Effects of Volatile Anesthetic Choice on Hospital Length-of-stay

A Retrospective Study and a Prospective Trial

Tatyana Kopyeva, M.D.,* Daniel I. Sessler, M.D.,† Stephanie Weiss, Ph.D., M.D.,‡ Jarrod E. Dalton, Ph.D.,§ Edward J. Mascha, Ph.D.,‖ Jae H. Lee, B.A.,# Ravi P. Kiran, M.D.,** Belinda Udeh, Ph.D.,†† Andrea Kurz, M.D.‡‡

ABSTRACT

Background: Volatile anesthetic prices differ substantially. But differences in drug-acquisition cost would be inconsequential if hospitalization were prolonged by more soluble anesthetics. The authors tested the hypothesis that the duration of hospitalization is prolonged with isoflurane anesthesia.
Methods: Initially, the authors queried their electronic records and used propensity matching to generate homogeneous sets of adults having inpatient noncardiac surgery who were given desflurane, sevoflurane, and isoflurane. The authors then conducted a prospective alternating intervention trial in which adults (mostly having colorectal surgery) were assigned to isoflurane or sevoflurane, based on protocol.
Results: In the retrospective analysis, 2,898 matched triplets were identified among 43,352 adults, each containing one patient receiving isoflurane, desflurane, and sevoflurane, respectively. The adjusted geometric mean (95% CI) hospital length-of-stay for the isoflurane cases was 2.85 days (2.78–2.93); this was longer than that observed for both desflurane (2.64 [2.57–2.72]; $P < 0.001$) and sevoflurane (2.55 [2.48–2.62]; $P < 0.001$). In the prospective trial (N = 1,584 operations), no difference was found; the adjusted ratio of means (95% CI) of hospital length-of-stay in patients receiving isoflurane *versus* sevoflurane was 0.98 (0.88–1.10), $P = 0.77$, with adjusted geometric means (95% CI) estimated at 4.1 (3.8–4.4) and 4.2 days (3.8–4.5), respectively.
Conclusions: Results of the propensity-matched retrospective analysis suggested that avoiding isoflurane significantly reduced the duration of hospitalization. In contrast, length-of-stay was comparable in our prospective trial. Volatile anesthetic choice should not be based on concerns about the duration of hospitalization. These studies illustrate the importance of following even the best retrospective analysis with a prospective trial.

What We Already Know about This Topic

- Retrospective analyses of large databases are subject to unknown biases
- When possible, results from retrospective studies should be validated with randomized controlled trials

What This Article Tells Us That Is New

- Results from a retrospective review of choice of volatile anesthetic and length-of-stay were not replicated in an alternating intervention trial
- Consider the limitations of retrospective analyses before using the results to change the practice

ANESTHETIC drugs account for 5–13% of hospital drug expenditures, with volatile anesthetics contributing approximately 20%.[1–4] Desflurane generally costs the most and isoflurane the least. It is difficult to estimate individual cost of volatile anesthetics because vaporizers are shared and uptake varies depending on dosage requirements and obesity; usage also varies within patients as a function of the fresh-gas flow rate and duration.[3–5] For an hour, volatile anesthesia costs between $0.20 and $6.45 at a flow rate of 0.5 l/min and between $2.45 and $77.90 at a flow rate of 6 l/min.[6] The cost of sevoflurane with a fresh-gas flow of 2 l/min is approximately 20 times that of isoflurane at a

◇ This article is featured in "This Month in Anesthesiology." Please see this issue of ANESTHESIOLOGY, page 3A.

* Professional Staff, Department of General Anesthesia, † Michael Cudahy Professor and Chair, # Research Assistant, †† Health Economist, ‡‡ Professor and Vice-chair, Department of Outcomes Research, § Senior Biostatistician, Departments of Quantitative Health Sciences and Outcomes Research, ‖ Associate Staff of Biostatistics, Departments of Quantitative Health Sciences and Outcomes Research, ** Head, Research Section, Department of Colorectal Surgery, Cleveland Clinic, Cleveland, Ohio. ‡ Medical Student, Lerner College of Medicine, Cleveland Clinic. Current position: Resident, Department of Emergency Medicine, University of Massachusetts Medical School, Worchester, Massachusetts.

Received from the Departments of General Anesthesia, OUTCOMES RESEARCH, Quantitative Health Sciences, and Colorectal Surgery, Cleveland Clinic, Cleveland, Ohio. Submitted for publication August 10, 2012. Accepted for publication March 12, 2013. Support was provided solely from institutional and/or departmental sources.

Address correspondence to Dr. Sessler: Department of OUTCOMES RESEARCH, The Cleveland Clinic—P77, Cleveland, Ohio 44195. ds@or.org. On the world wide web: www.OR.org. This article may be accessed for personal use at no charge through the Journal Web site, www.anesthesiology.org.

 Anesthesiology 2013; 119:61-70

fresh-gas flow of 0.5 l/min.[7] Reductions in drug-acquisition cost, which do not compromise safety or provoke other costs, represent direct savings and are thus well worth considering.

Some years ago, the Cleveland Clinic eliminated desflurane, a cost-saving measure by Supply Chain Management that was not based on clinical data. Recently, there was an analogous proposal to similarly eliminate sevoflurane in favor of use of isoflurane in adults. The assumption was that drug cost would be reduced without consequent cost or harm to patients. The validity of this assumption is nonobvious: available volatile anesthetics probably are comparably safe, but they are not identical, differing most obviously in blood and tissue solubility.

