

Will benchmarking ICUs improve outcome?

Douglas Woodhouse^a, Marc Berg^a, Joris van der Putten^a and Judith Houtepen^b

^aPlexus, Straatweg 68, 3621 BR Breukelen and
^bTRAG-Performance Intelligence in Healthcare,
Bergweg 151, 3707 AC Zeist, The Netherlands

Correspondence to Dr Marc Berg, PhD, MD, Plexus,
Straatweg 68, 3621 BR Breukelen, The Netherlands
Tel: +31 0 6 424 42 834; e-mail: berg@plexus.nl

No funding received for this work.

Current Opinion in Critical Care 2009,
15:450–455

Purpose of review

In this article we discuss our experiences benchmarking eight ICUs in The Netherlands. Benchmarks must be carefully designed and implemented to generate meaningful results. We define prerequisites that we have identified for successful benchmarking and discuss the development, implementation and results of ICU benchmarks that we have completed.

Recent findings

Previous articles have discussed benchmarking ICUs, but there are still few studies of significant size and appropriate design that measure the impact of benchmarking on outcomes. Perhaps the most well known, and still best example of a benchmarking study designed to measure outcome improvements is the work of Pronovost *et al.* in Michigan ICUs.

Summary

Benchmarking is an increasingly common activity, however it is difficult to prove that benchmarks result in improved outcomes. Concurrent with our benchmarking activities the Standardized Mortality Ratio in Dutch ICUs has decreased. We have been able to show that larger ICUs in our benchmarks generally had improved outcomes despite a higher average patient severity. Quality assurance in healthcare is maturing and benchmarks will become an increasingly useful way of comparing performance between institutions.

Keywords

assessment, benchmark, comparison, measurement, quality

Curr Opin Crit Care 15:450–455
© 2009 Wolters Kluwer Health | Lippincott Williams & Wilkins
1070-5295

Introduction

Browsing current literature related to quality assurance in ICUs, you find many articles about ICU benchmarks. On the basis of this reading, you may get the impression that ICU benchmarking makes sense, is possible, and results in ‘better’ ICUs.

The reality of benchmarking is, however, more nuanced. It is difficult to design benchmarks that are comparable between ICUs and implementation is fraught with challenges. Although adherence to clinical bundles has been shown to improve outcomes [1,2], clear evidence that benchmarking results in improvement in outcomes is scarce [3**]. However, with the right prerequisites and the implementation of appropriate measurements, benchmarking is a powerful enabler of healthcare quality improvement. Our experience benchmarking eight ICUs in The Netherlands indicates that benchmarking in ICUs is possible (assuming certain organizational prerequisites are met) and indicates a trend toward improved patient outcomes.

Benchmarking in medicine

A benchmark can refer to many things; therefore a discussion about benchmarking should ensure that the definition is clear. We refer to an ICU benchmark as a quantitative, standardized measurement that allows comparison of performance between ICUs.

Other industries have led the development of quality improvement programs. The main components being variability reduction, increased efficiency and measurement [4]. Powerful tools such as statistical process control and six sigma can be used to identify and reduce process variability. Lean and constraint theory help to improve efficiency by increasing throughput and/or reducing costs. Process measurement is an integral component of the Plan-Do-Study-Act cycle. The development of reliable benchmarks has since allowed organizations to compare the performance of their processes against their competitors.

Healthcare has been somewhat late in terms of adopting and adapting quality assurance tools from other industries,

but increasingly this is occurring [5,6,7*]. These tools supplement traditional medical quality assurance tools such as continuing education, case reports, clinical trials, journal clubs and professional responsibility. Currently, the overwhelming effort in Western healthcare systems relates to standardization of processes [8]. Developing and implementing order sets, clinical guidelines and ventilator assisted pneumonia (VAP) bundles are all methods of standardizing the way that clinical care is delivered. This results in variability reduction (and, if the standards are evidence based, also in higher quality care). Increasingly, the need to deliver care at a low cost is resulting in greater focus on efficiency improvement. The challenge here lies in aligning the goals of cost reduction and quality improvement – an alliance that often seems counter-intuitive, but has proven to be possible and fruitful [9]. In addition, tools for collecting information and measuring healthcare performance (such as electronic medical records) are increasingly available. For institutions that have successfully achieved these goals, benchmarking performance against other institutions is the next step.

