarking may help leaders identify what they want to find: areas in which there is room for improvement. In addition, it is also important to realise that striving for quality is a never-ending journey. Once more, we quote David T. Kearns: *"We're far from finished with our drive to improve…The pursuit of quality is a race with no finish line"* (Kearns 1990).

Published on : Tue, 29 Sep 2015