Translational clinical trials: an entropy-based approach to sample size

Steven Piantadosi

Translational clinical trials are small studies of therapies emerging from the laboratory. These trials are essential for generating early evidence regarding the effects of treatment on specific targets in the disease pathway and for guiding the next studies to be done. The statistical properties of such studies have been neglected, in part, because they do not fit the well-known clinical trials developmental paradigm. This paper discusses the translational trial setting, and presents an information (entropy) based approach to understanding the properties and use of these trials. The combination of biological knowledge with a designed experiment (albeit a small one) is a powerful device for resolving much of the considerable uncertainty surrounding an emerging therapeutic concept. An approach to motivating the sample size for translational trials is presented. Clinical Trials 2005; 2: 182–192. www.SCTjournal.com

1 Introduction

The purpose of this paper is to help qualitatively and quantitatively define translational clinical trials, a class of study designs that has not yet been well formulated in the literature. Translational studies are the bridge between therapeutic ideas emerging from the laboratory and traditional clinical development [1–5]. There is intense interest in translational research among investigators, funding agencies, and administrators, because there are so many promising directions arising from basic research, and every new therapy depends on a successful step from laboratory to clinic. Translational studies are among the most common types of clinical trials performed, especially in fields where therapeutic development is highly active. Although general methods for translational studies are still evolving, they have stimulated some unique questions of ethics [6].

It has often been stated informally that the dividing line between laboratory research and clinical development is the “phase I” study. In fact, it is translational studies that bridge this gap and often result in either additional laboratory experiments or clinical trials being done before any traditional clinical development begins. “Phase I” is better characterized as investigation of the relationship between dose and safety. Use of this term too broadly may have hindered the formalization of translational trial design. Translational questions can sometimes be nested within later developmental trials provided the appropriate subjects and biological measurements can be obtained.

Despite the importance of getting new therapeutic ideas out of the laboratory and into human trials, there is very little written about translational methodology in the statistical literature. For example, a computerized search of 15 statistics journals for the past five years found no papers cross referenced under translational clinical trials or translational research methods. In the clinical and biological sciences, there is much discussion of applied translational research findings but relatively little regarding clinical trial methodology. In 2003, the open source Journal of Translational Medicine was started [7], but currently has no papers on trial methodology. One of the best examples of organized translational research is the National Cancer Institute’s Specialized Program of Research Excellence (SPORE) grants, which explicitly target the laboratory—clinic interface. Although such programs have been in place for over 10 years, there has not yet been a formulation of translational clinical trial methods arising from them.

Johns Hopkins School of Medicine
Address for correspondence: Steven Piantadosi, Oncology Biostatistics, Johns Hopkins School of Medicine, 550 N Broadway, Suite 1103, Baltimore, MD 21205, USA.

© Society for Clinical Trials 2005

10.1191/1740774505cn078oa
In this article, I will characterize the interface between basic science and traditional clinical development using the concept of translational trials. My discussion of the quantitative properties of these trials is predicated on the idea that investigators design and interpret translational research studies predominantly on the basis of reduction in overall uncertainty regarding possible outcomes. This is in contrast with more familiar clinical trial designs (e.g., randomized trial) where the objective is often to attain moderate to high statistical precision in the estimate of a particular parameter (e.g., hazard ratio). It is clear from the size of most translational trials that statistical precision is not a primary design objective. It is also clear that investigators make reasonable decisions on the basis of such studies. We would have to conclude, based on numerous translational successes, that the studies in this class are informative whatever the basis for design and interpretation.

I begin by describing translational research studies and offering some formal structure and definition for them. Next I review some properties of information (entropy) and discuss why it might represent a fundamental principle for investigators' behavior in translational research. I then discuss simple statistical properties of an empirical approximation to information, and why relatively small sample sizes might be appropriate for such trials. Some theoretical and numerical results that can be used to guide sample size are presented. These are illustrated using hypothetical examples. A computer program at the author’s website (www.cancerbiostats.onc.jhmi.edu/software.cfm) is available for all calculations and simulations.

2 Character and definition

After developing a new therapeutic concept in the laboratory, a need arises for studies that assess the effects of treatment on a biological target in humans. The purposes of such studies are to assess the potential targeting and activity of the therapy, guide subsequent clinical and laboratory experiments, and perhaps to validate the target [5]. The treatment is likely to be modified based on the results of this early trial.

