

Lack of Association between Blood Pressure Management by Anesthesia Residents and Competence Committee Evaluations or In-training Exam Performance

A Cohort Analysis

Daniel I. Sessler, M.D., Natalya Makarova, M.S., Ricardo Riveros-Perez, M.D., David L. Brown, M.D., Stephen Kimatian, M.D.

ABSTRACT

Background: Prompt treatment of severe blood pressure instability requires both cognitive and technical skill. The ability to anticipate and respond to episodes of hemodynamic instability should improve with training. The authors tested the hypothesis that the duration of severe hypotension during anesthesia administered by residents correlates with concurrent adjusted overall performance evaluations by the Clinical Competence Committee and subsequent in-training exam scores.

Methods: The authors obtained data on 70 first- and second-year anesthesia residents at the Cleveland Clinic. Analysis was restricted to adults having noncardiac surgery with general anesthesia. Outcome variables were in-training exam scores and subjective evaluations of resident performance ranked in quintiles. The primary predictor was cumulative systolic arterial pressure less than 70 mmHg. Secondary predictors were administration of vasopressors, frequency of hypotension, average duration of hypotensive episodes, and blood pressure variability.

Results: The primary statistical approach was mixed-effects modeling, adjusted for potential confounders. The authors considered 15,216 anesthesia care episodes. A total of 1,807 hypotensive episodes were observed, lasting an average of 32 ± 20 min (SD) per 100 h of anesthesia, with 68% being followed by vasopressor administration. The duration of severe hypotension (systolic pressure less than 70 mmHg) was associated with neither Competence Committee evaluations nor in-training exam scores. There was also no association between secondary blood pressure predictors and either Competence Committee evaluations or in-training exam results.

Conclusions: There was no association between any of the five blood pressure management characteristics and either in-training exam scores or clinical competence evaluations. However, it remains possible that the measures of physiologic control, as assessed from electronic anesthesia records, evaluate useful but different aspects of anesthesiologist performance.

(*ANESTHESIOLOGY* 2016; 124:473-82)

THE Accreditation Council of Graduate Medical Education (ACGME) requires residency programs to evaluate residents across six core competencies. "Residents will be evaluated in six aspects: patient care, medical knowledge, problem-based learning and improvement, interpersonal and communication skill, professionalism, and system-based practice." The ACGME also requires residency programs to give residents semiannual feedback.¹

The competencies provide specific knowledge, skills, behaviors, attitudes, and the appropriate educational experiences required of residents to complete Graduate Medical Education programs with the ultimate goal of creating competent, self-reflective physicians who are lifelong learners. In 1999, Siker² suggested that "competent clinicians should possess the intellectual capacity to make valid medical

What We Already Know about This Topic

- Evaluation of clinical performance by anesthesia residents is suboptimal, guided primarily by subjective faculty evaluations and infrequent quantitative knowledge assessments
- Electronic anesthesia records provide detailed measures of hemodynamic variables that could provide performance quality information

What This Article Tells Us That Is New

- In a cohort of 70 anesthesia residents, there was no association between five measures of blood pressure control obtained from electronic anesthesia records and either faculty evaluations of clinical competence or quantitative knowledge testing
- Although negative, this study provides a novel and important initial attempt to use electronic anesthesia records to evaluate clinical performance

Corresponding article on page 259.

Submitted for publication December 16, 2014. Accepted for publication September 2, 2015. From the Departments of Outcomes Research (D.I.S., N.M.), Department of Quantitative Health Sciences (N.M.), Department of Pediatric Anesthesia (R.R.-P., S.K.), and Anesthesiology Institute (D.L.B.), Cleveland Clinic, Cleveland, Ohio.

Copyright © 2015, the American Society of Anesthesiologists, Inc. Wolters Kluwer Health, Inc. All Rights Reserved. *Anesthesiology* 2016; 124:473-82

judgment and the technical expertise in their own special field or endeavor to implement such judgment.” There are several ways to objectively assess the competency of medical knowledge including nationally standardized and validated written and oral examinations. But objectively assessing other competencies is far more difficult.

Developing skills in administering anesthesia and maintaining patient homeostasis are competencies of anesthesia that require an integration of knowledge, skill, and judgment. Malik *et al.*³ performed a systematic review from nine national educational surveys that comprise the opinion of 1,076 program directors. The two competences receiving the highest priority for assessment were patient care and medical knowledge. Several tools have been developed for evaluating these aspects of resident performance including health-care matrixes, report cards, evaluation by faculty members, 360-degree evaluations, and self-evaluations. Most are subjective, subjected to bias, and required long periods of observation to produce accurate and reliable results.

Residency programs use various methods to evaluate the six ACGME competences. By using the data collected during the ACGME site visits over 2 yr, Holt *et al.*⁴ analyzed the methods and evaluation approaches for assessing resident performance in each competency. The most common assessment methods were direct observation and global assessment, mostly by program directors and attendings. A limitation of global assessments is that standards for evaluation vary by context, level of training, and across evaluators.⁵ The authors thus encouraged educators to incorporate specific methods of evaluating patient safety and quality in evaluating residents’ clinical performance.

Automated anesthesia information management systems are revolutionary tools that are becoming a routine fixture of the operating room. Benefits of anesthesia information management systems include improved accuracy and legibility of clinical documentation, access to previously unsearchable perioperative data for clinical research, generation of statistical benchmarks for quality improvement programs, and auditable evidence of compliance with the documentation requirements of regulatory authorities, and third-party payers. Many electronic anesthesia record-keeping systems, including the Cleveland Clinic’s, provide an unmodifiable record of vital signs.

Immediate detection and prompt treatment of severe blood pressure instability is a key element of anesthesia practice that requires both cognitive and technical skill. As cognitive and technical skills mature, the ability to anticipate and respond to episodes of hemodynamic instability should improve. We thus tested the primary hypothesis that the duration of severe hypotension (cumulative minutes of systolic pressure less than 70 mmHg/h) during anesthesia administration by residents correlates with concurrent adjusted overall performance evaluations by the Clinical Competence Committee and subsequent in-training exam scores.