Prerequisites for benchmarking

From our experience in ICU benchmarking we have identified prerequisites that increase the chance of developing and implementing successful benchmarks:

- (1) Intention to initiate quality improvement based on benchmarking results
- (2) Willingness to be compared with other ICUs
- (3) Presence of organizational factors for successful change management
- (4) Mature data collection systems
- (5) Stable processes (processes with minimal variation over time)
- (6) Comparable measurements (similar processes or processes with similar outcomes)
- (7) Standard measurement definitions

The most important of these is a desire to use the results to improve processes. Assuming this is the case, there are three main prerequisites for developing a reliable benchmark. The benchmarked process must be stable, the benchmark must compare the same process (in different locations) and the measurement must be clearly defined. For instance, benchmarking APACHE scores in an ICU the day before and after a postoperative care unit opens would not provide useful comparative data over this period. A length of stay benchmark including five neonatal ICUs and one adult ICU would be challenging to interpret because the processes are not necessarily comparable. Benchmarking staff satisfaction by using staff surveys in one ICU and measuring retention rates in another would create unreliable data.

The better that these prerequisites are met, the more reliable the benchmark. Of course, no process is perfectly stable, identical between locations or measured exactly the same way. A decision regarding acceptable adherence to benchmarking prerequisites must be made, with the understanding that greater adherence to the prerequisites leads to a higher quality benchmark. These prerequisites are relatively onerous for many ICUs, which should not be interpreted as a criticism but as an acknowledgment of the importance of other quality initiatives (such as variability reduction, efficiency improvement and measurement) as prerequisites for benchmarking.

ICU benchmarking in The Netherlands

Recognizing the value of benchmarking as a tool for initiating and enabling process improvement, our group began a process of benchmark development for ICUs in The Netherlands approximately 3 years ago.

Fortunately, many of the prerequisites for benchmarking have been present in Dutch ICUs for several years. Data collection (often using hospital information software systems) has been a requirement for hospital accreditation for many years. In addition NICE (National Intensive Care Evaluation), an organization involved in evaluation of ICUs in The Netherlands, has been active in benchmark development and participating ICUs collect a minimal dataset every month.

There is also increasing development and adoption of standardized clinical protocols within Dutch hospitals, which helps to ensure that the processes that are measured are comparable.

A final factor, which contributed to a climate conducive to benchmarking, was a government requirement that a subset of hospital performance and outcome data be publicly available. This led to a significant amount of discussion and activity related to data validity and its comparability between hospitals. This work on standardized reporting tools, in conjunction with similar work in other countries, was an important prerequisite for benchmark development.

Benchmark development

Our goal was to develop benchmarks relevant for management, professionals and patients that would be comparable between ICUs.

As a first step, we identified important features of benchmark measurements. Benchmark measurements should be:

- (1) Clinically relevant

- (2) Valid (a change in the measurement should result in a change in the outcome of interest)
- (3) Robust (a small change in the measurement should not result in a large change in the benchmark)
- (4) Comparable (the measurement is taken the same way, and means the same thing in compared ICUs)
- (5) Standardizable (it is possible to compare the measurement between different hospitals)
- (6) Anonymous (measurements should not identify specific patients or staff)
- (7) On the basis of available, accurate, regularly collected data
- (8) Minimal added work

Our work resulted in the identification of three benchmark domains: quality, efficiency and availability. A well functioning ICU delivers high quality care, has beds available when needed and uses its expensive resources efficiently (less than 10% of a hospital's beds generate more than 1/3 of its costs) [10,11]. Within each domain, component benchmarks were developed based on available research, available data and expert opinion.

The three benchmarking domains provided a means of aggregating data in such a way that significant trends could be recognized even if individual component benchmarks were conflicting. In addition, the goal of our benchmarking development was to provide a strategic tool for hospitals – not to meet external reporting requirements or to serve as a disciplinary tool. Therefore benchmark data was anonymized (for hospital, employees and patients) and was not publicly available (unless hospitals wished to publish their own data).