The outcome used in such a study has to be measurable with a fair degree of accuracy, and the absence of change in an appropriate direction can be taken as reliable evidence of inactivity of the treatment. This is not a surrogate outcome because it is not used to make inferences about clinical benefit. At most, a translational trial might anticipate later questions of clinical benefit, but its principal goal is to define the next experimental steps to be taken.

The outcome measure chosen provides definitive mechanistic evidence of effect, or lack of effect. Within the working paradigm of disease and treatment, the biological assessment is a reliable signal that has to show promising changes in direction and magnitude for proof of principle. For example, a good biological signal might be a change in levels of a protein or gene expression or enzyme activity in the causal pathway of the disease. A positive effect on the target does not prove clinical benefit, but might be a prerequisite for continued development.

Explicit characteristics of a translational trial are as follows.

- The trial is predicated on promising preclinical developments, which create a need to evaluate the new treatment in human subjects.
- The treatment and/or its algorithm is changeable, perhaps following additional laboratory experiments.
- The treatment and its method of evaluation are formally and fully specified in a written protocol, as with any clinical trial.
- The evaluation relies on one or more biologic targets or outcomes that provide definitive evidence of mechanistic effect within the working paradigm of disease and treatment.
- The outcome is measurable with small uncertainty relative to the effect size, often soon after treatment. Target validation is sometimes also an objective, in which case imprecision in the outcome measurement may also represent a failure.
- Large effects on the target are sought.
- There is an explicit, unambiguous definition of “lack of effect” or failure to demonstrate the hoped for effect on the target.
- The study protocol specifies the next experimental step(s) to be taken for any possible outcome of the trial. Consequently, regardless of results, the trial will be informative with respect to a future experiment.
- Studies are structured and sized to provide sufficient reliable information to guide additional experiments, but not necessarily to yield strong statistical evidence.
- The study is typically undertaken by a small group of investigators with limited resources.

These considerations lead to the following formal definition of a translational trial:

A clinical trial where the primary outcome: 1) is a biological measurement (target) derived from a well-established paradigm of disease, and 2) represents an irrefutable signal regarding the intended therapeutic effect. The design and purposes of the trial are to guide further experiments in the laboratory or
Translational trials imply circularity between the clinic and the laboratory, with continued experimentation as the primary immediate objective. Many therapeutic ideas will prove useless or not feasible during this cycle. The laboratory—clinic iteration may eventually beget the familiar linear development of a new therapy, perhaps after numerous false starts.

The decision parameters for a translational trial are: 1) does the therapy affect the target in the intended direction? 2) is the magnitude of effect promising? and 3) are unanticipated safety concerns evident? Very generally, these questions imply that individual outcomes might be classified into mutually exclusive categories or on an ordinal scale. For example, the outcome could be classified as: a) promising versus b) not. More elaborate classifications would be needed in many circumstances. The uncertainty that needs to be resolved by a translational trial is the frequency in each outcome category. A translational trial will have met its objectives when the data contain sufficient information about those frequencies to guide the next studies needed.

It is not obvious from this characterization and definition that translational studies can routinely be small. But in this state of high uncertainty, a relatively small experiment increases information sufficiently to guide subsequent studies. The combination of biological knowledge with information from formal observation is a powerful proven tool for guiding early experiments. The reduction of uncertainty, or gain in information (to use this term in its formal sense), can be quantified as discussed in the next section.

3 Information, entropy, and empirical entropy

3.1 Information and entropy in statistics

In statistics, the term “information” is used in several ways, but all are interconnected and can all be linked to changes in entropy. Fisher information [8], the matrix of mixed second partial derivatives of the log-likelihood function, is most familiar to statisticians. Loosely speaking, it is the inverse variance in the context of a parameterized probability model. A different definition of information and its use in the context of coding theory was given by Shannon [9]. Shannon’s information is entropy, and is identical to the measure of disorder used in the physical sciences. The Shannon definition of information (entropy) has an axiomatic basis and is free of distributional or model assumptions. Properties of entropy can be extended to joint and conditional random variables in a natural way. In a sense, it is simpler and more pervasive than either of the definitions of information that arise in classical statistical inference.

