Air Medical Journal ■■ (2018) ■■–■■

Contents lists available at ScienceDirect

Air Medical Journal

journal homepage: http://www.airmedicaljournal.com/

Original Research

Setting the Benchmark for the Ground and Air Medical Quality in Transport International Quality Improvement Collaborative

Christine R. Aspiotes, DO [1], Matthew Q. Gothard [2], M. David Gothard, MS [2,3], Rollie Parrish, RN, BSN [4], Hamilton P. Schwartz, MD [5], Michael T. Bigham, MD [6,*]

[1] *Department of Pediatrics, Division of Medical Education, Akron Children's Hospital, Akron, OH*
[2] *BIOSTATS, Inc, East Canton, OH*
[3] *Akron Children's Hospital, Rebecca D. Considine Research Institute, Akron, OH*
[4] *Flightweb, Spokane, WA*
[5] *Department of Pediatrics, Division of Emergency Medicine, Cincinnati Children's Hospital Medical Center, Cincinnati, OH*
[6] *Department of Pediatrics, Division of Critical Care Medicine, Akron Children's Hospital, Akron, OH*

A B S T R A C T

Objective: Critical care transport (CCT) supports regionalization of medical care. Focus on the quality of CCT care prompted the development of the Ground and Air Medical qUality in Transport (GAMUT) Quality Improvement collaborative database which tracks consensus quality metrics. The Institute of Medicine recommends benchmarking of comparative data to accelerate improvement. Herein, we report the strategies and rationale for GAMUT QI Collaborative benchmarking.

Methods: The GAMUT database includes >350 programs internationally with >200,000 annual patient contacts. Evidence-based literature review performed in May 2016 and October 2017 identified benchmarking strategies were evaluated and summarized, specific to the GAMUT metrics. Statistical analyses include simple statistics and weighted expectation calculations for benchmark examples (Pearson chi-square with Bonferroni adjusted post-hoc z tests).

Results: Evidence-based literature search yielded 70 articles, and 31 were selected for inclusion in our evidence table. 5 evidence-based benchmark strategies were considered: average (mean), average (median), adjusted benchmark (based on expected outcome), Achievable Benchmark of Care (ABC), and Delphi. ABC threshold establishes a higher target (90th percentile) forcing more programs to achieve higher performance.

Conclusion: Benchmarking is not well-suited for a single strategy and requires customized consideration based on each metric, though adjusted benchmark and ABC generally set higher performance benchmarks.

Historically, the field of critical care transport has primarily focused on the task of safely moving a patient from one location to another in search of definitive medical care. Until only recently, priority has been placed on the single outcome measure of patients' survival during transport. Much less attention has been spent on other process and outcome measures concerning the quality of the medical care provided en route to the patients' destinations. In accordance with the medical field's renewed focus on high-quality care, the field of critical care medical transport has evolved. In 2007, the Institute of Medicine described a 3-step strategy to accelerate improvement in health care quality measurement focusing on measuring, reporting, and improving. This improvement acceleration paradigm has served as the foundation for the recent critical care transport–specific initiatives to improve the quality of care. It has also highlighted the importance of identifying and applying a benchmarking strategy to distinguish high-performing organizations.[1]

Benchmarking is the process of creating a standard that can be measured and then analyzed in comparison with others.[2] Comparative benchmarking has been used in business and manufacturing

Author Disclosures: The authors have no conflicts of interest or financial relationships to disclose relative to this manuscript.

Author Affirmation: CA, MG, MG, RP, HS and MB were involved in study design, data collection, data analysis, data review, and manuscript preparation.

* Address for correspondence: *Michael T. Bigham, MD, FAAP, Akron Children's Hospital, One Perkins Square, Akron, OH 44308-1066.*

E-mail address: mbigham@chmca.org (M.T. Bigham).

1067-991X/$36.00

https://doi.org/10.1016/j.amj.2018.03.002

Table 1
Medical Subject Heading Search Terms Used in Recent PubMed Queries

Query	Items Found
Search benchmark* AND (establish* OR develop* OR defin*) AND (quality OR performance OR improvement) Filters: Publication date from 2014/01/01 Sort by: [pubsolr12]	1,406
Search benchmark* AND (establishing OR establishment OR establish OR established) AND (quality OR performance) Filters: Systematic Reviews Sort by: [pubsolr12]	142
Search benchmark* AND (establishing OR establishment OR establish OR established) AND (quality indicators) Filters: Systematic Reviews Sort by: [pubsolr12]	21
Search benchmark* AND (establishing OR establishment OR establish OR established) Filters: Systematic Reviews Sort by: [pubsolr12]	299
Search benchmark* AND (establishing OR establishment OR establish OR established) Sort by: PublicationDate	3,404
Search benchmark* AND (establishing OR establishment OR establish OR established) AND database* Sort by: PublicationDate	423
Search benchmarking AND (establishing OR establishment OR establish OR established) Sort by: PublicationDate	1,292
Search establishing benchmarks transport Sort by: PublicationDate	4
Search establishing benchmarks Sort by: PublicationDate	345
Search "Benchmarking/methods"[Majr] OR "Benchmarking/standards"[Majr] Sort by: PublicationDate	1,198
Search "Benchmarking"[Majr] AND (Review[ptyp]) Sort by: PublicationDate	398
Search ("Benchmarking/methods"[Majr] OR "Benchmarking/standards"[Majr]) AND (Review[ptyp])) Sort by: PublicationDate	144
Search benchmark* AND (establish* OR develop* OR defin*) AND (quality OR performance OR improvement)	70

