READERS’ TOOLBOX
Understanding Research Methods

Understanding Research Methods: Up-and-down Designs for Dose-finding

Assaf P. Oron, Ph.D.,
Michael J. Souter, M.B., Ch.B., F.R.C.A., F.N.C.S.,
Nancy Flournoy, Ph.D.

SUMMARY

For the task of estimating a target benchmark dose such as the ED50 (the dose that would be effective for half the population), an adaptive dose-finding design is more effective than the standard approach of treating equal numbers of patients at a set of equally spaced doses. Up-and-down is the most popular family of dose-finding designs and is in common use in anesthesiology. Despite its widespread use, many aspects of up-and-down are not well known, implementation is often misguided, and standard, up-to-date reference material about the design is very limited. This article provides an overview of up-and-down properties, recent methodologic developments, and practical recommendations, illustrated with the help of simulated examples. Additional reference material is offered in the Supplemental Digital Content.

(ANESTHESIOLOGY 2022; 137:137–50)

Up-and-down designs were developed for dose–finding experiments in which patient responses can be dichot–omized; for example, whether an analgesic drug achieved sufficient pain relief, or whether an experimental drug triggered an unacceptable toxic reaction. Without loss of generality, we use the term "positive" to denote the response that occurs more frequently with increasing dose and *vice versa* for "negative." The objective of dose-finding is to estimate a target dose that would trigger positive responses in a prespecified proportion of the population. The most commonly sought target dose is the dose that produces a positive response in half the population, referred to in certain anesthetic contexts as the ED50 (the dose that would be effective for half the population), or the minimum alveolar concentration (MAC). In statistics this target dose is sometimes called the median response threshold. Other targets encountered often are the ED90 (the dose that would be effective for 90% of the population) and the TD30 (the dose that would be unacceptably toxic for 30% of the population).

Observations at doses far from the target dose are not very useful for the dose-finding task (fig. 1). Therefore,

---

Box 1. What to Look for in Research Using This Method

- **Does the binary endpoint have a consistent and sensible definition?**
- **Is the response probability F(x) expected to trend in the same direction throughout the range of doses used?**
- **Is the study design simple and straightforward?** Up-and-down designs are compatible with simple problems, such as finding a single threshold or possibly comparing a few groups. Anything more complex may require a different design.
- **Does the sample size seem sufficient?** n < 20, prevalent in many fields, is generally quite insufficient. If the article cites a sample-size calculation formula, it is likely an outdated reference.
- **Does the study present the complete sequence of doses and responses? Is there a dose–response plot showing observed rates and estimates?** These foster transparency and facilitate interpretation and evaluation of results.
- **Are the point and interval estimation methods adequate, up to date, clearly described, and explained?** New improved estimation methods have been published, and they should make their way to implementation.
- **Was any estimate reported for an "off-target" dose, such as estimating the ED95 from ED50-finding up-and-down design data?** These would be biased and rely on very little relevant information.
- **If the article compares up-and-down designs with other dose-finding methods (*e.g.*, *via* simulation), did it use a targeted up-and-down design as described in this article, and was it used properly in the comparison?** Many comparison articles conflate up-and-down designs with other designs, or use inadequate estimation methods.

when the goal is dose-finding, dividing patients equally across a broad range of doses is wasteful and potentially unethical. Adaptive dose-finding designs do not allocate all doses *a priori*. Rather, they aim to concentrate dose allocations around the target, using information from prior patients at each stage of the experiment.

Up-and-down designs, first described in the 1940s, are among the earliest adaptive dose-finding designs.[1–3] They remain the most broadly used dose-finding design overall[4] and in particular in anesthesiology.[5,6] Up-and-down designs specify clear and simple rules for assigning doses to patients and have robust properties that enhance efficiency and effectiveness.[7,8]

Improvements to up-and-down design theory and implementation have been developed over the years, keeping the design competitive and up to date with recent methodologic knowledge. In this journal, the 2007 article by Pace and Stylianou[5] has helped share developments with anesthesiologists, introducing novel up-and-down design variants and target-dose estimation approaches. That article has played an important role, because up-and-down designs

suffer from lack of accessible reference material written by methodologic experts.

Since 2007, substantial additional progress has been made but has yet to be shared effectively beyond the methodologic literature.[9–11] This Readers' Toolbox aims to close the information gap and also to provide a broader and generally accessible intuitive understanding of up-and-down design basics. We believe that such an understanding will promote better up-and-down design implementation practices and broader adoption. We begin by describing the basic principles and properties of up-and-down designs. Next we address target-dose estimation methods, followed by practical design recommendations. The discussion section reflects back on the article's main points and explores the conceptual and practical differences between up-and-down designs and other dose-finding approaches. The Supplemental Digital Content (http://links.lww.com/ALN/C867) contains complementary supporting information, in particular about estimation methods and software.

Up-and-down Basics

Principles and Behavior

Up-and-down designs share five simple elements:

1. The response is simplified to a binary endpoint (*e.g.*, analgesia effective/ineffective).
2. Potential treatments must be ordered as a discrete set of increasing doses of the same treatment or drug. Preferably, the allowed doses are uniformly spaced in an algebraic or geometric sequence.
3. For up-and-down designs described here, the probability of positive response must maintain the same direction of change (increasing or decreasing) with increasing dose. For notational simplicity, we assume that it is increasing and denote the relationship between dose and positive-response probabilities by the function F(x), where x is the dose-magnitude variable (fig. 1).
4. Doses are allocated to patients sequentially and only allow for increasing the dose by one level, decreasing by one level, or repeating the same dose. Hence, the design's name "up-and-down," or (in sensory studies and materials testing) the "Staircase Method."[12–14]
5. The dose-transition rules are based on the doses and responses of the last patient or several patients rather than on all patient data going back to the beginning of the experiment. Furthermore, the rules do not use any estimated quantity that changes during the study.

These elements make up-and-down design dose-allocation sequences, or trajectories, part of a family known as random walks.[15] Unlike simple random walks for which doses are equally likely to increase or decrease, with up-and-down design, these probabilities depend on the dose. When the current dose is above the target, a dose decrease is more likely than an increase—and more extremely so the further
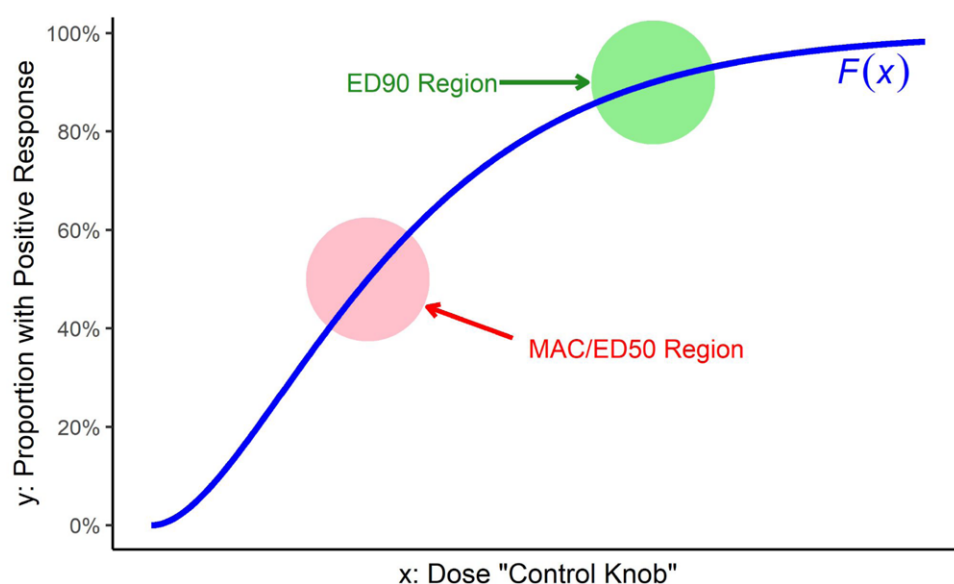
Fig. 1. Illustrating the context of dose-finding designs such as up-and-down. Observations are dichotomized to positive/negative, and the probability of positive response increases with increasing dose. Estimating the entire dose–response curve F(x) requires numerous observations throughout the dose range, including doses that might be too low or too high for patient needs. Dose-finding designs concentrate most observations in a region of interest; illustrated are two often-targeted regions: the ED50 or minimum alveolar concentration (MAC) and the ED90.

one ventures above the target. The opposite takes place below the target, resulting in a target-centered random walk. During up-and-down designs' early years, method developers experimented with various intuitive designs, some producing target-centered random walks and others not.[16] In general, the former type has enjoyed broader and longer-lasting adoption by practitioners. In the 1990s, Durham and Flournoy[7,17] discovered the criteria that govern whether a dose-finding design would generate a target-centered random walk. In this article, we reserve the term "up-and-down" only for designs meeting these criteria, the accepted terminology nowadays among up-and-down design experts. However, there is some confusion in the literature regarding what constitutes an up-and-down design, and the reader may encounter the term "up-and-down" used less carefully elsewhere.

