Adaptive Designs
IMS Lecture Notes - Monograph Series (1995) Volume 25

UP-AND-DOWN DESIGNS II: EXACT TREATMENT MOMENTS

BY STEPHEN D. DURHAM*, NANCY FLOURNOY AND ALI A. MONTAZER-HAGHIGHI

The University of South Carolina, The American University and Benedict College

Abstract

Consider a sequence of experiments in which a treatment is applied at a finite number of levels or dosages, and the cumulative number of responses at each level is observed after each trial. We consider such experiments in which the primary objective is to estimate the unknown dose μ that has a probability of response equal to a fixed value Γ, $0 \leq \Gamma \leq 1$. We restrict the unknown distribution of treatments so as to avoid treatment levels that are associated with high probabilities of response. When treatment levels are sequentially assigned to subjects in a way that forms a random walk, we give the exact expectation (and variance) of giving treatments that have high probabilities of response.

1. Introduction. Consider a sequence of experiments in which a treatment is applied at a finite number of levels or dosages, and the cumulative number of responses at each level is observed after each

Received November 1992; revised July 1994.

*Work supported in part by NSF Grant EPSCoR#OSR-9108772 and by the Westinghouse Savannah River Company, through the South Carolina University Research and Education Foundation.

AMS 1991 subject classification. Primary 62L15; secondary 62L05.

Key words and phrases. Dose-response, quantile targeting, random walks, toxicity study.

trial. We consider such experiments in which the primary objective is to estimate the unknown dose μ that has a probability of response equal to a fixed value Γ, $0 \leq \Gamma \leq 1$, we call Γ the *target probability of response,* and we call μ the *target quantile.* We restrict the unknown distribution of treatments so as to avoid treatment levels that are associated with high probabilities of response. Without loss of generality, we focus on target quantiles that are below the median, i.e. $0 < \Gamma \leq 0.5$; by symmetry, analogous results can be obtained for $0.5 \leq \Gamma < 1$. When treatment levels are sequentially assigned to subjects in a way that forms a random walk, we give the exact expectation (and variance) of giving treatments that have high probabilities of response.

Up-and-down rules are procedures that specify the treatment for the next trial to be one level higher, one level lower, or the same as the treatment selected for the current trial. Let $n = 0$ indicate the initial trial for which the treatment may be either fixed or random. Now suppose that the rule for allocating treatments to subjects produces a sequence of treatments $X(n)$, $n = 0, 1, 2, \ldots$ that forms a random walk on $\Omega_{\mathbf{x}}$, where $\Omega_{\mathbf{x}} = \{x_1, x_2, \ldots, x_K\}$ is a finite sample space of ordered treatments. Since $X(0)$ is the initial treatment, $X(n)$, $n \geq 1$, is the nth treatment after the initial trial. Let the treatment probability distribution of $X(n)$ be represented by the K dimensional vector $\mathbf{p}(n) = (p_1(n), p_2(n), \ldots, p_k(n))$, $n = 0, 1, 2, \ldots$, where $p_k(n) = P\{X(n) = x_k\}$. If the initial treatment is fixed at x_i, then $p_i(0) = 1$. Now $\mathbf{p}(n)$ depends on the initial treatment and the transition probabilities given by the treatment allocation rule. For up-and-down designs, the exact moments of the treatment distribution are given for the accumulating trials $n = 0, 1, 2, \ldots$. Then these moments are used to study the *biased coin up-and-down designs* given in Durham and Flournoy (companion manuscript in this volume).

Let $Y(n)$, $n = 0, 1, 2, \ldots$ be Bernoulli random variables with $Y(n) = 1$ indicating no response. This notation derives form thinking of the response as being toxicity, so that a *response* connotes *failure* and *no response* connotes *success.* Arbitrary outcomes and treatments may be denoted by Y and X, respectively, without explicit mention of their position in the sequence of trials. The probability of response at x is denoted by $Q(x) \equiv P\{Y = 1 \mid X = x\}$, with $P(x) \equiv 1 - Q(x)$. The *response function* $Q(x)$ is taken to be strictly increasing in x, but given x, $Q(x)$ is assumed to be constant over all trials $n = 0, 1, 2, \ldots$.

Exact expressions are obtained in Section 3 for the expected number and proportion of trials at each treatment level; the corresponding

covariances are derived in Section 4. Asymptotic results for the expectations are included in Section 3, and limiting results for the covariances are derived in Section 4. In Section 5, the probability of treating within a specific range of levels is derived. Assuming that responses follow an extreme value function or a logistic function and that a treatment allocation rule from Flournoy and Durham (companion manuscript in this volume) is used, these expectations and one standard deviation bands are graphed as a function of the trials $n = 0, 1, \ldots 99$ (see Figures 1-3). For these particular response functions models and treatment allocation rules, these figures demonstrate convergence to asymptotic values. Moreover, it demonstrates how the exact statistics can be used to study a wide class of treatment allocation rules and alternative model assumptions.

2. Transitions between treatments. From treatment x_i, the *transition probabilities* p_{ik} in one trial and $p_{ik}(n)$ in n trials, respectively, are

(1)
$$p_{ik} \equiv P\{X(n) = x_k \mid X(n-1) = x_i\};$$
$$p_{ik}(n) \equiv P\{X(n) = x_k \mid X(0 = x_i)\};$$

$n = 0,1,2,\ldots$; $k = 1,2,\ldots,K$. Of course, $p_{ik}(0) = \delta_{ik}$, where δ_{ik} is Kronecker's delta function, and since the transitions are assumed to follow a random walk, $p_{ik} = 0$ for $|i-k| > 1$. Thus the random walk is determined by the probabilities $p_{k,k+1}$, $p_{k,k-1}$ and p_{kk} that the treatments will move up from level k to $k+1$, down from level k to $k-1$, and stay at level k, respectively, for $k = 1, \ldots, K$, with $p_{k,k-1} + p_{kk} + p_{k,k+1} = 1$ for all k. Strict boundaries on the dosages are fixed by requiring that $p_{10} = 0$ and $p_{K,K+1} = 0$. Transition functions for the two biased coin up-and-down designs, BCD I and BCD II, are given in (1) and (4) of Durham and Flournoy (companion manuscript in this volume). These designs are used throughout this paper to make explicit the usefulness of the statistics that are derived.