Applebaum [10] discusses information in the context of probability theory. The Fisher and Shannon definitions of information are connected by Kullback’s information measure [11], which is also likelihood based and is the mean information for discriminating between two hypotheses. Similarly in Bayesian theory, scoring rules are used as loss functions for certain decision problems [12]. Under appropriate constraints, the minimum expected loss is the entropy of the distribution, and the difference between the expected loss and its minimum is the Kullback–Liebler divergence [13]. When making inferences regarding the entire posterior distribution, the Bayesian information of the data is the difference between the prior and posterior entropy [12].

The relationship between the expected value of sample information and cost can also be used to design sample surveys [14]. Shannon’s entropy has also been used in Bayesian experimental design [15, 16]. The best design strategy is to use the experiment that maximizes the expected gain in information, defined as the difference between the entropy of the prior and the posterior distributions. The maximum entropy principle will be helpful for certain sample size questions, as discussed below.

Thus, information and entropy are familiar in statistics and experimental design. Applicability of entropy in the setting of translational trials is based on 1) the axiomatic equivalence of information and entropy, 2) properties of entropy that seem well suited to the nature of translational trials and the behavior of investigators, 3) the need to provide a quantitative framework for assisting the sample size question with a minimum of assumptions, and 4) the useful role of entropy in decision problems. The approach to sample size proposed here is not Bayesian, nor is it a formal procedure of any type. The goal is to suggest an appropriate quantitative tool that might be useful in a very general way. In this setting, it may be difficult to construct specific representations of prior knowledge or ignorance, making it hard to use an overall framework for inferences that depends on them.

3.2 Empirical entropy

Suppose a translational test of a new treatment yields a multinomial response that represents categorical outcomes on a target biological marker as outlined in the previous section. The uncertainty in such a situation can be quantified formally using
Entropy-based approach to sample size

185

the entropy of the system, defined as

\[ H = - \sum_{i=1}^{m} p_i \log p_i \]  

where there are \( m \) possible states, each with probability \( p_i \) and \( \sum p_i = 1 \). If \( p = 0 \), then \( p \log (p) \) is defined to be 0. It can be shown that \( H \) is maximal when \( p_i = p_j \) for all \( i, j \). In information theory, it is customary to use base-2 logarithms, but this is arbitrary. Natural logarithms will be used here. For an outcome with \( S \) categories, the maximal uncertainty state is characterized by \( H = - \log(0.2) = 1.61 \).

Gain in information is measured by a change (reduction) in entropy. If we acquire information and revise the probabilities \( (q_i) \), the gain is

\[ \Delta H = \sum_{i=1}^{m} q_i \log q_i - \sum_{i=1}^{m} p_i \log p_i \]  

where the direction has been taken to make \( \Delta H > 0 \) correspond to an information increase (reduction in uncertainty) relative to the maximal entropy case. At the conclusion of our clinical trial, suppose we take the observed frequencies of outcomes as the true \( q_i \), temporarily ignoring the problem of small sample size. I will refer to the entropy calculated on the basis of the observed frequencies as the “empirical entropy”, denoted by \( H^* \), and the information apparently gained from the trial, \( \Delta H^* \).

For example, if 15 subjects were treated and yielded the multinomial outcome frequencies \([1, 7, 4, 3, 0]\), the empirical gain in information (reduction in entropy) relative to the state of highest uncertainty is

\[ \Delta H^* = 1.61 - \left[ \frac{1}{15} \log \left( \frac{1}{15} \right) + \frac{7}{15} \log \left( \frac{7}{15} \right) + \ldots \right] \]

\[ = 0.399 \]  

For \( \Delta H^* \) to make sense, the reference entropy has to be selected carefully. If we choose conditions of some certainty before performing our experiment, the results could show that uncertainty has increased (\( \Delta H^* < 0 \)). This does not seem to make sense, and could indicate simply that our apparent certainty before the experiment was unwise. However, if the reference entropy were based on earlier data, two possibilities arise. Conditions might have changed such that the outcome is, in fact, less certain now than before. Alternatively, we must recognize that \( \Delta H^* \) is a random variable and realizations based on small series of observations may appear inconsistent with one another. In any case, we know that our collective data must reduce uncertainty.

This example illustrates some features of information and entropy and why it might be useful in

the translational research setting. If the observed data had been \([0, 3, 4, 7, 1] \), the gain in information would be the same, but the biological inference would be dramatically different. Many qualitatively different results yield the same amount of information. Furthermore, large sample sizes will not produce proportionate gains in information because the point estimates tend to stabilize after modest sample sizes. Additional observations increase the precision in parameter estimates, but overall uncertainty depends more on the parameters and less on their precision. This represents only a change in perspective, not a paradox.