Secondarily, we tested the hypotheses that elapsed time between the onset of severe hypotension and vasopressor

administration, the frequency of hypotensive episodes, the average duration of hypotensive episodes, and blood pressure stability represented by the SD of mean arterial pressure (MAP) during anesthesia administration by residents correlates with concurrent adjusted overall performance evaluations by the Clinical Competence Committee and subsequent in-training exam scores. Confirming our hypotheses would suggest an objective and unbiased method for evaluating the competency of patient care and identify residents who might benefit from early intervention, giving them the best opportunity to improve.

Materials and Methods

With approval of the Cleveland Clinic Institutional Review Board (Cleveland, Ohio), the study was conducted with waived consent. To protect residents, only deidentified summary statistics were provided to the investigators and residency officials. It was, therefore, impossible for residency officials or any other faculty to link our novel measures of blood pressure management to any particular resident. Results of the study were thus unable to influence official or unofficial evaluations of participating residents.

We obtained data on 70 first-year (clinical anesthesia [CA]-1) and second-year (CA-2) CA residents at Cleveland Clinic, between July 1, 2011, and June 30, 2013. Our analysis was restricted to residents who started CA training in July and who did not have previous anesthesia residency training. Intraoperative blood pressure information was obtained on 15,216 anesthesia episodes (10,065 unique surgeries) from the Cleveland Clinic Perioperative Health Documentation System and Anesthesia Record Keeping System registries. Anesthesia episodes refer to continuous periods of time during which a single resident was signed into a case.

We considered only noncardiac, nonemergency surgeries in adults performed under general anesthesia or combined general and regional anesthesia. We excluded anesthesia episodes with missing induction or emergence time stamps, incorrect time stamps, and episodes with residents in operating room less than 45 min between anesthetic induction and emergence. Only times during which residents were electronically signed into a case were considered. We excluded operations in which deliberate hypotension was used. And finally, we excluded the first 2 months of the CA-1 year because residents are paired with senior residents on a one-to-one basis during this period (fig. 1).

Outcome Variables

American Board of Anesthesiologists In-training Examination Scores. For each resident, we used the percentile score reported by the American Board of Anesthesiologists in-training examination (ABA-ITE) from the relevant year.

Resident Evaluations by the Clinical Competency Committee (Competency Committee Evaluation). The committee evaluates each resident in narrative form at approximately 3-month intervals using all available information including

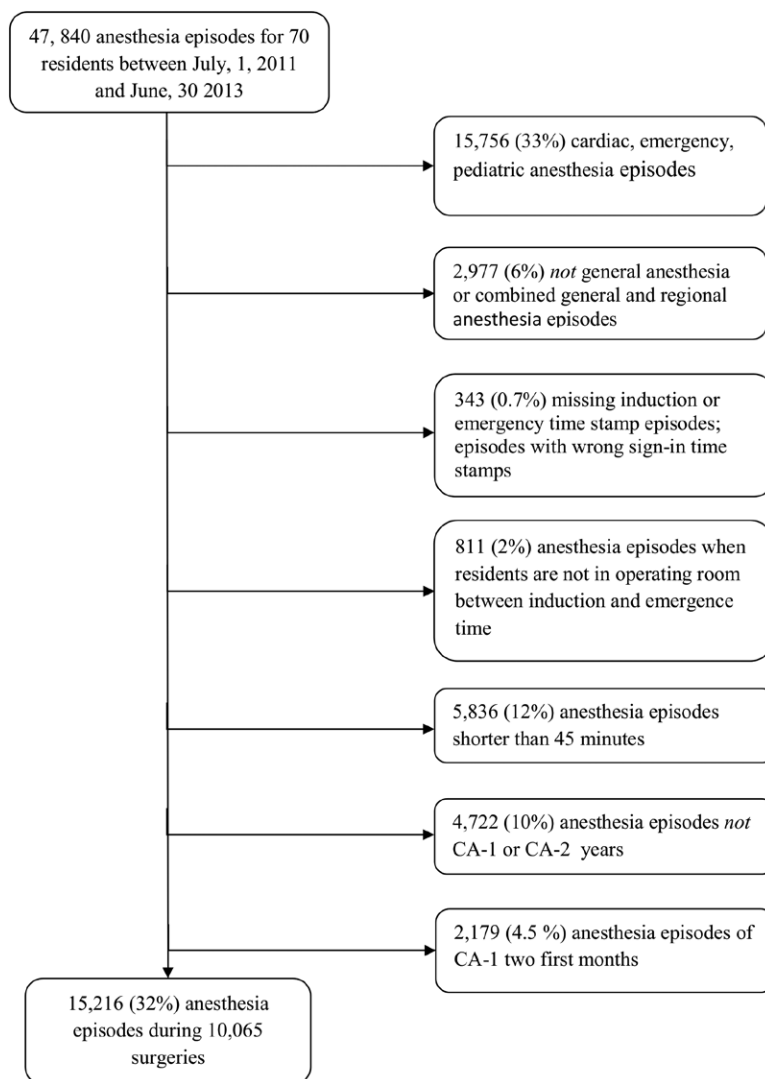


Fig. 1. Flow chart of patient selection.

attending evaluations of residents' performance relative to the ACGME competencies. Two investigators (R.R.-P. and S.K.) independently reviewed all the committee's reports for each resident in each academic year and divided the residents into ordered performance quintiles within each class and year. Differences were resolved by consensus. The investigators assigning performance quintiles did not consider in-training scores. The evaluation was thus meant to be an independent measure of performance.

Blood Pressure Intraoperative Measures: Predictor Variables

All intraoperative blood pressure measures were based on recordings while residents were signed into an operating room between induction and emergence. Pressures were stored every minute when an arterial catheter was used (the median of values obtained at approximately 2-s intervals) and every 1 to 5 min when noninvasive oscillometric blood pressure monitoring was used.