Let $\mathbf{P} \equiv [p_{ik}]$be the $K \times K$ dimensional matrix of transition probabilities from the treatment given in one trial to the treatment given in the next trial, and note that the nth power of $\mathbf{P}$, $\mathbf{P}^n = [p_{ik}(n)]$, is the $K \times K$ dimensional matrix of transition probabilities from the treatment given in one trial to another that is n trials ahead. $\mathbf{P}$ is regular for BCD I and BCD II, that is, for some n that is large enough,

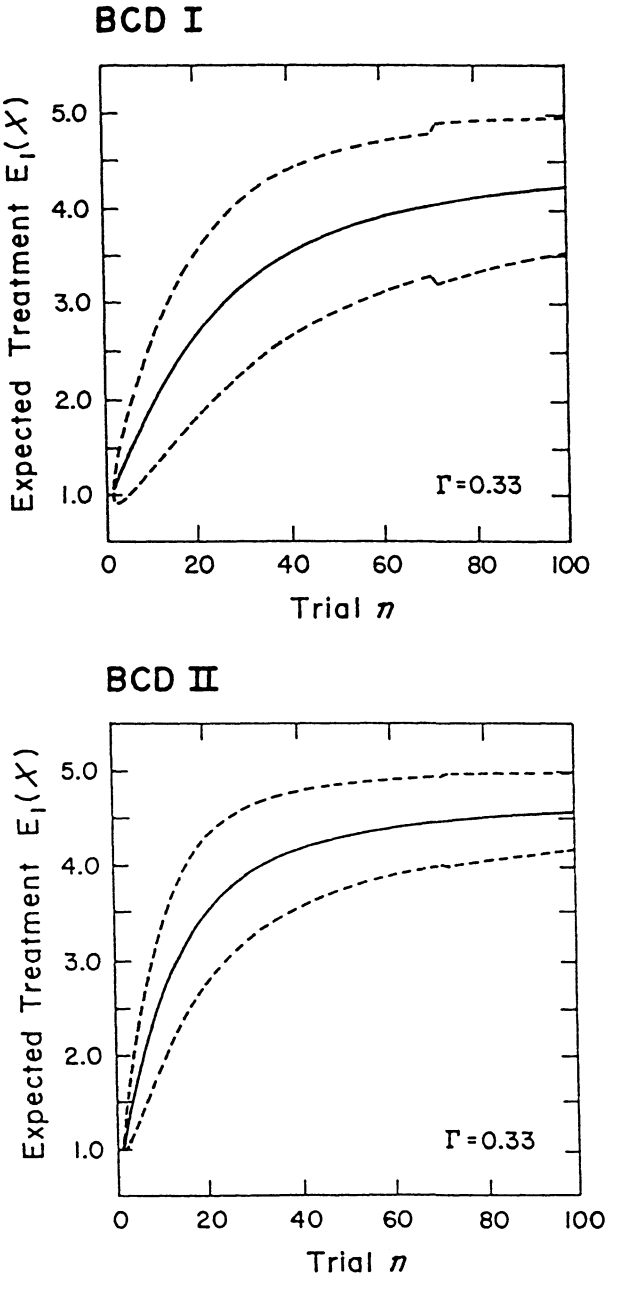

FIG. 1. The expected treatment, $E_1(X)$ ± 1 standard deviation, when targeting the 33rd percentile of $Q(x) = 1-\exp\{-\exp\{(x-6.931)/1.97\}\}$.

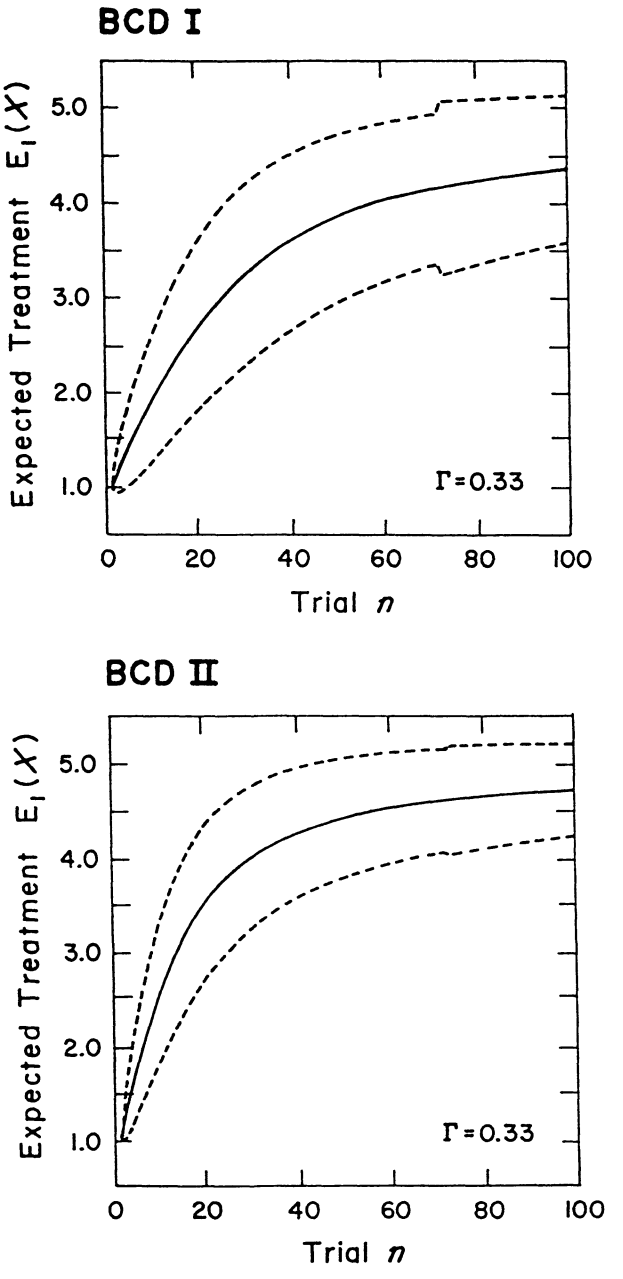

FIG. 2. The expected treatment, $E_1(X)$ ± 1 standard deviation, when targeting the 33rd percentile of $Q(x) = 1 - [1 + \exp(-3.569 + 0.549x)]^{-1}$.