We begin our description of up-and-down designs from the original 1940s design, which targets the ED50. While anesthesiologists often seek a dose that would be effective most of the time, the ED50 can be estimated most quickly and reliably, is often used in anesthesiology as a benchmark quantity (*e.g.*, the MAC), and is particularly useful for comparing the required dosage between different patient groups or different drugs.[18–21]

We use simulated data to illustrate the interplay between occasional idiosyncrasies of individual up-and-down design random walks and the guaranteed nature of their overall behavior. Simulations play a central role in dose-finding research and study design, because no closed-form formulae exist for design attributes such as sample size, power, or expected estimation error. Figure 2A displays data from an ED50-finding up-and-down design simulation. Shown are two virtual up-and-down design experiments with sample size n = 30, assuming the same F(x) (in real life, F(x) remains unknown to the researchers). The two experiments differ only by the random draw of patient responses. Black and white squares represent positive and negative responses, respectively. The target dose, known to us because this is a simulation, is marked with a horizontal red line. Even though the two random walks bear some similarity and both start from dose level 6, they are different. In the top panel, the most commonly allocated dose level is 8, while in the bottom, it is level 7. The randomness of individual experiments notwithstanding, using the same F(x) and the dose-transition rules, we can calculate exactly, based on up-and-down design theory, the *average* number of patients expected to be treated at each dose, were we to run a large collection or *ensemble* of n = 30 experiments (fig. 2B). On average, the most commonly allocated dose level is 7, which is also closest to target. It is followed by levels 8 and 6, respectively; nearly 24 of the first 30 patients are expected to receive one of these three doses, as opposed to only about 8 patients under a traditional design that splits the 30 patients evenly between the 11 available dose levels.
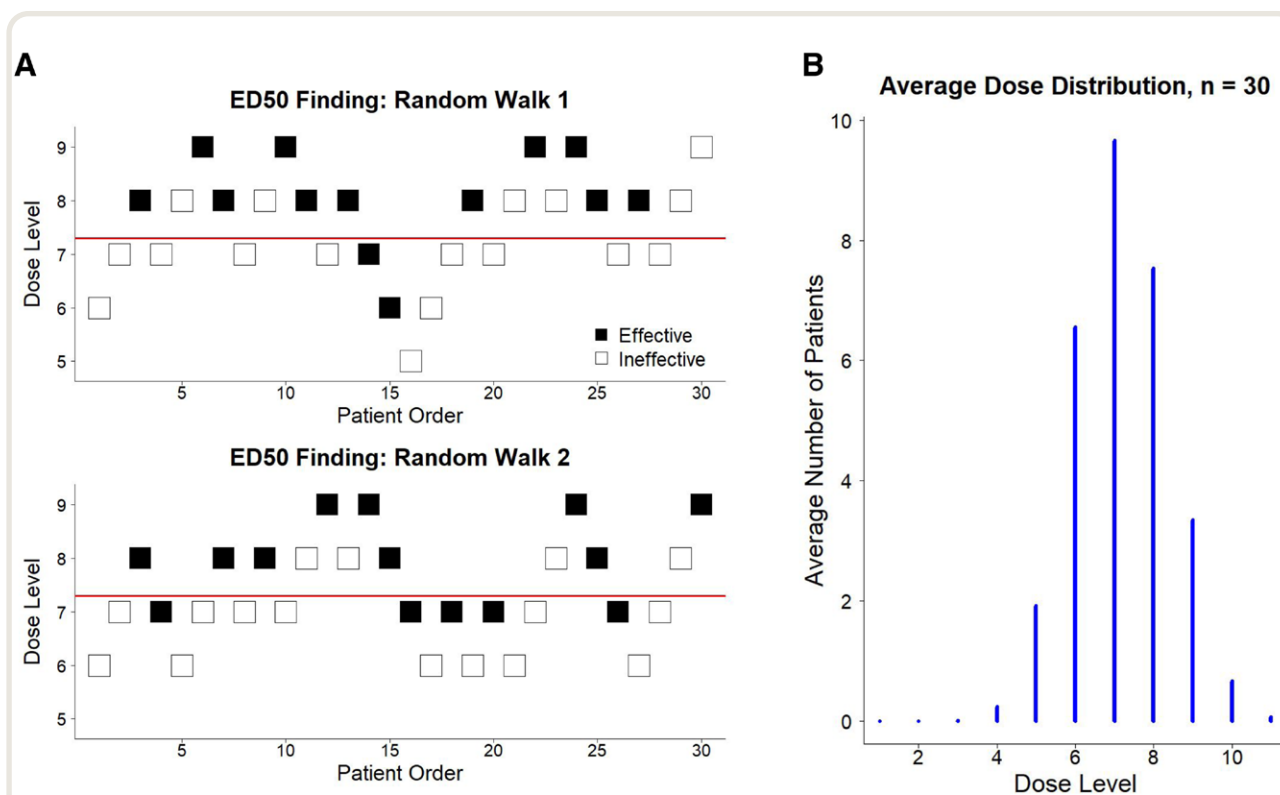
Fig. 2. (*A*) Two simulated experimental trajectories of an ED50-finding up-and-down designs with a sample size of n = 30 and with 11 dose levels, obtained *via* different random draws from the same F(x). (*B*) Expected cumulative number of patients allocated to each dose, at a total sample size of n = 30, calculated theoretically using knowledge of F(x).

Transition Rules of Common Up-and-down Design Variants

Under this ED50-targeting classical up-and-down design (fig. 2), dose assignments:

- Increase after a negative response.
- Decrease after a positive response.

A note about dose boundaries: usually, doses are not allowed to increase or decrease without bound, whether due to physical (*e.g.*, a dose of 0), logistical, or ethical restrictions. Whenever transitioning outside a boundary would be mandated, it is standard to repeat the boundary dose. This is the boundary rule we assume in the article.

Up-and-down designs have also been developed for other targets. The most common one in practice is the k-in-a-row design,[16,22,23] owing its popularity to sensory studies, in which it is known as the "fixed-staircase method" or "transformed up-and-down."[24] To target a highly effective dose with k-in-a-row design, dose assignments:

- Increase after a negative response.
- Decrease upon a positive response, but only after observing *k* consecutive positive responses at the same dose.
- Otherwise, remain the same dose.

Using *k* = 2 will bring the random walk near the 70th percentile, and *k* = 3 will center it just shy of the 80th percentile. To target the ED90, we recommend using *k* = 6, while targeting the ED95 would require *k* = 13.

Figure 3A shows two randomly simulated n = 50 experiments with *k* = 6. The trajectories are very asymmetric, exhibiting rapid dose escalations and slow descents. Nevertheless, the average dose allocations for n = 50 experiments (fig. 3B) are only mildly asymmetric around the target. Roughly half the doses allocated to the first 50 patients are expected to be at dose levels 4 and 5, immediately adjacent to the target.