More potent and soluble anesthetics have a slower onset and they last longer.[1,5] Isoflurane is approximately five times more potent than desflurane, and approximately five times more soluble.[8] The solubility of sevoflurane is only slightly less than that of isoflurane, but sevoflurane is partially metabolized and thus has a shorter duration-of-action than would otherwise be expected.[9] Even a few minutes delayed emergence has financial implications because institutional costs of operating room time are at least $10.00 per minute.[10]

A more serious question is whether more soluble volatile anesthetics delay hospital discharge. Delayed discharge seems somewhat unlikely given that nearly all volatile anesthetics have left the body by the first postoperative morning. Nonetheless, a potential mechanism by which hospitalization could be prolonged by residual anesthetic concentrations of isoflurane may be related to the drug being hyperalgesic at sub-anesthetic concentrations.[11,12] If isoflurane increases postoperative pain, it may augment opioid consumption, in turn provoking nausea and vomiting and aggravating ileus—both of which might contribute to prolonged hospitalization.[11] It is also possible that delayed awakening promotes atelectasis or other subtle respiratory sequelae.

Delayed hospital discharge would have substantial financial implications as the hospital expenses for an additional inpatient day exceeds $1,900.00.§§ If isoflurane use were clearly associated with prolonged hospitalization, total cost to the institution would far exceed any savings from the purchase of the alternative drug, sevoflurane. We thus used two approaches to evaluate the effects of volatile anesthetic choice on duration of hospitalization.

Initially, we queried electronic records and compared nearly homogeneous patients given desflurane, sevoflurane, and isoflurane. We tested the primary hypotheses that the duration of hospitalization is greatest in patients given isoflurane and shortest in those given desflurane, and that postoperative pain scores are greatest in patients given isoflurane and least for those given desflurane.

Even the best retrospective studies are subjected to residual selection bias and confounding, and cannot assign causality. We, therefore, also conducted a prospective alternating intervention trial in which patients were assigned to isoflurane or sevoflurane, based on protocol. Specifically, we tested the primary hypothesis that the duration of hospitalization is longer with isoflurane than sevoflurane. We also tested the secondary hypotheses that the pain scores are greater in patients recovering from isoflurane than sevoflurane anesthesia.

Materials and Methods

Retrospective Study

The Cleveland Clinic Perioperative Health Documentation System is a clinical registry that includes the entire electronic anesthesia record, data from various administrative databases, and portions of the electronic medical record. Perioperative variables were prospectively collected concurrently with patient care from our electronic anesthesia record and other electronic systems. Use of the perioperative registry for this retrospective cohort analysis was approved by the Institutional Review Board, Cleveland Clinic, Cleveland, Ohio.

We included adults who had noncardiac surgery at the Cleveland Clinic Main Campus; had received anesthesia using desflurane, sevoflurane, or isoflurane; and were admitted to the hospital for at least 24 h. The most recent surgical procedure was used for patients in whom multiple surgeries were included in our registry. We excluded patients given more than one of the volatile anesthetic; having emergency surgery; or already hospitalized for a medical condition at the time of surgery.

The primary outcomes for the study were hospital length-of-stay and postoperative pain. Length-of-stay was measured as date and time from start of operation to date and time of hospital discharge. Pain scores were evaluated by floor nurses. Polytomous logistic regression—the analogue of logistic regression when the response variable has three or more levels—was used to estimate bivariate propensity score vectors (*i.e.,* pairs of propensity scores, respectively estimating the probability of receiving desflurane and sevoflurane) based on the available patients' baseline and intraoperative covariables (see the Results section). Volatile anesthetic exposure was defined in terms of minimum alveolar concentration times exposure hours (MAC hours), using 6.6% as the MAC of desflurane, 1.8% as the MAC of sevoflurane, and 1.17% as the MAC of isoflurane.[8]

Risk Stratification Index,[13] American Society of Anesthesiologists (ASA) Physical Status,[14] year of surgery, age, total intraoperative MAC hours, intraoperative propofol, intraoperative morphine, intraoperative hydromorphone, intraoperative fentanyl, and intraoperative midazolam were all coded as continuous variables. Sex, race, ASA class, administration of nitrous oxide, regional anesthesia modality, and attending anesthesiologist and surgeon were coded as categorical variables. Attending anesthesiologists and surgeons for whom there were fewer than 100 cases in our sample after inclusion and exclusion criteria were combined into one category to ensure model stability; affecting 4.2 and 13.3% of cases, respectively.

§§ Annual Hospital Association Annual Survey. Available at: http://www.statehealthfacts.org/profileind.jsp?cat=5&sub=68&rgn=37. Accessed July 31, 2012.

Triplets of propensity-matched patients were obtained by a three-step procedure. First, each patient who received isoflurane was matched to a patient who received desflurane. Then, each patient who received isoflurane was again matched, this time to a patient who received sevoflurane. Finally, a filter was applied to include only those isoflurane patients who were successfully matched to both a desflurane patient and a sevoflurane patient, thus resulting in the matched triplets. The caliper width for the respective propensity scores was 0.05 (on the probability scale). Successfully matched triplets were restricted to those for which all three patients had the same type of surgery. Type of surgery was characterized using the Clinical Classifications Software for procedures, which was obtained from the U.S. Agency for Healthcare Research and Quality.[||||]

Multivariable linear regression was used to model each outcome within the propensity-matched cohort of patients; hospital length-of-stay was log-transformed during the modeling process. Any of the aforementioned covariables that were univariably significant at the 0.1 significance level, were included in these multivariable models. In addition, we adjusted for the number of postoperative pain measurements recorded for each patient in the analysis of numerical rating scale pain scores. Pairwise comparisons among the three anesthetic groups for the two outcomes were made using Wald tests for model contrasts. These tests were implemented under a Bonferroni-adjusted significance criterion of 0.0083 in order to maintain an overall significance level for the primary outcome at 0.05.

Because our retrospective study evaluated patients undergoing surgery in any of our operating rooms and our prospective study evaluated patients undergoing surgery in select operating rooms, we conducted a sensitivity analysis in an attempt to evaluate length-of-stay and verbal rating scale pain score for more comparable populations. For this analysis, we used only the matched triplets for which the type of surgery was within the top 10 for the prospective trial. The same methods were used for this analysis as in the primary analysis for the retrospective study.