Clinicians could mark blood pressures as artifactual but had no ability to alter recorded values. Blood pressure readings were assumed to be artifacts and removed using the following sequential rules: (1) documented by clinicians as artifact; (2) systolic pressure 300 mmHg or greater or less than 20 mmHg; (3) systolic pressure was less than diastolic pressure plus 5 mmHg; or (4) diastolic pressure was less than 5 mmHg or more than 225 mmHg.

Severe hypotension was expressed as minutes per 100 h of anesthesia, with duration being defined by the period between an initial systolic blood pressure less than 70 mmHg and the first subsequent pressure exceeding 70 mmHg. We assumed that a given pressure was maintained until a new one was registered.

Time to vasopressor administration was defined as the average time elapsed between the onset of severe hypotension and administration of a vasopressor (ephedrine, epinephrine, dopamine, dobutamine, isoproterenol, metaraminol, milrinone, norepinephrine, phenylephrine, or vasopressin).

When the time exceeded 10 min, vasopressor administration was not considered responsive to the preceding hypotension. Only cases with responsive vasopressor administration were included in the average. The frequency of severe hypotension was defined as the average number of episodes per 100 h of anesthesia. Blood pressure stability was defined by the SD of MAP over the entire monitored period.

Potential confounding variables included patient factors, such as age, sex, race, body mass index (BMI), American Society of Anesthesiologists (ASA) physical status, and the present-on-admission risk (POARisk) of hospital mortality measure,⁶ and residents factors, specifically residents' year of graduation and CA training year.

Because the study included the data from 2 yr, some residents were considered during both their CA-1 and CA-2 years. The end-of-year in-training examination score percentile and Clinical Competence Committee performance evaluations were included separately for each year for each of these residents.

Primary Hypotheses

We estimated strength of the linear relationship between percentile ABA-ITE scores and the total duration of severe hypotension in a mixed-effects model⁷ with percentile ABA-ITE score as the continuous outcome variable. We accounted for potential confounding factors and possible intrasubject correlation (assumed unstructured covariance matrix) of repeated exam scores within a resident; each resident might have one or two exam scores. The potential confounding adjustment included resident factors and summarized patient factors. For modeling purposes, all the patients' potential confounders were summarized for each resident with the following summary measures: average patients' age, percentage of female patients, percentage of Caucasian patients, percentage of ASA status III and greater, average BMI, and average risk of in-hospital mortality based on the present-on-admission version of the Risk Stratification Index.

The direction and strength of the relationship between exam performance and total duration of severe hypotension were assessed with the slope of the regression model along with corresponding confidence limits. The slope was tested against zero with the Wald test.

To evaluate the additional primary hypothesis of linear association between Competence Committee evaluation and the total duration of severe hypotension, we developed a proportional odds logistic regression with repeated-measures model. The model allowed us to accommodate the ordinal nature of the multilevel response variable (*i.e.*, first quintile better than the second quintile and so on), adjusted for potential confounding factors, and the possible correlation among observations within a resident.

The proportional odds model estimated the association between the duration of hypotension and resident evaluations *via* odds ratio. The resulting odds ratio estimated the

relative odds of being ranked into a higher quintile with each extra 1 min of severe hypotension per 100 h of anesthesia. The odds ratios were tested against one with a model-based Wald test.

A Bonferroni correction for multiple tests was applied, and a significance criterion of $0.05/2 = 0.025$ was used to control the type I error of the primary hypothesis at 5%.

Secondary Hypotheses

To estimate the association between four secondary blood pressure measures and the two outcomes (exam scores and evaluation), we built separate models for each blood pressure measure and outcome as described in the Primary Hypotheses section above. Model-based Wald tests were used to formally test the hypotheses.

The Bonferroni correction for multiple inferences was applied, and a significance criterion of $0.05/(2 \times 4) = 0.006$ at each blood pressure measure was used to control overall type I error of the secondary analyses at 5%. We used multivariable clustering analysis to identify the potential "resident outliers" based on their five averaged blood pressure measurements. Multidimensional Euclidean distance and single linkage algorithm (nearest neighbor) were used for clustering. The distance between residents was graphically displayed by dendrogram. By using Cook's distance,⁸ we additionally determined whether potential resident outliers unduly influenced the results. Cook's distance exceeding $4/70 = 0.06$ was considered indicative of potentially influential observations.⁹

Sensitivity Analysis

The in-training exam and Competence Committee evaluations happened toward the end of the academic year and were chosen as the outcomes in primary analysis. As a sensitivity analysis, we modeled the associations of interest at the level of individual case by switching the outcomes and predictors to adjust for the listed covariables at the case level (instead of averaged over all residents' cases over the academic year). In each model, we adjusted for patient age, BMI, sex, race, ASA physical status, POARisk score, year of the surgery, and possible intrasubject correlation of cases done by the same resident. To account for clinical development across anesthesia training, we also included "month in anesthesia training" for adjustment at the time when the case occurred.

First, we assessed whether the incidence of hypotension was associated with the in-training exam percentile (or Competence Committee evaluation) using hypotension as a binary outcome and resident exam percentile (or Competence Committee evaluation) as a predictor based on all 10,065 surgeries. Then, we estimated the association between exam score and hemodynamics (five blood pressure characteristics) among the 1,196 (12%) surgeries with at least one hypotension episode. Separate models were developed with a hemodynamic measure for the cases as an outcome and exam percentile (or Competence Committee

evaluation) as a predictor with the same adjustments used for the primary outcome. We used log transformations for the hemodynamic outcomes to resemble normal distributions. In addition, we assessed whether the duration in anesthesia training was independently associated with hemodynamic outcomes by fitting five separate models with each hemodynamic measure for the case as an outcome and number of months in training as a predictor. Each association was adjusted for the patient's characteristics such as age, sex, race, BMI, ASA physical status, and POARisk score.