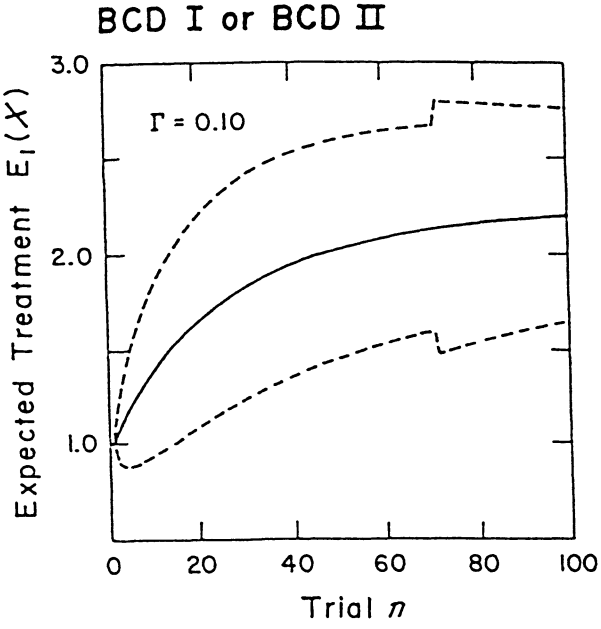

FIG. 3. The expected treatment, $E_1(X)$ ± 1 standard deviation, when targeting the 10th percentile of $Q(x) = 1-\exp\{-\exp\{(x-6.931)/1.97\}\}$ or $Q(x) = 1-[1+\exp(-3.569+0.549x)]^{-1}$.

the elements of $\mathbf{P}^n$ are all positive. Throughout this paper, we take $\mathbf{P}$ to be regular. Theorem 1 summarizes some well-known asymptotic properties of regular transition matrices. [See Kemeny and Snell (1960, pp. 70-71) for a proof].

THEOREM 1. *If* $\mathbf{P}$ *is a regular transition matrix, then*

(i) *the powers* $\mathbf{P}^n$ *approach a probability matrix* $\mathbf{\Pi}$;

(ii) *each row of* $\mathbf{\Pi}$ *is the same probability vector* $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)$;

(iii) *the components of* $\boldsymbol{\pi}$ *are all positive;*

(iv) *for any probability vector* $\mathbf{p}(0)$, $\mathbf{p}(0)\mathbf{P}^n = \mathbf{p}(n) \to \boldsymbol{\pi}$ *as* $n \to \infty$;

(v) $\mathbf{P\Pi} = \mathbf{\Pi P} = \mathbf{\Pi}$.

Note from *(iv)* that the vector $\boldsymbol{\pi}$ is the asymptotic treatment probability distribution, that is, $\pi_k = \lim_{n\to\infty} P(X(n) = x_k)$, $k = 1, \ldots, K$.

Suppose that the transition probabilities

$$\{p_{12} > \cdots > p_{K-1,K} > p_{K,K+1} = 0\}$$

to higher levels are monotone decreasing, and the transition probabilities

$$\{p_{10} = 0 < p_{21} < \cdots < p_{K,K-1}\}$$

to lower levels are monotone increasing, with $p_{12} > 0$ and $0 < p_{K,K-1}$. Thus as the sequence of treatments moves higher, it becomes increasing less likely that this trend will continue, and similarly, as the sequences of treatments moves lower, the chances of continuing this trend diminish. One might suspect that these opposing forces would cause the treatments to concentrate in some particular region. Denote by x_κ the mode of the asymptotic treatment distribution that results from using such an up-and-down design. Durham and Flournoy (1993, 1994a) showed that the asymptotic treatment distribution centers around the unknown quantile μ, in the sense that the limiting treatment distribution either has a single mode at x_κ, or it has a mode at both x_κ and $x_{\kappa-1}$. Furthermore, they show that if BCD I or BCD II is used and treatments are equidistant, then $|\mu - x_\kappa| \leq \triangle$, where $\triangle$ is the spacing between treatments.

3. Expected trial frequencies. Define a function to indicate whether or not the nth treatment is x_k:

$$I\left(X\left(n\right) = x_k\right) \equiv \left\{ \begin{array}{ll} 1 & if\ X\left(n\right) = x_k; \\ 0 & if\ X\left(n\right) \neq x_k. \end{array} \right\}$$

Then the frequency with which treatment x_k is used during trials 0 through n is

(2)
$$N_k\left(n\right) \equiv \sum_{m=0}^{n} I\left(X\left(m\right) = x_k\right),\ k = 1,\ 2,\ \ldots,\ K,$$

with $\sum_{k=1}^{K} N_k\left(n\right) = n + 1$. For notational simplicity, let the subscript i in $P_i\left(\cdot\right)$ and $E_i\left(\cdot\right)$ indicate that the probabilities and expectations,

respectively, are conditional on a fixed initial treatment $X(0) = x_i$, whereas the subscript $\mathbf{p}$ will indicate that the probabilities and/or expectations are taken with respect to a probability distribution $\mathbf{p}(0)$ of $X(0)$. Now from the law of total probability,

(3)

$$\begin{aligned} E_i(N_k(n)) &= E\left(\sum_{m=0}^{n} I(X(m) = x_k) \mid X(0) = x_i\right) \\ &= \sum_{m=0}^{n} p_{ik}(m) = \delta_{ik} + \sum_{m=1}^{n} p_{ik}(m), \ k = 1, 2, \ldots, K. \end{aligned}$$

Let $\mathbf{N}(n) \equiv [N_1(n), \ldots, N_K(n)]$ denote the vector of observed numbers treated at each level after the nth trial, and let $\mathbf{M}(n)$ denote the $K \times K$ dimensional matrices with ith row vector $E_i(\mathbf{N}(n))$, $n = 0, 1, 2, \ldots$. Then since $E_{\mathbf{p}}(N_k(n)) = \sum_{i=1}^{K} E_i(N_k(n)) p_i(0)$,

(4)

$$E_{\mathbf{p}}(\mathbf{N}(n)) = \mathbf{p}(0)\mathbf{M}(n)$$

when $X(0)$ is random with distribution given by the probability vector $\mathbf{p}(0)$. When the initial distribution is the stationary distribution, (4) becomes $E_{\boldsymbol{\pi}}(\mathbf{N}(n)) = (n+1)\boldsymbol{\Pi}$. This can be seen by taking the expectation of (3) with respect to $\boldsymbol{\pi}$ and using the *balance equations* for a stationary distribution, that is, for $k = 1, \ldots, K$, use $\sum_{i=1}^{K} \pi_i p_{ik}(m) = \pi_k$ to obtain

(5)

$$E_{\boldsymbol{\pi}}(N_k(n)) = \sum_{i=1}^{K} (\pi_i \sum_{m=0}^{n} p_{ik}(m)) = \sum_{m=0}^{n} \pi_k = (n+1)\pi_k.$$

Let $\bar{P}_k(n) = N_k(n)/(n+1)$ denote the proportion of trials at x_k in $n+1$ trials. Then (5) is equivalent to $E_{\boldsymbol{\pi}}\left(\bar{P}_k(n)\right) = \pi_k$, $k = 1, \ldots, K$.

