Case Report

Development and usability testing of an audit and feedback tool for anesthesiologists

Atilio Barbeito[1] and Noa Segall[2]

[1]Department of Anesthesiology, Duke University Health System & Veterans Affairs Healthcare System, Durham, North Carolina, USA and [2]Department of Anesthesiology, Duke University Medical Center, Durham, North Carolina, USA

Corresponding Author: Atilio Barbeito, MD, MPH, Department of Anesthesiology, Duke University Medical Center & Veterans Affairs Healthcare System, Box 3094, Durham, NC 27710, USA (atilio.barbeito@duke.edu).

ABSTRACT

Background: We describe the creation and evaluation of a personal audit and feedback (A&F) tool for anesthesiologists.

Methods: A survey aimed at capturing barriers for personal improvement efforts and feedback preferences was administered to attending anesthesiologists. The results informed the design and implementation of 4 dashboards that display information on individual practice characteristics as well as comparative performance on several quality metrics. The dashboards' usability was then tested using the human-centered design framework.

Results: Anesthesiologists listed lack of information on current practice as the main barrier for improvement. Regarding usability, participants gave the dashboards an average score of 3.8 (scale 1–5) on consistency, learnability, and information organization, and performed the assigned tasks well, with an average score of 89% (range 79–100%).

Conclusions: We describe the design, implementation, and usability testing of an innovative tool that utilizes data derived from the electronic health record (EHR) system to provide A&F to anesthesiology providers.

Key words: feedback (learning), electronic health records, anesthesiology, quality improvement

INTRODUCTION

Audit and feedback (A&F), the summary of clinical performance (audit) over a specific period of time, and the provision of that summary (feedback) to individual practitioners, teams, or health-care organizations,[1,2] is based on the belief that health-care professionals are prompted to modify their practice when feedback shows that their clinical practice is not consistent with a desired target.[3] In order to be most effective, feedback should be timely, intensive, originate from a trustworthy source, be confidential and nonjudgmental, be supported by the broader organization, be supplied continuously over time, and be integrated within a broader quality improvement (QI) framework.[4]

Business intelligence tools are being increasingly used in health care to drive improvement activities. Specifically, dashboards summarize and display meaningful information in a way that is easily consumed by clinicians.[5] Dashboards have the potential to leverage the vast amounts of data generated through the electronic medical record to provide A&F to providers. While others have reported on the utilization of dashboards for operating room management,[6] this concept has not been to our knowledge leveraged for personal quality improvement purposes within the specialty of anesthesiology, except in one report where physicians' efficiency was profiled and tied to financial incentives.[7] In this report, we describe that the development and usability testing of a series of web-based, interactive QI dashboards for anesthesiology providers that includes near real-time information about their practice characteristics and performance.

METHODS

The development and implementation of the dashboards were deemed a quality improvement activity. The usability testing portion

of the study was approved by the Duke University Medical Center Institutional Review Board (IRB), and written informed consent was obtained. The Department of anesthesiology at Duke University is currently composed of 154 anesthesiologists (113 at the time of the survey) who provide clinical care across nine main operating room locations. The department is organized in 10 clinical divisions according to subspecialty and work location.

Determinants of practice and feedback preferences

All clinical faculties in the department were surveyed using an online tool. Questions were designed to capture information in 2 domains: (1) determinants of practice (barriers and facilitators) for improvement activities; and (2) feedback preferences regarding quality domains and metrics. The survey included questions such as "what is preventing you from improving?" and is shown in Supplementary Appendix A. We also wanted to assess respondents' level of self-knowledge regarding their practice as a potential barrier for improvement. We asked questions such as "On average, how many minutes do your patients stay in postanesthesia care unit (PACU)?" In order to compare perceived versus actual practice, respondents' answers were compared to their respective divisional average values for the month when the survey was administered.

Creation of clinical dashboards for anesthesiologists

The survey results were analyzed in the context of institutional and departmental priorities. With this, the 4 prototype dashboards were produced.

Accessing the data

The Duke University Health System's Epic Systems enterprise data management solution is currently being leveraged to obtain, store, clean, and normalize the EHR system data of interest to anesthesiologists through the Duke Department of Anesthesiology's data mart project. We used this data mart to compose the quality metrics and populate the QI dashboard prototypes.