These properties echo investigator behavior regarding translational trials. Many clinical outcomes are equally informative, though not equally promising. Definitive clinical inferences are neither made nor needed from a translational trial. The information gained can be used to guide subsequent experiments. Investigators habitually avoid large translational experiments, likely reacting to reduced uncertainty (gain in information), as opposed to statistical precision, when they design and implement translational trials.

A minimalist translational trial will help illustrate this. Imagine that we had high uncertainty regarding the effect of a new oral drug targeted to inhibit an enzyme. A small number of independent samples would be useful to assess 1) the magnitude and range of serum levels, 2) a reasonable estimate of the mean level, 3) the degree of enzyme inhibition and fraction of subjects who achieve it, and 4) the best choice for additional studies (e.g., dose optimization). Measurement of serum levels in a few subjects could meet these needs without proving definitive clinical benefit. More generally, we could classify the effect on the enzyme target (or any outcome) on an ordinal scale: harmful or negative effect, no effect, slightly beneficial effect, definitely beneficial effect, for example. A discrete classification such as

<table>
<thead>
<tr>
<th>Table 1 True response probabilities for some hypothetical translational trials. These cases are used as examples in the tables and figures</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m )</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>-----</td>
</tr>
</tbody>
</table>

*Based on Gibbs distribution with mean = 2.
this is theoretically simple but allows nearly any outcome to fit this framework.

Some example probabilities that might represent ordinal classifications for outcomes of translational studies are shown in Table 1. These are meant to illustrate a range of both types of classification and true probability of outcome. Some cases from Table 1 will be used for illustration later. Qualitative behavior of entropy is not a strong function of the multinomial probabilities, but the quantitative behavior is, as seen in the next section.

3.3 Bias, variance, and MSE of empirical entropy

Equation (2) was applied as though the \( q_i \) from the trial were the true frequencies. The true frequencies will be closely approximated if the sample size is large. However, we can anticipate that entropy calculated from the outcome frequencies of a small study will be biased. This could cause Equation (2) to overestimate the gain in information from the study. The expected value of \( H^* \) using data from \( n \) subjects with \( r_i \) responses in each of \( m \) categories is

\[
E[H^*] = E \left( -\sum_{i=1}^{m} \frac{r_i \log r_i}{n} \right) = -\sum_{i=1}^{m} \frac{1}{n} (E[r_i \log r_i] - E[r_i] \log n)
\]

In a multinomial outcome, each category singly is binomial so that,

\[
E[H^*] = -\sum_{i=1}^{m} \left( \frac{1}{n} \sum_{k=0}^{n} k \log k \binom{n}{k} p_i^k (1-p_i)^{n-k} - p_i \log n \right)
\]

where the \( p_i \) are the true classification probabilities. \( E[H^*] \) is a function of sample size, the number of categories, and the true \( p_i \).

The bias in \( H^* \) is \( H - E[H^*] \), the difference between Equations (4) and (1),

\[
b[H^*] = -\sum_{i=1}^{m} p_i \log p_i + \sum_{i=1}^{m} \left( \frac{1}{n} \sum_{k=0}^{n} k \log k \binom{n}{k} p_i^k (1-p_i)^{n-k} - p_i \log n \right)
\]

Thus, the systematic error in using the empirical entropy depends on both the sample size and the unknown true multinomial probabilities. It can be shown that \( H^* \) underestimates \( H \), that is, \( b[H^*] > 0 \). Because the function \( p \log p \) is convex, Jensen's inequality yields

\[
E \left[ \frac{r_i \log r_i}{n} \right] \geq E \left[ \frac{r_i}{n} \right] \log E \left[ \frac{r_i}{n} \right] = p_i \log p_i
\]

where the rightmost equality uses the fact that \( r_i/n \) is unbiased for \( p_i \). Therefore,

\[
\sum_{i=1}^{m} E \left[ \frac{r_i \log r_i}{n} \right] - \sum_{i=1}^{m} p_i \log p_i \geq 0
\]

or \( H - EH^* \geq 0 \). In other words, the results of a small trial may not fully capture all the uncertainty in a system because low-frequency events may not be observed.