Another nonmedian up-and-down design, the biased-coin up-and-down design[7,8,17] is currently popular in anesthesiology,[25–29] possibly owing to its introduction by Pace and Stylianou. Under the biased-coin up-and-down design:

- Increase the dose after a negative response.
- Upon a positive response, "toss a biased coin" (draw a random number) and then either:

  - Decrease the dose with probability inverse to the odds of positive response at the target.
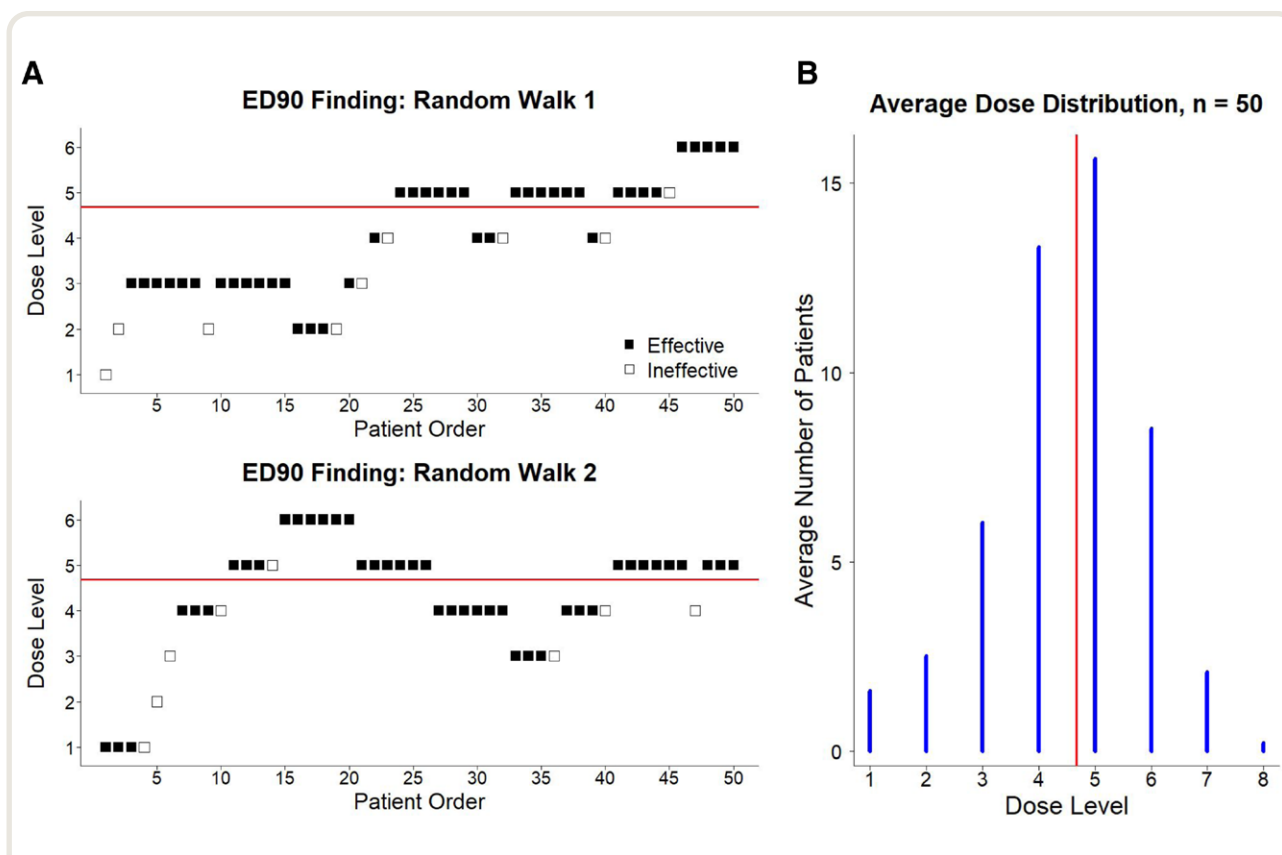 - Otherwise, remain the same dose.

Fig. 3. (*A*) Two simulated experimental trajectories of an ED90-finding k-in-a-row design ($k = 6$) with a sample size of n = 50 and with 8 dose levels, obtained *via* different random draws from the same F(x). (*B*) Expected cumulative number of patients allocated to each dose, at total sample size n = 50, calculated theoretically using knowledge of F(x).

We illustrate the term "odds" used in the rules above *via* an example. At the ED90, take the ratio between 90% and the remainder from 100%, *i.e.* 10%, obtaining an odds of 9. Therefore, under a biased-coin up-and-down design targeting the ED90, the probability for dose decrease after a positive response will be 1/9 (the inverse of these odds). Since the "coin" probability is so small, the random walk will gravitate toward doses with high positive response rates. Due to the randomization, during the experiment the number of consecutive positive responses before each dose decrease will vary randomly. Targeting the ED90, the average will be 9 patients, 3 more than under the analogous k-in-a-row design. A biased-coin up-and-down design targeting the ED95 would use a "coin" probability of 1/19, meaning that on average 19 consecutive positive responses will be required for each dose decrement.

With both k-in-a-row design and biased-coin up-and-down design, inverting the transition rules (mandating a decrease after every positive response, while requiring several negative responses for an increase) will cluster dose allocations around targets lower than the ED50. Another common up-and-down design that facilitates treating groups of patients simultaneously is described in the Supplemental Digital Content (http://links.lww.com/ALN/C867).

Estimating the Target Dose

As demonstrated in figures 2B and 3B, up-and-down design rules effectively concentrate doses around the target. However, the original up-and-down design estimation methods from the 1940s to the 1960s made simplifying assumptions about F(x) and did not consider the full implications of up-and-down designs' random-walk behavior. They only work well under narrow, specialized conditions. When these conditions are violated, the historical estimates perform very poorly[4,30,31] (see also the Supplemental Digital Content, http://links.lww.com/ALN/C867). The fact that such outdated up-and-down design estimation methods are still in very common use has been a concern to up-and-down design experts.[9,31]

Two decades ago, a robust alternative was developed, using a simple and standard statistical algorithm called isotonic regression.[5,31] More recently, an upgraded, more efficient version of this method was published.[10,11,32] Our discussion of UDD estimation begins with the original estimation approach, followed by the isotonic regression approach, and ending with CIs.

Dose-averaging Estimates: The Old and the New

Since up-and-down designs concentrate doses in roughly symmetrical fashion around the target (figs. 2B and 3B), method developers realized early on that averaging the sequence of allocated doses, possibly with some empirical correction, may serve as a reasonable target estimate. This is what Dixon and Mood suggested in 1948,[1] and to this day, most published up-and-down design experiments use some form of dose averaging as their estimate. The original Dixon–Mood averaging method, sometimes called Dixon–Massey, is still encountered in anesthesiology.[21] However, across the many research fields using up-and-down designs, the most popular averaging method was introduced by

Wetherill *et al.*[33] in the 1960s. It identifies and averages only the doses at reversal points (*i.e.*, a positive response observed immediately after a negative one or vice versa). In figure 4C, the first three reversals (sometimes called "crossovers"[6]) are marked with blue circles.