Theorem 2 gives a law of large numbers for the trial frequencies.

THEOREM 2. *Consider an up-and-down design with limiting treatment distribution* $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_k)$. *For any initial treatment probability distribution* $\mathbf{p}(0) = (p_1, p_2, \ldots, p_K)$, $E_{\mathbf{p}}\left(\bar{P}_k(n)\right) \to \pi_k$, *and for any* $\epsilon > 0$,

$$\lim_{n\to\infty} P_{\mathbf{p}}\left(\left|\bar{P}_k(n) - \pi_k\right| < \epsilon\right) \to 0.$$

The proof follows from $E_{\mathbf{p}}\left(\bar{P}_k(n) - \pi_k\right)^2 \to 0$ as $n \to \infty$, which in turn follows from $E_i\left(\bar{P}_k(n) - \pi_k\right)^2 \to 0$ as $n \to \infty$ for every i. Details of this proof can be found in Kemeny and Snell (1960, pp. 73).

To illustrate the usefulness of these statistics, suppose that treatments $\Omega_{\mathbf{x}} = \{1, 2, \ldots, 9\}$ are available to target $\Gamma = 0.33$. We know from Theorem 1 of Durham and Flournoy (companion manuscript in this volume) that asymptotically, using BCD I or BCD II, the mode of the treatment distribution will be as close as possible to μ given the discreteness of $\Omega_{\mathbf{x}}$, and we know that the asymptotic treatment distribution is unimodal. Therefore, starting the experiment at the lowest dosage, $X(0) \equiv 1$, as is commonly done in toxicity studies involving humans, it is of interest to study the performance of the sample mean of the treatment distribution,

$$E_1(X) = \sum_{k=1}^{K} x_k E_1\left(N_k(n)\right),$$

as a simple estimate of μ. We know from Tables 1 and 2 of Durham and Flournoy (companion manuscript found in this volume) that, asymptotically, such estimates have negative bias of size less than $\triangle$, and if one is willing to make parametric assumptions regarding the underlying response function, an exact amount can be added to the sample mean to correct for this bias.

However, the sample treatment mean initially will have larger negative bias. To determine the magnitude of the bias of the sample mean as an estimate of μ when $\Gamma = 0.33$, $E_1(X)$ is plotted as a function of n in Figures 1 and 2 assuming extreme value ($\alpha = 6.931$, $\beta = 1.97$) and logistic responses ($\alpha = -3.569$, $\beta = 0.549$), respectively.

When targeting $\Gamma = 0.10$ with these underlying extreme value and logistic response functions, the plots of $E_1(X)$ as a function of n are indistinguishable (see Figure 3). Observing when the curves in Figures 1, 2, and 3 asymptote provides an estimate of sample size for which asymptotic results provide good approximations. For smaller sample

sizes, if parametric models are assumed, an exact bias correction to the sample mean may be made for estimating μ.

4. Covariances between trial frequencies. For the case in which $X(0)$ is fixed at x_i, Theorem 3 states that the covariances between trial frequencies are sums over trials of the covariances between treatments. Also, the covariances between treatments are given in terms of this transition probabilities.

THEOREM 3. *After n+1 trials when X(0) has been fixed at x_i, covariances between trial frequencies at levels $k, l = 1, \ldots, K$ are*

$$Cov_i\left(N_k(n), N_l(n)\right) = \sum_{s=0}^{n}\sum_{t=0}^{n} Cov_i\left(I\left(X(s)=x_k\right), I\left(X(t)=x_l\right)\right),$$

$n = 0, 1, 2, \ldots$, where the covariance between using treatment x_k at trial s and treatment x_l at trial t is

(6) $$Cov_i\left(I\left(X(s)=x_k\right), I\left(X(t)=x_l\right)\right)$$

$$= \begin{cases} p_{lk}(s-t)\,p_{il}(t) - p_{ik}(s)\,p_{il}(t), & if\ s>t, \\ p_{kl}(t-s)\,p_{ik}(s) - p_{ik}(s)\,p_{il}(t), & if\ t>s, \\ p_{ik}(t)\left(\delta_{lk} - p_{il}(t)\right), & if\ s=t. \end{cases}$$

PROOF. Because the two transition probabilities of a random walk do not depend on n, the joint probability of two treatments can be written as

(7)

$$P_i = \left(X(s)=x_k,\ X(t)=x_l\right) = \begin{cases} p_{lk}(s-t)\,p_{il}(t), & if\ s>t, \\ p_{kl}(t-s)\,p_{ik}(s), & if\ t>s \\ p_{il}(t)\,\delta_{lk}, & if\ s=t. \end{cases}$$

Furthermore, since the expectation of the product of indicators is a joint probability, inserting (1) and (7) into the right hand side of the expression

$$Cov_i\left(I\left(X(s)=x_k\right), I\left(X(t)=x_l\right)\right)$$

$$= P_i\left(X\left(s\right) = x_k, X\left(t\right) = x_l\right) - P_i\left(X\left(s\right) = x_k\right) P_i\left(X\left(t\right) = x_l\right)$$

yields (6). Using (7), the second moment of trial frequencies given $X\left(0\right) = x_i$ can be written as

$$(8)\ E_i\left(N_k\left(n\right) N_l\left(n\right)\right) = E_i\left(\sum_{s=0}^{n} I\left(X\left(s\right) = x_k\right) \sum_{t=0}^{n} I\left(X\left(t\right) = x_l\right)\right)$$

$$= \sum_{s=0}^{n} \sum_{t=0}^{n} P_i\left(X\left(s\right) = x_k,\ X\left(t\right) = x_l\right)$$

$$= \sum_{s=0}^{n} \sum_{t=0}^{s-1} p_{lk}\left(s-t\right) p_{il}\left(t\right) + \sum_{t=0}^{n} \sum_{s=0}^{t-1} p_{kl}\left(t-s\right) p_{ik}\left(s\right) + \sum_{s=0}^{n} p_{il}\left(s\right) \delta_{lk}.$$

