



Effect of audit and feedback on physicians' intraoperative temperature management and patient outcomes: a three-arm cluster randomized-controlled trial comparing benchmarked and ranked feedback

Effet de la vérification et rétroactions de la gestion peropératoire de la température par les médecins et l'impact sur les patients: essai contrôlé randomisé par grappes à trois bras comparant une rétroaction selon performance et classement

Sylvain Boet, MD, PhD · Gregory L. Bryson, MD · Monica Taljaard, PhD · Ashlee-Ann Pigford, MSc · Daniel I. McIsaac, MD, MPH · Jamie Brehaut, PhD · Alan Forster, MD, MSc · Karim Mohamed, MD · Natalie Clavel, MD · Christopher Pysyk, MD · Jeremy M. Grimshaw, MBChB, PhD on behalf of the Canadian Perioperative Anesthesia Clinical Trials Group

Received: 5 March 2018 / Revised: 19 May 2018 / Accepted: 20 May 2018 / Published online: 29 August 2018
© Canadian Anesthesiologists' Society 2018

Abstract

Purpose *Audit and feedback can improve physicians' practice; however, the most effective type of feedback is unknown. Inadvertent perioperative hypothermia is associated with postoperative complications and remains common despite the use of effective and safe warming devices. This study aimed to measure the impact of targeted audit and feedback on anesthesiologists' intraoperative temperature management and subsequent patient outcomes.*

This article is accompanied by an editorial. Please see Can J Anesth 2018; 65: this issue.

S. Boet, MD, PhD (✉) · G. L. Bryson, MD · A.-A. Pigford, MSc · D. I. McIsaac, MD, MPH · K. Mohamed, MD · C. Pysyk, MD
Department of Anesthesiology and Pain Medicine, The Ottawa Hospital, University of Ottawa, 501 Smyth Rd, Critical Care Wing 1401, Ottawa, ON K1H 8L6, Canada
e-mail: sboet@toh.ca

S. Boet, MD, PhD
Department of Innovation in Medical Education, University of Ottawa, Ottawa, ON, Canada

S. Boet, MD, PhD · G. L. Bryson, MD · M. Taljaard, PhD · D. I. McIsaac, MD, MPH · J. Brehaut, PhD · A. Forster, MD, MSc · J. M. Grimshaw, MBChB, PhD
Clinical Epidemiology Program, The Ottawa Hospital Research Institute, Ottawa, ON, Canada

Methods *This study was a three-arm cluster randomized-controlled trial. Anesthesiologists' intraoperative temperature management performance was analyzed in two phases. The first was a baseline phase with audit but no feedback for eight months, followed by an intervention phase over the next seven-month period after participants had received interventions according to their randomized group allocation of no feedback (control), benchmarked feedback, or ranked feedback. Anesthesiologists' percentage of hypothermic patients at the end of surgery (primary endpoint) and use of a warming device were compared among the groups.*

M. Taljaard, PhD · J. Brehaut, PhD
School of Epidemiology, Public Health and Preventative Medicine, Ottawa, ON, Canada

A. Forster, MD, MSc · J. M. Grimshaw, MBChB, PhD
Department of Medicine, University of Ottawa, Ottawa, ON, Canada

N. Clavel, MD
Department of Anesthesia and Pain Management, Toronto Western Hospital, University of Toronto, Toronto, ON, Canada

Results Forty-five attending anesthesiologists who took care of 7,846 patients over 15 months were included. The odds of hypothermia (temperature $< 36^{\circ}\text{C}$ at the end of surgery) increased significantly from pre- to post-intervention in the control and ranked groups (control odds ratio [OR], 1.27; 95% confidence interval [CI], 1.03 to 1.56; $P = 0.02$; ranked OR, 1.26; 95% CI, 1.01 to 1.56; $P = 0.04$) but not in the benchmarked group (OR, 1.05; 95% CI, 0.87 to 1.28; $P = 0.58$). Between-arm differences in pre- to post-intervention changes were not significant (benchmark vs control OR, 0.83; 95% CI, 0.62 to 1.10; $P = 0.19$; ranked vs control OR, 0.99; 95% CI, 0.73 to 1.33, $P = 0.94$). No significant overall effect on intraoperative warmer use change was detected.

Conclusion We found no evidence to suggest that audit and feedback, using benchmarked or ranked feedback, is more effective than no feedback at all to change anesthesiologists' intraoperative temperature management performance. Feedback may need to be included in a bundle to produce its effect.

Trials registration www.clinicaltrials.gov (NCT02414191). Registered 19 March 2015.

Résumé

Objectif La vérification et la rétroaction peuvent améliorer la pratique des médecins. Néanmoins, nous ne savons pas quel type de rétroaction est la plus efficace. Une hypothermie périopératoire accidentelle est associée à des complications postopératoires et cela reste fréquent en dépit de l'utilisation de dispositifs de réchauffement efficaces et sécuritaires. Cette étude a eu pour objectif de mesurer les répercussions d'une vérification ciblée et d'une rétroaction sur la gestion peropératoire de la température par les anesthésiologistes et leurs impacts sur les patients.

Méthodes Cette étude était un essai contrôlé randomisé par grappes (cluster) à trois bras. La performance de gestion de la température peropératoire par les anesthésiologistes a été analysée en deux phases. La première était une phase initiale avec vérification, mais sans rétroaction pendant huit mois, suivie d'une phase d'intervention au cours des sept mois suivants après que les participants recevaient un suivi de leur résultat selon une affectation randomisée: pas de rétroaction (groupe contrôle), rétroaction de performance ou rétroaction de classement. Le pourcentage de patients hypothermiques des anesthésiologistes à la fin de la chirurgie (critère d'évaluation principal) et l'utilisation d'un dispositif de réchauffement ont été comparés entre les groupes.

Résultats Quarante-cinq anesthésiologistes ayant pris soin de 7 846 patients sur une période de 15 mois ont été inclus dans l'étude. Les probabilités d'hypothermie

(température $< 36^{\circ}\text{C}$ à la fin de la chirurgie) augmentaient significativement entre avant et après l'intervention dans le groupe contrôle et le groupe classement (rapport de cotes [RC] des contrôles: 1,27; intervalle de confiance [IC] à 95 %: 1,03 à 1,56; $P = 0,02$; RC classement: 1,26; IC à 95 %: 1,01 à 1,56; $P = 0,04$), mais pas dans le groupe avec rétroaction de performance (RC: 1,05; IC à 95 %: 0,87 à 1,28; $P = 0,58$). Les différences entre changements avant/après intervention entre les bras n'étaient pas significatives (RC performance contre contrôle: 0,83; IC à 95 %: 0,62 à 1,10; $P = 0,19$; RC classement contre contrôle: 0,99; IC à 95 %: 0,73 à 1,33; $P = 0,94$). Aucun effet global significatif sur un changement d'utilisation peropératoire de dispositifs de réchauffement n'a été détecté.

Conclusion Nous n'avons trouvé aucune donnée probante suggérant qu'une vérification et une rétroaction, utilisant soit une rétroaction de performance ou par classement, sont plus efficaces que l'absence totale de rétroaction pour changer la performance de gestion de la température peropératoire par les anesthésiologistes. Il pourrait être nécessaire d'inclure la rétroaction dans un plus vaste ensemble de mesures pour obtenir un effet.

Enregistrement de l'essai www.ClinicalTrials.gov (NCT02414191). Enregistré le 19 mars 2015.