Our benchmarks were designed to complement but not replace the NICE benchmarks, and therefore differ in several ways. First, the NICE program collects data continuously, whereas our benchmarks are designed to collect more detailed data over a defined period of time (usually 6 weeks) to provide a detailed 'snapshot' of performance. Our benchmarks are also more detailed in several areas, for instance we include component benchmarks for line sepsis, VAP and sedation level that are absent from the NICE benchmarks. We also include more component benchmarks that capture process parameters and efficiency performance such as productivity of physicians and nurses.

Finally, we felt that it was important that benchmarks provide information that could be used to guide quality improvement efforts. To facilitate this, we identified correlations between component benchmarks (based on available evidence) that were reported back to participating ICUs. For example, sedation level and VAP incidence were correlated with the expectation that this

information may lead to better adherence to sedation guidelines, and a decrease VAP incidence.

Benchmark components

A full list of our benchmark components is listed (Table 1). The main components in each domain are described briefly here.

Quality

The most important component benchmark of quality is the Standardized Mortality Ratio (SMR). This measurement allows comparison between expected and actual mortality based on the illness severity of patients. Measurement of complication rates also plays an important role in our quality benchmark.

Availability

Availability is made up of two parameters, the availability of a bed for patients with an ICU indication (based on the APACHE score) and the reduction of bed use for patients without an ICU indication. Refusal rates for referrals from within the hospital, and from other institutions are important measurements of availability.

Efficiency

We benchmark ICU efficiency using measurements such as the length of stay and duration of ventilation. Another component of efficiency is the bed utilization, with the understanding that very high bed utilization may impact negatively on the availability of the ICU for a new admission. Employee productivity is also benchmarked.

Results

Since the introduction of SMR benchmarking in The Netherlands, the average countrywide SMR has decreased from approximately 90 to 70% (a 3% per year decrease) (Fig. 1) [12].

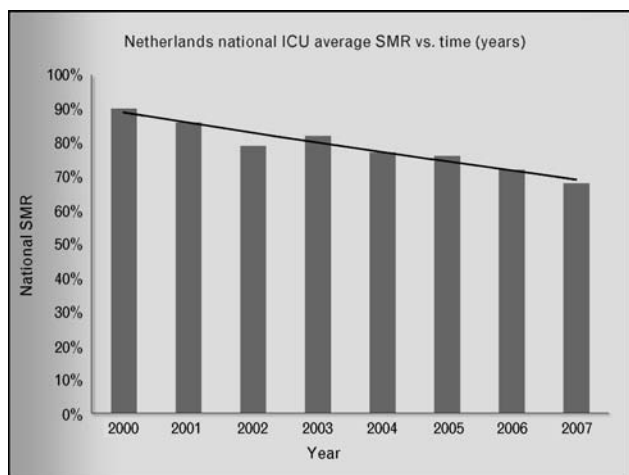
We have seen a large variance in the incidence of VAP in Dutch ICUs (ranging from 13 to 37 per 1000 ventilator-days), which is relatively higher than in other countries. We have seen a similarly large variance between ICUs in incidence of catheter-related sepsis (ranging from 0 to 37 per 1000 catheter-days), decubitus ulcers > Grade 1 (ranging from 3.7 to 10.9 per 1000 admission days) and severe sepsis (ranging from 0 to 11 per 1000 admission days) [13]. These large variances indicate that there may be opportunity to implement best practices from ICUs that have lower complication rates.

A reduction in sedation level has been achieved in several ICUs associated with a decrease in VAP per 1000 ventilation days (Fig. 2) [14]. Although not evidence of a causal relationship, this pattern fits with the expected