Sample Size Considerations

We planned to use information on all available and qualifying CA-1 and CA-2 residents at Cleveland Clinic between July 2011 and June 2013. Given approximately 50 residents and 7 potential confounders, we had approximately 90% power to detect a significant partial correlation at least as large as of 0.55 between the outcome and the duration of severe hypotension, while restricting type I error to 0.025.

SAS 9.3 software (SAS Institute, USA) and R statistical software version 2.7.2 (The R Foundation for Statistical Computing, Austria) were used for all analyses and graphics.

Results

A total of 79 first-year (CA-1) and second-year (CA-2) CA residents at Cleveland Clinic who were eligible for the study during the period from 2011 to 2012 and 2012 to 2013 academic years. We excluded nine of them because they started mid-year or had previous anesthesia training; therefore, we enrolled 70 anesthesia residents in the study. There were 31

residents who participated both as CA-1 and CA-2 in the study, all of whom started residency in June 2010. There were 19 residents who started residency on June 2009 and, therefore, participated in the study as CA-2 residents for the academic year 2011 to 2012. There were 20 residents who started residency on June 2011 and took part in the study as CA-1 residents for academic year 2012 to 2013.

An arterial catheter was used for blood pressure monitoring in 32% of the included anesthesia episodes.

We considered 15,216 anesthesia episodes representing 10,065 unique surgeries. Among the 10,065 surgeries, hypotension episodes were recorded in 1,196 (12%) including 290 (3%) surgeries where more than one hypotensive episode occurred. A total of 1,807 hypotension episodes were observed while residents were in the operating room with 1,220 (68% of hypotension episodes) being followed by vasopressor administration within 10 min from the onset of hypotension. Blood pressures, examination scores, and Competence Committee evaluations of the residents are presented in table 1, with the distributions shown in figure 2. There was no evidence of selective case assignment as might occur if weaker residents were "protected"; in other words, the distribution of case severities (adjusted for month of training) was similar across the entire study cohort.

Adjusting for potential confounding, we found no evidence that duration of severe hypotension was association with either Competence Committee evaluations or in-training exam results for the residents (table 2 and fig. 3). Similarly, there was no association between secondary blood pressure predictors and either Competence Committee evaluations or in-training exam results (table 2).

Table 1. Summary Description of Raw Hemodynamic Measures, Examination Scores, and Faculty Evaluations Based on 101 Resident-years

Measures	Mean \pm SD	Median (Lower Quartile, Upper Quartile)
Hemodynamic intraoperative measures		
Primary predictor		
Duration of severe hypotension, minutes per 100h of anesthesia	32 \pm 20	27 (18, 41)
Secondary predictors		
Average duration of severe hypotension episode, min	2.7 \pm 0.7	2.6 (2.2, 3.2)
Time elapsed between the onset of severe hypotension and administration of a vasopressor, min	1.9 \pm 0.9	1.8 (1.4, 2.4)
Number of episodes per 100h of anesthesia	11 \pm 6	10 (8, 14)
SD of MAP, mmHg	11.7 \pm 0.5	11.7 (11.4, 12.0)
Formal performance measures		
End-of-year in-training examination, percentile	34 \pm 6	34 (30, 39)
Clinical Competence Committee ranks*	N (%)	
First quintile	16 (15%)	
Second quintile	21 (21%)	
Third quintile	22 (22%)	
Fourth quintile	22 (22%)	
Fifth quintile	20 (19%)	

* The Clinical Competence Committee evaluation was an ordered categorical outcome with five categories: "first quintile" better than "second quintile" better than "third quintile" better than "fourth quintile" better than "fifth quintile."

MAP = mean arterial pressure.