Inserting (3) and (8) into $E_i\left(N_k\left(n\right), N_l\left(n\right)\right) - E_i\left(N_k\left(n\right)\right) E_i\left(N_l\left(n\right)\right)$, the exact covariance between the trial frequencies at x_k and x_l can be written as

$$(9)\ Cov_i\left(N_k\left(n\right), N_l\left(n\right)\right) = E_i\left(N_k\left(n\right), N_l\left(n\right)\right) - E_i\left(N_k\left(n\right)\right) E_i\left(N_l\left(n\right)\right)$$

$$= \sum_{s=0}^{n} \sum_{t=0}^{s-1} p_{lk}\left(s-t\right) p_{il}\left(t\right) + \sum_{t=0}^{n} \sum_{s=0}^{t-1} p_{kl}\left(t-s\right) p_{ik}\left(s\right) + \sum_{s=0}^{n} p_{il}\left(s\right) \delta_{lk}$$

$$- \sum_{s=0}^{n} p_{ik}\left(s\right) p_{il}\left(s\right) - \sum_{s=0}^{n} \sum_{t=0}^{s-1} p_{ik}\left(t\right) p_{il}\left(s\right)$$

$$= \sum_{s=0}^{n} \sum_{t=0}^{n} Cov_i\left(I\left(X\left(s\right) = x_k\right), I\left(X\left(t\right) = x_l\right)\right). \quad \square$$

COROLLARY 1. *After each trial n=0, 1, 2, ... when* $X\left(0\right)$ *is selected with probability* $\mathbf{p}\left(0\right)$,

$$Cov_{\mathbf{p}}\left(N_k\left(n\right), N_l\left(n\right)\right) = \sum_{s=0}^{n} \sum_{t=0}^{n} Cov_{\mathbf{p}}\left(I\left(X\left(s\right) = x_k\right), I\left(X\left(t\right) = x_l\right)\right),$$

where the covariances between using treatment x_k *at trial s and using treatment* x_l *at trial t are*

(10) $Cov_{\mathbf{p}}\left(I\left(X\left(s\right)=x_k\right),I\left(X\left(t\right)=x_l\right)\right)$

$$=\begin{cases} p_{lk}\left(s-t\right)p_l\left(t\right)-\sum\limits_{i=1}^{K}p_{ik}\left(s\right)p_{il}\left(t\right)p_i\left(0\right) & if\ s>t, \\ p_{kl}\left(t-s\right)p_k\left(s\right)-\sum\limits_{i=1}^{K}p_{ik}\left(s\right)p_{il}\left(t\right)p_i\left(0\right) & if\ t>s, \\ \sum\limits_{i=1}^{K}p_{ik}\left(t\right)\left(\delta_{lk}-p_{il}\left(t\right)\right)p_i\left(0\right) & if\ s=t. \end{cases}$$

COROLLARY 2. *When the initial distribution is the stationary distribution,* (10) *becomes*

(11) $Cov_{\boldsymbol{\pi}}\left(I\left(X\left(s\right)=x_k\right),I\left(X\left(t\right)=x_l\right)\right)$

$$=\begin{cases} p_{lk}\left(s-t\right)\pi_l & - & \sum\limits_{i=1}^{K}p_{ik}\left(s\right)p_{il}\left(t\right)\pi_i & if\ s>t, \\ p_{kl}\left(t-s\right)\pi_k & - & \sum\limits_{i=1}^{K}p_{ik}\left(s\right)p_{il}\left(t\right)\pi_i & if\ t>s, \\ \pi_k\delta_{lk} & - & \sum\limits_{i=1}^{K}p_{ik}\left(t\right)p_{il}\left(t\right)\pi_i & if\ s=t. \end{cases}$$

It is useful to note that evaluating $Cov_i\left(I\left(X\left(s\right)=x_k\right),I\left(X\left(t\right)=x_l\right)\right)$ in (6) at $k=l$ and $s=t$ yields

$$\begin{aligned} Var_i\left(I\left(X\left(s\right)=x_k\right)\right) &= p_{ik}\left(s\right)\left(1-p_{ik}\left(s\right)\right); \\ Var_{\mathbf{p}}\left(I\left(X\left(s\right)=x_k\right)\right) &= \sum_{i=1}^{K}p_{ik}\left(s\right)\left(1-p_{ik}\left(s\right)\right)p_i\left(0\right); \\ Var_{\boldsymbol{\pi}}\left(I\left(X\left(s\right)=x_k\right)\right) &= \pi_k-\sum_{i=1}^{K}\left(p_{ik}\left(s\right)\right)^2\pi_i. \end{aligned}$$

COROLLARY 3. Let $\bar{P}_k\left(n\right)=N_k\left(n\right)/\left(n+1\right)$ *denote the proportion of trials at* x_k *in* $n+1$ *trials. Then*

$$\left(n+1\right)Cov_i\left(\bar{P}_k\left(n\right),\bar{P}_l\left(n\right)\right)$$

$$=\left(n+1\right)^{-1}\sum_{s=0}^{n}\sum_{t=0}^{n}Cov_i\left(\mathrm{I}\left(X\left(s\right)=x_k\right),\mathrm{I}\left(X\left(t\right)=x_l\right)\right);$$

$$(n+1)\,Cov_{\mathbf{p}}\left(\bar{P}_k(n),\bar{P}_l(n)\right)$$

$$=(n+1)^{-1}\sum_{s=0}^{n}\sum_{t=0}^{n}Cov_{\mathbf{p}}\left(\mathrm{I}\left(X(s)=x_k\right),\mathrm{I}\left(X(t)=x_l\right)\right).$$

We now investigate some asymptotic properties. Let $\mathbf{C}$ be a $K\times K$ matrix with elements

$$(12)\quad c_{kl}\equiv\lim_{n\to\infty}(n+1)\,Cov_{\boldsymbol{\pi}}\left(\bar{P}_k(n),\bar{P}_l(n)\right)$$

$$=\lim_{n\to\infty}(n+1)^{-1}\,Cov_{\boldsymbol{\pi}}\left(N_k(n),N_l(n)\right).$$

Since $\lim_{n\to\infty}E_{\mathbf{p}}(\cdot)=\lim_{n\to\infty}E_{\boldsymbol{\pi}}(\cdot)$ for all initial distributions $\mathbf{p}(0)$, for n sufficiently large it is sufficient to consider $X(0)$ to be random with stationary probability distribution $\boldsymbol{\pi}$. Thus c_{kl} provides a large sample approximation to

$$(n+1)^{-1}\,Cov_{\mathbf{p}}\left(N_k(n),N_l(n)\right)$$

that is independent of $\mathbf{p}(0)$.

Theorem 4 makes c_{kl} tractable by expressing c_{kl} in terms of the stationary distribution and the fundamental matrix for regular Markov chains, $\mathbf{Z}=(z_{ij})\equiv\left(\mathbf{I}-(\mathbf{P}-\boldsymbol{\Pi})\right)^{-1}$, where $\mathbf{I}$ is the identity matrix. Let $\mathbf{D}_{\boldsymbol{\pi}}$ be a diagonal $K\times K$ matrix with diagonal elements π_1, π_2, $\ldots$, π_K. A proof can be found in Kemeny and Snell (1960).