Prospective Trial

Our prospective trial was designed to address the suggestion from Supply Chain Management that isoflurane be substituted for sevoflurane as a cost-saving measure. It was thus primarily a quality-cost-improvement initiative. The process was approved by the Cleveland Clinic Institutional Review Board; because isoflurane and sevoflurane are both routinely used at the Clinic and are thought to be similar, individual consent was waived. Desflurane was no longer used at the Clinic and was thus not evaluated.

Our study was restricted to a physically isolated suite of 10 operating rooms that are normally staffed by a small group of anesthesiologists. The rooms are primarily used for colorectal surgery and to a lesser extent, thyroid and parathyroid surgery; most cases are therefore substantial and require postoperative hospitalization. The test periods had a duration of 2 weeks, and alternated between all operating rooms in the test suite using either sevoflurane or isoflurane during a given period. The first period used sevoflurane, the second used isoflurane, and so on, for a total of 12 cycles (24 weeks).

During the study, only the designated vaporizers were mounted on the anesthesia machines. Anesthesiologists were encouraged to use the designated anesthetic, but vaporizers for the alternate anesthetic remained available if deemed necessary for a particular patient. There were no other restrictions on anesthetic management, and practitioners were free to use regional anesthesia, intravenous anesthetics, and nitrous oxide per their preference.

Demographic and morphometric characteristics were recorded, including age, sex, race, and body mass index. On the basis of procedure for the specific operation of interest, types of surgery were characterized from the International Classification of Diseases, Ninth Revision codes using Agency for Healthcare Research and Quality Clinical Classifications Software.[|||] Procedures having fewer than 10 occurrences were rolled up into an "other" category.

From our Perioperative Health Documentation System, we extracted the attending anesthesiologist who ended the case, date of surgery, duration of surgery, time-weighted average end-tidal volatile anesthetic concentration, use of nitrous oxide, total intraoperative opioid use, and time-weighted Bispectral Index as available. We also calculated each patient's time-weighted average pain score over the first 72 h after surgery as recorded by nurses at approximately 4-h intervals. Postoperative opioid use was not consistently available and therefore not reported. Length of hospital stay was calculated as the difference between the date and time of start of the operation and date and time of hospital discharge.

In our main analyses, we adjusted for potential confounding using inverse propensity score weighting. A logistic regression propensity score model predicting exposure group (sevoflurane = 1 *vs.* isoflurane = 0) from all of the potentially confounding variables (including demographics, baseline diagnoses, procedure, date of surgery, length of surgery, administration of nitrous oxide, anesthesia provider, and operating room number) was constructed (all variables forced into the model). In the analyses below for association between volatile anesthetic and response variable, each observation was inversely weighted by the estimated propensity score unless noted otherwise. Specifically, sevoflurane patients were weighted by 1/(propensity score) and isoflurane patients by 1/(1 – propensity score).[15]

We assessed the effect of sevoflurane *versus* isoflurane on the primary outcome of log-transformed hospital length-of-stay using a generalized estimating equation model in which we adjusted for the within-patient correlation across multiple operations on different visits (exchangeable correlation

[||||] HCUP CCS. Healthcare Cost and Utilization Project (HCUP), March 2012. Agency for Healthcare Research and Quality, Rockville, MD. Available at: www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp. Accessed July 31, 2012.

structure) and inversely weighted each observation by its estimated propensity score. Log-transformed length-of-stay was fairly normally distributed within each group (Shapiro–Wilks statistics of 0.95–0.96), thus justifying our use of ratio of geometric means estimated as the difference in log-transformed length-of-stay. Because coefficient of variation for length of-stay was similar (112% for sevoflurane and 115% for isoflurane), we are effectively reporting the ratio of means.[16] Results are therefore reported as the estimated ratio of mean length-of-stay and 95% CI.

We conducted four distinct sensitivity analyses to assess the robustness of our primary analysis method. First, we used a multivariable generalized estimating equation model with identity link to adjust for confounding instead of inverse propensity score weighting when assessing the relative effects of sevoflurane *versus* isoflurane on length-of-stay; all available potentially confounding variables (see the Results section) were included in the model (a backward stepwise model using conservative P value criteria of 0.80 to stay gave the same results). In a second analysis, we adjusted for the propensity score as a covariate instead of using inverse weighting. In a third sensitivity analysis, we considered only the first visit for a patient during the study period. And in a fourth sensitivity analysis, we used Cox proportional hazards regression to assess the effect of sevoflurane *versus* isoflurane on days to discharge alive, censoring the N = 12 patients who died before hospital discharge at the maximum length-of-stay of any patient (*i.e.,* 120 days) and adjusting for confounding using inverse propensity score analysis as described earlier.

Sample-size Considerations. In our retrospective study, we had observed a covariable-adjusted estimated geometric mean (CI) hospital length-of-stay of 2.60 (2.53–2.67) for sevoflurane and 2.83 (2.79–2.94) for isoflurane (mean ± SD of 0.95 ± 0.80 and 1.05 ± 0.80, respectively, on the natural log scale), corresponding to a ratio of geometric means (95% CI) of 0.91 (0.87–0.94). We designed the prospective study to have 90% power at the 0.05 significance level to detect a reduction of 10% in median hospital length-of-stay; that is a ratio of geometric means of 0.90 or stronger, between sevoflurane and isoflurane. We assumed a coefficient of variation of 0.8 on the log scale for both groups, as observed in our retrospective analysis, thus requiring a sample size of 1,876 total operations (938 per group). To assure adequate sample size and account for potentially higher coefficient of variation, we accrued patients for 24 weeks, during which we expected to enroll several thousand cases.

Results

Retrospective Study

We obtained data from our Perioperative Health Documentation System for 43,352 ASA class I–IV inpatients who had surgery between January 2005 and March 2009. Of these, 32,342 patients were given a single volatile anesthetic and were thus eligible for propensity matching. A total of 2,898 triplets were matched on isoflurane, desflurane, or sevoflurane (N = 8,694 patients). Of these, at least one postoperative pain measurement was available for 6,747 (77%).