Dose-averaging estimates are extremely simple and, when used appropriately under the right conditions, can also be the most efficient option. Unfortunately, the symmetry illustrated in figures 2B and 3B can be broken if the starting dose is far removed from the target or if the target is close to a boundary, rendering the dose average far less useful (see also Supplemental Digital Content, fig. S6, http://links.lww.com/ALN/C867).
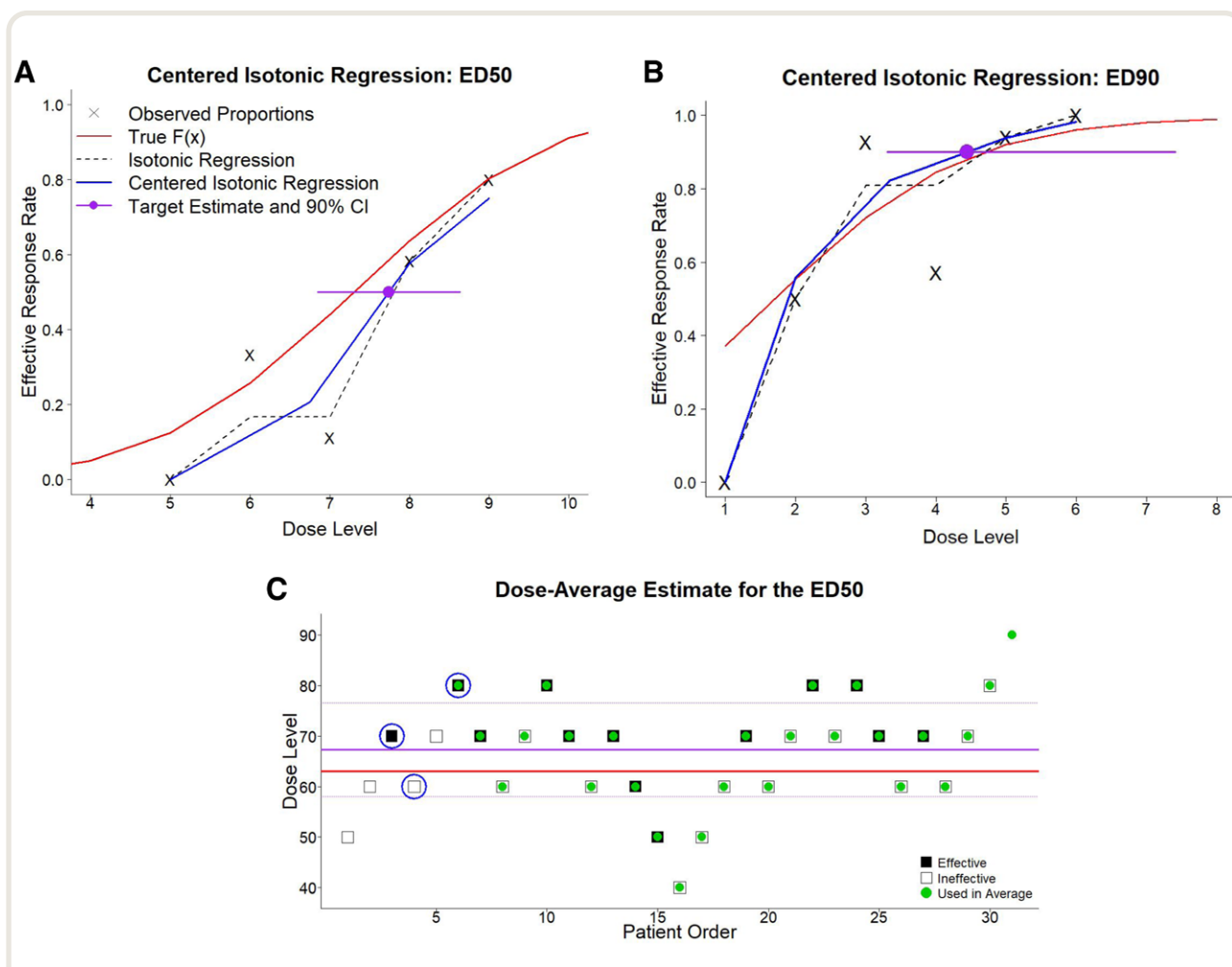
Fig. 4. (*A*) Dose–response plot and target estimates for the first simulated experimental trajectory from figure 2A. Shown are the observed response rates versus dose levels (× marks), the true F(x) used to simulate the data (*red curve*), the estimated isotonic regression (*black dashed line*) and centered isotonic regression (*blue line*) curves, and the centered isotonic regression ED50 estimate with its 90% CI (*purple dot and line*). (*B*) Similar, but for the first trajectory from figure 3A, and the centered isotonic regression estimate is for the ED90. (*C*) The first trajectory from figure 2A replotted to illustrate dose-averaging estimation. The first three reversal points are circled in *blue*. The observations participating in our recommended dose-averaging estimate are marked in *green*. Note that this includes a "virtual" 31st dose allocation, determined by the 30th patient's response. The *horizontal lines* show the true ED50 (*red*), the averaging estimate (*purple*), and the 90% CI (*dotted purple*).

For this and other reasons, we generally do not recommend dose-averaging up-and-down design estimates. If one wishes to use them, then they should be restricted only to ED50 finding and secondary to the main estimate, which should be centered isotonic regression. A safe dose-averaging method (relatively speaking) would be to average *all* doses starting from the third reversal, rather than the Dixon–Mood formula or Wetherill's reversal-only averages. The Supplemental Digital Content (http://links.lww.com/ALN/C867) provides additional information on the properties and limitations of dose-averaging estimates, including R code and comparisons with the isotonic regression methods we describe now.

Centered Isotonic Regression and Other Regression Estimates

Up-and-down experiments generate binary (positive/negative) response data. We can calculate and plot the proportion of positive responses at each dose on a dose–response plot. Figure 4 (A and B) shows dose–response plots for the first random walk in each of the example pairs from figures 2 and 3. In other words, the simulated experiments in figure 4 (A and B) target the ED50 and ED90, respectively. From the dose–response observation pairs (× marks), regression methods are used to estimate the dose–response curve. The point where this curve crosses the target response probability marks the target dose estimate (purple points).

While there are many regression methods in use, given up-and-down design's typically modest amount of data concentrated at a few doses, the most viable general-purpose option is isotonic regression, a standard nonparametric method that assumes only that F(x) is nondecreasing, making no further assumptions about its shape. However, isotonic regression (dashed black curves in the top panels of fig. 4) has a practical disadvantage: the curve it generates tends to have "flat" constant intervals, which are unrealistic in most contexts, and reduce estimation precision. Oron and Flournoy[10] developed a simple modification of the algorithm that eliminates most flat intervals. This centered isotonic regression estimate was shown to incur smaller estimation errors than the original isotonic regression, and is publicly available together with a confidence-interval method in a package named "cir" for the open–source statistical programming language R (see Supplemental Digital Content, http://links.lww.com/ALN/C867).[32] The purple circles and horizontal segments in figure 4 (A and B) indicate the centered isotonic regression point estimates and 90% CI for each target.

Many researchers have used parametric regression methods (most commonly logistic or probit) for up-and-down design estimation, a practice that antedates isotonic regression's introduction to the field and is still encountered in publications.[21] Parametric regression makes more specific assumptions about the shape of F(x). We recommend strongly *against* using parametric regression on up-and-down design data: the data are usually too limited and sparse to properly evaluate the underlying assumptions about F(x). In addition, parametric estimates can become nonsensical or nonexistent, far more often than isotonic estimates.[34]

A note of caution regarding the use of any regression in dose-finding: regressions assume that the observed proportions plotted on the dose–response plane are unbiased estimates of the values of F(x) at the assigned doses. However, recently Flournoy and Oron[11] showed that all adaptive dose-finding designs, including up-and-down design, induce some bias on observed response proportions. This bias is minimal near the target and therefore has little effect upon centered isotonic regression and isotonic regression target estimates. In addition, our R package implementing centered isotonic regression offers an empirical bias correction.

Because this bias increases as the dose gets further from target, it is ill advised to estimate percentiles far from the designated target (*e.g.*, estimating the ED95 using ED50-finding up-and-down design data, a practice that has become prevalent in recent anesthesiology studies).[28,35–40] Such estimates are likely biased in the direction of the ED50 (*i.e.*, downward). Furthermore, an ED50-centered up-and-down design would collect few observations near the ends of the dose–response curve, and therefore estimates in that region would rely upon very little direct data. The problematic nature and likely bias of estimating extreme percentiles from ED50-finding up-and-down design data has been noted previously by some commentators and researchers in anesthesiology.[40–42] For practical alternatives, see the section about estimating two distinct target doses.