Creating prototype dashboards

We used commercially available analytics and data visualization software (Tableau Desktop Professional 10.0, Seattle, WA, USA) to create 4 interactive dashboard prototypes. Three of the 4 prototypes, together with descriptions of the metrics displayed in each, are shown in Figures 1–3. The fourth prototype, "My Cases," displays detailed information on individual cases and is not shown.

The visual display types chosen to present the different metrics were selected using classic data visualization principles.[8,9] For example, we chose line graphs to display time series of continuous data (eg, transfusion rate over time), as the line connecting the data implies a connection between the data points, making the display intuitive and easy to interpret. Horizontal bar graphs were chosen for categorical data (eg, transfusion rates by procedure) because they are common and easy to interpret, always making sure that they have a zero baseline. A treemap (a type of area chart) was chosen to visualize data in nested categories where numbers may have vastly different magnitudes (eg, number of cases by specialty and by surgeon). Hue was used strategically to focus the user's attention to the data point of interest (eg, to compare personal transfusion rates to that of peers). When appropriate, we used simple numerical displays (transfusion rate for the current period) or stars (3 star efficiency rating) to summarize performance. The ultimate goal was to *communicate* performance data in different quality domains accurately and

efficiency using sound visualization principles such as (1) providing context, (2) selecting the best display for each type of data, (3) eliminating unnecessary detail (clutter), and (4) focusing the provider's attention on the important aspects of the data.

Distribution

These dashboard prototypes were iteratively refined based on informal feedback provided by prospective users and made available to each anesthesiology attending provider through the Duke Anesthesiology Department's intranet webpage. The dashboards are configured to update daily. Users may interact with the data in all of these dashboards, sorting, or filtering according to patient groups, surgery dates, or case types, among other categories.

Usability testing

Approximately 3 months following the implementation of the last dashboard, we examined their usability with 7 participants using the human-centered design (HCD) framework. An experienced human factors engineer performed the evaluation and proposed design changes. Volunteers were attending anesthesiologists who were asked to complete a total of 18 tasks on the 4 dashboards. Participants were asked to "think aloud," or describe their thoughts, as they carried out these tasks (Supplementary Appendix B).

Impact assessment

The perceptions of providers regarding the QI dashboards were assessed using surveys and interviews. The usability testing subjects were interviewed about what they liked most and least about the dashboards and how they would improve existing dashboards. They also completed a background survey and a satisfaction survey (adapted from the previously validated Questionnaire for User Interface Satisfaction[10]) eliciting their reactions to the dashboards.

RESULTS

Determinants of practice and feedback preferences

A total of 88 faculty members from 9 of 10 divisions responded to the initial survey, for a response rate of 78%. Regarding barriers to personal improvement activities, lack of information on current practice was listed as the main barrier for improvement (mean score 2.46, SD 0.95, on a 1–4 scale). Other barriers that were identified included not having enough time to keep up with the scientific literature, lack of improvement goals, administrative duties, research projects, and teaching responsibilities, in decreasing order of importance (Table 1).

Regarding A&F preferences, responders listed safety and effectiveness as the two quality domains of most importance, followed by patient satisfaction, efficiency, and compliance, in decreasing order. Adverse events, patient satisfaction scores, opioid use, and transfusions were listed as the most desirable metrics (Table 1).

We found discrepancy between estimated and actual practice among providers in the 8 divisions whose data were available for comparison. The average, estimated divisional case volume was 24.7% (−20.4, 54.0%) lower than the actual case volume. "Prep" time and PACU length of stay (LOS) had a discrepancy of 3.1% (−65.1, 33.4%) and 45.9% (21.3, 65.0%), respectively (estimated PACU LOS and Prep time were shorter than actual times). Other