Table 1 shows the group characteristics after propensity score matching; most of the characteristics considered were no longer significantly different among the three groups, with the exceptions of year of surgery, intraoperative MAC hours, intraoperative nitrous oxide, intraoperative propofol, intraoperative fentanyl, and total intraoperative opioids. The remaining differences were not clinically important. In our primary analyses comparing length-of-stay and pain between the three matched groups, we adjusted for year of surgery, Risk Stratification Index, ASA Physical Status, sex, intraoperative nitrous oxide, intraoperative propofol, intraoperative fentanyl, total intraoperative opioids, and intraoperative MAC hours.

The adjusted geometric mean (95% CI) hospital length-of-stay for the isoflurane cases was 2.85 days (2.78–2.93); this was longer than that observed for both desflurane (2.64 [2.57–2.72]; ratio of means [95% CI] of 1.08 [1.04–1.12]; $P < 0.001$) and sevoflurane (2.55 [2.48–2.62]; ratio of means 1.12 [1.08–1.16]; $P < 0.0001$; table 2 and fig. 1). Sevoflurane cases were associated with a slightly shorter duration of hospitalization than desflurane cases (ratio of means of 0.96 [0.93–1.00]; $P = 0.009$). We did not find an interaction between age and anesthetic choice ($P = 0.49$).

For the secondary outcome of patient-mean 72-h verbal response scale pain scores, adjusted means for the isoflurane, desflurane, and sevoflurane groups were 3.71 (3.61–3.81), 3.86 (3.76–3.96), and 3.71 (3.61–3.82), respectively. Means for both isoflurane and sevoflurane were lower (*i.e.,* better) than for desflurane ($P = 0.005$ and $P = 0.008$, respectively; table 2). No interaction with age was found ($P = 0.81$).

Sensitivity Analysis. When analysis was restricted to the top 10 procedures used in the prospective trial (mainly colorectal and gastrointestinal surgery: 85.4% of patients in the prospective trial and 33.1% of patients in the retrospective study), ratios of mean length-of-stay among the three anesthetics were very similar to the primary complete data analysis (not reported). Actual length-of-stay was increased approximately 0.5 days over our entire retrospective sample, approaching the length-of-stay observed in the prospective trial.

Prospective Trial

During the study period, there were 2,322 operations that met basic criteria for the study (surgery in the designated operating rooms during the period July 18, 2011 to December 31, 2011). After including only the first qualifying procedure for each hospital visit, 1,584 operations on 1,501 patients remained, and were used for this study; 1,424 patients had one operation, 71 patients had two operations, and 6 had three operations. Fifty-five percent of the procedures (N = 870) used only sevoflurane and 45% (N = 714) used isoflurane. The discrepancy is largely due to the last 2-week period (isoflurane) being at the end of December during which fewer patients presented for surgery than during other study periods.

Table 1. Retrospective Study: Characteristics after 1:1:1 Propensity Score Matching

Factor	Level	Desflurane (N = 2,898)	Isoflurane (N = 2,898)	Sevoflurane (N = 2,898)	*P* Value
Year of surgery		2,006.7 ± 0.8	2,006.7 ± 1.0	2,006.8 ± 1.1	<0.001
Risk Stratification Index		−0.6 [−1.1, −0.2]	−0.6 [−1.2, −0.2]	−0.6 [−1.1, −0.2]	0.90
ASA Physical Status	I	4	4	4	0.89
	II	47	47	45	
	III	45	45	46	
	IV	4	4	4	
Age, yr		56 ± 16	56 ± 16	56 ± 16	0.24
Male sex		44	46	44	0.16
Race	African American	12	10	11	0.22
	Caucasian	85	86	85	
	Other	3	4	4	
Intraoperative nitrous oxide		46	51	44	<0.001
Intraoperative propofol, mg		160 [120, 200]	160 [120, 200]	160 [120, 200]	0.08
Intraoperative fentanyl, μg		200 [150, 250]	200 [150, 250]	200 [150, 250]	0.03
Intraoperative midazolam, mg		1.4 ± 0.9	1.3 ± 0.9	1.4 ± 0.9	0.45
Intraoperative morphine, mg		0.0 [0.0, 0.0]	0.0 [0.0, 0.0]	0.0 [0.0, 0.0]	0.23
Intraoperative hydromorphone, mg		0.0 [0.0, 0.5]	0.0 [0.0, 0.6]	0.0 [0.0, 0.6]	0.20
Total intraoperative opioids (mg IV morphine)		28 [20, 39]	25 [19, 35]	27 [20, 38]	0.008
Total intraoperative MAC hours		1.9 [1.2, 2.9]	1.8 [1.2, 2.8]	2.0 [1.3, 3.1]	<0.001
Regional anesthesia	None	92	92	92	0.89
	Epidural	4	3	4	
	Femoral	1	1	1	
	Interscalene	1	1	1	
	Supraclavicular	1	1	0	
	Other	1	1	1	
Surgeon*	A	2	2	2	0.43
	B	2	2	2	
	C	3	3	3	
	D	3	3	4	
	E	2	3	2	
	F	2	2	2	
	G	2	3	3	
	H	2	2	2	
	I	2	2	2	
	J	2	2	3	
	Other	77	76	75	
Anesthesiologist*	A	3	3	3	0.69
	B	3	3	3	
	C	3	4	3	
	D	2	3	3	
	E	3	3	4	
	F	3	3	3	
	G	3	3	3	
	H	4	3	4	
	I	3	3	3	
	J	3	4	4	
	Other	70	69	69	

Statistics given are means ± SDs (symmetric continuous measures–*P* value from overall one-way ANOVA), median [1st quartile, 3rd quartile] (asymmetric continuous measures–Kruskal–Wallis one-way ANOVA by ranks), or N (%) (factors–Pearson chi-square test). Variables significant at the 0.1 level were adjusted for within the primary multivariable regression models.

