David C. Warltier, M.D., Ph.D., Editor

Anesthesiology 2007; 107:144–52

Advances in and Limitations of Up-and-down Methodology

A Précis of Clinical Use, Study Design, and Dose Estimation in Anesthesia Research

Nathan L. Pace, M.D., M.Stat., Mario P. Stylianou, Ph.D.†*

Sequential design methods for binary response variables exist for determination of the concentration or dose associated with the 50% point along the dose–response curve; the up-and-down method of Dixon and Mood is now commonly used in anesthesia research. There have been important developments in statistical methods that (1) allow the design of experiments for the measurement of the response at any point (quantile) along the dose–response curve, (2) demonstrate the risk of certain statistical methods commonly used in literature reports, (3) allow the estimation of the concentration or dose—the target dose—associated with the chosen quantile without the assumption of the symmetry of the tolerance distribution, and (4) set bounds on the probability of response at this target dose. This article details these developments, briefly surveys current use of the up-and-down method in anesthesia research, reanalyzes published reports using the up-and-down method for the study of the epidural relief of pain during labor, and discusses appropriate inferences from up-and-down method studies.

A *DRUG tolerance distribution* is the distribution among a number of subjects of the critical level of intensity (dose or concentration) at which the drug will just produce or just prevent a reaction in each subject. Experiments may be designed to characterize all or part of the tolerance distribution. Experimental methods for binary response variables exist for determination of the dose associated with the 50% point along the dose–response curve or tolerance distribution if the response of the patient can be observed immediately.[1] The earliest use by the anesthesia research community of one particular sequential design—the up-and-down method (UDM)—seems to have been in studies of the effective concentration for a 50% response (EC_{50} or minimum alveolar concentration) of inhalation anesthetics.[2,3] Columb and Lyons[4] used the same design and reported the EC_{50}—labeled by them the minimum local analgesic concentration (MLAC)— of epidural bupivacaine and lidocaine for pain relief during labor. The use of this experimental design is now commonly seen in studies of intrathecal local anesthetics,[5] intrathecal opioids,[6] intravenous opioids,[7] intravenous α_2 agonists,[8] intravenous anesthetics,[9] and inhalation anesthetics[10]—all studies determining the concentration or dose (EC_{50} or ED_{50}) of these drugs.

Since the first reported use of the UDM in the anesthesia literature, there have been important developments in statistical methods that (1) allow the design of experiments for the measurement of the response at any point (quantile g) along the dose–response curve, (2) demonstrate the risk of certain statistical methods commonly used in literature reports, (3) allow the estimation of the concentration or dose associated with the chosen quantile g (EC_g or ED_g) without the assumption of the symmetry of the tolerance distribution, and (4) set bounds on the probability of response at the EC_g or ED_g. This article will detail these developments, briefly survey current use of the UDM in anesthesia research, reanalyze published reports of MLAC for epidural relief of pain during labor to demonstrate newer methods, and discuss appropriate inferences from UDM studies.

This article is accompanied by an Editorial View. Please see: Fisher D: What if half of your patients moved (or remembered or did something else bad) at incision? ANESTHESIOLOGY 2007; 107:1–2.

Materials and Methods

Experimental Design

The following notation and nomenclature will be used. A study uses K ordered dose or concentration levels that are denoted by the set $\Omega_x = \{x_1 < x_2 < \cdots < x_k < \cdots < x_K\}$; e.g., if five concentrations of bupivacaine (0.07, 0.08, 0.09, 0.10, 0.11) are used, then $\Omega_x = \{0.07, 0.08, 0.09, 0.10, 0.11\}$. The purpose of the study is to determine the dose level that has a probability of effect Γ; most commonly the EC_{50} or ED_{50} is

* Professor of Anesthesiology, University of Utah. † Mathematical Statistician, Office of Biostatistics Research, National Heart, Lung, and Blood Institute, National Institutes of Health.

Received from the Department of Anesthesiology, University of Utah, Salt Lake City, Utah, and the Office of Biostatistics Research, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland. Submitted for publication July 19, 2006. Accepted for publication January 24, 2007. Dr. Pace received salary support from the Department of Anesthesiology, University of Utah, and Dr. Stylianou received salary support from the Office of Biostatistics Research, National Heart, Lung, and Blood Institute, National Institutes of Health. Presented in part at the Annual Meeting of the American Society of Anesthesiologists, Atlanta, Georgia, October 25, 2005.

Address correspondence to Dr. Pace: University of Utah, Department of Anesthesiology, 30 North 1900 East, 3C444, Salt Lake City, Utah 84132-2304. n.l.pace@utah.edu.

desired, and thus the assignment $\Gamma = 0.5$ is made so that a response is produced in $100 \times \Gamma = 50\%$ of the target population. This will be denoted the target dose. N denotes the total sample size, N_k and R_k are the number of subjects and the number of responses at dose level x_k, and $R_k/N_k = \hat{p}(x_k) = \hat{p}_k$ denotes the observed response rate at the dose x_k; the observed response rate $\hat{p}_k$ is an estimator of the true probability of response $p(x_k) = p_k$ at dose x_k. Let $X(i)$ denote the dose level administered to the ith patient. The target dose is designated μ—with $\hat{\mu}$ being the estimator of μ. The probability of effect at $\hat{\mu}$ is $p(\hat{\mu})$; the estimator of $p(\hat{\mu})$ will be denoted $\hat{p}(\hat{\mu})$.

The first description of the performance of a UDM experiment is attributed to Dixon and Mood.[11] To perform the study at the quantile $\Gamma = 0.5$, the K dose levels are chosen, the sample size of patients from a target population is selected, and the definition of a positive response is specified; the K doses are chosen with a fixed, constant interval between levels. The first patient is exposed to one of the K dose levels; at the discretion of the investigators, this may be chosen to be the lowest dose, may be the dose thought to be closest to μ, or may be selected arbitrarily. The response of the first patient determines the dose level for the second patient. If the first patient has a positive response at dose x_k, then the second patient is exposed to the next lower dose x_{k-1}. If the first patient has a negative response at dose x_k, then the second patient is exposed to the next higher dose x_{k+1}. Successive patients are assigned a dose similarly by the response of the previous patient. At the lowest and highest doses of Ω_k, appropriate adjustments for the next dose keep all doses within the specified dose range. Although Dixon proposed a modified UDM for very small samples ($N \leq 6$),[12] anesthesia trials using the UDM typically have 20–40 patients.

