

The Construction of Learning Curves for Basic Skills in Anesthetic Procedures: An Application for the Cumulative Sum Method

Getúlio Rodrigues de Oliveira Filho, MD

Department of Anaesthesiology, Hospital Governador Celso Ramos, Florianópolis, Brazil

This study aimed at constructing learning curves for basic procedural skills in anesthesiology using the cumulative sum method. We recorded 1234 peripheral venous cannulations, 895 orotracheal intubations, 688 spinal, and 344 epidurals performed by residents during the first 10 mo of training. Learning curves for each procedure were constructed by using the cusum method. The number of procedures performed until attainment of acceptable failure rates was calculated. All residents mastered peripheral venous cannulation after 79 ± 47 procedures. Four of 7 residents attained acceptable failure rates at orotracheal intubation after 43 ± 33 proce-

dures. Seven of 11 residents attained acceptable failure rates at spinal anesthesia after 36 ± 20 procedures. At epidural anesthesia, 5 of 11 residents attained acceptable failure rates after 21 ± 11 procedures. The cusum method is a useful tool for objectively measuring performance during the learning phase of basic procedures. The wide interindividual variability in the number of procedures required to be performed before attaining acceptable failure rates suggests that performance should be followed on an individual basis.

(Anesth Analg 2002;95:411-6)

Control charts are statistical tools for analyzing data during production or research in industry on which values of the quality characteristics being analyzed are plotted in sequence. These charts consist of a central line and limit lines spaced above and below. The distribution of the plotted values in relation to the control limits provides statistical information on the process under study. Cumulative sum (cusum) charts are one kind of such control charts. On a cusum chart, the cumulative differences of the quality characteristic from a target level are plotted in sequence, leading to tighter control of a given process and allowing detection of deviations from preestablished standards (1). Cusum charts have been used in medicine for detecting trends in temperature (2) and neutrophil count (3). They also have been used for constructing learning curves for colonoscopic examination (4), obstetric epidural anesthesia, spinal anesthesia, and central venous and arterial cannulation (5).

Variables for the construction of a cusum chart are the acceptable (p_0) and the unacceptable (p_1) failure rates and reasonable probabilities of type I and II

errors (α and β). From these, two decision limits (h_1 and h_0) and the variable s are calculated. The chart starts at zero. For each success, the amount s is subtracted from the previous cusum score. For each failure, the amount $1 - s$ is added to the previous cusum score. Thus, a negative trend of the cusum line indicates success, whereas a positive trend indicates failure at the procedure under analysis (1,4,5).

When the line crosses the upper decision limit (h_1) from below, the actual failure rate is significantly greater than the acceptable failure rate, with a probability of type I error equal to α . When the line crosses the lower decision limit (h_0) from above, the true failure rate does not differ significantly from the acceptable failure rate, with a probability of type II error equal to β . When the cusum line is kept between the decision limits, no statistical inference can be made, indicating that more observations are necessary.

Training in anesthetic procedures is made under regressive supervision, that is, the more proficient the resident becomes at a given technique, the less the amount of supervision provided by the instructors. Therefore, the turning point at which supervision is to be decreased or withdrawn has to be assessed objectively by a reliable method of performance measurement. This study aimed at constructing learning

Accepted for publication April 23, 2002.

Address correspondence and reprint requests to Getúlio R. de Oliveira Filho, MD, Rua Luiz Delfino 111/902, 88015-360-Florianópolis, SC, Brazil. Address e-mail to grof@th.com.br.

DOI: 10.1213/01.ANE.0000021360.03491.B7

curves for basic anesthetic procedures using the cumsum method.

Methods

During February (start of the training period) through October 2000 and 2001, data on 668 spinal and 344 epidurals performed by 11 first-year residents were collected. During the same period of 2001, data were collected on 1179 peripheral venous cannulations and 895 orotracheal intubations performed by 7 first-year residents. Data were collected for every procedure performed by each resident during the period of observation. Pediatric, cardiac, and obstetric procedures were not included.

Procedures were performed under an instructor's supervision. Remarks about technique were allowed. At the beginning of the training, residents were instructed about criteria for failure and success at each procedure. The attending resident completed a data collection form immediately after the procedure. At the same time, forms were reviewed and signed by the supervising instructor.