* Top 10 levels shown here. In our analysis, each surgeon or anesthesiologist was modeled as a separate factor level (those with fewer than 100 operations represented in the registry after inclusion or exclusion criteria were combined).

ASA = American Society of Anesthesiologists; MAC = minimum alveolar concentration.

Table 2. Retrospective Study: Ratios of Means for Adjusted Length-of-stay and Differences in Means for 72-hour Postoperative Pain Scores

Comparison	Ratio (95% CI) of Geometric Mean Length-of-stay	*P* Value	Difference (95% CI) in Mean VRS Pain	*P* Value
Isoflurane *vs.* sevoflurane	1.12 (1.08–1.16)	<0.0001*	–0.01 (–0.15 to 0.14)	0.73
Isoflurane *vs.* desflurane	1.08 (1.04–1.12)	<0.0001*	–0.15 (–0.30 to –0.01)	0.005*
Sevoflurane *vs.* desflurane	0.96 (0.93–1.00)	0.009*	–0.15 (–0.29 to –0.00)	0.008*

VRS pain score is pain on a 11-point verbal response scale (0–10). CIs adjusted using the Bonferroni correction. *P* values for significant independent associations after the Bonferroni correction are asterisked. Median length-of-stay assuming log-normal distribution; equivalent to geometric mean length-of-stay.

VRS = verbal rating scale.

Table 3 compares the sevoflurane and isoflurane cases on all variables adjusted for in the propensity score model, including demographic and morphometric characteristics, ASA Physical Status score, past medical history variables, length of surgery, operating room, procedure, and anesthesiologist. Due to the alternating intervention study design, balance between the groups on potentially confounding variables was very good; only ASA status, use of nitrous oxide, procedure, anesthesiologist, and weight loss showed imbalance (absolute standardized differences >0.10). Our logistic regression propensity score model included 110 parameters representing 35 variables, with a resulting a c-statistic of 0.70. The last column of table 3 shows that, after inverse weighting by the propensity score, all but one of the standardized differences was less than 0.10. The only exception was an anesthetic provider with standardized difference of 0.13, which is still excellent. We thus had excellent baseline balance when comparing sevoflurane and isoflurane on pain and duration of hospitalization.

Length-of-stay. Unadjusted median [quartiles] hospital length-of-stay of isoflurane *versus* sevoflurane were 4.9 [2.0–8.4] and 4.4 days [1.5–7.8], respectively (univariable $P = 0.24$, adjusting for within-patient correlation). In our primary analysis generalized estimating equation model using inverse propensity score weighting to adjust for confounding, the estimated ratio of means (95% CI) of hospital length-of-stay in patients receiving isoflurane *versus* sevoflurane was 0.98 (0.88–1.10); $P = 0.77$; table 4; figure 1.

The effect of isoflurane *versus* sevoflurane on hospital length-of-stay did not depend on age ($P = 0.71$), sex ($P = 0.56$), ASA status ($P = 0.13$), or surgical service (colorectal, general, or other; $P = 0.60$), assessed through tests of interaction in a generalized estimating equation model weighted by propensity score.

Our four sensitivity analyses for the primary hypothesis all reached the same conclusion of no difference. When only the first visit per patient was included (N = 1,424), the estimated ratio of means (95% CI) was 0.98 (0.89–1.09), $P = 0.76$, very similar to the primary results mentioned earlier. When the treatment effect was assessed using a multivariable generalized estimating equation model to adjust for confounding variable (instead of inverse weighting by the propensity score), the estimated ratio of means (95% CI) was 1.0 (0.94–1.05), $P = 0.86$. Adjusting for the propensity score as a covariate instead of using inverse weighting yielded an estimated ratio of means (95% CI) of 1.0 (0.90–1.10), $P = 0.94$. A Cox proportional hazards regression model on days to discharge alive using inverse propensity score weighting resulted in a hazards ratio (95% CI) of isoflurane *versus* sevoflurane of 1.0 (0.91–1.11), $P = 0.95$.

Pain. Intraoperative use of opioid in morphine equivalents did not differ among the groups, with confounder-adjusted mean (SEM) of 32 mg (1.3) for isoflurane and 33 mg (1.6) for sevoflurane, for a mean difference of –0.69 (–4.7 to 3.3), $P = 0.74$. Postoperative pain measurements were available for N of 682 isoflurane operations and N of 850 sevoflurane operations. Mean of time-weighted average pain scores through the first 72 h did not differ between the groups, with estimated isoflurane–sevoflurane difference (95% CI) of –0.02 (–0.23 to 0.19), $P = 0.87$, using inverse propensity score weighting in a generalized estimating equation model (table 5).

Discussion

Our retrospective analysis used propensity matching and exact matching on the type of surgery. It thus included sophisticated protections against bias and confounding. And the results were encouraging: the duration of hospitalization with sevoflurane or desflurane was significantly less than that of isoflurane and duration of hospitalization with sevoflurane was less than that of desflurane. Furthermore, the reduction in the length-of-stay associated with avoiding isoflurane was clinically important, at approximately 0.2 days. Given the cost of a hospital day, this analysis suggested that switching to isoflurane would substantially augment hospital cost even though the alternative volatile anesthetics are more expensive.