While there are several types of sequential designs for determining the target dose, all outperform nonsequential designs by having a smaller mean square error (MSE) for the same sample size[1]; often a nonsequential design will require two to three times as many patients to have the same MSE. The UDM has been generalized as the biased coin design (BCD) by Durham *et al.*[13] to find ED_g for other quantiles g. For example, suppose the ED_{95} is to be determined ($\Gamma = 0.95$); the probability $B = \frac{1-\Gamma}{\Gamma} = \frac{1-0.95}{0.95} = \frac{1}{19} \approx 0.05$ is defined. The trial is initiated similarly as described in the second paragraph of this section on Experimental Design. If no response is observed in the previous patient, the dose is stepped up in the next patient. If a positive response is observed, the next patient is randomized with probability $B = 0.05$ to the next lower dose and with probability $1 - B = 0.95$ to the same dose. If a quantile less than 0.5 is selected, the probability $B = \frac{\Gamma}{1-\Gamma}$ is defined analogously; if a positive response is observed, the dose is stepped down; if no response is observed, the next patient is randomized with probability B to the next higher dose and with probability $1 - B$ to the same dose. Intervals between doses do not have to be equal in the BCD. The BCD is optimal among the general class of up-and-down designs in the sense that the distribution of administered doses is most peaked around μ. At $\Gamma = 0.5$, the BCD reduces to the Dixon and Mood UDM. No studies in the anesthesia literature have attempted to study other quantiles on the dose–response curve using a sequential design.

Graphical Display

An ingenious graphical display is now almost universal in reports using the UDM; the first report of UDM methodology by Dixon and Mood[11] included this graphical display. It simultaneously displays the doses/concentrations administered (Ω_x), the sequencing of administered doses ($X(i)$), and the positive or negative response for each subject at their assigned dose/concentration. As all experimental data are displayed, a reanalysis is easily possible. (Reanalysis is also possible from a tabular plotting of dose, number of events, and number treated.) Figure 1 shows a replotting of data from two published UDM studies.[14,15] Each study reported the MLAC from two UDM experiments—one experiment performed for levobupivacaine and one experiment performed for ropivacaine. (If two drugs are being compared in the same study, patients should be randomly allocated to the one or the other drug.) Long sequences of steadily increasing or decreasing doses with repeated ineffective or effective responses are common. In such circumstances, some assigned doses or concentrations must vary considerably from the underlying true ED_g or EC_g.

Target Dose Estimation

Besides the goal of estimating $\hat{\mu}$ from a UDM study, it is also necessary to specify the precision of $\hat{\mu}$ with a 95% confidence interval (CI). Furthermore, in designing an UDM study, a quantile Γ was selected. While $\hat{\mu}$ is an estimate of the true value of the dose or concentration μ at Γ, the $\hat{\mu}$ estimator may be at some distance from μ. If the dose–response curve is steep at the quantile of interest, then the probability of response at the recommended dose $p(\hat{\mu})$ may differ appreciably from the desired quantile. Although it is not possible to know the true probability of effect $p(\hat{\mu})$ at $\hat{\mu}$, an estimator $\hat{p}(\hat{\mu})$ with a CI should be obtained.

Target dose estimation from UDM data are complicated by three considerations: (1) the assigned doses $X(i)$, $i = 1, \ldots, n$, are not independent variables because each dose is dependent on the response to the previous dose; (2) because UDM studies are in-

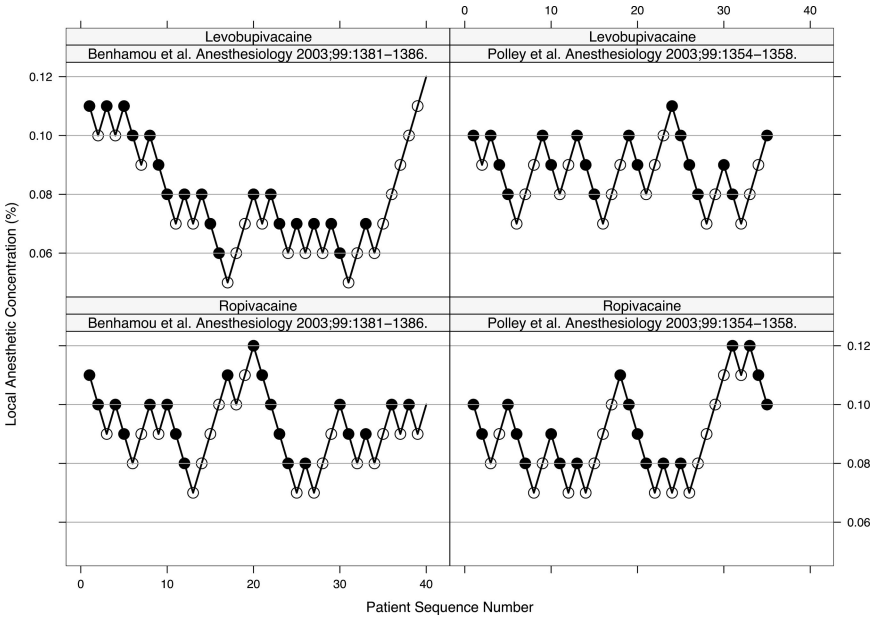

Fig. 1. Determination of minimum local analgesic concentration for epidural analgesia in labor. The patient sequence number (*x-axis*) is the ordering of patient exposures using the up-and-down design. The assigned concentrations (*y-axis*) are (0.07, 0.08, 0.09, 0.10, 0.11, 0.12) percent for ropivacaine and (0.05, 0.06, 0.07, 0.08, 0.09, 0.10, 0.11, 0.12) percent for levobupivacaine. An effective dose (analgesia achieved) is denoted by a *solid circle*; an ineffective dose (analgesia not achieved) is denoted by an *open circle*. Benhamou *et al.*[14] do not report the response of the last patient in the sequences for ropivacaine and levobupivacaine. From Benhamou *et al.*[14] and Polley *et al.*[15]; used with permission.