Table 1 ICU benchmark components

Type	Availability	Efficiency	Safety
Size	Always beds available	Ratio ventilator days/admission days	SMR (Standardized mortality ratio)
Number of admissions	External refusal rate	% admissions > 7d	Patients with ICU indications
Length of stay	Internal refusal rate	Plannability	Medical patients
Type of patients	Operation cancellation rate b/c ICU availability	Unplanned admissions OR Ward Internal	Surgical
Severity score (APACHE, SOFA)	Postponed admissions	After planned OR	Planned admissions
Preadmission location (Surgery, Medical, Other)	Admission when no beds available (redundant admissions)	Ratio POSSUM planned/unplanned (Physiologic and operative severity score for the enumeration of mortality and morbidity)	Unplanned admissions
Meeting ICU Admission criteria	Discharges outside daytime hours	Scheduling of planned admissions	>65 years
Source	Improper use	Bed utilization	<65 years
ICU diagnosis	ICU indication for admission	Overall (optimal = 80%)	Complications
Preadmission service	% patients ventilated <24 h	Personnel	At admission
Preadmission location	Severity (APACHE)	Nursing productivity	Infections
Discharge location	Discharged patients leave ICU	Patients/diploma nurse	Decubitus
	Postponed discharges	TISS (Therapeutic Intervention Score System) Patient/TISS nurse	Line sepsis
	Average SOFA score	Productivity per physician	New sepsis
	Readmissions within 48 h	Presence in ICU	Other infections
		% of procedures done	Decubitus
		Revenue/physician	Readmissions <48 h
		Patient/physician	Sedation
		Length of stay/physician	Glucose Control (NVIC guidelines)
		Ventilator days/physician	Glucose (IHI guidelines)
			Safety culture
			Cooperation between department
			Teamwork inside department
			Good shift changeovers
			Reporting frequency
			Nonpunitive reporting
			Feedback and learning from mistakes
			Expect action from managers
			Management support
			Adequate manpower resources

Figure 1 Average Standardized Mortality Rates on Dutch ICU's participating in the NICE registration show a yearly 3% reduction from 2000–2007



effect of decreasing sedation levels to shorter ventilation duration.

An interesting finding has been that there are significant differences in average length of stay between ICUs. Six ICUs with a comparable illness severity (APACHE) had an average length of stay of 2.3–6.3 days (Fig. 3) [14]. We also found a trend of lower infection rates per 100 admission days in larger ICUs [15]. This appears to show a quality benefit from larger scale.

At the beginning of our benchmarking program we theorized that the result of benchmarking between hospitals would be a gradual convergence on ‘best practices’. Interestingly, we have instead seen an overall

Figure 2 Average sedation level correlated to incidence of VAP per 1000 ventilation days

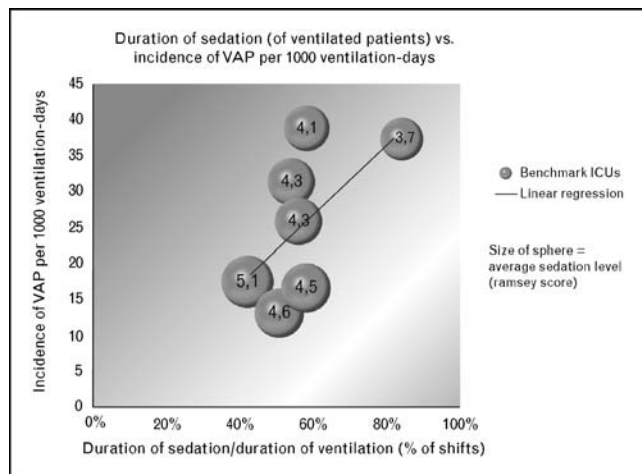
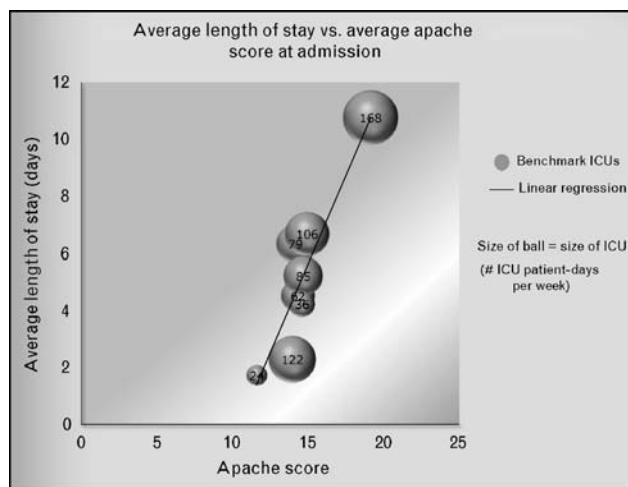


Figure 3 Average length of stay vs. average APACHE score at admission



shift with time, in that all ICUs appear to improve performance in such a way that the relative difference between the ‘best’ and ‘worst’ performing ICUs is generally preserved. This effect has also been noted previously by Pronovost during his quality improvement work in ICUs [3**]. Whether this trend will continue, remains to be seen.