Even the best retrospective analysis can suffer from omitted variable bias (to say nothing of a host of other types of bias and confounding) and cannot evaluate causality. It is thus conventional to consider most retrospective analyses to be exploratory, generating results that are best confirmed in prospective trials. The results of our prospective trial highlight the importance of following retrospective analyses with prospective trials: in our 12-fold alternative intervention trial involving nearly 1,600 operations in more than 1,500 patients, the mean durations of hospitalization with isoflurane and sevoflurane anesthesia

Table 3. Prospective Trial: Baseline and Operative Potential Confounders

Factor*	Sevoflurane (N = 870)	Isoflurane (N = 701)	*P* Value	Raw STD	Adjusted STD
BMI, kg/m^2	27.6 ± 7.0	27.5 ± 6.7	0.58	0.03	0.00
Patient age, yr	53.6 ± 16.6	53.4 ± 16.5	0.83	0.01	0.01
Duration of surgery, h	2.5 [1.7–3.7]	2.5 [1.6–3.9]	0.52	–0.03	0.03
White, n (%)	763 (88)	632 (89)	0.62	–0.03	0.01
Male, n (%)	368 (42)	306 (43)	0.82	–0.01	0.02
ASA Physical Status			0.039	0.15	0.00
I	17 (2)	13 (2)			
II	424 (49)	297 (42)			
III	388 (45)	366 (51)			
IV	41 (5)	38 (5)			
Operating room number			0.95	0.09	0.05
40	105 (12)	76 (11)			
41	92 (11)	86 (12)			
42	66 (8)	63 (9)			
43	38 (4)	38 (5)			
45	107 (12)	84 (12)			
46	106 (12)	86 (12)			
47	115 (13)	89 (12)			
48	116 (13)	92 (13)			
49	18 (2)	15 (2)			
50	107 (12)	85 (12)			
Nitrous oxide, n (%) yes	248 (29)	269 (38)	<0.001	–0.20	0.02
Emergency, n (%)	19 (2)	14 (2)	0.76	0.02	0.03
Past medical history, n (%)					
Congestive heart failure	24 (3)	22 (3)	0.70	–0.02	–0.01
Valvular disease	11 (1)	10 (1)	0.81	–0.01	–0.03
Pulmonary circulation disease	8 (1)	4 (1)	0.41	0.04	0.00
Peripheral vascular disease	19 (2)	11 (2)	0.35	0.05	0.01
HTN	263 (30)	203 (28)	0.43	0.04	0.00
Chronic pulmonary disease	74 (9)	55 (8)	0.56	0.03	0.01
Diabetes w/o chronic comps	79 (9)	67 (9)	0.84	–0.01	0.00
Hypothyroidism	74 (9)	74 (10)	0.21	–0.06	0.00
Renal failure	32 (4)	23 (3)	0.62	0.03	0.01
Liver disease	17 (2)	17 (2)	0.56	–0.03	0.03
Metastatic cancer	69 (8)	77 (11)	0.051	–0.10	0.00
Solid tumor w/out metastasis	132 (15)	111 (16)	0.84	–0.01	0.02
Coagulopathy	29 (3)	32 (4)	0.24	–0.06	0.00
Obesity	76 (9)	61 (9)	0.89	0.01	–0.02
Weight loss	92 (11)	101 (14)	0.031	–0.11	0.02
Fluid, electrolyte disorders	160 (18)	152 (21)	0.15	–0.07	0.00
Deficiency anemias	67 (8)	42 (6)	0.15	0.07	–0.02
Alcohol abuse	0 (0)	7 (1)	0.004	N/A	N/A
Drug abuse	7 (1)	4 (1)	0.76	0.03	0.00
Psychoses	15 (2)	15 (2)	0.58	–0.03	0.00
Depression	83 (10)	65 (9)	0.77	0.02	–0.01

(*continued*)

Table 3. (*Continued*)

Factor*	Sevoflurane (N = 870)	Isoflurane (N = 701)	*P* Value	Raw STD	Adjusted STD
Procedure†			0.47	0.23	0.09
Other lower GI therapeutic procedure	245 (28)	190 (27)			
Colorectal resection	202 (23)	146 (20)			
Thyroidectomy; partial or complete	84 (10)	69 (10)			
Other therapeutic endocrine procs	65 (7)	48 (7)			
Other OR GI therapeutic procs	52 (6)	56 (8)			
Small bowel resection	42 (5)	32 (4)			
Other hernia repair	27 (3)	22 (3)			
Other OR therapeutic procs; female organs	23 (3)	19 (3)			
Excision; lysis peritoneal adhesions	17 (2)	23 (3)			
Other small abdominal	19 (2)	10 (1)			
Anesthesia provider‡			<0.001	0.55	0.13
A	89 (10)	72 (10)			
B	58 (7)	46 (6)			
C	46 (5)	36 (5)			
D	39 (4)	23 (3)			
E	23 (3)	30 (4)			
F	35 (4)	16 (2)			
G	23 (3)	26 (4)			
H	18 (2)	29 (4)			
I	26 (3)	16 (2)			
J	13 (1)	25 (4)			
Surgeon			0.43	0.29	0.08
A	60 (7)	53 (7)			
B	64 (7)	46 (6)			
C	62 (7)	49 (7)			
D	64 (7)	45 (6)			
E	70 (8)	38 (5)			
F	41 (5)	34 (5)			
G	41 (5)	28 (4)			
H	32 (4)	33 (5)			
I	48 (6)	18 (3)			
J	32 (4)	28 (4)			

* All factors were included in propensity score model. † Top 10 frequency procedures (*P* value and SD statistics based on all categories). ‡ Top 10 frequency providers (*P* value and SD statistics based on all providers).

Adjusted STD = standardized difference after weighting by propensity score; ASA = American Society of Anesthesiologists; BMI = body mass index; GI = gastrointestinal; HTN = hypertension; N/A = standardized difference not appropriate due to zero in 2×2 table; OR = operating room; Procs = procedures; Raw STD = unadjusted standardized difference; STD = standardized difference (difference in means or proportions divided by pooled SD).

were virtually identical. In distinct contrast to our observational analysis, there was thus no statistically significant or clinically important difference in length-of-stay as a function of volatile anesthetic choice in our prospective trial.