Audit and feedback is effective to improve physicians' professional practice.¹ The literature provides specific guidance on how to make feedback effective outlining that it should be specific, timely, relevant, individualized, delivered with specific suggestions for improvement in writing, and given more than once.²⁻⁶

Comparison between different types of feedback has been identified as a key knowledge gap in the literature.^{1,7} The most common type of feedback is benchmarked feedback, in which an individual's performance is reported with respect to a set benchmark (i.e., a norm) based on minimum standards or best practice guidelines.⁶ Benchmarked feedback has been shown to dramatically improve anesthesiologists' performance in the administration of perioperative antibiotics.⁸ Alternately, ranked feedback allows someone to compare their individual performance with peer performance. The premise of ranked feedback⁹ holds that viewing personal performance within the context of peer performance is a powerful motivator for change.¹⁰⁻¹² Nevertheless, the impact of ranked feedback on performance within and outside of healthcare shows conflicting results.¹³⁻¹⁵

An important clinical issue that could benefit from audit and feedback is intraoperative temperature management.

Even mild hypothermia immediately before, during, or after surgery can have a wide range of adverse health effects including: cardiac events, blood loss, increased need for blood transfusion, increased length of stay, and mortality.^{16–23} Intraoperative hypothermia is intimately linked with surgical site infection, which is one of the most common and costly postoperative complications.^{16–20} Effective temperature maintenance during surgery has been linked to a significant, three-fold decrease in the rate of surgical site infection.²⁴ Unintended hypothermia can be prevented by the use of safe, effective, and inexpensive methods of warming.²⁵ Nevertheless, regardless of warming technologies and the existence of global²⁶ and Canadian guidelines,²⁷ perioperative hypothermia remains common and affects up to 20% of patients globally.²⁸

Only a few studies have investigated the impact of feedback on anesthesiologists' clinical performance, and only one has recently examined perioperative temperature management.²⁹ In addition, our group has identified barriers and facilitators to perioperative temperature management by anesthesiologists.³⁰ The lack of a commonly accepted temperature target and the absence of feedback on performance were among the main perceived barriers to improving temperature management by anesthesiologists. With about 230 million surgeries around the world annually, intraoperative temperature management optimization has the potential to greatly improve patient outcome.³¹

This study aimed to assess the impact of benchmarked and ranked feedback, compared with control (no feedback), on anesthesiologists' perioperative temperature management and subsequent patient outcomes. We hypothesized that, when compared with no feedback controls, both benchmarked and ranked feedback would improve anesthesiologists' perioperative temperature management and patient outcomes. For the secondary head-to-head comparison between benchmarked feedback vs ranked feedback, there was no prior evidence on which to base a hypothesized direction of effect.

Methods

Ethics and registration

This prospective, three-arm, cluster randomized-controlled trial is reported according to the cluster extension of The Consolidated Standards of Reporting Trials (CONSORT) guidelines.³² Institution approval was granted on 7 April 2014 from the Ottawa Health Science Network Research Ethics Board (OHSN-REB) (20140117-01H). The study was registered at ClinicalTrials.gov (NCT02414191).

Participants, setting, randomization

The study recruited staff anesthesiologists from the General and Civic campuses of The Ottawa Hospital (TOH), a large Canadian health sciences centre. A total of 79 anesthesiologists provide care to over 26,000 surgical patients annually. A research assistant enrolled participants. After informed consent, volunteer staff anesthesiologists ($n = 45$) were randomly allocated to one of the three study arms, using computer-generated random numbers. An independent statistician not otherwise involved in the study performed the randomization to protect allocation concealment. Allocation was stratified by site (General/Civic) to account for variability in surgical specialties specific to either site. Each staff anesthesiologist was considered to be a cluster. Each cluster included individual patient surgical events.

Study design

The study was designed with three phases: 1) a six-month baseline phase consisting of auditing whereby anesthesiologists' intraoperative temperature management performance was recorded with no feedback given; 2) a six-month intervention phase during which participants' performance was audited and monthly feedback was provided to anesthesiologists according to group allocation (i.e., six feedback emails per anesthesiologist); and 3) a three-month post-intervention phase consisting of an audit of anesthesiologists' temperature management with no feedback provided to any of the three groups. This is outlined in the CONSORT per protocol flow chart (Fig. 1). A monthly feedback frequency has been identified as optimal by the largest Cochrane systematic review on audit and feedback.¹

Due to practical issues relating to the availability of data and generation of individual monthly reports, it took five weeks following data collection to send feedback reports to clinicians. For example, performance data collected in February 2015 were processed in March 2015 and feedback emails were sent on 8 April 2015, which may produce its full effect on performance in May 2015. This required us to change our study phases as there was less time available for the follow-up period. Therefore, we analyzed the data in two phases: 1) an eight-month baseline phase (August 2014 to March 2015 inclusive) and 2) a seven-month intervention phase (April 2015 to December 2015 inclusive). Operationally, this meant that the *per protocol* post-intervention months of August, September, and October 2015 (i.e., phase 3) were included in the “intervention” phase of analysis. Figure 2 shows the CONSORT *per analysis* flow chart. Full details and the corresponding dates between the planned phases (per