Figure 1. *My Practice* dashboard. The top panel shows a column chart of the total daily case volume for the user over a preselected time interval (from the drop down menu on top). Colors represent the proportions of general versus nongeneral anesthesia cases for each day. The middle left panel, "My Procedures," shows a bar chart of volume of interventional procedures performed over the selected time interval. The bottom two left panels show a frequency histogram of the age distribution of total cases over the selected time interval, with colors representing proportions of male/female for each age category, and a horizontal stacked bar chart with the proportion of patients in each American Society of Anesthesiologists (ASA) class. The panel on the right displays a treemap of the cases performed, grouped by surgical specialty (different color) and surgeon (different box). The size of the individual boxes is determined by the volume of cases performed with that particular surgeon. The specific details of each group of cases may be displayed by hovering over the square with the pointer. (Data are for demonstration purposes only and do not represent Duke practice.)
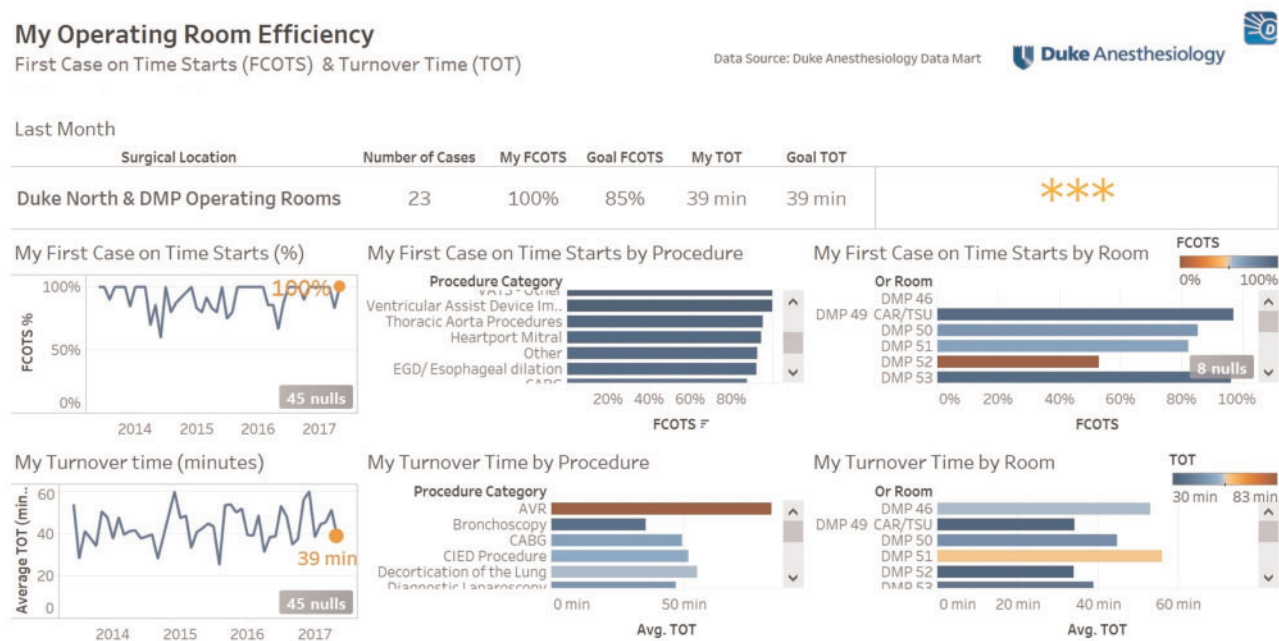