Challenge of Confidence Intervals

CIs are essential to research: point estimates always have a degree of uncertainty, which tends to be substantial with small samples. Conveying the amount of uncertainty promotes responsible decision-making in the interpretation and incorporation of study data.

A property of CIs rarely discussed outside of statistical literature is interval coverage: in brief, whether the CI performs "as advertised." To examine coverage, statisticians simulate a large ensemble of experiments under known conditions and calculate the CI for each. The coverage is the proportion of simulated experiments in which the CI indeed contains the true value. A CI with substantially deficient coverage is misleadingly optimistic about the true amount of information provided by the study and about the location of the target itself. Conversely, excessive coverage, also known as conservative CIs, provide too little information about target location.

The CIs in most published up-and-down design experiments tend to be deficient in coverage. In fairness, producing a CI with the correct coverage is rather challenging

for dose-finding in general. Therefore, our first recommendation regarding up-and-down design CIs is to provide 90% rather than 95% intervals. Unless the sample size is far greater than those typically used in up-and-down designs, 90% is probably the highest level of confidence that the experiment can promise while remaining both reliable and informative. Furthermore, per our investigations, bootstrap-based CIs as recommended 15 yr ago by Pace and Stylianou[5] tend to have insufficient coverage regardless of target. Instead, the CIs implemented in the "cir" package are analytically derived.[10,11] These centered isotonic regression CIs can guarantee sufficient coverage up to the ED90 at typical up-and-down design sample sizes. However, for ED95 with n = 50, the 90% CIs we have simulated achieve less than 80% coverage.

CIs for dose-averaging estimates are even more challenging and problematic. Due to space limitations, we do not discuss them here. A relatively sound method to calculate them for ED50 finding is presented in the Supplemental Digital Content (http://links.lww.com/ALN/C867), together with R code for both point and interval dose-averaging estimates.

Practical Design Considerations

Despite the lack of simple, "off-the-shelf" sample size calculation formulae and the widespread reliance upon computer simulations to investigate design properties, for most run-of-the-mill up-and-down design experiments, investigators should be able to yield satisfactory results by following the general guidelines summarized in box 2. We explore here a few topics worthy of more detailed discussion. Information regarding other recommendations is in the Supplemental Digital Content (http://links.lww.com/ALN/C867).

Midexperiment Design Changes

Dose-finding study designers often face a long wishlist of desired properties and a short sample size. To meet all these demands, it is tempting to have the experiment "learn" quickly and change its own rules. Such attempts include halving the dose spacing midexperiment to more precisely center the random walk, or stopping as soon as enough information appears to have been collected rather than after a fixed sample size.

The Achilles' heel of these innovations is that the uncertainty early in the experiment is too great for making such decisions effectively. This is a matter of quantity rather than principle: after 40 to 50 patients, there should be sufficient information regarding the approximate location of the target to guide a beneficial placement of the next dose upon halving the step size. However, this is more than the entire sample size of many up-and-down design experiments. This perspective should also be applied to the practice of stopping after a fixed number of reversals rather than

a fixed sample size.[6] The Supplemental Digital Content (http://links.lww.com/ALN/C867) discusses one exception to this guideline: a short quick-start stage that might help in certain situations (see also box 2, design recommendation 10).

Choosing the Sample Size

We note that up-and-down is not strictly a small-sample design. To obtain a very narrow and accurate 95% CI for the target dose will require n > 100. Even with that many participants, as long as the main goal is estimating the target dose rather than the entire dose–response curve, up-and-down designs should deliver better precision than nonadaptive designs and are therefore still preferable. If one is able to afford such larger samples, then splitting the sample into stages whose sizes are more typical of published up-and-down design studies (n = 30 to 60) allows for design-change decisions such as changing the step size and the boundaries in a way that improves overall performance. As long as the target, treatment, and response evaluation protocols remain the same, one can pool data from all stages at the experiment's end into a single centered isotonic regression estimate.

That said, up-and-down design in practice remains predominantly a small-sample design, mostly due to logistical constraints and researcher expectations. Our general recommendation of using at least n = 30 to 40 for ED50 finding and n = 50 to 60 for the ED90 is a compromise indicating the minimal sample sizes that would still provide a useful estimate, albeit not a very precise one. Since mishaps and inconclusive experimental trajectories cannot be avoided, consider adding to the study protocol a list of conditions that would mandate a sample-size increase with possible design modification. This may include a relatively low patient count at the most visited dose or triggering the dose boundary rules more than once.

Keep in mind that the choice of target, sample size, dose levels, and starting dose are all interrelated. For example, if the experiment's safety board requires starting at one end of the dose range, then sample-size considerations should allow ample contingency for a scenario in which the target is near the other end. Otherwise, the entire experiment may be spent traversing the dose range.[44]

Another, poignant example for this interdependence is the sharp increase in sample size requirements when shifting the target from ED90 to ED95. As mentioned when presenting k-in-a-row design and biased-coin up-and-down design, each dose decrease with both designs requires more than twice as many positive responses when targeting the ED95 as when targeting the ED90. We also reported the lack of sufficient CI coverage for ED95 estimates after n = 50, a sample size generally sufficient for ED90 CI coverage. Thus, we recommend considering carefully whether an ED95 estimate is indeed required or whether the ED90 would suffice. If the former is still

Box 2. Basic Guidelines for Up-and-Down Design and Estimation

Stronger recommendations are in boldface. For guidelines marked with an asterisk, most of the supporting information is provided in the Supplemental Digital Content (http://links.lww.com/ALN/C867).

Design choices:

1. *Do not impose dose boundaries, unless mandated by physical, feasibility, or safety constraints.
2. **Use a fixed prespecified sample size rather than a random stopping rule.** Note: "counting" the sample using reversals is also a random stopping rule.
3. **Do not target percentiles more extreme than the 10th to 90th, unless using a substantially larger sample size (n > 100).**
4. Typical sample sizes for a moderate-confidence estimate are n = 30 (ED50) to 60 (ED90). **The more extreme the target percentile, the larger the recommended sample size.** If you can afford a larger sample than that, split it up by prespecifying design reevaluation points (dose-spacing changes, boundary modification, and so forth) every 30 to 60 patients.
5. Consider adding to the study protocol, conditions under which the sample size would automatically increase with potential design modifications. See text for details.
6. **Dose spacing: employ a dose range based on known clinical effectiveness and split it into 8 to 12 dose levels.** See article text for specific guidance.
7. Starting dose: barring safety constraints, for ED50 finding, start near the presumed target. For high percentiles, start about 1 to 2 dose levels *below* the presumed target, and *vice versa*.
8. *When a k-in-a-row up-and-down design is suitable, it is more efficient than the biased-coin design; however, both are viable.
9. *For experiments to compare target doses between two different groups, designing two separate up-and-down designs and testing for an overlap of 83% CIs is acceptable.
10. ***Do not incorporate adaptive rules that change design parameters after a few patients.** The only exception is for nonmedian targets, for which an initial ED50-targeting stage may help make the trial more efficient in some cases.

Estimation choices:

1. ***Use centered isotonic regression.** Include all data in the estimate and report a CI.
2. Do not use data from a dose-finding experiment to make off-target estimates (*e.g.*, using ED50-finding up-and-down design to estimate the ED95).
3. Do not use parametric regression for target-dose estimates.
4. ***Dose-averaging estimates are germane only for ED50-finding experiments. Do not use Dixon and Mood's original method or reversals-only estimates.** The dose-averaging estimate we recommend uses all doses starting from the third reversal.
5. Unless using larger sample sizes (n > 100), do not report CIs at the 95% confidence level; use 90% or less.

---

desired, then absent strong prior knowledge regarding the ED95's approximate location, one should be willing to plan for n > 100 with up-and-down and even more with a nonadaptive design.