From a clinical perspective, the retrospective analysis alone may have prompted clinicians to avoid isoflurane. In contrast, results from our prospective trial indicate that hospital length-of-stay is not different between isoflurane and sevoflurane and should not be a basis for choosing one anesthetic over the other. Our observational study was based on approximately 2,900 matched triplets selected from more than 32,000 patients, whereas our prospective study involved fewer than 1,600 patients. Nonetheless, the prospective results are more likely to be accurate because excellent balance was achieved on potentially confounding variables and because including large numbers of patients in

Table 4. Prospective Trial: Comparing Sevoflurane and Isoflurane on Hospital Length-of-stay

Analysis	Sevoflurane (N = 870)	Isoflurane (N = 701)	Ratio* of Geometric Means (95% CI)	*P* Value
Univariable, median [q1, q3]	4.4 [1.5, 7.8]	4.9 [2.0, 8.4]	1.06 (0.97–1.16)	0.24
Propensity score-adjusted† geometric mean (CI)	4.2 (3.8–4.5)	4.1 (3.8–4.4)	0.98 (0.88–1.10)†	0.77

* Ratio of geometric means: isoflurane/sevoflurane (assuming log-normal distribution) and adjusting for within-patient correlation across visits using generalized estimating equation model assuming exchangeable correlation structure. † Adjusted for confounding using inverse propensity score weighting.

observational analyses does not much reduce error induced by bias or confounding.[17,18]

We recognize that duration of hospitalization and direct drug cost are hardly the only considerations when selecting a volatile anesthetic. Other important factors that were not evaluated in our studies include time to extubation, duration of recovery, and pulmonary irritation. Furthermore, although it requires considerable skill to awaken patients promptly after prolonged isoflurane anesthesia, sevoflurane is more "forgiving" and can thus be valuable when the duration of surgery is unpredictable and in teaching institutions. Sevoflurane is also far easier to use for inhalational inductions.

Because our prospective (and most reliable) results do not suggest that hospitalization is prolonged by the use of isoflurane, we did not pursue a formal cost analysis. Assuming no adverse effects of isoflurane on unmeasured outcomes such as extubation and recovery times, drug acquisition becomes the only cost factor. The 2010 average wholesale price of sevoflure is $222.00 per 250 ml (Baxter, Deerfield, IL), whereas the same volume of isoflurane is only $30.00 (Rx Elite, Eagle, ID).[6] The difficulty is that actual use per patient-hour of anesthesia is difficult to determine because it depends on solubility, metabolism (in the case of sevoflurane), duration of the case, and most importantly on fresh-gas flow. That said, a *very* rough estimate might be that drug-acquisition cost for typical cases conducted with sevoflurane exceeds that of isoflurane by perhaps $20.00. This is tiny compared with other surgical costs and roughly equal to a couple of minutes of operating room time.[19] A potential mechanism for prolonged hospitalization with isoflurane is hyperalgesia promoted by low residual concentrations of volatile anesthetics[11,12]—which would be greater with the more soluble isoflurane. However, pain scores in the observational analysis did not support this theory, being significantly greater only in the patients given desflurane. Sevoflurane was thus associated with reduced hospital length-of-stay but not reduced pain—a disparity suggesting that increased pain was not the mechanism explaining prolonged hospitalization in the observational study. Furthermore, the reduction in pain scores with desflurane was small and probably not of a clinically important magnitude. In the prospective trial, we observed no difference whatsoever between sevoflurane and isoflurane on either mean pain score or 72-hr opioid consumption.

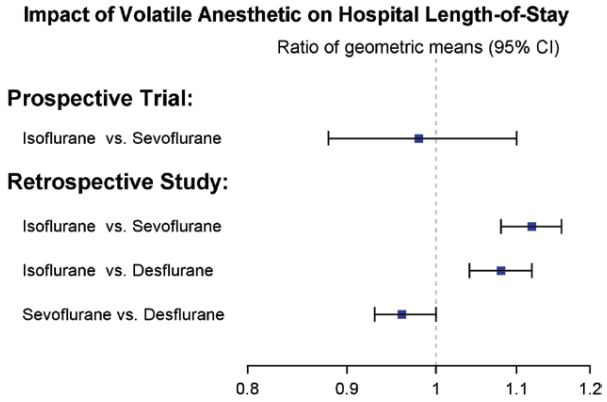

Fig. 1. Forest plot of isoflurane to sevoflurane ratio of geometric means and 95% CI hospital length-of-stay for prospective and retrospective studies. Prospective study results are confounder-adjusted using inverse propensity score weighting, whereas the retrospective study used propensity score matching. CIs for retrospective study are Bonferroni-adjusted for making three comparisons and for having two primary outcomes (hospital length-of-stay and pain score), maintaining the overall significance level at 0.05. CIs that do not contain the null hypothesis of 1.0 indicate statistically significant findings.

Analysis of nonrandomized single interventions ("before-and-after" studies) are common in quality improvement and health services research. Although sometimes unavoidable, this approach is inherently weak because it provides little protection against practice changes over time and the Hawthorne effect. We avoided both problems by alternating between our two anesthetic types 12 times during our 24-week protocol. Compliance with the assigned volatile anesthetic was excellent with only 13 procedures in the analysis dataset being conducted with the alternative volatile anesthetic (<1%).

A limitation of our study is that anesthetic type was not randomly assigned to each 2-week administration period, instead being alternated throughout the study. It seems unlikely, though, that patient or provider characteristics would change as a function of anesthetic selection. Furthermore, analysis of baseline and basic intraoperative characteristics and patient management details demonstrated that excellent balance on baseline characteristics was achieved by the design. An additional limitation is that although patients were not generally told which anesthetic they would receive, anesthesia providers were not blinded. And although surgeons were not specifically

Table 5. Prospective Trial: Intraoperative and Postoperative Outcomes (Univariable)

Factor	Sevoflurane (N = 870)	Isoflurane (N = 701)	*P* Value‡	STD
MAC hours	2.3 [1.4–3.5]	2.0 [1.1–3.1]	<0.001	0.20
Intraoperative morphine equivalents, mg	25 [20–38]	25 [18–36]	0.32	0.05
Pain measurement interval, h*	67 [26–70]	68 [30–70]	0.40	–0.04
VRS pain–patient mean*	3.8 ± 1.8	3.7 ± 1.9	0.42	0.04
VRS pain–patient TWA†	3.6 ± 1.8	3.6 ± 1.9	0.87	0.02

*, † 9 and 10 missing values. ‡ *P* values from Wilcoxon rank sum test if median [quartiles] reported and generalized estimating equation model weighted by inverse propensity score if mean ± SD reported.