Figure 2. *My Operating Room Efficiency* dashboard. The top section (last month) displays the number of qualifying cases, and the provider's efficiency performance, as measured by first case on-time starts (FCOTS, the percentage of first cases that entered the operating room before or at the scheduled start time) and turnover times (TOT, the average time in minutes between in-OR time and out-of-OR time for the prior case). It also displays institutional efficiency goals for each anesthetizing location, as determined by hospital leadership A 3 star rating is assigned depending on whether the provider achieves both, one, or no goals that month. The time series below display historical trends in FCOTS and TOT, and the FCOTS and TOT by procedure category and by operating room. (Data are for demonstration purposes only and do not represent Duke practice.)
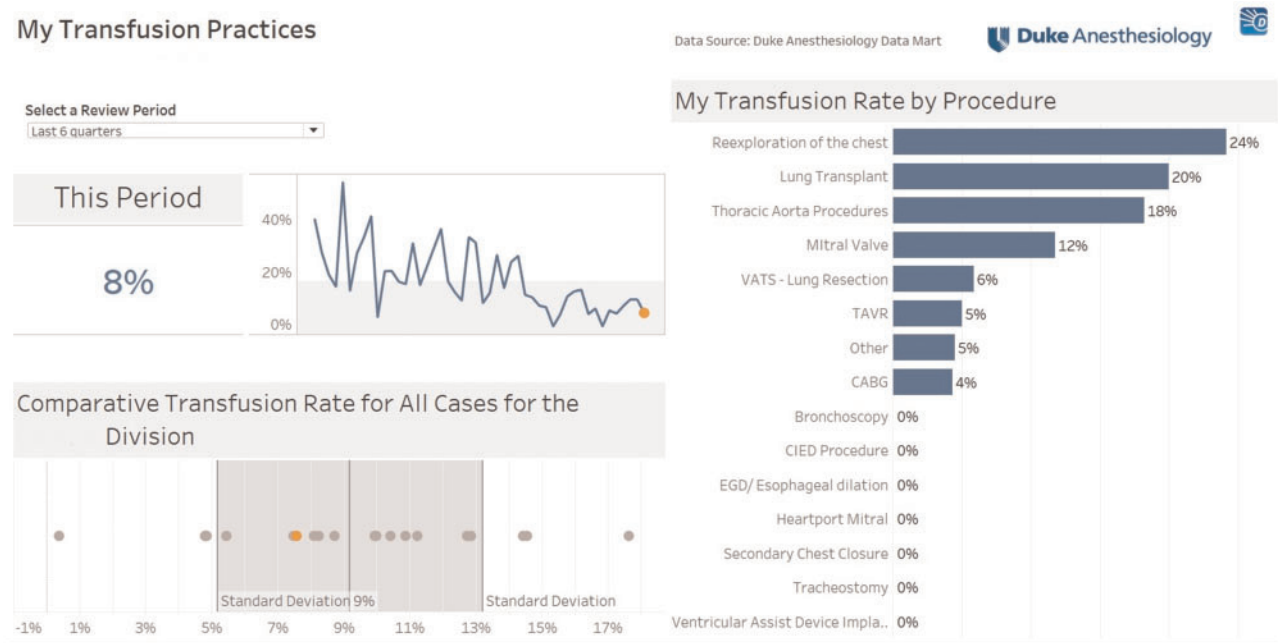
Figure 3. *My Transfusion Practices* dashboard. Provides the user's overall transfusion rate (the percentage of patients who received an intraoperative transfusion of at least one blood product—red blood cells, platelets, plasma, and/or cryoprecipitate) over time for the time period selected, together with a historical trend of the transfusion rate. Below, the graph displays the user's transfusion rate (orange point) in relation to the divisional average ($\pm$ 1 SD) and to his divisional peers (grey points). To the right, the transfusion rate by procedure for the user's most commonly performed procedures is displayed using a horizontal bar chart. The bars for each procedure category act as filters, allowing the user to compare personal transfusion rates for individual procedures against other divisional members. (Data are for demonstration purposes only and do not represent Duke practice.)

discrepancies between estimated and actual practice are shown in Supplementary Appendix C.

Clinical dashboards—usability testing

Participants

Data are available from 7 attending anesthesiologists. Volunteers ranged in age from 37 to 64 and 6 were males. Participants spent 2–4 days in the operating room (OR) per week. Their frequency of using the dashboards ranged from "once" to "monthly."

Dashboard functionality

In general, participants were pleased with the dashboards. They liked being able to see their own performance data and found them useful. They appreciated the ability to sort and filter data in an effort to find opportunities for improvement. Six of seven participants somewhat or strongly agreed with the statements, "Using the dashboards could help me improve my clinical performance" and "I would prefer to receive performance feedback through the dashboards than through other means."

Two issues were of concern to participants. The first was trust in the data. Six of seven participants explicitly mentioned this, that is, "If I don't trust some of the data, I lose faith in all of it." Distrust occurred when participants perceived data to be inaccurate. This made participants "question the data, and could make people frustrated and less willing to use the dashboards." Most commonly, participants questioned the data presented in the all-encompassing "Other" category, used to capture data for those operations that did not fit the more commonly performed surgical procedures done in each specialty. In addition, some metrics were perceived to be driven by factors beyond the anesthesiologist's control, that is, LOS and

turnover times, which are affected by other clinicians, and division transfusion averages, which are affected by anesthesiologists with a different case mix.