Dose Spacing and Boundaries

A dose boundary near the target dose disrupts the allocation symmetry of figures 2B and 3B, inducing large biases on dose-averaging estimates and potentially preventing centered isotonic regression from obtaining a target estimate at all. Therefore, we recommend against setting any dose boundaries narrower than those dictated by physical and safety constraints. After the experiment's data are collected, if the target appears to be very close to a boundary or the target estimate sits on the edge of the range of doses administered, the reliability of the estimate may be questionable. The appropriate resolution might be, if applicable, to extend the boundary, or halve the step size to allow for more symmetry—or simply to increase the sample size. Exact remedies would vary by study context, and therefore we cannot recommend a single course of action.

The closely related question of step size, also known as dose spacing, has been addressed as early as the work of Dixon and Mood in 1948. Their recommendation, which still holds sway over many fields, was a spacing of approximately 1 SD of the response-threshold distribution, which they assumed to be Normal. Here, we offer an alternative approach, informed by more recent understanding of up-and-down design random walk and reflecting the limited knowledge researchers possess about F(x) before the study.

We begin by noting that even in the absence of "hard" dose boundaries in the study rules, up-and-down design experiments face "soft" boundaries generated by F(x) and by the dose-transition rules. Using classical up-and-down as an example, if the experiment reaches doses for which F(x) is very close to 0 or 1, it is unlikely to reach them often and even less likely to transition beyond them. Therefore, for the classical up-and-down design, consider the range of F(x) from 0.02 to 0.98 to define the effective dose range, containing the dose levels that have reasonable probability of being used during the experiment. Under the Normal distribution (mentioned for illustrative purposes rather than

---

Box 3. Where to Find More Information on This Topic

There is very little up-to-date, published reference information about up-and-down designs. In particular, there is yet not a single book dedicated to up-and-down designs, neither for method developers nor for practitioners. **Some relatively recent accessible reviews:**

Görges M, Zhou G, Brant R, Ansermino JM: Sequential allocation trial design in anesthesia: An introduction to methods, modeling, and clinical applications. Paediatr Anaesth 2017; 27:240–7

Notes: This may be the most recent review in the spirit of Pace and Stylianou.[5] It also presents a moderate-sized simulation study. Some perspectives in that review are discouraged in the current article (*e.g.*, using reversal count as a stopping rule).

Flournoy N, Oron AP: Up-and-down designs for dose-finding, Handbook of Design and Analysis of Experiments. Edited by Dean A, Morris M, Stufken J, Bingham D. London, CRC Press/Chapman Hall, 2015, pp 862–98

Notes: This is the most detailed up-and-down design reference we are aware of. It is somewhat more technical than the current article. Since writing that text, we have acquired more insights and results about up-and-down designs and have considered more carefully practical guidelines for researchers. These more recent results are shared in the current article.

Saranteas T, Finlayson RJ, Tran DQH: Dose-finding methodology for peripheral nerve blocks. Reg Anesth Pain Med 2014; 39:550–5

Notes: A clear, concise, and highly informative review of standard, classical up-and-down, biased-coin up-and-down, and continual reassessment method designs.

Pace NL, Stylianou MP: Advances in and limitations of up-and-down methodology: A précis of clinical use, study design, and dose estimation in anesthesia research. Anesthesiology 2007; 107:144–52

Notes: Most of the essential content in this seminal review is covered or referred to in the current article. However, revisiting it is still valuable.

Columb MO, D'Angelo R: Up-down studies: Responding to dosing! Int J Obstet Anesth 2006; 15:129–36

Notes: A lively debate article between a leading popularizer of up-and-down designs in anesthesiology and a colleague critical of the design. Interestingly, the latter had identified in his experimental work a point we make here: ED50-targeted up-and-down design is ill suited for estimating extreme percentiles.

Simulation studies from other fields, exemplifying the siloed nature of up-and-down design implementation paradigms and practices. In each study, the authors diligently attempt to fix the problems of a "legacy" estimation approach. Our own perspective is that these approaches have outlived their useful life and should simply be replaced with centered isotonic regression, or at least with a more robust dose-averaging estimate. See the Supplemental Digital Content (http://links.lww.com/ALN/C867) for more detailed comparison and discussion of estimation approaches.

Müller C, Wächter M, Masendorf R, Esderts A: Accuracy of fatigue limits estimated by the staircase method using different evaluation techniques. International Journal of Fatigue 2017; 100:296–307

Pollak R, Palazotto A, Nicholas T: A simulation-based investigation of the staircase method for fatigue strength testing. Mechanics of Materials 2006; 38:1170–81

Garcìa-Perez MA: Forced-choice staircases with fixed step sizes: Asymptotic and small-sample properties. Vision Research 1998; 38:1861–81

---

as an assumption), this effective range comprises 4 standard deviations. The tradeoff between the benefit of a symmetric well defined up-and-down design dose distribution (figs. 2B and 3B) and the risk of the experiment having to traverse too many dose levels en route to the target region suggests dividing this effective range into 8 to 12 dose levels. Therefore, our recommendation is analogous to about half of the step size suggested by Dixon and Mood.

For non–ED50 up-and-down designs, one must consider the transition rules, which cause the effective range to be narrower and also asymmetric over F(x). Specifically for k–in–a–row design with $k = 6$ to estimate the ED90, the upper end of the range would still be around F(x) = 0.98, because the dose-increase rule is identical to classical up–and–down design. However, the point with dose-decrease probability of 2% is near F(x) = 0.60. Thus, a rough

guideline for the "soft" lower boundary could be the ED50, in case information is available regarding its approximate magnitude. We reiterate that no hard boundaries need to be set to ensure desirable behavior. If your assumptions regarding the effective rate are reasonably correct, then the strong central-tendency random walk will keep nearly all dose allocations to the 8 to 12 dose levels you have chosen; usually only to a subset of them.

Estimating Two Distinct Target Doses

Many recent anesthesiology up–and–down studies have reported two target estimates after a single ED50-finding experiment: the ED50 estimate and another, "off-target" high-percentile estimate. As explained earlier, there is far less confidence in the latter estimate. In view of this, some

of these studies added a confirmatory stage enrolling n = 30 to 100 patients and treating them at the estimated higher percentile.[37,40] While some studies reported results that corroborate the off-target estimate, a better use of the additional sample might have been to follow upon the classical up-and-down design with a high-target up-and-down design, using the first stage to refine the dose spacing as suggested earlier in this section.

Discussion

There is established precedent for the use of up-and-down design in anesthesiology to improve the efficiency of experimental design. Here, we have attempted to offer further insight into this adaptive strategy that concentrates patient exposure in regions of interest along the dose–response curve F(x) (fig. 1). The design's rules generate target-centered random walk behavior. The development of asymmetric up-and-down design rules provides an ability to focus upon different regions of F(x).

We have noted caveats in the application of different methods for target-dose estimation from up-and-down design data, including appropriate and inappropriate use of dose-averaging estimates. We recommend the method of centered isotonic regression, which improves the reliability of estimates, is informed by recent research, provides a reasonable CI, and has a publicly available R package.[32] Basic application notes appear in the Supplemental Digital Content (http://links.lww.com/ALN/C867). A longer tutorial is found in the package's vignette.

Sample sizes and other design decisions can be explored beforehand using simulation to inform pragmatic planning of study conduct. We have provided simple design recommendations that may obviate the need for planning simulations in most typical applications, along with concise insights into more elaborate strategies. The Supplemental Digital Content (http://links.lww.com/ALN/C867) contains further material that may be relevant for design choice.