MAC = minimum alveolar concentrations; STD = standardized difference (difference in means or proportions divided by pooled SD); TWA = time-weighted average; VRS = verbal rating scale.

told about selection of anesthetic, they were not blinded either as the anesthetic type was recorded in our electronic records. It nonetheless seems unlikely that anesthetic type influenced timing of hospital discharge by the surgical team. Finally, although intraoperative opioid usage was measured, postoperative opioid consumption was not available for either study because it is not routinely entered into our research database.

Analytic methods were not the same for our retrospective and prospective studies, as might be expected with distinct designs and statisticians. For example, the retrospective study controlled for available confounding variables using propensity score matching, whereas the prospective study used the inverse of the propensity score as a weight. Both are valid methods for confounding adjustment, and with both methods the resulting balance on baseline characteristics was demonstrated.

In summary, our propensity-matched retrospective analysis suggested that the duration of hospitalization was significantly less with sevoflurane or desflurane as compared with isoflurane. In contrast, length-of-stay with isoflurane and sevoflurane was comparable in our prospective trial. Volatile anesthetic choice should thus not be based on concerns about the duration of hospitalization. These studies demonstrated the potential consequences of relying on even a highly robust retrospective analysis without doing a follow-up prospective trial. Had we relied on the initial retrospective study, an inappropriate decision would have been made.

The authors gratefully acknowledge the generous assistance of Anesthesia Institute technical staff, especially Marc McConnell, B.S., Cer.A.T. (Department Supervisor, Department of General Anesthesiology, Anesthesiology Institute, Cleveland Clinic, Cleveland, Ohio), Stephen Halton, Jr., Cer.A.T. (Certified Anesthesia Technician, Department of General Anesthesiology, Anesthesiology Institute, Cleveland Clinic), and Donna Keaton, Cer.A.T. (Certified Anesthesia Technician, Department of Colorectal/General Surgery, Cleveland Clinic).

References

1. Stachnik J: Inhaled anesthetic agents. Am J Health Syst Pharm 2006; 63:623–34
2. Chernin EL: Pharmacoeconomics of inhaled anesthetic agents: Considerations for the pharmacist. Am J Health Syst Pharm 2004; 61(suppl 4):S18–22
3. Smith I: Cost considerations in the use of anaesthetic drugs. Pharmacoeconomics 2001; 19(5 Pt 1):469–81
4. Odin I, Feiss P: Low flow and economics of inhalational anaesthesia. Best Pract Res Clin Anaesthesiol 2005; 19:399–413
5. Sakai EM, Connolly LA, Klauck JA: Inhalation anesthesiology and volatile liquid anesthetics: Focus on isoflurane, desflurane, and sevoflurane. Pharmacotherapy 2005; 25:1773–88
6. Drug Topics Red Book, Montvale, NJ, Thomson Healthcare, 2010
7. Weinberg L, Story D, Nam J, McNicol L: Pharmacoeconomics of volatile inhalational anaesthetic agents: An 11-year retrospective analysis. Anaesth Intensive Care 2010; 38:849–54
8. Nickalls RW, Mapleson WW: Age-related iso-MAC charts for isoflurane, sevoflurane and desflurane in man. Br J Anaesth 2003; 91:170–4
9. Kharasch ED, Armstrong AS, Gunn K, Artru A, Cox K, Karol MD: Clinical sevoflurane metabolism and disposition. II. The role of cytochrome P450 2E1 in fluoride and hexafluoroisopropanol formation. Anesthesiology 1995; 82:1379–88
10. Dexter F: Impact on operating room efficiency of reducing turnover times and anesthesia-controlled times. Ann Surg 2007; 245:336–7
11. Zhang Y, Eger EI II, Dutton RC, Sonner JM: Inhaled anesthetics have hyperalgesic effects at 0.1 minimum alveolar anesthetic concentration. Anesth Analg 2000; 91:462–6
12. Flood P, Sonner JM, Gong D, Coates KM: Isoflurane hyperalgesia is modulated by nicotinic inhibition. Anesthesiology 2002; 97:192–8
13. Sessler DI, Sigl JC, Manberg PJ, Kelley SD, Schubert A, Chamoun NG: Broadly applicable risk stratification system for predicting duration of hospitalization and mortality. Anesthesiology 2010; 113:1026–37
14. Saklad M: Grading of patients for surgical procedures. Anesthesiology 1941; 2:281–4
15. Curtis LH, Hammill BG, Eisenstein EL, Kramer JM, Anstrom KJ: Using inverse probability-weighted estimators in comparative effectiveness analyses with observational databases. Med Care 2007; 45(10 suppl 2):S103–7
16. Ledolter J, Dexter F, Epstein RH: Analysis of variance of communication latencies in anesthesia: Comparing means of multiple log-normal distributions. Anesth Analg 2011; 113:888–96
17. Ioannidis JP, Haidich AB, Pappa M, Pantazis N, Kokori SI, Tektonidou MG, Contopoulos-Ioannidis DG, Lau J: Comparison of evidence of treatment effects in randomized and nonrandomized studies. JAMA 2001; 286:821–30
18. Ioannidis JP: Contradicted and initially stronger effects in highly cited clinical research. JAMA 2005; 294:218–28
19. Dexter F, Epstein RH: Typical savings from each minute reduction in tardy first case of the day starts. Anesth Analg 2009; 108:1262–7