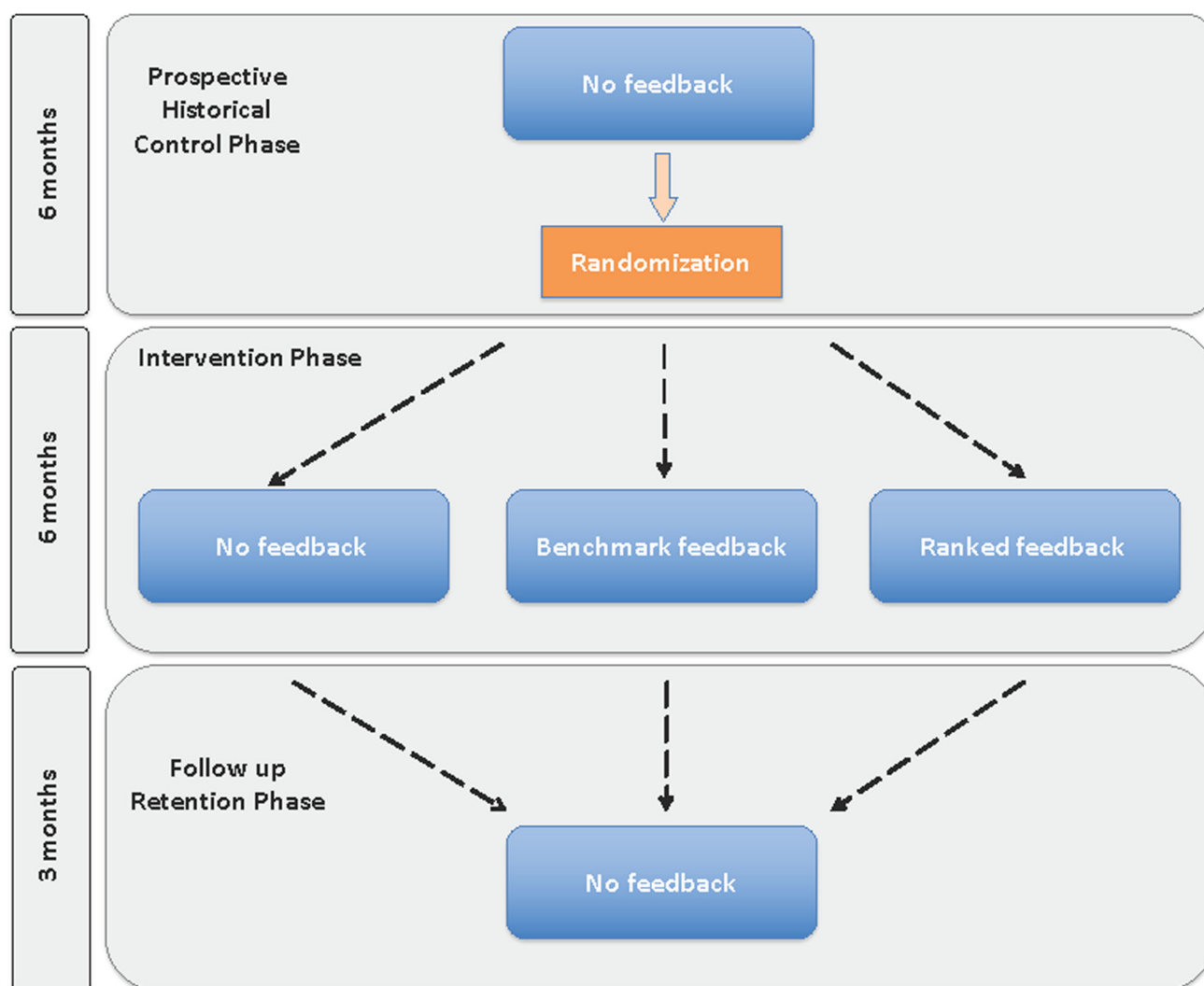


Fig. 1 CONSORT per protocol flow chart. Note: As described in the methods (study design section), the analysis has followed two phases: baseline (8 months) and intervention (7 months), see Figure 2 and Appendix 1

protocol) and the analyzed phases are shown in Appendix 1 (showing the study timeline and various phases).

Intervention

For the intervention phase, participants were randomized to one of three study arms: no feedback (control), benchmarked feedback, and ranked feedback. The control group had their performance audited, but no feedback was provided as per current hospital practice. The benchmarked feedback group had their performance audited and monthly benchmarked feedback was provided by email. Feedback included their individual performance outcomes and a reminder of the target temperature expected to achieve. The ranked feedback group had their performance audited and monthly ranked feedback was provided by email. Feedback included their individual performance and their

corresponding ranking regarding temperature management within the anesthesiology department. Both intervention groups (benchmarked and ranked feedback) were provided with basic information and strategies to improve their temperature management in each monthly email (Appendix 2: benchmarked and ranked feedback emails).

The feedback emails used in this study were designed based on the best practices for feedback interventions.⁵ We also conducted a pilot study to refine the feedback design and test the study logistics. Several volunteer clinical fellows of the anesthesiology department at The Ottawa Hospital were recruited for the pilot. They individually read aloud the feedback emails in front of a research assistant to convey their thought process in real time when reading the feedback email for the first time. Based on this pilot, the research team refined the design of the email to ease understanding of emails and to confirm readability on

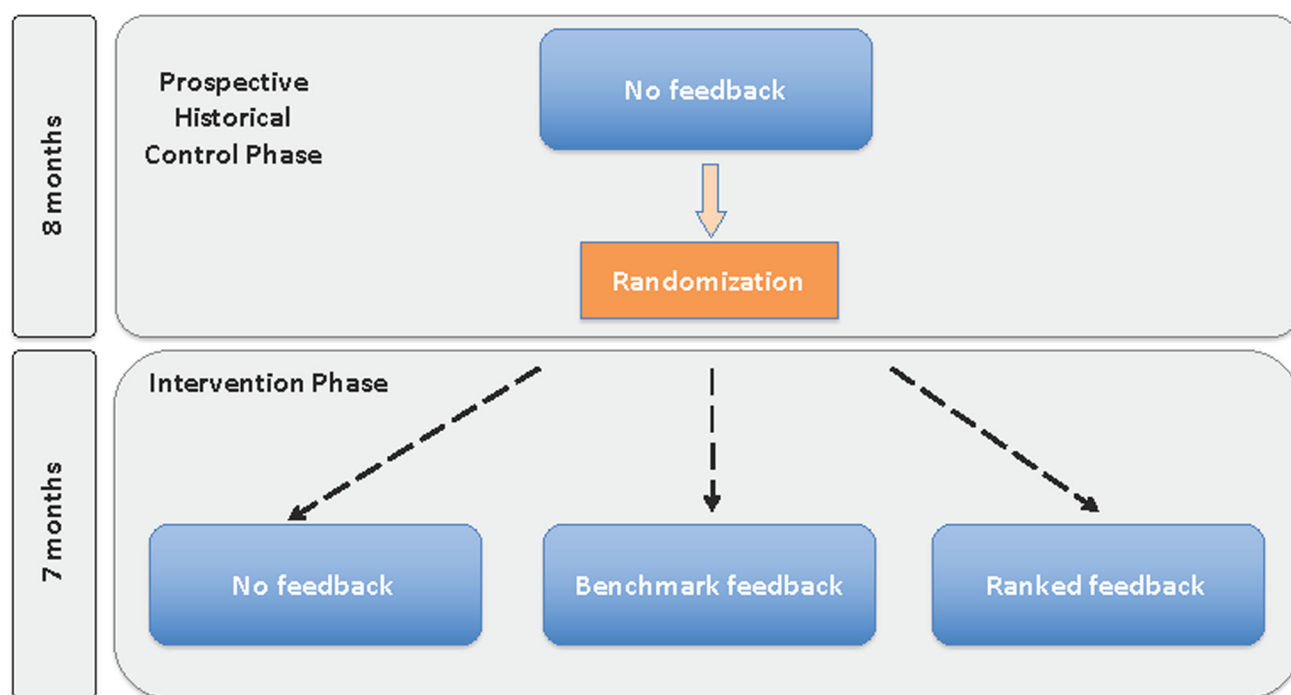


Fig. 2 CONSORT per analysis flow chart. Note: As described in the methods (study design section), the protocol has followed three phases: baseline (6 months), intervention (6 months) and post-intervention (3 months), see Figure 1 and Appendix 1

the most common e-devices used (e.g., tablet, phone, computer). Emails were sent from an institutional research group email account on behalf of the principal investigator (S.B.). A link requesting a read receipt was included in each feedback email sent to anesthesiologists to confirm that they had read the feedback.

Data collection and outcomes

Data were prospectively collected using our hospital's electronic medical records system (vOaxis, Telus Health, Canada) and our anesthesia information management system (Surgical Information System [SIMS]). Our hospital's electronic medical records system was the same for operating room and postanesthesia care unit (PACU). Patient data for all patients anesthetized by each anesthesiologist who participated in the study were included. Key indicators related to physician performance relevant to both *patient outcome* (temperature) and the *process of care* (warmer use) were collected. Data collected included the patient characteristics, perioperative patient temperature, use of warming devices, perioperative times, type of surgery, and healthcare provider identity. A fluid warmer and forced air were the active warming devices included in data collection; other active means were not considered as only forced air and fluid warming were systematically recorded in SIMS. The thermistor probes used were the autoclavable

YSI401AC (Measurement Specialties Inc., Hampton, VA, USA) to measure the patient's temperature intraoperatively (nose or mouth) and Filac TM 3000AD EZ (Covidien, Toronto, ON, Canada) in the PACU (accuracy $\pm 0.1^{\circ}\text{C}$ for both probes). The patient's temperature in the operating room was automatically captured electronically each minute when the temperature probe was connected to the monitor. Oral temperature in the PACU was measured manually within 15 min from arrival in the unit. The nurse transcribed the temperature into the electronic health record in real time. The temperature probe was used as per manufacturer recommendation. Briefly, it was placed in the patient's mouth and the patient was asked to close their mouth until the temperature appeared on the measurement tool. Oral temperature measurement was the standard practice in our PACU and only temperature considered in our study. If the maximum temperature measurement was $< 30^{\circ}$, then we considered it as artefact.