Various design modifications and extensions inspired by up-and-down have been developed.[12,16,44–46] In particular, escalation designs used in toxicity trials appear to have originated from up-and-down designs but do not produce a targeted random walk. Most prominent among them is the "3 + 3" escalation algorithm, to date the most popular Phase I cancer trial design and very often misrepresented as an up-and-down design.[46] Escalation designs stop the experiment after a prespecified, usually modest, amount of toxicity has been observed at a given dose. They serve an important role in approaching the general vicinity of doses at which toxicities may begin to appear, without placing participants at unacceptable risk. This is particularly adequate for first-in-human studies with healthy volunteers. However, escalation designs do not provide reliable target-dose estimates and should not be used for estimation.[47]

At the other end of the risk-taking continuum, prominent statisticians have promoted the use of sophisticated Bayesian model-based designs, in particular the continual reassessment method for Phase I cancer trials.[47–49] These designs have "long memory," in contrast to up-and-down designs, the dose-transition rules of which require only the last several subjects. Anesthesiology articles utilizing the continual reassessment method appeared at least as early as 2000, and their numbers have grown recently.[50–55] Other long-memory designs used in anesthesiology, such as the modified Narayana design, the origins of which date back to the 1950s,[56] also show similar overall behavior despite not being perceived as associated with the Bayesian approaches.[57,58]

We note that up-and-down designs, while using only recent observations for dose allocation, do use the entire experiment's data during the estimation stage. Therefore, they achieve dose-finding performance similar to leading long-memory designs.[8,34,59,60] Where the two approaches differ markedly is the dose-allocation distribution: unlike up-and-down design's target-centered random walk, long-memory designs promise to dedicate *all* allocations to the dose closest to target and to do so quickly. This often leads to detrimental experimental behavior,[59,61–63] because as written regarding midexperiment design changes, early on there is insufficient information to justify dramatic decisions. That said, long-memory designs may be warranted when the study objective is more complicated than straightforward dose-finding. This would likely also require larger sample sizes. The Supplemental Digital Content (http://links.lww.com/ALN/C867) contains further discussion of long-memory designs.

These three families—escalation designs, up-and-down designs, and long-memory designs—offer different philosophical approaches to uncertainty. Uncertainty and randomness in research are inevitable; in the domain of small-sample, binary-endpoint experiments, the uncertainty is particularly substantial. Escalation designs aim to preempt uncertainty by stopping upon the first encounter with a predefined risk level. Long-memory designs and in particular Bayesian ones promise to overcome the uncertainty with the aid of special models, zooming in swiftly and exclusively on the correct dose. The "swiftly" part of this promise is not backed by proven theory.[59] In our view, among established approaches, up-and-down designs offer the most realistic path for managing uncertainty, channeling it into a target-centered random walk with clear theoretical properties, in a manner that generally conforms with ethical and clinical expectations.

Overall, up-and-down design remains an established yet intriguing experimental design approach. Up-and-down design are competitive with other significantly more complicated alternatives, and their simplicity and transparent behavior are popular features that keep them in vogue. There are, however, caveats in their application that may confound the unwary, and the new user is strongly advised to seek statistical help to guide their use within study designs.

Acknowledgments

The authors thank the reviewers whose meticulous comments and willingness to discuss our inquiries directly have helped make this article substantially better. Special thanks go to Dustin R. Long, M.D. (Division of Critical Care Medicine, Department of Anesthesiology and Pain Medicine, University of Washington, Seattle, Washington), who read the revised draft and provided important insights.

Research Support

Competing Interests

The authors declare no competing interests.

Correspondence

Address correspondence to Dr. Oron: Hans Rosling Center, Guthrie Annex 3, Seattle, Washington 98105. assaf@uw.edu. ANESTHESIOLOGY's articles are made freely accessible to all readers on www.anesthesiology.org, for personal use only, 6 months from the cover date of the issue.

Supplemental Digital Content

Up-and-down Designs: Methodological Supplement, http://links.lww.com/ALN/C867.

References

1. Dixon WJ, Mood AM: A method for obtaining and analyzing sensitivity data. J Am Stat Assoc 1948; 43:109–26
2. von Békésy G: A new audiometer. Acta Otolaryngol (Stockh) 1947; 35:411–22
3. Anderson TW, McCarthy PJ, Tukey JW: "Staircase" Method of Sensitivity Testing. Princeton, Statistical Research Group, Princeton University, 1946
4. Oron AP: Up-and-Down and the Percentile-Finding Problem, Ph.D. thesis. Seattle, University of Washington, 2007. https://arxiv.org/abs/0808.3004. Accessed June 29, 2022.
5. Pace NL, Stylianou MP: Advances in and limitations of up-and-down methodology: A précis of clinical use, study design, and dose estimation in anesthesia research. ANESTHESIOLOGY 2007; 107:144–52
6. Görges M, Zhou G, Brant R, Ansermino JM: Sequential allocation trial design in anesthesia: An introduction to methods, modeling, and clinical applications. Paediatr Anaesth 2017; 27:240–7
7. Durham SD, Flournoy N: Random walks for quantile estimation. Statistical Decision Theory and Related Topics V (West Lafayette, IN, 1992). Berlin, Springer, 1994, pp 467–76
8. Durham SD, Flournoy N, Rosenberger WF: A random walk rule for phase I clinical trials. Biometrics 1997; 53:745–60
9. Flournoy N, Oron AP: Up-and-down designs for dose-finding, Handbook of Design and Analysis of Experiments. Edited by Bingham D, Dean AM, Morris M, Stufken J. London, CRC Press/Chapman Hall, 2015, pp 862–98
10. Oron AP, Flournoy N: Centered isotonic regression: Point and interval estimation for dose–response studies. Stat Biopharm Res 2017; 9:258–67
11. Flournoy N, Oron AP: Bias induced by adaptive dose-finding designs. J Appl Stat 2020; 47:2431–42
12. García-Pérez MA: Forced-choice staircases with fixed step sizes: Asymptotic and small-sample properties. Vision Res 1998; 38:1861–81
13. Pollak R, Palazotto A, Nicholas T: A simulation-based investigation of the staircase method for fatigue strength testing. Mech Mater 2006; 38:1170–81
14. Müller C, Wächter M, Masendorf R, Esderts A: Accuracy of fatigue limits estimated by the staircase method using different evaluation techniques. Int J Fatigue 2017; 100:296–307.
15. Hughes BD: Random Walks and Random Environments, Vol. 1. Oxford, Clarendon Press/Oxford University Press, 1995.
16. Wetherill GB: Sequential estimation of quantal response curves. J R Stat Soc Ser B 1963; 25:1–48
17. Durham SD, Flournoy N: Up-and-down designs: I. Stationary treatment distributions, Adaptive Designs, Vol. 25. Edited by Flournoy N, Rosenberger WF. Hayward, CA, Institute of Mathematical Statistics, 1995, pp 139–57
18. Capogna G, Parpaglioni R, Lyons G, Columb M, Celleno D: Minimum analgesic dose of epidural sufentanil for first-stage labor analgesia: A comparison between spontaneous and prostaglandin-induced labors in nulliparous women. ANESTHESIOLOGY 2001; 94:740–4
19. 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
20. Camorcia M, Capogna G, Lyons G, Columb M: Epidural test dose with levobupivacaine and ropivacaine: Determination of ED_{50} motor block after spinal administration. Br J Anaesth 2004; 92:850–3
21. Camorcia M, Capogna G, Columb MO: Effect of sex and pregnancy on the potency of intrathecal bupivacaine: Determination of ED_{50} for motor block with the up–down sequential allocation method. Eur J Anaesthesiol 2011; 28:240–4
22. Gezmu M: The Geometric Up-and-Down Design for Allocating Dosage Levels, Ph.D. Thesis. Washington, D.C., American University, 1996

23. Oron AP, Hoff PD: The k-in-a-row up-and-down design, revisited. Stat Med 2009; 28:1805–20

24. Levitt H: Transformed up-down methods in psychoacoustics. J Acoust Soc Am 1971; 49:467–77

25. Dosani M, McCormack J, Reimer E, Brant R, Dumont G, Lim J, Mark Ansermino J: Slower administration of propofol preserves adequate respiration in children. Paediatr Anaesth 2010; 20:1001–8