Conclusion

Although there is great potential value in benchmarking of ICUs, there are currently significant challenges to implementation of appropriate benchmarks in most ICUs. When benchmarking is considered, the primary role of benchmarks as a comparative indicator of process performance between ICUs requires involvement of multiple ICUs and the choice of appropriate benchmarks.

The evidence of improved outcomes in benchmarked ICUs is not yet conclusive, although early trends look promising. Although the benchmarks themselves do not increase safety, they enable committed teams to better target quality improvement initiatives and measure the relative success of their efforts. Our work in developing and implementing benchmarks in ICUs has shown that most ICUs improve in the majority of the component benchmarks over time, and that the relative gap between the highest and lowest performing ICUs remains approximately stable.

As the prerequisites for ICU benchmarking are more widely achieved in ICUs, we expect that benchmarking will play an increasingly important role as a tool to identify variation between ICU performances both within and between countries. We expect that quality

improvement initiatives based on this data will help ICUs to increase efficiency, availability and patient safety.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest

Additional references related to this topic can also be found in the Current World Literature section in this issue (p. 465).

- 1 Pronovost PJ, Needham D, Berenholtz S, *et al.* An intervention to decrease catheter-related bloodstream infections in the ICU. *N Engl J Med* 2006; 355:2725–2732.
- 2 Bonello RS, Fletcher CE, Becker WK, *et al.* An intensive care unit quality improvement collaborative in nine departments of Veterans Affairs hospitals. *J Comm J Qual Patient Saf* 2008; 34:639–645.
- 3 Pronovost PJ, Berenholtz SM, Goeschel C, *et al.* Improving patient safety in intensive care units in Michigan. *J Crit Care* 2008; 23:207–221.
This paper indicated that a large-scale improvement in safety culture can be achieved using a standardized and feasible safety program, and that improvements in safety culture occurred concurrently with improved patient outcomes.
- 4 Wikipedia. W. Edwards Deming [document on the internet]. http://en.wikipedia.org/wiki/W._Edwards_Deming. [Accessed 17 April 2009]
- 5 Peek N, Goud R, Abu-Hanna A. Application of statistical process control methods to monitor guideline adherence: a case study. *AMIA Annu Symp Proc* 2008; 581–585.
- 6 Messahel F, Al-Qhatani A. Benchmarking of World Health Organization surgical safety checklist. *Saudi Med J* 2009; 30:422–425.
- 7 Muder R, Cunningham C, McCray E, *et al.* Implementation of an Industrial systems-engineering approach to reduce the incidence of methicillin-resistant staphylococcus aureus infection. *Infect Control Hosp Epidemiol* 2008; 29: 702–708.
The results of this study indicate that quality improvement tools from other industries, in this case the Toyota Production System, can be adapted and applied to healthcare and that this can result in improved patient outcomes.
- 8 McMillan T, Hyzy R. Bringing quality improvement into the intensive care unit. *Crit Care Med* 2007; 35 (2 Supplement):S59–S65.
- 9 Kahn J, Rubenfeld G, Rohrbach J, Fuchs B. Cost savings attributable to reductions in intensive care unit length of stay for mechanically ventilated patients. *Med Care* 2008; 46:1226–1233.
- 10 Shorr AR. An update on cost-effectiveness analysis in critical care. *Curr Opin Crit Care* 2002; 8:337–343.
- 11 Dasta JF, McLaughlin TP, Moody SH, *et al.* Daily cost of an intensive care unit day: the contribution of mechanical ventilation. *Crit Care Med* 2005; 33: 1266–1271.
- 12 National ICU Registration Data. NICE (Nationale Intensive Care Evaluatie); 2006.
- 13 ICU benchmark database. Plexus; 2008.
- 14 Houtepen J, Van Bommel K, Berg M, Brouwer RML. Safety and efficiency on ICU. The ICU benchmark (in Dutch). In: *Op het Scherp van de Snede*. Edition 1. Amsterdam: Plexus; 2008; pp. 58–75.
- 15 ICU benchmark database. Plexus; 2008.