Decisions about eligible surgeries and patient outcomes were informed by the Physician Quality Reporting System (PQRS) stating that the "percentage of patients undergoing surgical procedures of 60 min or more, except cardiac bypass, for whom either active warming was used to maintain normothermia, or at least one body temperature was equal to or greater than 36.0°C recorded within the 30 min directly before or the 15 min directly after anesthesia end time".³³ Based on this definition of temperature management, surgeries lasting < 60 min and surgeries

including cardiac bypass were excluded. Since the most rapid temperature loss occurs immediately after induction of anesthesia,³⁴ patients were attributed to the anesthesiologist who induced their anesthesia, regardless of transfer of intraoperative care to other staff.

The *primary outcome measure* (i.e., patient outcome measure) was defined at the individual patient level as the presence of hypothermia (temperature < 36.0°C) at the end of surgery. In line with the PQRS definition of temperature management, the temperature considered for analysis was the highest temperature recorded within the 30 min immediately before or the 15 min immediately after anesthesia end time.

Secondary outcome measures (i.e., process of care measures) were measured at both the patient (i.e., as the use of fluid warmer and forced air warming devices) and provider level (i.e., as a pass/fail measure based on temperature management performance according to PQRS³³ definition). The PQRS defined good temperature management performance (i.e., pass) as > 80% of surgery patients (at > 60 min duration regardless of age or surgery) for whom either active warming was used intraoperatively for the purpose of maintaining normothermia or at least one body temperature $\geq 36^\circ\text{C}$ was recorded at the end of surgery. Among secondary outcomes, we had initially planned to collect surgical site infection data from the National Surgical Quality Improvement Program (NSQIP) database. Nevertheless, given the cost to obtain these data and the fact that we observed no effect on the primary outcome, we elected not to collect them. Indeed, should the surgical site infection rate show any difference, we could not have related it to our intervention since patients' temperatures were not different.

Statistical analysis

Sample size calculations were based on two pairwise comparisons of the proportions of hypothermic patients (primary outcome) between each intervention arm *vs* control over the six-month intervention period. The calculations accounted for clustering of patients by provider and for multiple pairwise comparisons using the Bonferroni correction. Prior to our study, we had explored data involving 22 providers and 400 of their most recent patients to inform our sample size. Based on these data, our estimate for the control arm proportion was 0.50 and the intraclass correlation coefficient for the primary outcome was 0.02. To detect a clinically relevant absolute difference of 10% between at least one of the intervention arms *vs* control, with 80% power and using a two-sided multiplicity-corrected α of 0.025, we required 14 physicians in each arm, assuming an average of 180 patients per physician over a six-month period (2,520

patients per arm). This calculation incorporated a further 10% inflation to account for variable cluster sizes. To account for 15% potential attrition, we planned to enrol 15 providers in each arm for a total of 45 providers. For the secondary head-to-head comparison of the two interventions, this sample size yielded 80% power to detect an absolute difference of 8% (i.e., 40% in the ranked *vs* 32% in the benchmark group.) These comparisons were conservative, as they did not account for the baseline proportions that were controlled for in the analysis.

For analysis, the baseline phase lasted eight months (August 2014 - March 2015) and the intervention phase seven months (April 2015 - December 2015) (Appendix 1). Data were summed across the months of each phase to obtain the average number of patients per physician in each study phase. Descriptive statistics were calculated to compare characteristics of patients and providers among the study arms before and after intervention. Continuous variables were described using means (standard deviation) and medians [interquartile range], with frequencies and proportions for dichotomous variables.

The dichotomous primary outcome (proportion of hypothermic patients) was analyzed at the patient-level using a mixed-effects repeated measures logistic regression analysis to estimate the effect of the intervention, adjusted for clustering by provider and for all the pre-specified covariates (patient age, surgery duration, American Society of Anesthesiologists physical status [ASA-PS] classification [ASA-PS I-II *vs* III-V], emergency nature of surgery, type of anesthesia). Continuous covariates were modeled using restricted cubic splines with three knots for age and five for surgery length, based on visual inspection of graphs of the associations. Time (defined as a categorical variable), study arm, and interaction between study arm and time were included as fixed effects, in addition to site. The physician and time were specified as random effects to account for clustering of patients by physician and to allow for a decline in the strength of correlation over time. Least square mean estimates from the model were used to compare the two intervention arms *vs* control in terms of change from baseline and during the intervention and post-intervention phases together with a 95% confidence interval (CI). The effect of the intervention was estimated as difference in change from baseline for each intervention arm *vs* control and expressed as a ratio of odds ratios. The method of estimation was pseudo-likelihood with degrees of freedom estimated using the Kenward-Roger method. The Bonferroni correction was used to account for multiple comparisons of the two intervention arms *vs* control according to our pre-specified protocol, with all *P* values judged at a significance level of 0.025.

Use of warming technology was analyzed in a similar way. Pass/fail temperature management performance in each arm was described using descriptive statistics. We also elected to conduct an *a priori* subgroup analysis to explore the change in performance for the primary outcome between campuses. These analyses were done by including two- and three-way interaction terms with the subgroup indicators into the models. Data were analyzed using SAS v.9.2 (SAS Institute Inc., Cary, NC, USA).

Results

The study participant flow chart is shown in Fig. 3. Forty-five anesthesiologists were enrolled in the study with 15 anesthesiologists randomized to each of the three study arms. Patients and surgery characteristics by study arm and group allocation are presented in Table 1. The baseline phase involved 4,072 patients and the intervention phase involved 3,774 patients, representing a total of 7,846 patients. The number of patients per arm of the study was comparable across each phase. The most prevalent surgical specialties in both phases were general surgery (49%) or orthopedic (46%). The median length of surgery was similar across arms and phases.

Primary outcome: proportion of hypothermic patients

The observed proportions of hypothermic patients at the end of surgery per physician are summarized in Table 2. Table 3 shows the results from the mixed effects logistic regression, with the intervention effect expressed as adjusted odds ratio for change from pre- to post-intervention for each group as well for the difference in performance between groups. First, the diagonal cells in Table 3 show that the adjusted odds of hypothermia significantly increased from baseline to intervention in the control and ranked groups (control odds ratio [OR], 1.27; 95% CI, 1.03 to 1.56; $P = 0.02$; ranked OR, 1.26; 95% CI, 1.01 to 1.56; $P = 0.04$) but did not change significantly in the benchmarked group (OR, 1.05; 95% CI, 0.87 to 1.28; $P = 0.58$). Second, the off-diagonal cells indicate that there were no significant between-arm differences in hypothermia change from pre to post (benchmark vs ranked OR, 0.84; 95% CI, 0.63 to 1.12; $P = 0.22$; benchmark vs control OR, 0.83; 0.62 to 1.10; $P = 0.19$; ranked vs control OR, 0.99; 0.73 to 1.34; $P = 0.95$). The comparison of the average of the two intervention arms vs control, expressed as difference in change from baseline, was not significant (OR, 0.91; 95% CI, 0.70 to 1.17; $P = 0.43$).

Overall, the odds of a hypothermic patient was significantly lower at the Civic than the General hospital

campus (OR, 0.69; 95% CI, 0.59 to 0.79; $P < 0.001$); however, the statistical test of interaction comparing differences in the intervention by campus was not significant ($P = 0.33$).