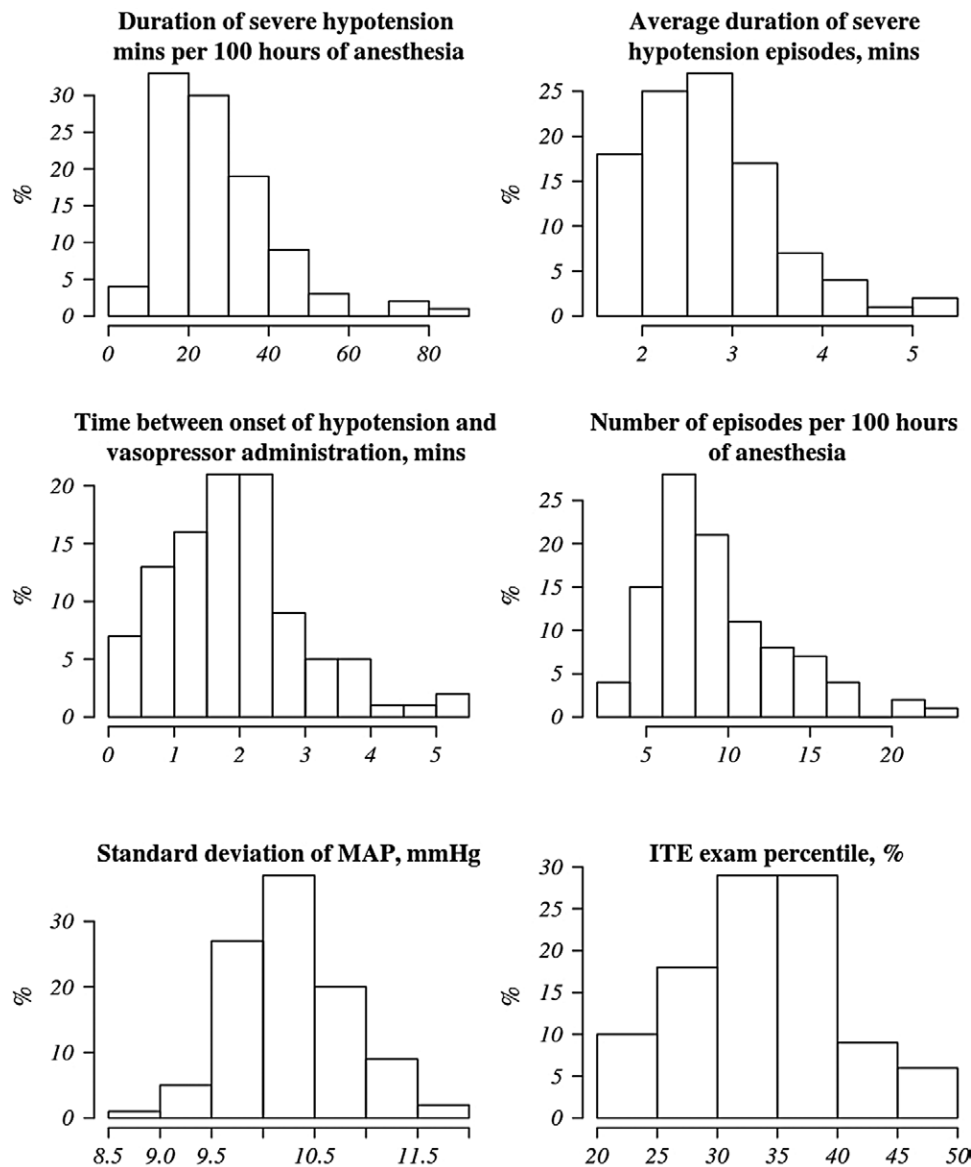


Fig. 2. The distributions of the hemodynamic intraoperative measures and in-training exam percentiles for 101 resident-years. ITE = in-training exam; MAP = mean arterial pressure.

We identified potential “resident outliers” based on the five averaged blood pressure measures. From the dendrogram (fig. 4), we see that most of the residents are at the low distance (high similarity) level. The residents on the top of the dendrogram are potential “residency outliers” who are less similar to the rest of the residents on five blood pressure measurements. None of the observed potential resident outliers was among the influential observations.

Sensitivity analyses were consistent with the primary and secondary results in not showing any significant association between the chosen hemodynamic measures and either in-training exam percentile or Competence Committee evaluation. None of the 10 associations changed over the months of anesthesia training (no interaction effect).

Months in anesthesia training was independently associated (after adjusting for patient characteristics) with three outcomes: “Duration of severe hypotension, minutes per 100 h of anesthesia,” “Average duration of severe hypotension episodes, minutes,” and “Number of episodes per 100 h of anesthesia.” All three associations were negative, suggesting that residents were better able to control each of these hemodynamic measures as their clinical experience increased.

Discussion

There was no association between any of the five blood pressure management characteristics we evaluated and either in-training exam scores or clinical competence evaluations. Our results thus do not provide evidence that measures of blood pressure control, a primary anesthetic responsibility,

Table 2. The Association between Hemodynamic Measures and In-training Exam Percentile in 70 Residents and 101 Resident-year Combinations

	Nonadjusted Mean Difference in ABA-ITE Percentile, % (CI)*	Mean Difference in ABA-ITE Percentile, % (Adjusted CI)†‡	P Value§	Nonadjusted Faculty Evaluation, Odds Ratio (CI)*	Faculty Evaluation, Odds Ratio (Adjusted CI)†	P Value§
Primary predictor						
Duration of severe hypotension, minutes per 100 h of anesthesia	0.08 (0.03 to 0.13)	0.04 (−0.02 to 0.10)‡	0.11	1.00 (0.98 to 1.01)	1.0 (0.98 to 1.02)	0.93
Secondary predictor						
Average duration of severe hypotension episodes, min	−0.67 (−2.76 to 1.43)	−0.58 (−2.54 to 1.39)	0.41	1.37 (0.82 to 2.30)	1.38 (0.75 to 2.54)	0.12
Time elapsed between the onset of severe hypotension and administration of a vasopressor, min	0.72 (−0.97 to 2.42)	0.49 (−1.08 to 2.07)	0.38	0.85 (0.56 to 1.30)	0.84 (0.52 to 1.36)	0.27
Number of episodes per 100 h of anesthesia	0.34 (0.16 to 0.52)	0.22 (−0.02 to 0.47)	0.01	0.97 (0.91 to 1.04)	0.96 (0.88 to 1.05)	0.19
SD of MAP, mmHg	−0.64 (−3.5 to 2.26)	−0.79 (−3.60 to 2.02)	0.43	1.23 (0.57 to 2.62)	1.44 (0.57 to 3.66)	0.24

* The results were not adjusted for confounding variables although the results were adjusted for possible intrasubject correlation of repeated records within a resident; 97.5% CI were reported for the primary predictor and 99.4% CI for the secondary predictors. † 97.5% CIs reported for the primary predictor and 99.4% CIs reported for the secondary predictors. Intervals were adjusted for possible intrasubject correlation of repeated records within a resident and the potential confounding factors including CA-1 or CA-2, year, and average patient characteristics such as age, sex, race, body mass index, American Society of Anesthesiologists physical status, and present-on-admission risk of hospital mortality. ‡ Mean difference in ABA-ITE percentile (%) quantifies the effect of primary and secondary predictors on the ABA-ITE for each unit increase in a predictor. For the primary predictor, for example, the ABA-ITE percentile increases insignificantly by 0.04% (−0.02 to 0.10%) with each 1 min/100h increase in the duration of severe hypotension. § P value significance criteria were 0.025 for the primary predictor and 0.006 for the secondary predictors. || Faculty evaluation odds ratio assesses the effect of primary and secondary predictors on faculty evaluation for each unit increase in a predictor. For the second secondary predictor, for example, the odds of being ranked into a higher quintile decreased (nonsignificantly) by 16% with each extra 1 min between the onset of severe hypotension and administration of a vasopressor. ABA-ITE = American Board of Anesthesiologists in-training examination; CA = clinical anesthesia; MAP = mean arterial pressure.

are associated with either of the two most commonly used measures of resident performance. Our hypotheses were thus disproved and it is clear that blood pressure control cannot be substituted for in-training exams or Competence Committee evaluations. Realistically, of course, blood pressure control measures would never be substituted for either in-training exam scores or clinical competence evaluations; but blood pressure measures obtained from electronic anesthesia records might have provided an early warning for residents who could profit from intervention.