Peripheral venous cannulations were performed with over-the-needle Teflon or Vialon catheters at 14-gauge ($n = 36$), 16-gauge ($n = 128$), 18-gauge ($n = 770$), 20-gauge ($n = 189$), 22-gauge ($n = 53$), or 24-gauge ($n = 3$). Orotracheal intubations were performed with cuffed tracheal tubes of the appropriate size under direct laryngoscopy with MacIntosh blades. Lumbar spinal blocks were performed with 25-gauge ($n = 289$) or 27-gauge ($n = 399$) Quincke needles. Lumbar epidural blocks were performed with 17-gauge Tuohy needles. No attempts were made to select cases according to predicted difficulty criteria.

Success at peripheral venous cannulation was defined as insertion of the catheter after identification of blood in the needle's hub, allowing free flow of the infusion fluid after a single skin puncture. Successful orotracheal intubation required confirmation by chest auscultation and capnometry after a single laryngoscopy. The subarachnoid space was identified by retrieval of cerebrospinal fluid and the epidural space by loss of resistance to saline. Success at spinal and epidural anesthesia was defined as the correct identification of the proper space at the interspace first chosen followed by adequate surgical anesthesia, which was defined as no need for opioid or general anesthetic supplementation during the surgery. Outcomes were measured on a binary variable, 1 representing success and 0 representing failure. Instructors took over procedures after two unsuccessful peripheral venous cannulations, two failed attempts at orotracheal intubation, after failed identification of the subarachnoid or epidural spaces at the interspace first chosen, or at any time if judged appropriate for the patients' comfort or

safety. In these cases, the outcome was rated as a failure.

Acceptable failure rates (p_0) for peripheral venous cannulation and tracheal intubation were arbitrarily set at 20%. Control samples of 459 spinal and 385 epidurals performed by 22 staff anesthesiologists from February through October 2000 were used to establish the acceptable failure rates (p_0) at the interspace first chosen for spinal (15%) and lumbar epidural (20%) blocks. The probability of type I (α) and II (β) errors was set to 0.1. Unacceptable failure rates (p_1) were set at 2 times p_0 .

Average sample sizes for runs having acceptable failure rates (p_0) of 20% and 15% were estimated as 19 and 28 procedures, respectively. For runs having unacceptable failure rates (p_1) of 40% and 30%, the average sample sizes were estimated as 17 and 24 procedures, respectively. Cumsum calculations and sample size estimations were performed according to the formulae provided in Table 1. Pooled success rates before and after crossing the lower decision limits were compared by χ^2 tests with continuity correction. The level of significance was set to 5%.

Because the total number of procedures per resident and the respective number of procedures performed until crossing h_0 followed the normal distribution (Shapiro-Wilk's $W > 0.05$), data are presented as mean \pm SD. Residents are represented by capital letters, which were randomly changed from one procedure to another to avoid identification.

Results

Peripheral Venous Cannulation

One thousand-four patients (85.15%) had 1 skin puncture, 159 (13.48%) had 2, and 16 (1.34%) had 3 to 5 skin punctures. Instructors took over 16 procedures (1.34%). The mean number of procedures per resident was 168.42 ± 67.37 (range, 87 to 269 procedures). All residents crossed the 20% acceptable failure rate line (h_0) after 56.85 ± 43.77 procedures (range, 19 to 146 procedures) (Fig. 1). Pooled success rates before and after reaching h_0 were 75.2% and 87.38% (χ^2 [1 df] = 26.53; $P = 0$).

Orotracheal Intubation

Tracheal intubation was performed after 1 attempt in 756 (84.46%) patients, after 2 attempts in 129 (14.41%) patients, and after 3 attempts in 7 (0.78%) patients. Intubation was impossible under direct laryngoscopy in 3 patients (0.33%). Instructors took over 10 procedures (1.11%). The mean number of procedures per resident was 127 ± 46.29 (range, 50 to 190 procedures). Only 4 residents (57.14%) (C, D, E, and F) crossed the 20% acceptable failure rate line (h_0). They did so after

Table 1. Calculations Involved in the Cusum Method

$$\begin{aligned}a &= \ln [(1 - \beta)/\alpha] \\b &= \ln [(1 - \alpha)/\beta] \\P &= \ln (p_1/p_0) \\Q &= \ln [(1 - p_0)/(1 - p_1)] \\s &= Q/(P + Q) \\h_0 &= -b/(P + Q) \\h_1 &= a/(P + Q)\end{aligned}$$

Average size of samples with failure rate = $p_0 = [(h_0(1 - \alpha) - \alpha h_1)/(s - p_0)]$. Average size of samples with failure rate = $p_1 = [(h_1(1 - \beta) - \beta h_0)/(p_1 - s)]$. p_0 = acceptable failure rate; p_1 = unacceptable failure rate; α = probability of Type I error; β = the probability of Type II error.