THEOREM 4.

$$\mathbf{C}=\mathbf{D}_{\boldsymbol{\pi}}-\boldsymbol{\pi}'\boldsymbol{\pi}+\mathbf{D}_{\boldsymbol{\pi}}(\mathbf{Z}-\mathbf{I})+\left(\mathbf{Z}'-\mathbf{I}\right)\mathbf{D}_{\boldsymbol{\pi}}.$$

To see the usefulness of these statistics, suppose again that treatments $\Omega_{\mathbf{x}}=\{1,\ 2,\ \ldots,\ \}$ are available to target $\Gamma=0.33$, and that an experiment begins at $X(0)=1$. Now we can compute

$$Var_1\left(X(n)\right)=\sum_{k=1}^{9}\left(x_k-E_1\left(X(n)\right)\right)^2 p_{1k}.$$

These calculations give the confidence bands

$$E_1(X(n)) \pm \sqrt{Var_1(X(n))}$$

shown in Figures 1 - 3. Computational issues relating to the production of these Figures are discussed in Durham, Flournoy and Montazer-Haghighi (1994). In particular, they discuss the discontinuities in the figures appearing at $n = 69$ that are caused by switching from the exact calculations described here to large sample approximations.

5. The chances of treating at high levels. In practice, it may be useful to predict the number of treatments above a specific level, or within a certain range of levels. For example, when targeting the *33*rd percentile, one may wish to choose a up–and-down rule for which the expected number of treatments above the *50*th percentile of the response function is small. With such motivation in mind, partition $\Omega_{\mathbf{x}}$ into two subsets ω_1 and ω_2. For notational convenience, the expression $k \in \{i\}$ will be used to mean $\{k\colon\ x_k \in \omega_i\}$. Define $N_{\omega_j}(n) \equiv \Sigma_{k\in\{j\}} N_k(n)$ to be the total number of trials given at levels within the subset ω_j, $j = 1, 2$, respectively. Exact expectations and variances of $N_{\omega_j}(n)$, $j = 1, 2$, are computed by summing (3) and (9), respectively, over those treatment levels k for which $x_k \in \omega_j$ given $X(0)$ is selected to be x_i :

(13)

$$E_i\left(N_{\omega_j}(n)\right) = \sum_{k\in\{j\}} E_i(N_k(n)) = \sum_{k\in\{j\}} \sum_{s=0}^{n} P_i(X(s) = x_k),$$

$$Var_i\left(N_{\omega_j}(n)\right) = \sum_{k,l\in\{j\}} \sum_{s=0}^{n} \sum_{t=0}^{n} Cov_i(I(X(s) = x_k), I(X(t) = x_l)),$$

$k = 1, \ldots, K$. Of course, it follows directly from (13) that, if the proportion of trials in region ω_j is $\bar{P}_{\omega_j}(n) \equiv (n+1)^{-1} N_{\omega_j}(n)$, then

(14)

$$E_i\left(\bar{P}_{\omega_j}(n)\right) = (n+1)^{-1} \sum_{m=0}^{n} \sum_{k\in\{j\}} P_i(X(m) = x_k);$$

$$(n+1)\, Var_i\left(\bar{P}_{\omega_j}(n)\right) = (n+1)^{-1} \sum_{s=0}^{n} \sum_{t=0}^{n} \sum_{k,l\in\{j\}} Cov_i(N_k(n), N_l(n)).$$

It also follows from (14) that when $X(0)$ is selected with probability distribution $\mathbf{p}(0)$, the mean and variance of $\bar{P}_{\omega_j}(n)$ is

(15)

$$E_{\mathbf{p}}\left(\bar{P}_{\omega_j}(n)\right) = (n+1)^{-1} \sum_{i=1}^{K} p_i(0) \left(\sum_{k \in \{j\}} E_i\left(N_k(n)\right) \right);$$

$$(n+1)\, Var_{\mathbf{p}}\left(\bar{P}_{\omega_j}(n)\right)$$

$$= (n+1)^{-1} \sum_{i=1}^{K} p_i(0) \left(\sum_{k,l \in \{j\}} Cov_i\left(N_k(n), N_l(n)\right) \right) \rightarrow \sum_{k,l \in \{j\}} c_{kl},$$

as $n \rightarrow \infty$. When the initial distribution is the stationary distribution, (15) becomes

(16)

$$E_{\boldsymbol{\pi}}\left(\bar{P}_{\omega_j}(n)\right) = (n+1)^{-1} \sum_{k \in \{j\}} E_{\boldsymbol{\pi}}\left(N_k(n)\right) = \sum_{k \in \{j\}} \pi_k;$$

$$(n+1)\, Var_{\boldsymbol{\pi}}\left(\bar{P}_{w_j}(n)\right)$$

$$= (n+1)^{-1} \sum_{i=1}^{K} \pi_i \left(\sum_{k,l \in \{j\}} Cov_{\pi}\left(N_k(n), N_l(n)\right) \right)$$

$$\longrightarrow \sum_{k,l \in \{j\}} c_{kl},$$

as $n \rightarrow \infty$.

If the cardinality of $\Omega_{\mathbf{x}}$ is large and the initial treatment is far from the target μ, it is possible that a long sequence of increasing or decreasing changes in the treatment levels will occur, depending on whether the initial treatment was far below or far above the target level, respectively. A long series of treatments far below or above μ would bias estimators based on the stationary treatment distribution, and therefore, it is recommended in practice that observations be truncated after the first change in the direction of the treatment sequence, that is, if the initial sequence of treatments is increasing, let $X(0)$ be the first treatment that is less than the preceding one. This truncation of the

initial sequence of observations was recommended by Brownlee, Hodges and Rosenblatt (1953) when sampling according to Dixon and Mood's (1948) up-and-down design for estimating the *50*th percentile.

Suppose again that treatments $\Omega_x = \{1, 2, \ldots, 9\}$ are available to target $\Gamma = 0.33$, and suppose the experimenter wishes to evaluate the chances of giving a treatment for which $Q(x) \equiv P(Y = 1 \mid x) \geq 0.50$. As is commonly done in toxicity studies involving humans, again let the initial treatment be set to the lowest possible treatment, i.e., $X(0) \equiv 1$. Since the probability of treating at high levels depends on the up-and-down rule that is used and on the unknown response function, a sensitivity analysis can be useful.