There was a considerable variation among residents in the degree of blood pressure control using each of our five definitions. Possibly, blood pressure is effectively not under the anesthesiologist's control and entirely random or a function of underlying patient physiology and surgical factors. However, this is an unattractive conclusion because maintaining hemodynamic control is generally considered an important anesthetic function. It is also a conclusion that is inconsistent with clinical experience, which strongly suggests that skilled anesthesiologists both prevent hemodynamic extremes and respond quickly and effectively when they do occur. Thus, it seems more likely that blood pressure control, averaged over 100 h of care, indeed reflects an anesthesiologist's skill. And consistent with this theory, the amount of hypotension decreased over training time after adjustment for patient characteristics. The question, then, is why none of our blood pressure control measures correlated with Competence Committee evaluations or in-training exams scores?

Clinical evaluations and in-training exams are blunt instruments. Clinical evaluations in general are poorly validated, and those used at the clinic are probably not much better or worse than those used in most residency programs. And like most, they have never been formally validated. The in-training exam is better validated but is designed to evaluate the knowledge rather than skill. Neither is thus an ideal reference, but these were the two quantitative measure of resident performance available to us.

There are a limited number of alternative explanations for our negative results: (1) blood pressure is not under anesthesiologist control—which seems clearly false; (2) blood pressure management does not matter—which seems unlikely given that hypotension is such a strong predictor of mortality¹⁰; (3) anesthesia residents prioritize other aspects of care above hemodynamic management—which is certainly not our impression from decades of teaching; (4) type 2 error or analysis errors; (5) clinical evaluations (which are perhaps excellent for professionalism and other aspects of performance) poorly assess anesthesia skills; and (6) the in-training exam assesses knowledge rather than clinical skill/performance. Among these, the last two seem most likely and thus represent the most probable reasons for false-negative results.

The underlying assumption for our study was that either Competence Committee evaluations or in-training exams are reliable estimates of residents' clinical performance. It remains possible that, although the correlation with blood

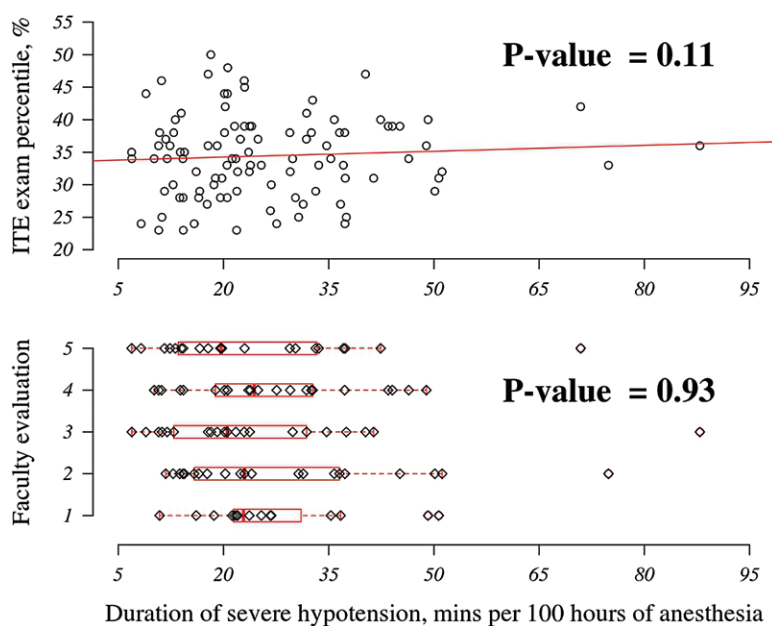


Fig. 3. Relationship between the duration of severe hypotension and faculty evaluations and in-training exam results (raw observations displayed). There was no association between duration of severe hypotension residents and either their in-training exam (ITE) results ($P = 0.11$) or faculty evaluations ($P = 0.93$). The analysis was adjusted for potential confounding patient factors such as age, sex, race, body mass index, American Society of Anesthesiologists physical status, and the present-on-admission risk of hospital mortality; it was also adjusted for resident factors such as year of graduation and clinical anesthesia training year.

pressure control was essentially nonexistent, the blood pressure control is nonetheless a valid indicator because it evaluates a *different dimension* of resident performance. Or to put it another way, our results may be “negative” because we asked the wrong question.

To the extent that what we as anesthesiologists do matters,¹¹ critical decisions and actions should be reflected in electronic anesthesia records. After all, if it is impossible to quantify anesthesia performance from records, then either something is wrong with our records, which seems unlikely given their breadth and density, or intraoperative decisions and actions do not actually matter—which also seems unlikely. Almost surely anesthetic decisions do matter and are perfectly well reflected in our electronic records.

Our study represents an initial attempt to use information contained in electronic records, specifically blood pressure control, to evaluate the performance quality. Although the attempt failed, there is almost surely performance information coded in anesthesia records—although it remains to be determined what measures best quantify the skill. Although “negative,” our study is important because it identifies an entirely novel approach to anesthesiologist evaluation. We expect that subsequent studies will eventually determine which measures best reflect the quality of intraoperative decision-making and performance. When they do, evaluations based on electronic records may well become standard supplements to Competence Committee evaluations and the in-training exam.

We fully recognized the limitations of comparing blood pressure control to Competence Committee evaluations or in-training scores while designing our study. Our difficulty, though, is that there are no other generally available measures of performance, and the two we chose are by far the most common measures of resident performance. Among other potential measures, performance in simulators seems most likely to correlate with physiologic control during practice.