26. Sotthisopha T, Elgueta MF, Samerchua A, Leurcharusmee P, Tiyaprasertkul W, Gordon A, Finlayson RJ, Tran DQ: Minimum effective volume of lidocaine for ultrasound-guided costoclavicular block. Reg Anesth Pain Med 2017; 42:571–4

27. Bittencourt R, Arzola C, Zakus P, Downey K, Ye XY, Carvalho JCA: A biased coin up-and-down sequential allocation trial to determine the optimum programmed intermittent epidural bolus time interval between 5 mL boluses of bupivacaine 0.125% with fentanyl 2 μg·mL^{-1}. Can J Anesth Can Anesth 2019; 66:1075–81

28. Kwon JH, Shin YH, Gil NS, Park J, Chung YJ, Hahm TS, Jeong JS: Effect-site concentration of remifentanil for smooth emergence from sevoflurane anesthesia in patients undergoing endovascular neurointervention. PLoS One 2019; 14:e0218074

29. Fang G, Wan L, Mei W, Yu HH, Luo AL: The minimum effective concentration (MEC90) of ropivacaine for ultrasound-guided supraclavicular brachial plexus block. Anaesthesia 2016; 71:700–5

30. Kershaw CD: A comparison of the estimators of the ED50 in up-and-down experiments. J Stat Comput Simul 1987; 27:175–84

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

32. Oron AP: cir: Centered isotonic regression and dose–response utilities, 2021. https://CRAN.R-project.org/package=cir. Accessed February 10, 2022

33. Wetherill GB, Chen H, Vasudeva RB: Sequential estimation of quantal response curves: A new method of estimation. Biometrika 1966; 53:439–54

34. Flournoy N, Moler J, Plo F: Performance measures in dose-finding experiments. Int Stat Rev 2020; 88:728–51

35. Choi EM, Park WK, Choi SH, Soh S, Lee JR: Smooth emergence in men undergoing nasal surgery: The effect site concentration of remifentanil for preventing cough after sevoflurane-balanced anaesthesia. Acta Anaesthesiol Scand 2012; 56:498–503

36. Chen L, Zhou Y, Cai Y, Bao N, Xu X, Shi B: The ED95 of nalbuphine in outpatient-induced abortion compared to equivalent sufentanil. Basic Clin Pharmacol Toxicol 2018; 123:202–6

37. Zhu X, Chen L, Zheng S, Pan L: Comparison of ED95 of butorphanol and sufentanil for gastrointestinal endoscopy sedation: A randomized controlled trial. BMC Anesthesiol 2020; 20:101

38. Kewlani A, Bhatia N, Makkar JK, Kumar V: Median effective volume of 0.5% ropivacaine for ultrasound-guided costoclavicular block. ANESTHESIOLOGY 2021; 134:617–25

39. Johnston DF, Sondekoppam RV, Giffin R, Litchfield R, Ganapathy S: Determination of ED50 and ED95 of 0.5% ropivacaine in adductor canal block to produce quadriceps weakness: A dose-finding study. Reg Anesth Pain Med 2017; 42:731–6

40. Guntz E, Latrech B, Tsiberidis C, Gouwy J, Kapessidou Y: ED50 and ED90 of intrathecal hyperbaric 2% prilocaine in ambulatory knee arthroscopy. Can J Anaesth 2014; 61:801–7

41. Columb MO, D'Angelo R: Up-down studies: Responding to dosing! Int J Obstet Anesth 2006; 15:129–36

42. Saranteas T, Finlayson RJ, Tran DQ: Dose-finding methodology for peripheral nerve blocks. Reg Anesth Pain Med 2014; 39:550–5

43. Olutoye OA, Yu X, Govindan K, Tjia IM, East DL, Spearman R, Garcia PJ, Coulter-Nava C, Needham J, Abrams S, Kozinetz CA, Andropoulos DB, Watcha MF: The effect of obesity on the ED$_{95}$ of propofol for loss of consciousness in children and adolescents. Anesth Analg 2012; 115:147–53

44. Ivanova A, Montazer-Haghighi A, Mohanty SG, Durham SD: Improved up-and-down designs for phase I trials. Stat Med 2003; 22:69–82

45. Ivanova A, Flournoy N, Chung Y: Cumulative cohort design for dose-finding. J Stat Plan Inference 2007; 137:2316–27

46. Ivanova A: Escalation, group and A + B designs for dose-finding trials. Stat Med 2006; 25:3668–78

47. Rogatko A, Schoeneck D, Jonas W, Tighiouart M, Khuri FR, Porter A: Translation of innovative designs into phase I trials. J Clin Oncol 2007; 25:4982–6

48. O'Quigley J, Pepe M, Fisher L: Continual reassessment method: A practical design for Phase I clinical trials in cancer. Biom J Biom Soc 1990; 46:33–48

49. Cheung YK: Dose Finding by the Continual Reassessment Method. London, Chapman and Hall/CRC, 2011

50. Dougherty TB, Porche VH, Thall PF: Maximum tolerated dose of nalmefene in patients receiving epidural fentanyl and dilute bupivacaine for postoperative analgesia. ANESTHESIOLOGY 2000; 92:1010–6

51. Beloeil H, Eurin M, Thévenin A, Benhamou D, Mazoit JX: Effective dose of nefopam in 80% of patients

(ED80): A study using the continual reassessment method. Br J Clin Pharmacol 2007; 64:686–93

52. Zeidan A, Mazoit JX: Minimal alveolar concentration of sevoflurane for maintaining bispectral index below 50 in morbidly obese patients. Acta Anaesthesiol Scand 2013; 57:474–9

53. Gupta PK, Chevret S, Zohar S, Hopkins PM: What is the ED95 of prilocaine for femoral nerve block using ultrasound? Br J Anaesth 2013; 110:831–6

54. Liu H, Huang Y, Diao M, Li H, Ma Y, Lin X, Zhou J: Determination of the 90% effective dose (ED90) of phenylephrine for hypotension during elective cesarean delivery using a continual reassessment method. Eur J Obstet Gynecol Reprod Biol 2015; 194:136–40

55. Le Gouez A, Bonnet MP, Leclerc T, Mazoit JX, Benhamou D, Mercier FJ: Effective concentration of levobupivacaine and ropivacaine in 80% of patients receiving epidural analgesia (EC80) in the first stage of labour: A study using the continual reassessment method. Anaesth Crit Care Pain Med 2018; 37:429–34

56. Narayana TV: Sequential Procedures in the Probit Analysis, Ph.D. thesis. Chapel Hill, NC, University of North Carolina, 1953

57. Tanaka M, Balki M, Parkes RK, Carvalho JC: ED95 of phenylephrine to prevent spinal-induced hypotension and/or nausea at elective cesarean delivery. Int J Obstet Anesth 2009; 18:125–30

58. Mohta M, Dubey M, Malhotra RK, Tyagi A: Comparison of the potency of phenylephrine and norepinephrine bolus doses used to treat post–spinal hypotension during elective caesarean section. Int J Obstet Anesth 2019; 38:25–31

59. Oron AP, Hoff PD: Small-sample behavior of novel phase I cancer trial designs. Clin Trials 2013; 10:63–80

60. Oron AP, Hoff PD: Small-sample behavior of novel phase I designs: Rejoinder. Clin Trials 2013; 10:88–92

61. Mathew P, Thall PF, Jones D, Perez C, Bucana C, Troncoso P, Kim SJ, Fidler IJ, Logothetis C: Platelet-derived growth factor receptor inhibitor imatinib mesylate and docetaxel: A modular phase I trial in androgen-independent prostate cancer. J Clin Oncol 2004; 22:3323–9

62. Neuenschwander B, Branson M, Gsponer T: Critical aspects of the Bayesian approach to phase I cancer trials. Stat Med 2008; 27:2420–39

63. Resche-Rigon M, Zohar S, Chevret S: Adaptive designs for dose-finding in non-cancer phase II trials: Influence of early unexpected outcomes. Clin Trials 2008; 5:595–606