Related to data reliability were questions about the "other" category of cases, which appeared in multiple dashboards (the "other" category lumps cases that have not been included into any specific case category, and cases done with less frequency by a particular provider). "Other" cases made the data frustrating and difficult to interpret to 5 of 7 participants.

Suggestions made by the participants regarding new features for the existing dashboards as well as suggestions for new dashboards are described in Table 2.

Dashboard usability

The average usability score was 3.82 on a scale of 1 (low) to 5 (high) on characteristics such as consistency, learnability, and information organization, with a standard deviation of 1.07. There was strong agreement with statements such as, "Learning to use the Dashboards is easy" and "System speeds are fast enough". They also performed the assigned tasks well, with an average score of 89% (range: 79–100%). The dashboards' usability was generally good, with only minor suggestions for improvement (Table 2).

DISCUSSION

We describe the design, implementation, and usability testing of an innovative tool that utilizes data derived from the EHR system to provide A&F to anesthesiology providers at a large academic medical center. In an initial survey to all faculty members, respondents listed lack of information on their current practice as the main

Table 1. Determinants of practice and feedback preferences survey results

Think about the type of information that might help you to improve your practice. Would it help you to know more about…(1 = not helpful to 4 = most helpful)	Mean score	SD
Patient satisfaction scores	2.74	0.90
Total opioid use	2.48	0.84
Intraoperative hemodynamics	2.42	0.85
Transfusions	2.41	0.86
Anesthesia "prep" time	2.34	0.89
First case on-time starts	2.24	0.99
Amount of anesthetic used	2.21	0.80
What is preventing you from improving (1 = not a barrier to 4 = a major barrier)?		
No information on my current performance	2.46	0.95
Not enough time to keep up with the scientific literature	2.22	0.88
Lack of improvement goals	2.19	0.91
Administrative duties	1.78	0.69
Research projects	1.68	0.73
Teaching responsibilities	1.33	0.58
Please rank from 1 to 5 based on the importance you give to each of these aspects of care in your practice (1 = most important to 5 = least important)		
Safety (adverse events such as medication errors, line complications, etc.)	1.54	0.61
Effectiveness (in treating pain, nausea and vomiting, resuscitation, etc.)	2.43	0.99
Patient satisfaction	2.83	1.15
Efficiency (in room time, turnover time, etc.)	3.89	0.71
Compliance with billing and documentation requirements	4.31	0.84

Table 2. Suggestions for current and new dashboards provided by the study subjects during the usability testing

Suggested enhancements to the current dashboards
Displaying data by provider (resident/CRNA)
In transfusions, showing how many patients of those who received blood had a hemoglobin >7
Adding comparisons to department/division averages (similar to the Comparative Transfusion Rate chart)
Customizing dashboards by division through input from division chiefs/representatives
Adding a help menu and an introduction
Allowing users to easily provide feedback about the dashboards
Showing data for procedures done outside the OR, like interventional radiology
Tying metrics to departmental safety priorities
Suggested topics and metrics for new dashboards
Surgical outcomes, such as surgical site infections, DVT, stroke, and other complications
Prolonged use of opioids following surgery
ICU length of stay, sorted by procedures
PONV and pain (eg, pain score, time to antiemetics)
PACU metrics, like naloxone administration and reintubation rate
Neuromuscular blocking drug use and reversal
Lung protective ventilation
Depth of anesthesia in elderly patients
Antibiotic administration and related prolonged LOS
Use of postoperative multimodal analgesia
Teamwork metrics
A dashboard for division chiefs

Abbreviations: CRNA = Certified Registered Nurse Anesthetist; DVT = deep venous thrombosis; ICU = intensive care unit; PONV = postoperative nausea and vomiting; PACU = postanesthesia care unit; LOS = length of stay.

impediment for improvement activities, emphasizing the need for A&F interventions. Using the same survey, we documented a discrepancy between perceived and actual practice. This is important, because accurate knowledge of one's practice is a pre-requisite for any improvement effort.