tended to minimize sample size and generally choose smaller sample sizes, large sample size assumptions (asymptotic properties) may not apply to the estimators; and (3) although the assumption of dose–response curve symmetry of certain mathematical models (*e.g.*, normal or logistic distribution) is typically made, this assumption is unverifiable, and estimators may have a large bias. Algorithms for the exact distribution of estimators have been discussed, but computational power for samples larger than $N = 15$ is lacking for their calculation.[16] Thus the derivation of the definitive statistical properties of estimators is an intractable problem; properties of estimators have been established by simulation studies in most statistical methods articles.

Target dose estimation may be classified as parametric or nonparametric depending on whether a probability distribution of data is assumed (parametric estimation) or is not assumed (nonparametric estimation). For example, because of the nonindependence of dose assignments, the typical assumption of normality of data is suspect. Some target dose estimators do assume normality based on a series of assumptions and approximations; typically, the small sample properties of these estimators are unknown; these will also be classed within nonparametric estimation.

Dixon and Mood Estimator (Nonparametric)

Although many estimators have been proposed for UDM studies, two different nonparametric estimators of μ are commonly used in anesthesia UDM studies. Because the assigned doses $X(i)$ cluster unimodally around μ, the first nonparametric estimator is the empirical average $\hat{\mu} = \sum_{i=1}^{N} X(i)$. There are several variants of the simple average, including truncation of the first run of all assigned doses with similar (either positive or negative) observed responses[17] and the use of only the assigned doses with positive or negative responses as suggested by Dixon and Mood.[11,18] The truncated version has been shown to be superior.[17] Most anesthesia UDM studies do not specify which variant of the simple average is being used.

Turning Point Estimator (Nonparametric)

Also commonly used is the turning point estimator of Choi.[19] The values of pairs of successive assigned doses at which the observed response changed direction (pair reversals) are averaged. The turning point estimator is commonly used in anesthesia UDM studies and has been used in a methods article on UDM in the anesthesia literature.[20] Choi[19] demonstrated by simulation that the turning point estimator was superior to the Dixon and Mood simple average, but only for an assumed normal tolerance distribution.

Logistic/Probit Estimators (Parametric)

Many anesthesia UDM studies also present a logistic or probit regression analysis of the data as a so-called sensitivity or backup analysis using commonly available logistic or probit regression software. Vågerö and Sundberg[21] specifically reexamined and criticized the use of such analysis in several published anesthesia UDM studies. By simulation, they demonstrated that the parameter estimate of the regression slope is biased and the CIs of the ED_{50} are unrealistically narrow. This seemed to be a consequence of the nonindependence of assigned dose values. A logistic regression estimation routine adapted to the nonindependence of up-and-down data has been developed,[22] but the requirement of tolerance distribution symmetry is necessary. In addition, the $\hat{\mu}$ estimator does not always exist for all UDM data sets.

Isotonic Regression Estimators (Nonparametric)

There is an implicit assumption in dose–response determinations: Drug effect increases with increasing dose. This assumption can be used in a technique known as isotonic regression to estimate $\hat{\mu}$ and $\hat{p}(\hat{\mu})$ with a minimum use of other unverifiable assumptions such as symmetry of the dose tolerance distribution. The plot of sequential dose responses (fig. 1) does not make obvious that the observed response rate is not necessarily monotonically increasing with increasing concentration. As tabulated in table 1 and displayed in figure 2, biologic and experimental variability may produce unexpected ups and downs in the observed response rate as dose increases.

Isotonic regression is a well-described variant of restricted least squares regression that constrains the point estimates to be either monotonic increasing (never decreasing) or monotonic decreasing (never increasing). Isotonic regression has favorable statistical properties.[23] An isotonic point estimate may be constant over some range of doses. At each assigned dose, an adjusted response probability p_k^* is easily calculated by the pooled-adjacent-violators algorithm (PAVA) (appendix 1). As shown in figure 2, the PAVA-adjusted response rate is monotonic.

Several target dose estimators of the EC_g or ED_g are possible with isotonic regression. These include the following:

- the largest dose with p_k^* less than or equal to Γ, $\hat{\mu}_1 = \max_{x_k \in \Omega_x} (x_k : p_k^* \leq \Gamma)$;
- the dose for which the difference between Γ and p_k^* is minimized, $\hat{\mu}_2 = (x_k : |p_k^* - \Gamma|$ is minimum among $x_k \in \Omega_x)$; and
- a linearly interpolated dose between the p_k^* and p_{k+1}^* bounding Γ, $\hat{\mu}_3 = \frac{\Gamma - p_k^*}{p_{k+1}^* - p_k^*}(x_{k+1} - \hat{\mu}_1) + \hat{\mu}_1$.

Simulation studies by Stylianou *et al.* have demonstrated that the isotonic estimator $\hat{\mu}_3$ has a smaller bias and MSE compared with the Dixon and Mood estimators.[22] Compared with the logistic regression estimator adjusted for nonindependence of data, the isotonic $\hat{\mu}_3$ estimator requires no symmetry assumption, always exists, and has similar MSE.[22] Simulation studies have also demonstrated that of the three isotonic regression estimators, the isotonic $\hat{\mu}_3$ estimator will have the narrowest CIs.[24] While the calculation of the isotonic $\hat{\mu}_3$ estimator is simple, the CIs are obtained by computer numerical methods. Stylianou[16] developed a parametric bootstrap routine calculating CIs by a bias corrected percentile method. This bootstrapping approach allows the estimation of CIs of $\hat{\mu}_3$ for an experimental design with any Γ, not just the 50th quantile. Also, bootstrapping is implemented to obtain an estimate of the bias of $\hat{p}(\hat{\mu})$ from Γ with CIs (appendix 2).