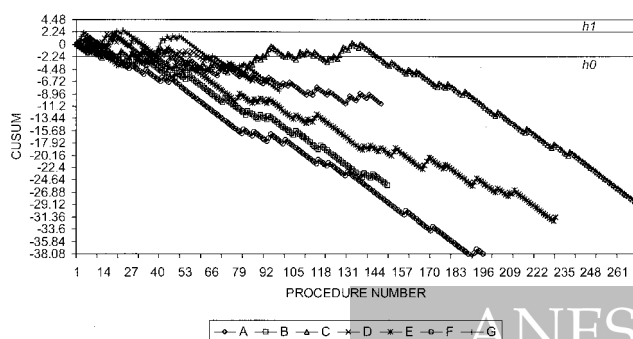


Figure 1. Cumulative sum (cusum) chart for peripheral venous cannulation. Lines A-G represent the learning curves of individual residents. Lines h_1 and h_0 represent upper and lower control limits. Y-axis marks represent multiples of h_1 (2.24) and h_0 (-2.24). All lines crossed h_0 from above after a variable number of procedures, indicating failure rates not significantly different from the acceptable failure rate (p_0).

43 \pm 33.49 procedures (range, 9 to 88 procedures). Curves relative to residents A, B, and E remained between the decision limit lines even after 104, 144, and 104 procedures, respectively (Fig. 2). Pooled success rates before and after reaching h_0 were 72.41% and 84.72% (χ^2 [1 df] = 18.21; $P = 0$).

Spinal Anesthesia

Successful identification of the subarachnoid space at the interspace first chosen occurred in 569 procedures (82.70%). A second interspace was used in the remaining 119 procedures (17.29%). This figure represents cases taken over by instructors. The median (25th and 75th percentiles) number of attempts at the interspace first chosen was 1 (1,2) attempts, with a maximum of 7 skin punctures. The overall median number of attempts was 1 (1,2) attempts, with a maximum of 10 skin punctures. Complete surgical anesthesia was obtained in 631 (91.71%) blocks, 12 (1.74%) required IV opioid supplementation, 21 (3.05%) required general anesthesia supplementation, and 24 (3.48%) were rated as impossible. The mean number of procedures per resident was 62.45 \pm 21.38 (range, 31 to 114 procedures).

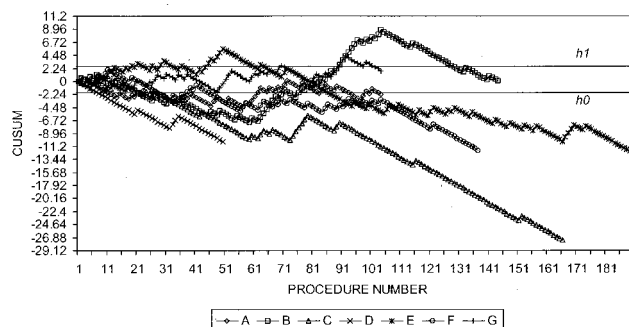


Figure 2. Cumulative sum (cusum) chart for orotracheal intubation. Lines A-G represent learning curves of individual residents. Upper and lower control limits are represented by lines h_1 and h_0 . Y-axis marks represent multiples of h_1 (2.24) and h_0 (-2.24). Lines crossing h_0 from above (C, D, E, and F) indicate failure rates not significantly different from the acceptable failure rate (p_0). No statistical inference can be made about lines lying between h_1 and h_0 (A, B, and E).

Seven residents (63.63%) (A, B, C, E, G, H, and K) crossed the 15% acceptable failure rate line (h_0). They did so after 36 \pm 20.16 procedures (range, 13 to 68 procedures). Cusum lines relative to residents F and I remained greater than the 30% unacceptable failure rate limit (h_1) after 48 and 66 procedures, respectively, whereas those referring to residents D and J remained between h_0 and h_1 after 31 and 76 procedures (Fig. 3), respectively. Pooled success rates before and after reaching h_0 were 79.82% and 87.57% (χ^2 [1 df] = 6.59; $P = 0.01$).