Of note, among the non-hypothermic patients, only three patients had a temperature $> 40^{\circ}\text{C}$ at the end of surgery; the three temperatures recorded were 40.1°C for two of these patients and 40.2°C for a third one.

Secondary outcomes

The observed proportions of intraoperative warmer use are shown in Table 2. There was a significant increase in intraoperative warmer use from baseline to intervention in each arm ($P < 0.001$) (Table 2). In particular, the odds of intraoperative warmer use increased in the benchmarked arm (OR, 2.26; 95% CI, 1.62 to 3.16; $P < 0.001$; ranked OR, 3.12; 95% CI, 2.11 to 4.62; $P < 0.001$; control OR, 2.12; 95% CI, 1.48 to 3.05; $P < 0.001$) (Table 3). Nevertheless, there was no significant overall effect of the intervention on intraoperative warmer use (benchmark vs ranked OR, 0.72; 95% CI, 0.43 to 1.21; $P = 0.21$; benchmark vs control OR, 1.06; 95% CI, 0.65 to 1.74; $P = 0.80$; ranked vs control OR, 1.47; 95% CI, 0.86 to 2.51; $P = 0.15$). The comparison of the average of the two intervention arms vs control, expressed as difference in change from baseline, did not show any significant difference (OR, 1.25; 95% CI, 0.80 to 1.95; $P = 0.31$).

The odds of intraoperative warmer use were much lower at the Civic campus (OR, 0.68; 95% CI, 0.48 to 0.97; $P = 0.03$) but the statistical test of interaction comparing differences in the intervention by campus was not significant ($P = 0.77$). The observed proportions of physicians who “pass” on temperature management performance according to Physician Quality Reporting System (PQRS) are summarized in Table 2. All physicians met the PQRS pass criteria (i.e., 80% of patients warm or warmed per physician) every month. Given that no physician failed according to the PQRS system, no statistical analysis of this outcome was conducted.

Discussion

Our study found no evidence to suggest that a formative assessment using benchmarked or ranked feedback was more effective than no feedback in influencing anesthesiologists’ performance related to patient temperature outcome in our clinical setting.

Our results contradict our hypothesis that was drawn from robust educational and implementation science literature. Feedback is one of the most recognized and powerful educational interventions for continuing

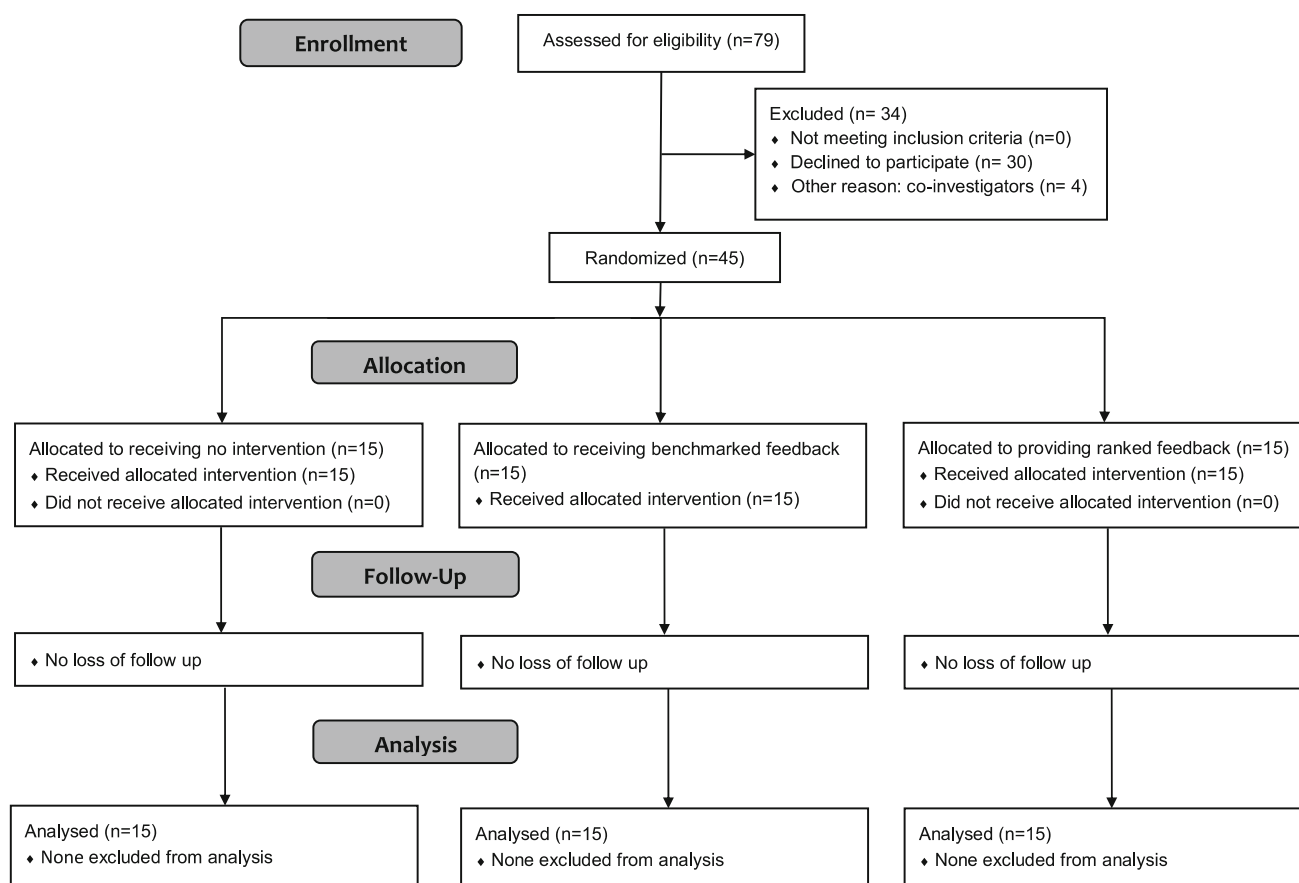


Fig. 3 Patient flow chart

professional development.² In their systematic review of 140 trials on audit and feedback in healthcare, Ivers *et al.* concluded that it is effective overall in improving professional performance.¹ Audit and feedback in anesthesiology practice has rarely been investigated. Nevertheless, benchmarked feedback has been shown to dramatically improve anesthesiologists' performance in the administration of intraoperative antibiotics.⁸ Another recent quality improvement initiative using audit and feedback for spine surgery resulted in shortening the time delay until instituting temperature monitoring without decreasing percentage of hypothermia time.²⁹

In our study, we relied on evidence for best practices for audit and feedback in healthcare to optimize the hypothesized effect of feedback on performance.⁵ Nevertheless, our study differs on several points from previous studies that found audit and feedback effective in anesthesiology. First, both O'Reilly *et al.* and Görges *et al.* fed back process measures, i.e., something that is done to patients, namely the administration of antibiotics within one hour of surgical incision and the time delay of temperature monitoring institution respectively.^{8,29} We elected to use the percentage of hypothermic patients as

the primary outcome, which is an outcome measure (i.e., the result of an action). The Donabedian structure-process-outcome conceptual model commonly used in audit and feedback classifies measures based on their distance between a variable such as an operator's action and its results such as patient's outcome.³⁴ Process measures may be directly actionable by feedback recipients while outcome measures are more distant and may not depend directly of only anesthesiologists' actions.³⁵ We note that Görges *et al.* did not observe any change in their outcome measure of percentage of time patients were hypothermic²⁹ while our process measure (percentage of intraoperative warmer use) improved in all groups despite little room for improvement at baseline (i.e., a possible ceiling effect). Second, it is often challenging to choose quality indicators in anesthesiology that truly only depend on anesthesiologists' performance. In their systematic review, Haller *et al.* found that about half of their 108 quality indicators were non-specific to anesthesia performance.³⁶ Both appropriate antibiotic administration and time delay for temperature monitoring were specific to anesthesia performance, potentially resulting in better perceived empowerment by feedback recipients.