To see how variations in the parametric form of the response function affect the treatment distribution, we compare the extreme value response function $Q(x) = 1 - \exp\{-\exp\{(x - \alpha)/\beta\}\}$ with the logistic response function $Q(x) = \exp(\alpha + \beta x)/(1 + \exp\alpha + \beta x)$. Suppose the researcher does not know the form of the response function but predicts an increasing response function with $Q(6.50) = 0.50$ and $Q(2.50) = 0.10$. Then if the responses follow an extreme value distribution, the predicted parameters are $\alpha = 6.931$ and $\beta = 1.97$, whereas if the responses follow a logistic distribution, the predicted parameters are $\alpha = -3.569$ and $\beta = 0.549$. Since $Q(6.50) = 0.50$ for both the extreme value and logistic response function, and since treatments are only given at integer values, finding the chances of treating at levels for which $Q(x) \geq 0.50$ is equivalent to finding $P_1(X(n) > 6.50)$, where now n is the intended number of trials. Now $\omega \equiv \{7, 8, 9\}$.

A large sample approximation to $P_1(X(n) > 6.50)$ is given by (16), namely $E_{\boldsymbol{\pi}}(P_w(n)) = \Sigma_{k=7}^{9}\pi_k$. This approximation can be calculated using the stationary probabilities given by equations (9) and (10) of Flournoy and Durham (companion manuscript found in this volume) for BCD I and BCD II, respectively, assuming an extreme value response function or by equations (11) and (12) for BCD I and BCD II, respectively, assuming a logistic response function. For an extreme value response function with $\alpha = 6.931$ and $\beta = 1.97$, $\Sigma_{k=7}^{9}\pi_k$ equals 0.14 and 0.11 using BCD I and BCD II, respectively, whereas for a logistic response function with $\alpha = -3.569$ and $\beta = 0.549$, $\Sigma_{k=7}^{9}\pi_k$ equals 0.18 and 0.15 using BCD I and BCD II, respectively.

Exact values of expected proportion trials above the 50th percentile, and its associated variance can be determined using (16):

(17)

$$P_1(X(n) \geq 7) = (n+1)^{-1} \sum_{m=0}^{n} \sum_{k=7}^{9} P_1(X(m) = x_k);$$

$$Var_1(I(X(n) \geq 7)) = (n+1)^{-2} \sum_{s=0}^{n} \sum_{t=0}^{n} \sum_{k,l=7}^{9} Cov_1(N_k(n), N_l(n))$$

Using (17), the expected proportions of subjects (plus and minus one standard deviation) receiving treatments greater than 6.50 are graphed as trials accumulate for $n = 0, 1, \ldots, 99$ in Figure 4 using BCD I and BCD II to target $\Gamma = 0.33$ when responses follow the extreme value function. Figure 5 displays the same graphs when responses follow the logistic function. Computational issues relating to the production of these Figures are discussed in Durham, Flournoy and Montazer-Haghighi (1994). In particular, again note that discontinuities in the figures appearing at $n = 69$ are caused by switching from exact calculations to large sample approximations.

In both Figures 4 and 5, the expected proportion of subjects receiving treatments greater than the $50th$ percentile is seen to approach its limit from below for both up-and-down designs and both response functions. This is a consequence of selecting the initial treatment to be $X(0) = 1$. Note that an estimate based on asymptotic theory will overestimate the true numbers for samples as large as 99. The exact theory can be used to correct for biases in moment estimates that are caused by starting at low dose levels. The expected proportions approach their asymptote faster if BCD II is used rather than BCD I for both the logistic and extreme value response functions. Thus the expected proportion treated at high levels increases faster using BCD II than BCD I.

The confidence bounds can be seen to be tighter for extreme value responses in Figure 4 than they are for logistic responses in Figure 5, and the probability of treating at levels $x \geq 7$ is greater for logistic responses than for extreme value responses.

Alternatively if the target is $\Gamma = 0.10$, graphs of the expected proportions of subjects treated at levels greater than the 50th percentile are indistinguishable as shown in Figure 6.

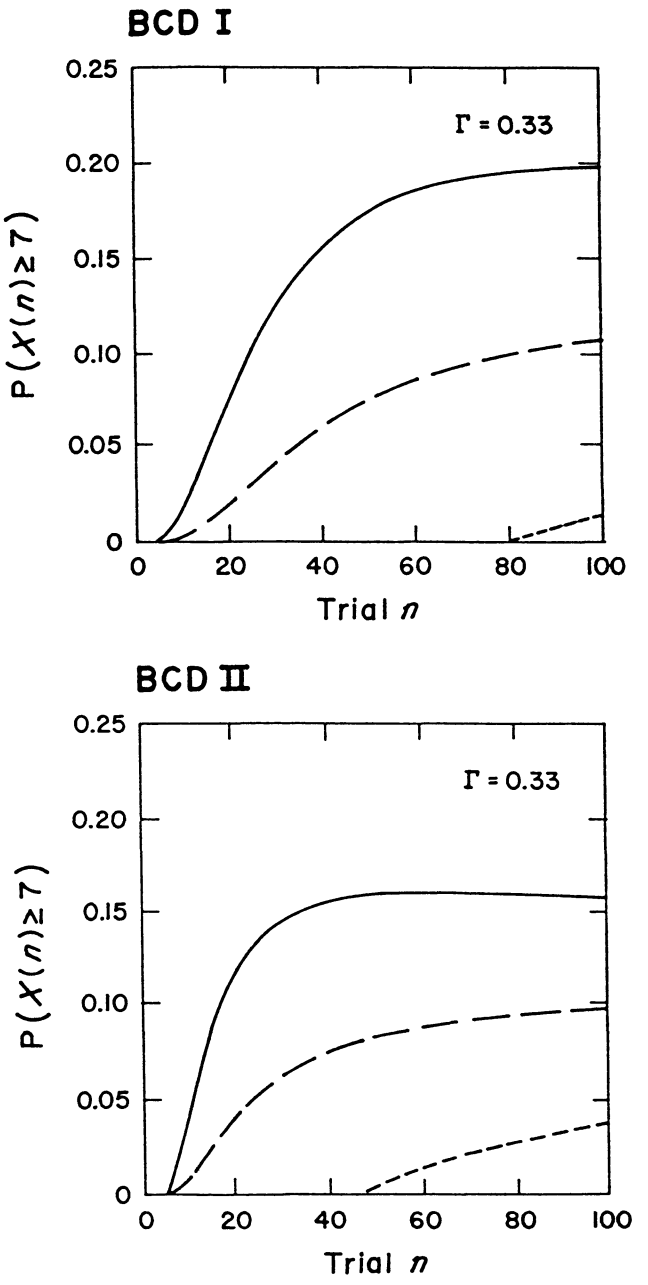

FIG. 4. The probability of treating at dosages for which the P{toxicity} >.50, ± 1 standard deviation, when targeting the 33rd percentile of $Q(x) = 1-\exp\{-\exp\{(x-6.931)/1.97\}\}$.