Epidural Anesthesia

Successful identification of the epidural space at the interspace first chosen occurred in 275 patients (79.94%). A second interspace was used in the remaining 69 cases (20.05%). This figure represents cases taken over by instructors. The median (25th and 75th percentiles) number of attempts at the interspace first chosen was 1 (1,2) attempts, with a maximum of 4 skin punctures. The overall median number of attempts was 1 (1,2) attempts, with a maximum of 7 skin punctures. Complete surgical anesthesia was obtained in 293 (85.17%) blocks, 10 (2.9%) required IV opioid supplementation, 18 (5.23%) required general anesthesia supplementation, and 23 (6.68%) were rated as impossible and converted to another anesthetic technique. Accidental vascular punctures occurred in 4 patients (1.16%), and there was 1 case of intravascular injection of local anesthetic (0.29%), resulting in convulsions. The mean number of procedures per resident was 31.27 \pm 16.93 (range, 10 to 65 procedures).

No statistical inference could be made about the performance of residents I and J because the number of procedures did not reach the estimated average

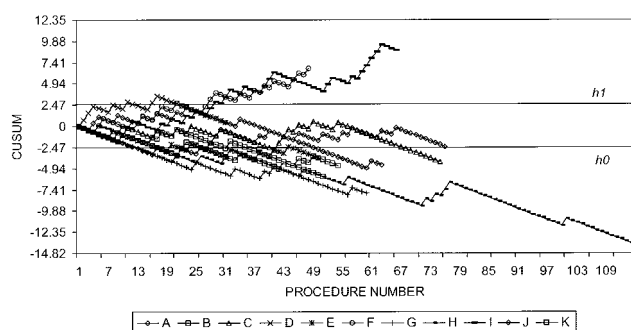


Figure 3. Cumulative sum (cusum) chart for spinal anesthesia. Lines A–K represent learning curves of individual residents. Upper and lower control limits are represented by the lines h_1 and h_0 . Y-axis marks represent multiples of h_1 (2.47) and h_0 (–2.47). Lines crossing h_0 from above (A, B, C, E, G, H, and K) indicate failure rates not significantly different from the acceptable failure rate (p_0). Lines crossing h_1 from below (F and I) indicate failure rates significantly larger than the acceptable failure rate (p_0). No statistical inference can be made about lines lying between h_1 and h_0 (D and J).

sample size, although resident I, having performed 16 procedures, crossed h_0 after 9 procedures. Only residents A, C, E, and H crossed the 15% acceptable failure rate limit (h_0). They did so after 21.4 ± 11.17 procedures (range, 9 to 36 procedures). Cusum line pertaining to resident D remained more than the 30% unacceptable failure rate limit (h_1), whereas those relative to residents F, G, and K remained between h_0 and h_1 after 21, 44, and 19 procedures, respectively (Fig. 4). Pooled success rates before and after reaching h_0 were 75.30% and 91.75% ($\chi^2 [1 \text{ df}] = 10.74$; $P = 0$).

Discussion

Several methods have been developed for measuring competence at important aspects of training, such as cognitive knowledge, judgment, communicative skills, and adaptability, by means of written and oral examinations. However, residents' aptitudes at procedural skills are not routinely quantified. As a consequence, how and when residents achieve their level of proficiency is not exactly known (6). Although instructors easily recognize residents having extreme difficulty or performing outstandingly at procedures, less obvious performances may not be recognized. An easily obtainable quantitative measure of performance would help objective evaluation of resident performance, contributing to a better training.

Lawler et al. (7), using a simple graphical method for detecting the number of successes and failures on a sequential basis, suggested that 20 consecutive successful tracheal intubations might be appropriate for solo anesthesia. However, this approach does not allow statistical inference.

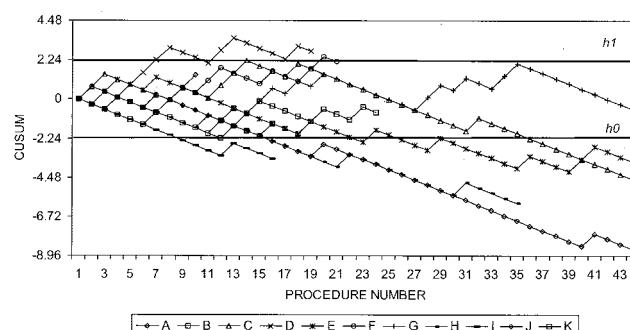


Figure 4. Cumulative sum (cusum) chart for epidural anesthesia. Lines A–K represent learning curves of individual residents. Upper and lower control limits are represented by the lines h_1 and h_0 . Y-axis marks represent multiples of h_1 (2.24) and h_0 (–2.24). Lines crossing h_0 from above (A, C, E, and H) indicate failure rates not significantly different from the acceptable failure rate (p_0). The line crossing h_1 from below (D) indicates a failure rate significantly larger than the acceptable failure rate (p_0). No statistical inference can be made about lines lying between h_1 and h_0 (F, G, and K). Lines I and J represent residents who did not meet the minimum number of blocks.