Table 1 Patient and surgical characteristics

	Control	Benchmarked	Ranked
Arm, <i>n</i> patients			
Baseline	1,384	1,466	1,222
Intervention	1,225	1,428	1,121
Average number of patients per anesthesiologist, mean (SD)			
Baseline	92.3 (38.7)	97.7 (36.5)	81.5 (37.2)
Intervention	81.7 (30.3)	95.2 (33.7)	74.7 (46.3)
Patient age, mean (SD)			
Baseline	55.6 (17.5)	54.4 (17.0)	55.3 (17.6)
Intervention	56.0 (17.3)	54.9 (17.1)	57.0 (17.4)
Male frequency, count (%)			
Baseline	616 (44.5%)	607 (41.4%)	561 (45.9%)
Intervention	523 (42.8%)	595 (41.7%)	477 (42.6%)
Length of surgery, median hours [IQR]			
Baseline	2.38 [1.72-3.54]	2.30 [1.63-3.32]	2.48 [1.77-3.70]
Intervention	2.28 [1.60-3.52]	2.22 [1.62-3.25]	2.23 [1.63-3.28]
Campus, count (%) for Civic campus			
Baseline	635 (45.9%)	612 (41.8%)	531 (43.5%)
Intervention	519 (42.4%)	532 (37.3%)	420 (37.5%)
Type of anesthesia, count (%) for general anesthesia			
Baseline	1,166 (84.3%)	1,246 (85.0%)	999 (81.8%)
Intervention	1,043 (85.1%)	1,174 (82.2%)	885 (79.0%)
Preoperative warmer use, count (%)			
Baseline	492 (35.6%)	520 (35.5%)	462 (37.9%)
Intervention	490 (40.0%)	562 (39.4%)	462 (41.2%)
Type of surgery, count (% per group per phase)			
General surgery (general, urology, gynecology)			
Baseline	624 (45.1%)	748 (51.0%)	561 (45.9%)
Intervention	522 (42.6%)	708 (49.6%)	554 (49.4%)
Orthopedic (including spine, plastic)			
Baseline	465 (33.6%)	484 (33.0%)	415 (34.0%)
Intervention	429 (35.0%)	484 (33.9%)	395 (35.2%)
Head and neck surgery (otolaryngology, dental, eye surgery, neurosurgery)			
Baseline	176 (12.7%)	150 (10.2%)	135 (11.0%)
Intervention	158 (12.9%)	151 (10.6%)	99 (8.8%)
Other (vascular, thoracic, etc.)			
Baseline	119 (8.6%)	84 (5.7%)	111 (9.1%)
Intervention	116 (9.5%)	85 (6.0%)	73 (6.5%)
Emergency procedure, count (%)			
Baseline	353 (25.5%)	265 (18.1%)	318 (26.0%)
Intervention	328 (26.8%)	316 (22.1%)	301 (26.9%)
ASA score ≥ 3 , count (%)			
Baseline	759 (54.8%)	724 (49.4%)	735 (60.1%)
Intervention	678 (55.3%)	711 (49.8%)	653 (58.3%)

ASA = American Society of Anesthesiologists; IQR = interquartile range; SD = standard deviation

Our intervention of audit and feedback was evidence-based as it resulted directly from a previous qualitative study that investigated barriers and facilitators for

temperature management by anesthesiologists.³⁰ This qualitative study used an evidence-based approach, relying on the Theoretical Domains Framework, which is

Table 2 Observed frequency and percentage of hypothermic patients, warmer use and pass performance

		Control	Benchmarked	Ranked
Patient level data	Hypothermic patients (%)			
	Baseline	335/1,384 (24.2%)	407/1,466 (27.8%)	314/1,222 (25.7%)
	Intervention	360/1,225 (29.4%)	425/1,428 (29.8%)	364/1,121 (32.5%)
	Warmer use (%)			
	Baseline	1,219/1,384 (88.1%)	1,252/1,466 (85.4%)	1065/1,222 (87.2%)
	Intervention	1,134/1,225 (92.6%)	1,301/1,428 (91.1%)	1052/1,121 (93.8%)
Physician level data	Pass performance according to PQRS (%)*			
	Baseline	15 (100.0%)	15 (100.0%)	15 (100.0%)
	Intervention	15 (100.0%)	15 (100.0%)	15 (100.0%)

*Variable could not be analyzed at multivariate level as no physician failed according to PQRS (Physician Quality Reporting System), i.e., all physicians met the PQRS pass criteria (80% of patients warm or warmed per physician) every month

Table 3 Adjusted odd ratios of proportion of (a) hypothermic patients and (b) warmer use

(a)	Control	Benchmarked	Ranked
Control	1.27 [1.03 to 1.56], $P = 0.02$		
Benchmarked	0.83 [0.62 to 1.10], $P = 0.19$	1.05 [0.87 to 1.28], $P = 0.58$	
Ranked	0.99 [0.73 to 1.33], $P = 0.94$	0.84 [0.63 to 1.12], $P = 0.22$	1.26 [1.01 to 1.56], $P = 0.04$
(b)	Control	Benchmarked	Ranked
Control	2.12 [1.48 to 3.05], $P < 0.001$		
Benchmarked	1.06 [0.65 to 1.74], $P = 0.80$	2.26 [1.62 to 3.16], $P < 0.001$	
Ranked	1.47 [0.86 to 2.51], $P = 0.15$	0.72 [0.43 to 1.21], $P = 0.21$	3.12 [2.11 to 4.62], $P < 0.001$

Odds ratios [95% confidence interval] were adjusted for the following variables: patient age, surgery duration, American Society of Anesthesiologists (ASA) classification (ASA I-II vs III-V), the emergency nature of surgery, type of anesthesia. P values comparing the two intervention arms vs control should be judged against a significance level of $\alpha = 0.025$ to account for multiplicity according to our pre-specified protocol. The table shows adjusted odds ratios for outcomes. Adjusted odd ratios of the change in performance from pre- to post-intervention for a given group are shown at the intersection cell of the same group name. Adjusted odds ratios of the difference in performance change between two groups are shown at the intersection of the different group names

a well-validated framework in implementation science.³⁷ One may only speculate on the reasons why our evidence-based feedback intervention did not show the anticipated results. Nevertheless, our previous qualitative study identified several barriers to optimized temperature management by anesthesiologists and our designed intervention accounted only for one of these identified barriers (i.e., lack of feedback on temperature management performance). It is possible that benchmarked and/or ranked feedback alone may not generate sustained improvements in anesthesiologists' performance and that a bundle approach tackling several of the identified barriers

is necessary to improve physicians' performance and patients' outcomes. This consideration is in line with literature that suggests that feedback interventions may be more effective if not isolated but integrated into a larger quality improvement agenda.^{4,35} Görges *et al.*²⁹ used a traditional quality improvement approach and O'Reilly *et al.*⁸ used a multitier approach by implementing both new practice guidelines and feedback.