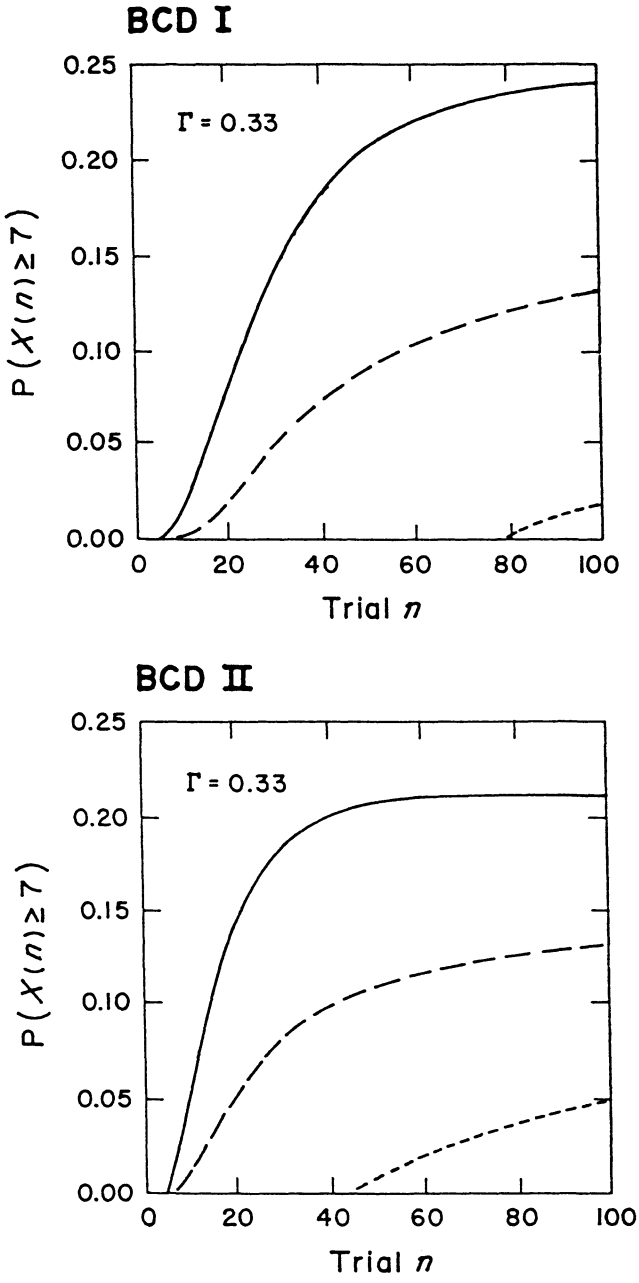

FIG. 5. The probability of treating at dosages for which the P{toxicity} > .50, ± 1 standard deviation, when targeting the 33rd percentile of $Q(x) = 1 - [1+\exp(-3.569 + 0.549x)]^{-1}$.

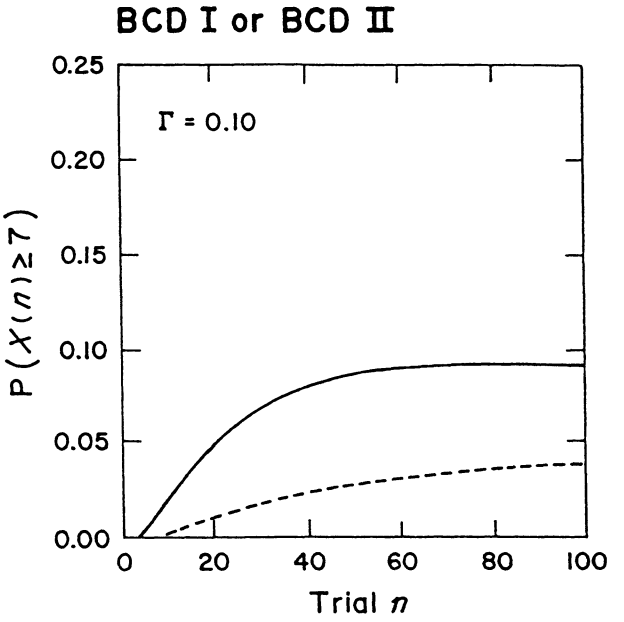

FIG. 6. The probability of treating at dosages for which the P{toxicity} > .50, ± 1 standard deviation, when targeting the 10th percentile of $Q(x) = 1-\exp\{-\exp\{(x-6.931)/1.97\}\}$ or $Q(x) = 1-[1+\exp(-3.569+0.549x)]^{-1}$.

References

BROWNLEE, K.A., HODGES, JR., J.L., and ROSENBLATT, M. (1953). An up-and-down method with small samples. *Journal of the American Statistical Association* **48** 262-277.

DIXON, W.J. and MOOD, A.M. (1948). A method for obtaining and analyzing sensitivity data. *Journal of the American Statistical Association* **43** 109-126.

DURHAM, S.D. and FLOURNOY, N. (1994). Random walks for quantile estimation. *Statistical Decision Theory and Related Topics V* (Berger, J. and Gupta, S., eds). New York: Springer-Verlag, 467-476.

DURHAM, S.D., FLOURNOY, N. and MONTAZER-HAGHIGHI, A.A. (1994). Up-and-down designs. *Computing Science and Statistics: Interface* (Tarter, M.E. and Lock, M.D., eds) Fairfax, VA: The Interface Foundation of North America, **25** 375-384.

KEMENY, J.D. and SNELL, J.L. (1960). *Finite Markov Chains.* Princeton: D. Van Nostrand Company.

STEPHEN D. DURHAM
DEPARTMENT OF STATISTICS
UNIVERSITY OF SOUTH CAROLINA
COLUMBIA, SC 29208

NANCY FLOURNOY
DEPARTMEN OF MATHEMATICS AND STATISTICS
THE AMERICAN UNIVERSITY
WASHINGTON, DC 20016-8050

ALI A. MONTAZER-HAGHIGHI
DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE
BENEDICT COLLEGE
COLUMBIA, SC 29204