Stopping Rules

Studies with sequential designs must have a stopping rule. The nonindependence and unknown distribution of data of a UDM study prevent the development of theoretically rigorous rules to calculate the necessary sample size for a prespecified precision of the estimation of EC_g or ED_g. Simulation studies suggest that including at least 20–40 patients will provide stable estimates of the target dose for most realistic scenarios.[16] Other sequential methods use computationally intensive methods that require reanalysis of all data after each patient.[25]

Survey of UDM Trials in the Anesthesia Literature

The contents of the journal ANESTHESIOLOGY were searched from January 2000 to June 2006 for clinical reports using the UDM. Discovered reports were assessed for (1) the presentation of raw data, (2) sample sizes, (3) stopping rules, (4) the estimator of EC_{50} or

Table 1. Observed and PAVA-adjusted Response Rates

Assigned Concentration	Number Analgesic	Number Tested	Observed Response Rate ($\hat{p}_k$)	PAVA-adjusted Response Rate (p_k^*)
Ropivacaine				
0.07	0	3	0.000	0.000
0.08	3	8	0.375	0.375
0.09	5	13	0.385	0.385
0.10	8	10	0.800	0.786
0.11	3	4	0.750	0.786
0.12	1	1	1.000	1.000
Levobupivacaine				
0.05	0	2	0.000	0.000
0.06	2	8	0.250	0.250
0.07	6	11	0.545	0.545
0.08	5	6	0.833	0.571
0.09	1	3	0.333	0.571
0.10	2	5	0.400	0.571
0.11	3	4	0.750	0.750

The data of Benhamou *et al.*[14] are tabulated to demonstrate that the observed response rates ($\hat{p}_k$) sometimes decrease as dose increases. The use of the pooled-adjacent-violators algorithm (PAVA) makes monotonic — never decreasing — the estimator of the response rate (p_k^*) as dose increases. The calculation of PAVA estimators by isotonic regression is demonstrated in appendix 1.

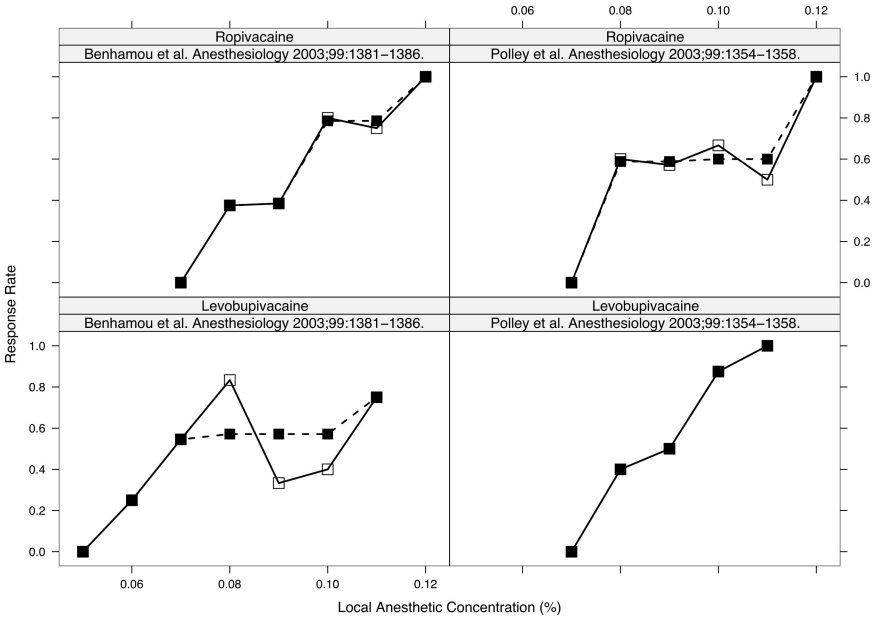

Fig. 2. Observed and pooled-adjacent-violators algorithm (PAVA) response rate. The assigned concentrations (*x-axis*) are (0.07, 0.08, 0.09, 0.10, 0.11, 0.12) percent for ropivacaine and (0.05, 0.06, 0.07, 0.08, 0.09, 0.10, 0.11, 0.12) percent for levobupivacaine. The estimate of the observed response rate $\hat{p}_k$ (*solid line*, *open square*) and the PAVA-adjusted response rate p_k^* (*dashed line*, *solid square*) at the local anesthetic concentrations are plotted on the *y-axis*. Although $\hat{p}_k$ and p_k^* are equal at some concentrations, in three of the four drug trials, $\hat{p}_k$ was not monotonically increasing at two or more concentrations.

ED_{50}, (5) the use of probit or logit regression analysis, and (6) the desired inferences.

Reanalysis of Published Studies

As new local anesthetics are introduced, their potencies in various types of clinical use should be determined. Benhamou *et al.*[14] and Polley *et al.*[15] each reported the MLAC for levobupivacaine and ropivacaine when administered for epidural analgesia in parturients during labor. In each study, patients who requested epidural analgesia were randomly allocated to receive either levobupivacaine or ropivacaine. Within each study, two independent UDM trials were conducted—one for levobupivacaine and the other for ropivacaine. The concentrations (levobupivacaine and ropivacaine) used in the first patient of each sequence were 0.10% (Polley *et al.*[15]) and 0.11% (Benhamou *et al.*[14]). Each study published figures (Benhamou *et al.*,[14] fig. 1; Polley *et al.*,[15] figs. 1 and 2) showing the sequential assignment of local anesthetic concentrations and the success or failure of the block. Data were extracted from these figures. A secondary analysis of results using these data were run using the previously described software; 9,999 bootstrap replications were simulated; the first concentration in the resampling algorithm was set to the first concentration used in the study. Comparison of MLACs for each study used the method of overlapping confidence intervals. If 83% CIs are nonoverlapping, then the null hypothesis of equal MLACs is rejected at an α of approximately 0.05.[26]