Clinical dashboards "provide clinicians with the relevant and timely information they need to inform daily decisions that improve the quality of patient care. They enable easy access to multiple sources of data being captured locally, in a visual, concise, and usable format."[11] Dashboards differ from more traditional performance reports in several ways: (1) they provide summary data on performance against benchmarks or performance goals; (2) they use data visualization techniques, which allow for effective and efficient communication of complex data; (3) they may allow users to interact with the data (eg, filters or drill-down capabilities); (4) they provide more timely—real time or near real time—feedback to end-users, in contrast with more traditional summative episodic reports.

The web-based QI dashboards described here are intuitive and easy to use, and contain meaningful QI metrics designed with provider input. The dashboards were tested and refined using a HCD approach. HCD is a product and system design philosophy that aspires to enhance human abilities, overcome human limitations, and foster user acceptance.[12] It achieves these objectives by designing products around users' characteristics, tasks, and workflow, as opposed to being driven by the available technology.[13,14] The application of HCD principles can beget systems that increase user productivity, acceptance, and satisfaction.[15,16]

Several groups have described the creation of dashboards for quality improvement in health care. Dashboards have been used to aid management of specific diseases,[17–19] specific hospital wards or units,[5,20] operating room suites,[6] clinical departments,[21] and health systems.[22] To our knowledge, this is the first report of the use of clinical dashboards for personal quality improvement in anesthesiology.

Versions of these dashboards can also be used as divisional, departmental, and operating room management platforms once the appropriate metrics are developed and tested. Additionally, similar methods could be used to generate detailed data on trainee clinical performance and completion of "educational milestones" as recently set by the Accreditation Council for Graduate Medical Education (ACGME).[23] Finally, the technology and processes described here could be employed to generate Ongoing Professional Performance Evaluation (OPPE) and Focused Provider Performance Evaluation (FPPE) reports as required by the Joint Commission, thereby significantly reducing the reporting burden and increasing operational efficiency.[24]

The A&F tool described here has several limitations. First, metrics displayed in the dashboards were constrained by the available data in our departmental data mart. We therefore could not provide feedback to clinicians on information they considered of high relevance, such as patient satisfaction scores or specific safety metrics. The incorporation of these metrics requires integration of EHR data with other hospital data sources and represents our next step in this line of work. Second, in an attempt to provide as comprehensive an

audit as possible, we included all cases for each provider, which required the display of an "Other" procedural category, which most users found confusing and inaccurate. It is likely that displaying performance on fewer, more commonly performed cases will retain the granularity of the feedback while eliminating the "noise." Third, we only surveyed attending anesthesiologists at a large academic center. Their practice and preferences may not fully represent that of other provider groups such as anesthesiology trainees or Certified Registered Nurse Anesthetists, or other practice settings. Lastly, the effectiveness of the dashboards as a behavior modification tool has not been tested.

In summary, we describe that the design, implementation, and usability testing of an innovative tool that utilizes data derived from the EHR system to provide A&F to anesthesiology providers at a large academic medical center. Our tool allows for near real-time, on-demand provision of A&F to anesthesiology providers on metrics that are important to them and the institution. Future research will focus on the impact of this A&F tool on provider behavior and patient outcomes.

FUNDING

AUTHOR CONTRIBUTION

A.B. designed the study, prepared the manuscript, and designed the figures. N.S. conducted the usability interviews, helped to design the study, and prepared the manuscript.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