Sidi *et al.*,¹² for example, was able to differentiate between cognitive and technical skills in 47 postgraduate years three and four anesthesia residents using simulation scenarios with specific checklists for trauma, operating room, and cardiac resuscitation. But to our knowledge, there is no operative assessment tool relevant to anesthesia and generalizable to the wide spectrum of clinical care and medical knowledge. We considered the fact that a simulation environment might provide a better environment for evaluation of endpoints relative to morbidity or mortality (such as severe hypotension) because it is highly likely that safeguards such as monitor alarms and faculty intervention are structured to specifically prevent these events in clinical practice. Although it might be interesting to examine much longer-term measures of performance such as success in practice or malpractice claims, neither seems likely to be a sensitive indicator of clinical skill. Furthermore, Ryan *et al.*¹³ found no correlation between ITE scores and clinical performance measured by faculty evaluation of medical knowledge and overall clinical competency in emergency medicine residents.

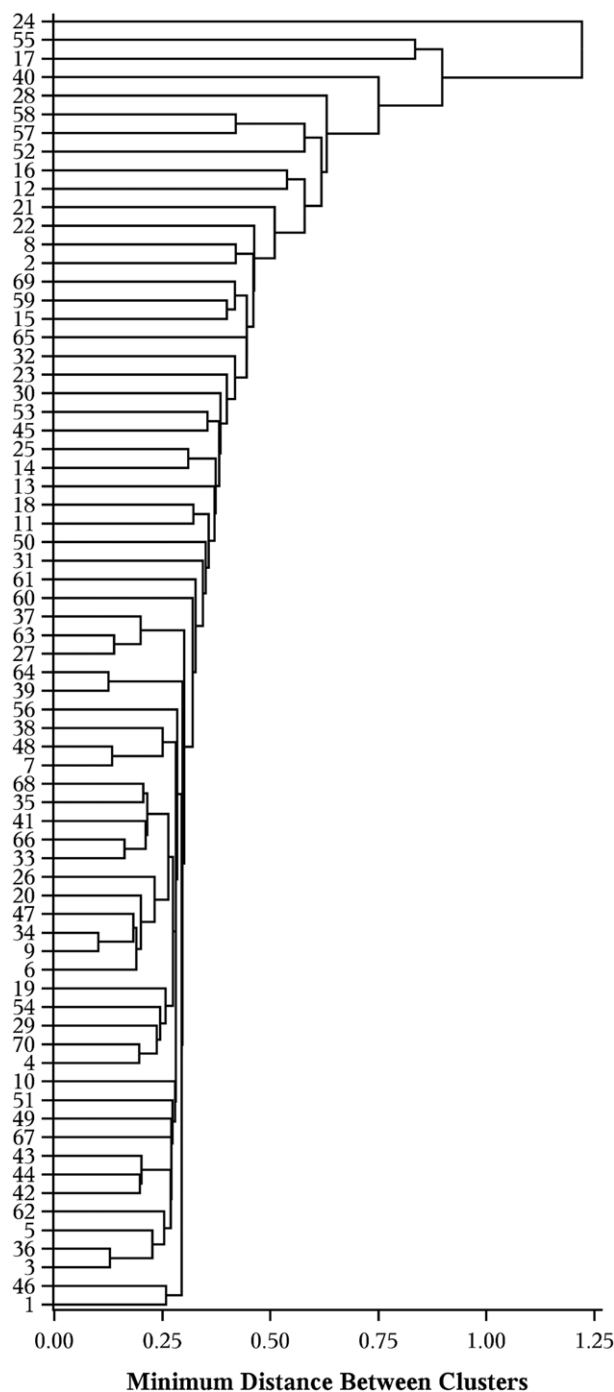


Fig. 4. Dendrogram for distance between residents based on five hemodynamic intraoperative measures. The residents on the top of the dendrogram are potential “residency outliers” who are less similar to the rest of the residence on five blood pressure measurements. None of the observed potential resident outliers was among the influential observations.

We considered five measures of blood pressure control. However, there is no reason to restrict analysis of electronic data to blood pressure, and it remains possible that heart rate or respiratory variables might be more useful measures

of performance. Similarly, there is no reason to restrict analysis of electronic data to residents because skill at anticipating and controlling other physiologic measures should apply equally well to practicing anesthesiologists. Thus, although our results disprove our hypotheses, we remain enthusiastic about the general concept of using measures of physiologic control as derived from electronic records as indicators of anesthesiologist skill.

A systolic blood pressure of 70 mmHg corresponds roughly to an MAP of 55 mmHg, which, since our current study was started, was shown to be strongly associated with acute kidney and myocardial injury.¹⁴ The blood pressure threshold we used thus appears to be highly clinically relevant. More importantly, a systolic pressure of 70 mmHg is a value that no anesthesiologist would normally tolerate in adults. Although it remains possible that another measure of blood pressure control would correlate better with independent evaluations of resident skill, the five we tested cover a fairly broad range. We did not evaluate other measures of clinical skill such as heart rate control, and it is possible that some other measure is superior.

Faculty evaluations of residents are, by their nature, subjective and thus potentially biased. However, each resident was evaluated by many faculty; furthermore, we averaged evaluations over 10 to 12 months. These evaluations, along with the Competency Committee’s experience, provided a reasonable estimate of the faculty’s impression of each resident’s performance. It thus seems likely that the resulting quintile assignments reasonably reflected the faculty’s evaluation of individual resident performance. Nonetheless, our quintile assignments are based on completely subjective evaluations and thus inherently lack the rigor of primary quantitative assessments.

Although our electronic anesthesia record is objective and blood pressure values cannot be modified by users, it is subjected to artifact and error. However, there is no reason to believe that artifact or error would be anything but randomly distributed among residents. Random error adds noise to the analysis and reduces precision, but it seems highly unlikely that registry error was sufficient to obliterate real associations between blood pressure control and other measures of resident performance. We considered a single value for each of the two dependent and five independent variables for each resident for each year, each based on averages over 10 to 12 months. This approach improves reliability of our assessments by virtue of averaging many faculty evaluations and much anesthetic management.