Results

During 6½ yr (January 2000 through June 2006), 16 clinical reports used the UDM in their research.[6,14,15,27–39] Only 1 report did not display the raw data in either graphical or tabular form.[31] In 5 reports, sample sizes were less than 20.[31–33,35,39] It seems that all studies used a fixed sample-size stopping rule. Among the 16 reports, the most commonly used estimators of ED_{50} or EC_{50} included the Dixon and Mood average,[6,14,15,28,30,34,36,37] the turning point estimator of pair reversals,[27,29,38,39] and the small sample Dixon estimate.[31,35] These empirical mean estimators were accompanied in 9 studies by a probit or logit regression analysis.[15,28,29,33–38] In addition, 2 studies calculated an ED_{95} by probit or logit regression.[27,29] Most studies of regional anesthesia using the UDM classified some patients as "rejects" not to be evaluated and used the same dose or concentration for the next patient in sequence.[14,15,28,29,34–38] Two studies pursued a study of drug interaction. In each study, three sequential trials were performed: drug A, drug B, and drugs A plus B in a fixed ratio; on an isobologram, the three ED_{50}s with confidence intervals were plotted; no statistical properties of this isobolographic analysis were provided. Most studies sought a sparing effect on the EC_{50} or ED_{50} of one drug by adding a second drug,[6,27,33,37,38] compared the potency of one drug in different patient groups or under different clinical conditions,[29,32,35,36,39] or compared the potency of two or more drugs.[14,15,28] Two studies explicitly recognized that an EC_{50} or ED_{50} is not a dose recommendation[29] or that the entire response surface of drug combinations across low to high quantiles of effect must be studied to map the interaction of two drugs.[31]

The MLACs for levobupivacaine and ropivacaine as reported by Benhamou *et al.*[14] and Polley *et al.*[15] are presented in table 2 along with MLACs obtained from isotonic regression reanalysis of the original data. The sample sizes are adequate for stable estimation of param-

Table 2. Reported and Reanalyzed MLAC for Epidural Analgesia in Labor

Reference, Drug, No. of Patients	Empirical Mean MLAC (95% Confidence Interval)	Isotonic Regression MLAC (95% Confidence Interval) (83% Confidence Interval)
Polley *et al.*,[15] levobupivacaine, n = 35	0.087% (0.08–0.09%)	0.090% (0.079–0.095%) (0.082–0.094%)
Polley *et al.*,[15] ropivacaine, n = 35	0.089% (0.08–0.10%)	0.079% (0.076–0.096%) (0.076–0.084%)
Benhamou *et al.*,[14] levobupivacaine, n = 40	0.077% (0.058–0.096%)	0.068% (0.058–0.095%) (0.059–0.081%)
Benhamou *et al.*,[14] ropivacaine, n = 40	0.092% (0.082–0.102%)	0.093% (0.080–0.100%) (0.087–0.097%)

The empirical mean minimum local analgesic concentration (MLAC) is reported in each reference citing the method of Dixon and Massey.[18] The isotonic regression MLAC is a reanalysis using the original data. Estimates are reported as percent concentration local anesthetic. The concentrations (levobupivacaine and ropivacaine) used in the first patient of each sequence were 0.10% (Polley *et al.*[15]) and 0.11% (Benhamou *et al.*[14]); the same first concentration was used in bootstrapping the confidence intervals of the isotonic regression MLAC.

eters. Some estimates of MLACs by empirical mean and isotonic regression are similar (*e.g.*, levobupivacaine 0.087% *vs.* 0.090%), whereas others are different by 0.10% (*e.g.*, ropivacaine 0.089% *vs.* 0.079%). The widths (upper bound minus lower bound) of the 95% CIs are the same or slightly larger for the isotonic regression MLACs. To compare the isotonic regression MLACs between levobupivacaine and ropivacaine within each study, 83% CIs were estimated (table 2). In the report by Benhamou, levobupivacaine was more potent than ropivacaine, but without statistical significance; using the 83% CIs from the bootstrap distribution, the difference between levobupivacaine and ropivacaine was statistically significant. In the report of Polley, there was no difference between levobupivacaine and ropivacaine.

Not possible with the methods of the empirical mean, isotonic regression estimated a precision range for the probability of effect at the MLAC. Although obtaining an EC_{50} ($\Gamma = 0.5$) was the goal of each study, the CIs on the nominal probability of effect (50%) were wide: levobupivacaine (Benhamou: 28–76%; Polley: 27–80%) and ropivacaine (Benhamou: 29–67%; Polley: 33–76%).

Some investigators exclude a patient from analysis because of protocol violation, doubts about drug administration, treatment failure, and so on; these patients are denoted as "rejects." The next patient in the study is substituted for the excluded patient and receives the same local anesthetic concentration as previously assigned. The levobupivacaine experiment of Polley *et al.*[15] analyzed the responses of 35 patients; an additional 29 patients were excluded. A sensitivity analysis of all 64 patients, those included and those rejected, revealed that imputing values of no analgesia for rejected patients estimates a larger MLAC, 0.097% (0.092–0.102%) *versus* 0.090% (0.079–0.095%).

Discussion

For anesthetics, the characterization of changes in drug effect as dose increases is of fundamental importance. As one of the elements in the pharmacodynamic description of a drug, this information is used to choose the initial dosing of local, intravenous, and inhaled anesthetics and can be used for dosing adjustments of anesthetics given repeatedly or continuously. To completely characterize the dose–response (tolerance distribution) of an anesthetic requires an experimental design in which multiple patients are exposed to several doses. This is true for both continuous response variables such as blood pressure and binary response variables such as anesthetized/not anesthetized. Such an experiment will predetermine the number of patients to be assigned at each of several doses and will randomly assign patients to these several different doses; analysis of data will typically use logit or probit regression methods.