to assist companies in evaluating their structure and processes against industry leaders. With recent health care initiatives focusing on quality improvement (QI), benchmarking is being used to help determine and prioritize QI opportunities in the health sector. One of the key criteria in determining an appropriate benchmarking strategy is the type of variable being measured. Typical benchmarking techniques in health care can be broadly divided into 2 separate categories: midleaning benchmarks and best performances.

There is currently no consensus on an appropriate benchmarking strategy to compare quality metrics in critical care transport. The Ground and Air Medical qUality in Transport (GAMUT) QI Collaborative is an international cohort founded in 2013 focused on accelerating improvements in transport medicine. Participants voluntarily submit monthly data on consensus quality metrics. The data are reported individually by program and as a cohort, allowing participating programs to see their relative performance compared with others. The intent is to spur better performance by each individual program in specific areas of care and, in doing so, improve overall quality in the field of critical care transport medicine. Current GAMUT biannual reports display an arithmetic mean as a matter of convenience. This review aims to describe different benchmarking techniques and establish the rationale for selecting a benchmarking strategy for the GAMUT QI Collaborative.

A literature review was performed in PubMed in May 2016 and again in October 2017. The goal was to select very relevant articles on the quality benchmarking strategies available. The key word "benchmarking" (all key words truncated to include as many results as possible) was searched for in combination with establishment/definition/development and quality/performance/improvement (Table 1). Reviews and systematic reviews were given higher priority as well as articles indexed with the main subject heading "benchmarking." We limited the search to articles from 2014 to present. This search yielded 70 articles that were included for review by the team (C.R.A. and M.Q.G.), and 31 were subsequently selected for inclusion in our evidence table based on their applicability (Table 2).[3-33]

The first and most conventional of the midleaning benchmarking categories is the average or arithmetic mean. The average defines the number as a measure of centrality, allowing organizations to determine whether they are above or below that middle value. The greatest weakness to using the mean is that extreme outliers may significantly shift this number and may lead to confusion interpreting the validity of the benchmark. Despite the midleaning nature and risk of skewness when using mean-based benchmarking, there is an ease to calculating the mean, which strengthens the appeal of this methodology. Multiple studies use the mean as a benchmark to measure and compare data.[3-10] The mean is not the only midleaning benchmark strategy.

A second midleaning benchmarking category, which is closely related to the arithmetic mean is the median. The median discounts the effects of skewness to improve on the mean as a benchmarking strategy and mitigates the effects of outliers. The median, like the mean, still proves difficult in delineating performance relative to the other programs; however, studies evaluating the generalization of care and outcomes are easily evaluated.[11,12,30-33] A study reviewing increased evidence-based treatment for community-acquired pneumonia made use of the median benchmark in regard to antibiotics chosen pre- and posthospitalization.[12] Both the average and median give the perception that mediocrity is a quality goal by not identifying the more rarely encountered, outstanding performance by certain programs. The midleaning benchmark also does not take into account factors that may affect performance evaluation such as varying populations, which could drastically influence outcomes.

The third midleaning benchmarking methodology is the adjusted benchmark. As the name suggests, this benchmark adjusts for different factors based on an expected outcome. One example could be the outcomes of patients cared for in a given hospital after adjustment for illness severity.[14] Each incremental increase in illness severity yields a different expected outcome, allowing hospitals to find their own expected outcome based on the known illness severity of their patients. It is similar to the previously described midleaning methods using mean or median calculations; however, this method compares programs to an adjusted average instead.[13-20] One such illustration can be seen in a recent GAMUT report of first-attempt intubation success.[21] The populations studied were adult, pediatric, and neonatal patients undergoing interfacility transport. This study suggested that adjustments are required when comparing program-level performance to account for the more technically difficult neonatal intubations. As such, the adjusted benchmark was used when comparing these data. Mathematically, the following weighted average equation was used to calculate the adjusted benchmark for transport intubations and accounts for different levels of intubation difficulty, yielding a better comparison among programs with differing populations: (.595 * neonate attempts) + (.817 * pediatric attempts) + (.87 * adult attempts).

Adjusted benchmarking improves on the mean and median yet still depicts a mediocre goal as the benchmark relative to how a typical program would perform after adjustment. Also, the adjusted benchmark requires appropriate identification of relevant factors that should be adjusted for and the effects of those factors to be determined a priori.

All 3 midleaning benchmark strategies create a binary performance scoring system of high and low performers. The result of a single midleaning target is a threshold that serves to kick-start the low performers and assure the high performers. Kick-starting low performers is desirable although above average–performing organizations must still strive to improve and to not simply accept what may still be mediocre care. Hence, the greater risk to the patients