Although our study did not support our hypothesis, our work has several strengths. We provide an example of how an electronic health record system can be used for practice improvement interventions in perioperative performance

on patient outcomes. This study targeted a well-identified evidence-practice gap (i.e., intraoperative temperature management) and investigated an evidence-based intervention (i.e., audit and feedback) to close the existing gap. Furthermore, we explored the impact of our intervention across the practitioner-process-outcome spectrum. Nevertheless, our study has several limitations. First, we only included temperature measure at the end of the surgery while Sun *et al.* recently showed that the patient hypothermia throughout the intraoperative period is related to postoperative outcome.³⁸ Our process measure only recorded whether the warming device was used or not, but the time delay of starting the warmer and the type of warmer might explain temperature management performance better than simply using it or not. Finally, our study was conducted in an academic center and we did not account for the presence of trainees in the operating room. It is possible that anesthesia trainees interfered with attending anesthesiologists' performance and therefore limited the impact of feedback administered.

We need to consider several possible limitations when interpreting the results of this study. A Hawthorne effect resulting from observation bias was always possible, although it seems unlikely that it occurred in our study because the proportion of hypothermic patients during the study was not lower than the years before. We cannot be certain that anesthesiologists who volunteered to participate in the study were representative of the whole population of anesthesiologists. Nevertheless, examining the external validity would have required us to analyze data from anesthesiologists who did not volunteer to participate in the study, which represents some significant ethical/consenting challenges. Our secondary outcomes were process measures (i.e., warmer use, PQRS) and showed high baseline compliance, which may suggest a possible ceiling effect. Nevertheless, with approximately 25–30% of patients being hypothermic at the end of surgery, we would not consider that participants in our study were high performers, despite a possible ceiling effect on documentation of warmer use. We believe that the process measures largely relying on documentation of warmer use may have been insufficient to reflect a closer link to outcome measure (i.e., patient's temperature). If a warmer is used late (e.g., > 60 min) after patient induction and the surgery is relatively short (e.g., < two hours), then the warmer is unlikely to bring the patient temperature up to normothermia. Accounting for *how* and *when* warmers are actually used may be important to better reflect the process that results in positive outcome. Thus, we note that the quality measure criterion for perioperative normothermia was modified in early 2016 by the United States' Centers for

Medicare and Medicaid Services PQRS from a quality measure (the "PQRS outcome" in our study) to an outcome measure (patient temperature). In the future, innovative technology such as video recording of the operating room may allow effective capturing of information about the entire process measure including when and how warmers are used. Video capture may also allow for the simultaneous addressing of the source and timing of the temperature being used and to relate it to warmer use.

Conclusions

This study did not determine the most effective form of feedback for improving perioperative temperature outcomes and management. Even though this study is inconclusive in terms of showing the most appropriate form of feedback, our findings may be helpful to others planning interventional studies involving physician feedback and quality improvement. It may be necessary for future interventions using feedback to bundle an assortment of interventions to see a substantial change.

Acknowledgements We thank Sarah Larrigan for checking data quality, the Data Warehouse at The Ottawa Hospital for its assistance with the extraction of audit data, the team of research assistants from the anesthesia education research group who assisted with the study, and the staff anesthesiologists who accepted to consent to this study.

Conflicts of interest The authors have no conflicts of interest to declare.

Editorial responsibility This submission was handled by Dr. Hilary P. Grocott, Editor-in-Chief, *Canadian Journal of Anesthesia*.

Author contributions All authors contributed to drafting or revising the article. Sylvain Boet, Monica Taljaard, Jeremy Grimshaw, Daniel I. McIsaac, Gregory L. Bryson, and Alan Forster contributed substantially to all aspects of this manuscript, including conception and design; acquisition, analysis, and interpretation of data. Jamie Brehaut, Karim Mohamed, Natalie Clavel, Christopher Pysyk, and Ashlee-Ann Pigford contributed substantially to the conception, design of the manuscript, and interpretation of data.

Funding This research was supported by three project research grants: (1) a 2014 Medical Council of Canada (MCC-6/1415); (2) a 2013–2014 The Ottawa Hospital Academic Medical Organization (TOHAMO) Innovation Fund provided by the Ontario Medical Association (OMA) and the Ministry of Health and Long-Term Care (MOHLTC); and a 2013 AIME Education Grant committee to fund the pilot phase of the project. This research project was also supported by the 2015 Canadian Anesthesiologists' Society Career Scientist Award in Anesthesia funded by the Canadian Anesthesia Research Foundation. Drs. Boet, Bryson, McIsaac, and Pysyk were supported by The Ottawa Hospital Anesthesia Alternate Funds Association. Funders had no involvement in any of the study phases. Jeremy Grimshaw is supported by a Canada Research Chair in Health Knowledge Transfer and Uptake.

Appendix 1 Study timeline and study phases

Month/Year	Aug-14	Sep-14	Oct-14	Nov-14	Dec-14	Jan-15	Feb-15	Mar-15	Apr-15	May-15	Jun-15	Jul-15	Aug-15	Sep-15	Oct-15
Study phase according to the planned protocol timeline	Baseline 1	Baseline 2	Baseline 3	Baseline 4	Baseline 5	Baseline 6	Intervention 1	Intervention 2	Intervention 3	Intervention 4	Intervention 5	Intervention 6	Post-intervention 1	Post-intervention 2	Post-intervention 3
Month of data that was fed back to subjects									Feb-15	Mar-15	Apr-15	May-15	Jun-15	Jul-15	
Date when feedback email was sent									2015-04-08 (April 2015)	2015-05-05 (May 2015)	2015-05-29 (June 2015)	2015-07-02 (July 2015)	2015-08-10 (Aug 2015)	2015-09-01 (Sept 2015)	
Month of the expected full effect of the feedback									May-15	Jun-15	Jul-15	Aug-15	Sep-15	Oct-15	
Actual study phase according to the expected full effect based on	Aug 2014 - Baseline 1	Sept 2014 - Baseline 2	Oct 2014 - Baseline 3	Nov 2014 - Baseline 4	Dec 2014 - Baseline 5	Jan 2015 - Baseline 6	Feb 2015 - Baseline 7	March 2015 - Baseline 8	April 2015 - Intervention 1	May 2015 - Intervention 2	June 2015 - Intervention 3	July 2015 - Intervention 4	Aug 2015 - Intervention 5	Sept 2015 - Intervention 6	Oct 2015 - Post-intervention 1
when the feedback emails were sent															
Study phase according to analysis	8 months of baseline data								7 months of intervention data						

Red cells represent the pre-intervention phase, green the intervention, and blue the post-intervention phase.

Appendix 2 Benchmarked and ranked feedback emails

Benchmarked feedback email template

Dear Dr.,

Summary of your perioperative temperature management performance during month year:

X cases ^a seen in month		Possible ways to optimize temperature management
	Results	
Hypothermic patients at end of case ^{b, c}	___%	• use a temperature probe • maximize Bair Hugger™ coverage • do not delay warming after induction • fluid warmer • use warm intravenous fluid • turn on PerfecTemp™ (warming table)
Intra-op warmer use	___%	

- a) Only surgeries lasting more than one hour and in which the patient's temperature was measured in the 30 min before leaving the OR and/or measured orally in the 15 min after PACU arrival
- b) Maximal temperature < 36.0°C in the 30 min pre- and 15 min post- "PACU arrival time"
- c) Regardless of anesthesia type (from anesthesia start time to anesthesia end time)

If you have any questions regarding the feedback provided, please contact us.

Thank you,

OR = operating room; PACU = postanesthesia care unit.

Ranked feedback email template

Dear Dr.,

Summary of your perioperative temperature management performance during **month year**:

- a) Only surgeries lasting more than 1 hour and in which the patient's temperature was measured in the 30 min before leaving the OR and/or measured orally in the 15 min after PACU arrival
- b) Maximal temperature < 36.0°C in the 30 min pre- and 15 min post- "PACU arrival time"
- c) Regardless of anesthesia type
- d) Rank 1 is the best performance; rank 79 is the worst performance

If you have any questions regarding the feedback provided, please contact us.