The dose–response is conceived as a curve in the cartesian plane with increasing dose represented along the x-axis and the response variable plotted along the y-axis—thus the descriptive term *dose–response curve*. The interpretation of this curve is dissimilar for continuous and binary responses. For continuous responses, this curve usually depicts the averaged observed values at different doses. For binary responses, this curve has a probabilistic meaning. The points along the curve represent the tolerance distribution of the population exposed to the anesthetic[40]; thus the dose at which 50% of the population responds is considered the median effective dose—the effective dose for a response by 50% of patients exposed (ED_{50}). It must be emphasized that a patient does not have an ED_{50}; if the probability distribution is known, then the ED_{50} is a parameter describing a population of patients. The determination of this dose–response curve is called *quantal bioassay*.

In designing clinical trials to determine the target dose, the experimenter must work to minimize bias and variability. Bias is the difference between the true value of ED_{50} and the estimated value obtained by experimentation. Variability or variance can be defined as the square of the deviations of observed sample values from the estimated ED_{50}. In comparing various experimental designs and estimators for determination of the ED_{50}, the accuracy of a particular estimator is defined by the MSE. The MSE is the sum of the squared bias and the variance. In comparing estimators, a smaller MSE is desired. Simulation studies have demonstrated that performing an experiment to characterize the ED_{50} by obtaining a fully

specified tolerance distribution (nonsequential design with preassigned doses at multiple quantiles from ED_1 to ED_{99}) is almost always less efficient than a sequential design to estimate only the ED_{50}.[17,41] For both sequential and nonsequential designs, MSE becomes smaller as sample size increases. But it may require two to three times as many experimental subjects in a nonsequential design to achieve the same lower MSE of a sequential design. Because clinical trials are always subject to patient availability and the economic constraints of minimizing the number of subjects, a sequential design is to be preferred.

Up-and-down method experiments are a simply performed type of sequential design for dose finding at the 50th quantile. The isotonic $\hat{\mu}_3$ estimator—as demonstrated by simulation studies— has very favorable properties compared with other UDM estimators for obtaining the target dose with a smaller bias and greater precision (tighter CIs). The precision of the estimate may be narrowed by increasing the sample size. Because in a UDM experiment the assigned doses are concentrated around the target dose for the 50th quantile, the relatively modest sample size prevents any precision in target dose estimators for the lower and upper tails of the tolerance curve. As noted by Dixon in discussing the estimation of small or large percentage points from a UDM experiment, "Nothing short of an extensive exploration of the distribution, involving perhaps thousands of observations, will suffice in that case."[18] With the generalization of the UDM process to BCD experiments, clinical trials may be planned for other quantiles along the tolerance curve. Simulation studies of BCD trials for quantiles far from the midpoint, *e.g.*, $\Gamma = 0.1$, have demonstrated that the isotonic $\hat{\mu}_3$ estimator will have low bias and variance.[16] Thus sufficient accuracy exists to justify BCD trials at high or low quantiles.

In the field of cancer chemotherapy, a common objective in a phase I clinical trial is to estimate the dose of a drug (the target dose) that will produce toxicity with a predetermined probability. Ethical concerns obligate the researchers to use the fewest subjects possible while still being confident that this dose estimate for toxicity is accurate; sequential dose-finding studies are attractive for these reasons. This has prompted a considerable expansion of statistical methods articles treating the sequential dosing of subjects and the estimation of toxicity. One of the types of sequential design for dose finding is known as the continual reassessment method.[25] Implementation of a continual reassessment method design requires the specification of the mathematical model of the dose–response curve and previous parameter values before initiation of the study; after each patient's response is observed, the totality of all patients' responses are reanalyzed to choose the drug dose for the next patient; the full involvement of a statistician is necessary to design, perform, and analyze the study.[25] By contrast, the BCD sequential study avoids unverifiable assumptions about the mathematical model of the dose–response curve, requires minimal computation during the study to assign successive doses, and has favorable statistical properties of the estimators of the target dose. New developments in BCD methods also are appearing.[42]

In the anesthesia literature, UDM studies have almost always been used to explore the performance of drugs already in common clinical use. Instead of seeking toxicity doses, these studies are typically exploring the performance of a drug or two drugs (EC_{50} or ED_{50}) in specifically defined clinical circumstances. For example, assuming the parallelism of the dose–response curves of two different anesthetics, the EC_{50}s or ED_{50}s alone will allow a comparison of the relative potencies.

New methods have advanced the performance of UDM studies considerably since the original methods proposed by Dixon and Mood.[11] There are particular reasons for the BCD to be of interest to anesthesia researchers. Although anesthesiologists discuss potency of inhaled anesthetics in terms of the EC_{50}, the potency dose of real interest is the EC_{95} or EC_{99}. A BCD study can be performed setting $\Gamma = 0.95$, permitting a direct estimation of EC_{95} and avoiding unverifiable extrapolations from the EC_{50} value.

The use of bootstrap methods to set a confidence interval on the probability of effect in the studies of Benhamou and Polley showed that these confidence intervals are broad—typically from 0.25 to 0.75 or more on the nominal probability of $\Gamma = 0.5$. These confidence intervals may seem unreasonably large but are a consequence of the small sample sizes. Suppose that only one dose, the true ED_{50}, was given to all patients. Because of random variation in patient responses and assuming a binomial distribution with $\pi = 0.5$, a sample size of 20 would be chosen to have a 95% CI extending from 0.28 to 0.72; even a sample size of 40 would only narrow the 95% CI to a range of 0.35 to 0.65. To achieve a 95% CI extending from 0.45 to 0.55 would require a sample size of 385 patients. When interpreting a UDM study, one should be cautious about the probability of effect at the estimated ED_{50} or EC_{50} because there is little precision.