Our residents are always supervised by staff anesthesiologists, nearly always with two residents per attending. Staff anesthesiologist’s decisions and plans thus considerably influence anesthetic care. But over a 1- to 2-yr period, distribution of staff was presumably at least roughly homogenous, leaving resident skill as the primary overall determinant of blood pressure control. A limitation of our

analysis is that we could not determine from our records when attending anesthesiologists were actually in a specific room. It thus remains possible that residents perceived to be weaker may have been better supervised and that blood pressure control in such cases may have more reflected attending than resident skill. It is also possible that residents thought to be weaker may have been assigned easier cases. We note though that the case difficulty is largely determined by baseline medical condition rather than the operation *per se*, a factor that was unknown by those assigning cases.

An arterial catheter was inserted in one third of the patients included in our analysis. Blood pressure was recorded at 1-min intervals in patients who had arterial catheters. Blood pressure was measured oscillometrically in the remaining patients. Although there can be potentially important differences in systolic or diastolic pressures with the two techniques, MAP is similar with each. But for the purpose of this study, it is resident *response* to systolic hypotension that is important rather than absolute accuracy of the blood pressure monitoring method. Furthermore, use of an arterial catheter is largely determined by the type of case and underlying patient comorbidity, factors that were presumably randomly distributed among residents.

A more important difference between direct and oscillometric measurements is that oscillometric pressures are usually obtained at 2- to 5-min intervals, whereas direct measures are continuously available. However, anesthesiologists can trigger a “stat” blood pressure at any time—and normally would request frequent readings during a period of critical hypotension. In fact, requesting a “stat” pressure is an appropriate response to hypotension and thus part of hemodynamic management. We, therefore, believe that an adequate number and frequency of blood pressure readings were (or could have been) available for analysis during the relevant periods, even in patients having oscillometric measurements.

In summary, there was no association between any of the five blood pressure management characteristics and either in-training exam scores or clinical competence evaluations. It remains likely though that some measures of physiologic control, as assessed from electronic anesthesia records, evaluate useful but different aspects of anesthesiologist performance.

Acknowledgments

Support was provided solely from institutional and/or departmental sources.

Competing Interests

The authors declare no competing interests.

Correspondence

Address correspondence to Dr. Sessler: Department of Outcomes Research, Cleveland Clinic, 9500 Euclid Avenue-P77, Cleveland, Ohio 44195. ds@or.org; www.OR.org. Information on purchasing reprints may be found at www.anesthesiology.org or on the masthead page at the beginning of this issue. ANESTHESIOLOGY's articles are made freely accessible to all readers, for personal use only, 6 months from the cover date of the issue.

References

1. The Accreditation Council for Graduate Medical Education and the American Board of Anesthesiology: The Anesthesiology Milestone Project, 2013. Available at: <http://www.acgme.org/acgmeweb/Portals/0/PDFs/Milestones/AnesthesiologyMilestones.pdf>. Accessed September 6, 2015
2. Siker ES: Assessment of clinical competence. *Curr Opin Anaesthesiol* 1999; 12:677–84
3. Malik MU, Diaz Voss Varela DA, Stewart CM, Laeeq K, Yenokyan G, Francis HW, Bhatti NI: Barriers to implementing the ACGME outcome project: A systematic review of program director surveys. *J Grad Med Educ* 2012; 4:425–33
4. Holt KD, Miller RS, Nasca TJ: Residency programs' evaluations of the competencies: Data provided to the ACGME about types of assessments used by programs. *J Grad Med Educ* 2010; 2:649–55
5. Huddle TS, Heudebert GR: Taking apart the art: The risk of anatomizing clinical competence. *Acad Med* 2007; 82:536–41
6. Dalton JE, Glance LG, Mascha EJ, Ehrlinger J, Chamoun N, Sessler DI: Impact of present-on-admission indicators on risk-adjusted hospital mortality measurement. *ANESTHESIOLOGY* 2013; 118:1298–306
7. Breslow NE, Clayton DG: Approximate inference in generalized linear mixed models. *J Am Stat Assoc* 1993; 88:9–25
8. Cook D: Detection of influential observations in linear regression: An expository treatment of outliers and influential cases. *Technometrics* 1977; 19:15–8
9. Bollen K, Jackman RW: Regression diagnostics: An expository treatment of outliers and influential cases, *Modern Methods of Data Analysis*. Edited by Fox J, Long JS. Newbury Park, Sage, 1990, pp 257–91
10. Mascha EJ, Yang D, Weiss S, Sessler DI: Intraoperative mean arterial pressure variability and 30-day mortality in patients having noncardiac surgery. *ANESTHESIOLOGY* 2015; 123:79–91
11. Glance LG, Kellermann AL, Hannan EL, Fleisher LA, Eaton MP, Dutton RP, Lustik SJ, Li Y, Dick AW: The impact of anesthesiologists on coronary artery bypass graft surgery outcomes. *Anesth Analg* 2015; 120:526–33
12. Sidi A, Baslanti TO, Gravenstein N, Lampotang S: Simulation-based assessment to evaluate cognitive performance in an anesthesiology residency program. *J Grad Med Educ* 2014; 6:85–92
13. Ryan JG, Barlas D, Pollack S: The relationship between faculty performance assessment and results on the in-training examination for residents in an emergency medicine training program. *J Grad Med Educ* 2013; 5:582–6
14. Walsh M, Devereaux PJ, Garg AX, Kurz A, Turan A, Rodseth RN, Cywinski J, Thabane L, Sessler DI: Relationship between intraoperative mean arterial pressure and clinical outcomes after noncardiac surgery: Toward an empirical definition of hypotension. *ANESTHESIOLOGY* 2013; 119:507–15