Thank you,

OR = operating room; PACU = postanesthesia care unit.

X cases ^a seen in month			Possible ways to optimize temperature management
	Results	Department Rank ^d	
Hypothermic patients at end of case ^{b, c}	___%	___/79	• use a temperature probe • maximize Bair Hugger™ coverage
Intra-op warmer use	___%	___/79	• do not delay warming after induction • fluid warmer • use warm intravenous fluid • turn on PerfecTemp™ (warming table)

References

- Ivers N, Jamtvedt G, Flottorp S, et al. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2012; 13: CD000259.
- Hattie J, Timperley H. The power of feedback. *Rev Educ Res* 2007; 77: 81-112.
- Ramani S, Krackov SK. Twelve tips for giving feedback effectively in the clinical environment. *Med Teach* 2012; 34: 787-91.
- 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: 80-91.
- Brehaut JC, Colquhoun HL, Eva KW, et al. Practice feedback interventions: 15 suggestions for optimizing effectiveness. *Ann Intern Med* 2016; 164: 435-41.
- Hysong S. Meta-analysis: audit and feedback features impact effectiveness on care quality. *Med Care* 2009; 47: 356-63.
- Foy R, Eccles M, Jamtvedt G, Young J, Grimshaw J, Baker R. What do we know about how to do audit and feedback? Pitfalls in applying evidence from a systematic review. *BMC Health Serv Res* 2005; 5: 50.
- O'Reilly M, Talsma A, VanRiper S, Kheterpal S, Burney R. An anesthesia information system designed to provide physician-specific feedback improves timely administration of prophylactic antibiotics. *Anesth Analg* 2006; 103: 908-12.
- Sirota D, Mischkind LA, Meltzer MI. The Enthusiastic Employee: How Companies Profit by Giving Workers What They Want. Wharton School Publishing; 2005.
- Jencks SF. Changing health care practices in Medicare's Health Care Quality Improvement Program. *Jt Comm J Qual Improv* 1995; 21: 343-7.
- Jencks SF, Wilensky GR. The health care quality improvement initiative. A new approach to quality assurance in Medicare. *JAMA* 1992; 268: 900-3.
- Hayes RP, Ballard DJ. Review: feedback about practice patterns for measurable improvements in quality of care-a challenge for PROs under the Health Care Quality Improvement Program. *Clin Perform Qual Health Care* 1995; 3: 15-22.
- Kiefe CI, Allison JJ, Williams OD, Person SD, Weaver MT, Weissman NW. Improving quality improvement using achievable benchmarks for physician feedback: a randomized controlled trial. *JAMA* 2001; 285: 2871-9.
- Wones RG. Failure of low-cost audits with feedback to reduce laboratory test utilization. *Med Care* 1987; 25: 78-82.
- Winickoff RN, Coltin KL, Morgan MM, Buxbaum RC, Barnett GO. Improving physician performance through peer comparison feedback. *Med Care* 1984; 22: 527-34.
- Mangram AJ, Horan TC, Pearson ML, Silver LC, Jarvis WR. *Guideline for prevention of surgical site infection*. Centers for Disease Control and Prevention (CDC) Hospital Infection Control Practices Advisory Committee. *Am J Infect Control* 1999; 1999(27): 97-134.
- Kirkland KB, Briggs JP, Trivette SL, Wilkinson WE, Sexton DJ. The impact of surgical-site infections in the 1990s: attributable mortality, excess length of hospitalization, and extra costs. *Infect Control Hosp Epidemiol* 1999; 20: 725-30.
- Poulsen KB, Bremmelgaard A, Sorensen AI, Raahave D, Petersen JV. Estimated costs of postoperative wound infections. A case-control study of marginal hospital and social security costs. *Epidemiol Infect* 1994; 113: 283-95.
- Vegas AA, Jodra VM, Garcia ML. Nosocomial infection in surgery wards: a controlled study of increased duration of hospital stays and direct cost of hospitalization. *Eur J Epidemiol* 1993; 9: 504-10.
- Boyce JM, Potter-Bynoe G, Dziobek L. Hospital reimbursement patterns among patients with surgical wound infections following open heart surgery. *Infect Control Hosp Epidemiol* 1990; 11: 89-93.
- Leslie K, Sessler D. Perioperative hypothermia in the high-risk surgical patient. *Best Pract Res Clin Anaesthesiol* 2003; 17: 485-98.
- Sessler DI. Mild Perioperative hypothermia. *N Engl J Med* 1997; 336: 1730-7.
- Sessler DI. Complications and treatment of mild hypothermia. *Anesthesiology* 2001; 95: 531-43.
- Kurz A, Sessler DI, Lenhardt R. Perioperative normothermia to reduce the incidence of surgical-wound infection and shorten hospitalization. Study of Wound Infection and Temperature Group. *N Engl J Med* 1996; 334: 1209-16.
- Kurz A. Thermal care in the perioperative period. *Best Pract Res Clin Anaesthesiol* 2008; 22: 39-62.
- Hooper VD, Chard R, Clifford T, et al. ASPAN's evidence-based clinical practice guideline for the promotion of perioperative normothermia: second edition. *J Perianesth Nurs* 2010; 25: 346-65.
- Dobson G, Chong M, Chow L, et al. Guidelines to the practice of anesthesia - revised edition 2018. *Can J Anesth* 2018; 65: 76-104.
- Harper CM, Andrzejowski JC, Alexander R. NICE and warm. *Br J Anaesth* 2008; 101: 293-5.
- Görge M, West NC, Whyte SD. Using physiological monitoring data for performance feedback: an initiative using thermoregulation metrics. *Can J Anesth* 2017; 64: 245-51.
- Boet S, Patey AM, Baron JS, et al. Factors that influence effective perioperative temperature management by anesthesiologists: a qualitative study using the Theoretical Domains Framework. *Can J Anesth* 2017; 64: 581-96.
- Weiser TG, Regenbogen SE, Thompson KD, et al. An estimation of the global volume of surgery: a modelling strategy based on available data. *Lancet* 2008; 372: 139-44.
- Campbell MK, Piaggio G, Elbourne DR, Altman DG; CONSORT Group. Consort 2010 statement: extension to cluster randomised trials. *BMJ* 2012; 345: e5661.
- National Archives and Records Administration. Federal Register. Department of Health and Human Services. Centers for Medicare and Medicaid Services. Rules and Regulations - December 2013. Available from URL: <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/Downloads/2014-Medicare-Physician-Fee-Schedule-Final-Rule.pdf> (accessed July 2018).
- Donabedian A. The quality of care. How can it be assessed? *JAMA* 1988; 260: 1743-8.
- Kaye AD, Okanlawon OJ, Urman RD. Clinical performance feedback and quality improvement opportunities for perioperative physicians. *Adv Med Educ Pract* 2014; 5: 115-23.
- Haller G, Stoelwinder J, Myles PS, McNeil J. Quality and safety indicators in anesthesia. *Anesthesiology* 2009; 110: 1158-75.
- Cane J, O'Connor D, Michie S. Validation of the theoretical domains framework for use in behaviour change and implementation research. *Implement Sci* 2012; 7: 37.
- Sun Z, Honar H, Sessler DI, et al. Intraoperative core temperature patterns, transfusion requirement, and hospital duration in patients warmed with forced air. *Anesthesiology* 2015; 122: 276-85.