Finally, anesthesiologists are in most circumstances truly interested in the $ED_{95/99}$ or $EC_{95/99}$ of a drug. For several reasons, the ED_{50} or EC_{50} determined in a UDM experiment cannot provide any reliable insight into the upper tail of the tolerance distribution. Although it is often assumed that the sigmoidal dose–response curves of all drugs are well fitted by a symmetric logistic curve, in fact that is an unverifiable assumption, especially when dealing with small sample sizes. Near the midpoint (median) of the tolerance distribution, multiple mathematical models (normal distribution, logistic distribution, Cauchy distribution, log, complementary log–log, and so on) fit the observed data equally well, but at the upper tail of the distribution, the estimates of ED_{95} or

EC_{95} may be quite different under these alternative mathematical models. When two different drugs are compared at the ED_{50} or EC_{50}, the equality or inequality of these median values do not necessarily allow inferences about the potency relation at the ED_{95} or EC_{95}. The tolerance distribution functions may not be the same; even if the same, there may be a lack of parallelism. The tolerance distributions also may have a lack of symmetry outside the median point.

Although a brief survey of the current use of UDM in anesthesia research showed most studies to use 20 or more patients in a sequential trial, some studies were smaller. Statistical methods articles using simulation methods recommend that studies have 20 or more patients. Also, researchers should consider the use of the isotonic regression estimator with CIs derived by bootstrapping. Because standard probit or logit regression is likely to produce biased estimators and considering the lack of familiarity, by many anesthesiologists, with probit or logit regression for nonindependent data, UDM reports should no longer include logit or probit regression analysis. No attempt should be made to extrapolate a high-quantile (95 or 99) effect dose/concentration from the median points of the tolerance distribution curve; instead, a BCD study should be planned with $\Gamma = 0.95$. Researchers of regional anesthesia should reconsider the practice of labeling some patients as nonevaluable and then repeating the same dose/concentration in the next patient in sequence; this will definitely downwardly bias the estimated value from the true value. The use of UDM Dixon ED_{50} or EC_{50} values for isobolographic analysis[30,31] should be approached cautiously because the statistical properties are unknown; a nonparametric approach using a bivariate isotonic estimator for two-dimensional dose finding should be considered.[43]

Commenting on the reports of Benhamou and Polley,[14,15] Lassie *et al.*[44] posed the question whether the MLAC method and by extension UDM design is robust enough. Using the restrictions and changes suggested in the Discussion, UDM maintains its advantages of smaller sample size and reasonably simple study performance. Caution is appropriate in other uses.

Appendix 1: Use of PAVA

1. Derive the naive estimate of response rate for each dose level, $\frac{R_k}{N_k} = \hat{p}(x_k) = \hat{p}_k$.
2. Starting at the lowest dose, find the first adjacent pair of response rates that violate the ordering restriction $\hat{p}_k \leq \hat{p}_{k+1}$. The weight for each response rate is the number of patients assigned to that dose, N_k.
3. Replace this pair by the average of $\hat{p}_k$ and $\hat{p}_{k+1}$ weighted by the number of patients: $p_k^* = p_{k+1}^* = \frac{N_k \hat{p}_k + N_{k+1} \hat{p}_{k+1}}{N_k + N_{k+1}}$, where the superscript star denotes a PAVA estimator.
4. Search the sequence for the next adjacent pair of doses with an order violation. Apply step 2 to create a weighted average.
5. Continue until no further order violations are noted.
6. If no such pair exists with an order violation, then assign the PAVA estimator $p_k^* = \hat{p}_k$ at each dose k.
7. PAVA creates unique estimates p_k^* no matter what the starting point in the dose sequence is.

Appendix 2: Software for PAVA Estimators and Confidence Intervals

Software has been written in GAUSS (Aptech Systems, Inc, Black Diamond, WA) by M.P.S.[24]; it is compatible at least to GAUSS version 3.6. The algorithms were translated into the public domain R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, Vienna, Austria) by N.L.P.; it is compatible at least to R version 2.4.1.‡

‡ Software is available from the authors: GAUSS code from M.P.S. (stylianm@nhlbi.nih.gov) and R code from N.L.P. (n.l.pace@utah.edu).

References

1. Morgan BJT: Analysis of Quantal Response Data. New York, Chapman & Hall, 1992, pp 340–69
2. Rampil IJ, Lockhart SH, Zwass MS, Peterson N, Yasuda N, Eger EI II, Weiskopf RB, Damask MC: Clinical characteristics of desflurane in surgical patients: Minimum alveolar concentration. ANESTHESIOLOGY 1991; 74:429–33
3. Scheller MS, Saidman LJ, Partridge BL: MAC of sevoflurane in humans and the New Zealand white rabbit. Can J Anaesth 1988; 35:153–6
4. Columb MO, Lyons G: Determination of the minimum local analgesic concentrations of epidural bupivacaine and lidocaine in labor. Anesth Analg 1995; 81:833–7
5. Ginosar Y, Mirikatani E, Drover DR, Cohen SE, Riley ET: ED_{50} and ED_{95} of intrathecal hyperbaric bupivacaine coadministered with opioids for cesarean delivery. ANESTHESIOLOGY 2004; 100:676–82
6. Stocks GM, Hallworth SP, Fernando R, England AJ, Columb MO, Lyons G: Minimum local analgesic dose of intrathecal bupivacaine in labor and the effect of intrathecal fentanyl. ANESTHESIOLOGY 2001; 94:593–8
7. Sztark F, Chopin F, Bonnet A, Cros AM: Concentration of remifentanil needed for tracheal intubation with sevoflurane at 1 MAC in adult patients. Eur J Anaesthesiol 2005; 22:919–24
8. Stapelfeldt C, Lobo EP, Brown R, Talke PO: Intraoperative clonidine administration to neurosurgical patients. Anesth Analg 2005; 100:226–32
9. Kodaka M, Okamoto Y, Koyama K, Miyao H: Predicted values of propofol EC_{50} and sevoflurane concentration for insertion of laryngeal mask Classic™ and ProSeal™. Br J Anaesth 2004; 92:242–5
10. Shim YH, Shin CS, Chang CH, Shin Y-S: Optimal end-tidal sevoflurane concentration for the removal of the laryngeal mask airway in anesthetized adults. Anesth Analg 2005; 101:1034–7.
11. Dixon WJ, Mood AM: A method for obtaining and analyzing sensitivity data. J Am Stat Assoc 1948; 43:109–26
12. Dixon WJ: The up-and-down method for small samples. J Am Stat Assoc 1965; 60:967–78
13. Durham SD, Flournoy N, Rosenberger WF: A random walk rule for phase I clinical trials. Biometrics 1997; 53:745–60
14. Benhamou D, Ghosh C, Mercier FJ: A randomized sequential allocation study to determine the minimum effective analgesic concentration of levobupivacaine and ropivacaine in patients receiving epidural analgesia for labor. ANESTHESIOLOGY 2003; 99:1383–6
15. Polley LS, Columb MO, Naughton NN, Wagner DS, van de Ven CJM, Goralski KH: Relative analgesic potencies of levobupivacaine and ropivacaine for epidural analgesia in labor. ANESTHESIOLOGY 2003; 99:1354–8
16. Stylianou M: Sequential Analysis of Durham and Flournoy's Biased Coin Design for Phase I Clinical Trials. Washington, DC, American University, Department of Mathematics and Statistics, College of Arts and Sciences, 2000, pp i–xiii, 1–178
17. Davis M: Comparison of sequential bioassays in small samples. J R Stat Soc Ser B 1971; 33:78–87
18. Dixon WJ, Massey FJ: Introduction to Statistical Analysis, 4th edition. New York, McGraw-Hill, 1983, pp 377–94
19. Choi SC: Interval estimation of the LD50 based on an up-and-down experiment. Biometrics 1990; 46:485–92
20. Paul M, Fisher DM: Are estimates of MAC reliable? ANESTHESIOLOGY 2001; 95:1362–70
21. Vågerö M, Sundberg R: The distribution of the maximum likelihood estimator in up-and-down experiments for quantal dose-response data. J Biopharm Stat 1999; 9:499–519