Four additional studies examined the learning processes of basic practical skills in anesthesiology using statistical approaches. Kopacz et al. (8), using the pooled cumulative success rate at groups of 5 attempts, demonstrated that a 90.8% success rate at tracheal intubation was achieved after 45 attempts, a 75.7% success rate at epidural anesthesia was achieved after 60 attempts, and a 79.3% success rate at spinal anesthesia was achieved after 41 attempts. Konrad et al. (9), using a least square fit model and Monte Carlo procedures, demonstrated that 90% success rates at tracheal intubation and spinal anesthesia were achieved after a mean of 57 and 71 attempts, respectively. An 80% success rate at epidural anesthesia was achieved after a mean of 90 attempts. Schuepfer et al. (10), using the same statistical approach, demonstrated that an 80% success rate at caudal epidural anesthesia was achieved after 32 blocks. Kestin (5), using the cusum method, analyzed individual learning curves for obstetric epidurals and spinal anesthesia. Only 2 of 8 residents (25%) attained the 10% acceptable failure rate (p_0) for spinal anesthesia after 39 to 67 blocks, whereas only 5 of 12 residents (41%) attained the 5% acceptable failure rate for epidural anesthesia after 29 to 185 procedures (5). The author suggests that the acceptable failure rates, taken from consensus among experienced anesthesiologists, might have been too stringent in the context of the study.

The cusum method consists of relatively simple calculations that can be easily performed on an electronic spreadsheet. Statistical inference can be made from the observed successes and failures. The method also provides both numerical and graphical representation of the learning process. There are some critical aspects that must be taken into consideration when applying the method for the construction of learning curves.

Variables of success must be clearly defined and represented by a binary variable. At peripheral venous cannulation, we rated as successes only procedures performed after a single skin puncture. Our criterion of success at tracheal intubation, requiring completion of the procedure after one single laryngoscopy maneuver, was more stringent than those who have used the absence of physical intervention by the supervising instructor as a measure of success, with or without limitation of the number of attempts (8-10). Success at spinal and epidural anesthesia was defined as successful surgical anesthesia after the location of the subarachnoid or epidural space via the interspace first chosen. Other investigators (11) have used this criterion, reporting an 87% success rate.

The planning process also requires the establishment of significant and, at the same time, feasible acceptable failure rates (p_0). These may be taken from control groups, actual or estimated institutional rates, previously published studies, or expert consensus (4,5,9,10). However, its feasibility depends on institutional characteristics such as the teaching method, the instructor-to-resident ratio, the time available for training, and the number of procedures to which residents are exposed.

In this study, acceptable failure rates at spinal and epidural anesthesia were taken from samples of procedures performed by instructors. For peripheral venous cannulation and tracheal intubation, acceptable failure rates were set arbitrarily at 20% given the stringency of our criterion of success. In practice, acceptable failure rates are initially set at larger values, allowing beginners to reach them after performing a small number of procedures. As residents achieve these initial rates, the cumsum line is recalculated at progressively more stringent failure rates until the acceptable failure rates reach the desired level.

Some practical issues influence the establishment of unacceptable failure rates for the cumsum method. The probabilities of the type I and II errors and the difference between the acceptable and unacceptable failure rates are major determinants of the adequate sample size and of the angle of the upward inclination of the cumsum curve at each failure. It has been recommended that for a given size of α and β , the difference between p_0 and p_1 should be adjusted to keep the angle of ascent of the cumsum line no more than 60° (1). Otherwise, because each failure will cause a steep ascent of the cumsum line, it will tend to cross the upper decision limit, even in the presence of an acceptable cumulative failure rate (p_0). Conversely, a long run of successful procedures will be required after a few failures for the cumsum line to return to the baseline or to cross the lower decision limit (h_0) and artificially distorting the learning curve (5). It is also desirable to keep the average number of procedures low to detect failure

rates corresponding to p_0 and p_1 so that corrective measures can be taken at relatively short intervals (1).