Conflict of interest statement. None declared

REFERENCES

1. Brehaut JC, Eva KW. Building theories of knowledge translation interventions: use the entire menu of constructs. *Implement Sci* 2012; 7: 114.
2. Ivers N, Jamtvedt G, Flottorp S, *et al*. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2012; (6): CD000259.
3. Colquhoun HL, Brehaut JC, Sales A, *et al*. A systematic review of the use of theory in randomized controlled trials of audit and feedback. *Implement Sci* 2013; 8: 66.
4. Benn J, Arnold G, Wei I, Riley C, Aleva F. Using quality indicators in anaesthesia: feeding back data to improve care. *Br J Anaesth* 2012; 109 (1): 80–91.
5. Render ML, Freyberg RW, Hasselbeck R, *et al*. Infrastructure for quality transformation: measurement and reporting in veterans administration intensive care units. *BMJ Qual Saf* 2011; 20 (6): 498–507.
6. Nagy PG, Konewko R, Warnock M, *et al*. Novel, web-based, information-exploration approach for improving operating room logistics and system processes. *Surg Innov* 2008; 15 (1): 7–16.
7. St Jacques PJ, Patel N, Higgins MS. Improving anesthesiologist performance through profiling and incentives. *J Clin Anesth* 2004; 16 (7): 523–8.
8. Few S. *Information Dashboard Design*. Sebastopol, CA: O'Reilly; 2006.
9. Knaflic CN. *Storytelling with Data*. Hoboken, NJ: John Wiley & Sons; 2015.
10. Chin JP, Diehl VA, Norman KL. Development of a tool measuring user satisfaction of the human-computer interface. In: Association for Computing Machinery, ed. ACM CHI''8 *Conference on Human Factors in Computing Systems (1988)*. New York, NY; 1988: 213–8.
11. Dowding D, Randell R, Gardner P, *et al*. Dashboards for improving patient care: review of the literature. *Int J Med Inform* 2015; 84 (2): 87–100.
12. Rouse WB. *Design for Success: A Human-Centered Approach to Designing Successful Products and Systems*. New York, NY: ACM: John Wiley & Sons, Inc.; 1991.
13. Bannon LJ. Issues in design: some notes. In: Norman DA, Draper SW, eds. *User Centered System Design: New Perspectives on Human-Computer Interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1986: 25–30.
14. Gould JD. How to design usable systems. In: Helander M, Landauer TK, Prabhu PV, eds. *Handbook of Human-Computer Interaction*. 2nd ed. Amsterdam: Elsevier; 1997: 231–54.
15. Johnson CM, Johnson TR, Zhang J. A user-centered framework for redesigning health care interfaces. *J Biomed Inf* 2005; 38 (1): 75–87.
16. Kujala S. User involvement: a review of the benefits and challenges. *Behav Inf Technol* 2003; 22 (1): 1–16.
17. Jung E, Schnipper JL, Li Q, *et al*. The coronary artery disease quality dashboard: a chronic care disease management tool in an electronic health record. In: AMIA Annual Symposium Proceedings/AMIA Symposium AMIA Symposium, 999; Chicago, IL. October 11, 2007.
18. Sebastian K, Sari V, Loy LY, Zhang F, Zhang Z, Feng M. Multi-signal visualization of physiology (MVP): a novel visualization dashboard for physiological monitoring of Traumatic Brain Injury patients. *Conf Proc IEEE Eng Med Biol Soc* 2012; (2012): 2000–3.
19. Cheng CK, Ip DK, Cowling BJ, Ho LM, Leung GM, Lau EH. Digital dashboard design using multiple data streams for disease surveillance with influenza surveillance as an example. *J Med Internet Res* 2011; 13 (4): e85.
20. Stone-Griffith S, Englebright JD, Cheung D, Korwek KM, Perlin JB. Data-driven process and operational improvement in the emergency department: the ED Dashboard and Reporting Application. *J Healthc Manag* 2012; 57 (3): 167–80; discussion 80–1.
21. McLaughlin N, Afsar-Manesh N, Ragland V, Buxey F, Martin NA. Tracking and sustaining improvement initiatives: leveraging quality dashboards to lead change in a neurosurgical department. *Neurosurgery* 2014; 74 (3): 235–43; discussion 43–4.
22. Harrison L. Using agency-wide dashboards for data monitoring and data mining: the Solano County Health and Social Services Department. *J Evid Based Soc Work* 2012; 9 (1–2): 160–73.
23. Ehrenfeld JM, McEvoy MD, Furman WR, Snyder D, Sandberg WS. Automated near-real-time clinical performance feedback for anesthesiology residents: one piece of the milestones puzzle. *Anesthesiology* 2014; 120 (1): 172–84.
24. Ehrenfeld JM, Henneman JP, Peterfreund RA, *et al*. Ongoing professional performance evaluation (OPPE) using automatically captured electronic anesthesia data. *Jt Comm J Qual Patient Saf* 2012; 38 (2): 73–80.