22. Stylianou M, Flournoy N: Dose finding using the biased coin up-and-down design and isotonic regression. Biometrics 2002; 58:171–7

23. Robertson T, Wright FT, Dykstra RL: Order Restricted Statistical Inference. New York, John Wiley & Sons, 1988, pp 1–40

24. Stylianou M, Proschan M, Flournoy N: Estimating the probability of toxicity at the target dose following an up-and-down design. Stat Med 2003; 22:535–43

25. Garrett-Mayer E: The continual reassessment method for dose-finding studies: A tutorial. Clin Trials 2006; 3:57–71

26. Payton ME, Greenstone MH, Schenker N: Overlapping confidence intervals or standard error intervals: What do they mean in terms of statistical significance? J Insect Sci 2003; 3:1–6

27. Albertin A, Casati A, Bergonzi P, Fano G, Torri G: Effects of two target-controlled concentrations (1 and 3 ng/ml) of remifentanil on MAC_{BAR} of sevoflurane. Anesthesiology 2004; 100:255–9

28. Camorcia M, Capogna G, Columb MO: Minimum local analgesic doses of ropivacaine, levobupivacaine, and bupivacaine for intrathecal labor analgesia. Anesthesiology 2005; 102:646–50

29. Capogna G, Parpaglioni R, Lyons G, Columb M, Celleno D: Minimum analgesic dose of epidural sufentanil for first-stage labor analgesia. Anesthesiology 2001; 94:740–4

30. Delage N, Maaliki H, Beloeil H, Benhamou D, Mazoit J-X: Median effective dose (ED_{50}) of nefopam and ketoprofen in postoperative patients. Anesthesiology 2005; 102:1211–6

31. Harris RS, Lazar O, Johansen JW, Sebel PS: Interaction of propofol and sevoflurane on loss of consciousness and movement of skin incision during general anesthesia. Anesthesiology 2006; 104:1170–5

32. Mingzheng M, Hu X, Liu J: The effect of hypothermia on isoflurane MAC in children. Anesthesiology 2001; 94:429–32

33. Nakata Y, Goto T, Ishiguro Y, Terui K, Kawakami H, Santo M, Niimi Y, Morita S: Minimum alveolar concentration (MAC) of xenon with sevoflurane in humans. Anesthesiology 2001; 94:611–4

34. Nelson KE, Rauch T, Terebuh V, D'Angelo R: A comparison of intrathecal fentanyl and sufentanil for labor analgesia. Anesthesiology 2002; 96:1070–3

35. Panni MK, Segal S: Local anesthetic requirements are greater in dystocia than in normal labor. Anesthesiology 2003; 98:957–63

36. Parpaglioni R, Frigo MG, Lemma A, Sebastiani M, Barbati G, Celleno D: Minimum local analgesic dose: Effect of different volumes of intrathecal levobupivacaine in early labor. Anesthesiology 2005; 103:1233–7

37. Polley LS, Columb MO, Naughton NN, Wagner DS, Dorantes DM, van de Ven CJM: Effect of intravenous *versus* epidural fentanyl on the minimum local analgesic concentration of epidural bupivacaine in labor. Anesthesiology 2000; 93:122–8

38. Polley LS, Columb MO, Naughton NN, Wagner DS, van de Ven CJM: Effect of epidural epinephrine on the minimum local analgesic concentration of epidural bupivacaine in labor. Anesthesiology 2002; 96:1123–8

39. Wadhwa A, Durrani J, Sengupta P, Doufas AG, Sessler DI: Women have the same desflurane minimum alveolar concentration as men. Anesthesiology 2003; 99:1062–5

40. Morgan BJT: Analysis of Quantal Response Data. New York, Chapman & Hall, 1992, pp 1–40

41. Brownlee KA, Hodges JL Jr, Rosenblatt M: The up-and-down method with small samples. J Am Stat Assoc 1953; 48:262–77

42. Stylianou M, Follmann DA: The accelerated biased coin up-and-down design in phase I trials. J Biopharm Stat 2004; 14:249–60

43. Ivanova A, Wang K: A non-parametric approach to the design and analysis of two-dimensional dose-finding trials. Stat Med 2004; 23:1861–70

44. Lacassie HJ, Lacassie HP, Muir HA: Is the minimum local analgesic concentration method robust enough? Anesthesiology 2004; 101:550