In this study, α and β were set to 0.1 and unacceptable failure rates were set by doubling p_0 . By using these variables, the upward slopes of the curves were kept at approximately 45° , and the average sample sizes were <30 procedures.

The construction of learning curves based on the cumsum method depends on honesty at self-reporting. We tried to overcome self-reporting biases by having the forms reviewed and signed by the supervising instructor. Only at peripheral venous cannulation did all participants attain the acceptable failure rate. At tracheal intubation, spinal, and epidural anesthesia, only 57%, 63%, and 45% of trainees crossed the lower decision limit, respectively. In addition, among those crossing the lower decision limit, there was a wide variability in the number of procedures required to do so. This is in agreement with other studies (5,6,8-10). Our data refer to a limited number of residents at a single institution. Because training may vary between different institutions, our results may not apply to other teaching environments.

The cumsum method does not allow weighting of the cumsum score according to the expected difficulty at an individual procedure. For this reason, we did not account for patient variables that may have influenced the success rates of individual residents.

The learning process for clinical management aspects has not been a subject of this study. It has been demonstrated that essential motor skills are achieved earlier than knowledge on spinal anesthesia (12). Other evaluation tools must thus be used to assess other domains of the learning process, such as knowledge and judgment.

We conclude that the cumsum method is a useful tool for objective evaluation of practical skills during the learning phase of basic anesthetic procedures. As acceptable failure rates may be progressively adjusted as residents acquire more proficiency, the cumsum method can evaluate the effects of training on an individual basis when applied to the teaching of procedural skills.

References

1. Davies OL, Goldsmith PL, eds. Statistical methods in research and production. London: Longman, 1976.
2. Wohl H. The cumsum plot: its utility in the analysis of clinical data. *N Engl J Med* 1977;296:1044-5.
3. Kinsey SE, Giles FJ, Holton J. Cumsum plotting of temperature charts for assessing antimicrobial treatment in neutropenic patients. *BMJ* 1989;299:775-6.
4. Williams SM, Parry BR, Schlup MMT. Quality control: an application of the cumsum. *BMJ* 1992;304:1359-61.
5. Kestin IG. A statistical approach to measuring the competence of anaesthetic trainees at practical procedures. *Br J Anaesth* 1995;75:805-9.
6. Kopacz D. QA in regional anesthesia training: quantity or quality? *Reg Anesth* 1997;22:209-11.

7. Lawler PG, Patla VR, Garcia E, Puttick N. Assessment of training in anaesthesia and related skills. *Anaesthesia* 1991;46:597.
8. Kopacz DJ, Neal JM, Pollock JE. The regional anesthesia "learning curve:" what is the minimum number of epidural and spinal blocks to reach consistency? *Reg Anesth* 1996;21:182-90.
9. Konrad C, Schüpfer G, Wietlisbach M, Gerber H. Learning manual skills in anesthesiology: is there a recommended number of cases for anesthetic procedures? *Anesth Analg* 1998;86: 635-9.
10. Schuepfer G, Konrad C, Schmeck J, et al. Generating a learning curve for pediatric caudal epidural blocks: an empirical evaluation of technical skills in novice and experienced anesthesiologists. *Reg Anesth Pain Med* 2000;25:385-8.
11. Sprung J, Bourke D, Grass J, et al. Predicting the difficult neuraxial block: a prospective study. *Anesth Analg* 1999;89: 384-9.
12. Sivarajan M, Miller E, Hardy C, et al. Objective evaluation of clinical performance and correlation with knowledge. *Anesth Analg* 1984;63:603-7.

ANESTHESIA

Attention & Authors!

ANALGESIA

Submit Your Papers Online

Journal of the International Anesthetic Research Society, the Society of Canadian Anesthesiologists, the Society for Pediatric Anesthesia, the Society for Ambulatory Anesthesia, the International Society for Anesthetic Pharmacology, and the Society for Technology in Anesthesia

You can now have your paper processed and reviewed faster by sending it to us through our new, web-based Rapid Review System.

Submitting your manuscript online will mean that the time and expense of sending papers through the mail can be eliminated. Moreover, because our reviewers will also be working online, the entire review process will be significantly faster.

You can submit manuscripts electronically via www.rapidreview.com. There are links to this site from the Anesthesia & Analgesia website (www.anesthesia-analgesia.org), and the IARS website (www.iars.org).

To find out more about Rapid Review, go to www.rapidreview.com and click on "About Rapid Review."