The scale of measurement for entropy is arbitrary. The state of maximal uncertainty could be chosen as a reference, for example. Equation (2) yields a biased estimate of the information gained, if we take the observed frequencies, \( q_i \), as the true probabilities,

\[
\Delta H^* = \sum_{i=1}^{m} q_i \log q_i + H_{ref}
\]

where \( H_{ref} \) represents the reference entropy. The bias in \( \Delta H^* \) is \( b[\Delta H^*] = \Delta H - E[\Delta H^*] \), which must be the same as \( b[H^*] \), because \( H_{ref} \) cancels from both terms.

We can investigate the bias in \( H^* \) quantitatively using Equation (5). Examples with varying response probabilities and sample sizes are shown in Figure 1. Some bias is present even for large sample sizes. However, the general behavior seems to be that the bias is reduced very substantially for moderate sample sizes. This behavior is not surprising when we consider that the individual estimates of the multinomial probabilities (\( r_i/n \)) are unbiased.

The variance of \( H^* \) is

\[
\text{Var}[H^*] = E[H'^2] - E[H^*]^2
\]

Figure 1 Bias versus sample size for empirically calculated entropy. The true probabilities used to generate each curve are: \( A = (0.20, 0.20, 0.20, 0.20, 0.20) \), \( B = (0.46, 0.26, 0.15, 0.08, 0.05) \), \( C = (0.01, 0.15, 0.80, 0.03, 0.01) \), \( D = 0.50, 0.30, 0.10, 0.05, 0.05 \).
Entropy-based approach to sample size

3.4 Exact sampling distribution

Because multiple outcomes result in the same information content, the distribution of $H^*$ is somewhat coarse for small sample sizes. For a given $n$ and $m$ and assuming specific true response probabilities, we can enumerate all possible outcomes of a relatively small experiment. For example, for $n = 10$ and $m = 3$, the only 14 possible cell frequencies disregarding order are shown in Table 2.

The complete sample space can be enumerated from the unique permutations of cell frequencies. For $n = 10$ and $m = 3$, the 66 possible outcomes are shown in Table 3 along with $H^*$ and the multinomial probability of each outcome when $p = [0.1, 0.85, 0.05]$. The number of possible outcomes becomes quite high for modest sample sizes, but still easily enumerable by computer. For 15 subjects and 5 categories, for example, there are 3876 possible outcomes. For 35 subjects and 5 categories, there are 82250 possible outcomes.

The exact sampling distribution of $H^*$ can be constructed from such enumerations. The effect of sample size on variability and bias can also be observed. An example of this is shown in Figure 2. The distributions of entropy become more refined and shifted to the right as the sample size increases. The shift to higher mean values is a manifestation of the bias due to small sample sizes.

### 3.5 Simulations

The bias and variance of $H^*$ are complicated functions of $n$, $m$, and the $p_i$. The theoretical formulas provided are also appropriate for computation as can be seen in the following small set of simulations covering typical cases for which sample size might be needed. All calculations were performed in Mathematica version 5 [17]. For each of 10000 replicates in a simulation, a multinomial sample of a given size was taken from the specified distribution. Then the empirical entropy was calculated. The mean and variance of values across the simulation were

www.SCTJournal.com
Table 3 All 66 possible outcomes of 10 subjects classified into 3 distinguishable categories $H$ is the entropy of the classification and $P$ the chance of that outcome assuming the true probabilities are $p = (0.20, 0.75, 0.05)$.

<table>
<thead>
<tr>
<th>$r_1$</th>
<th>$H^*$</th>
<th>$P$</th>
<th>$r_1$</th>
<th>$H^*$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10,0</td>
<td>0.00</td>
<td>1.02e-07</td>
<td>1.81</td>
<td>0.64</td>
<td>9.01e-02</td>
</tr>
<tr>
<td>0,10</td>
<td>0.00</td>
<td>5.63e-02</td>
<td>1.18</td>
<td>0.64</td>
<td>5.27e-10</td>
</tr>
<tr>
<td>0,10</td>
<td>0.00</td>
<td>9.77e-14</td>
<td>6.40</td>
<td>0.67</td>
<td>4.25e-03</td>
</tr>
<tr>
<td>9,10</td>
<td>0.33</td>
<td>3.84e-06</td>
<td>6.04</td>
<td>0.67</td>
<td>8.40e-08</td>
</tr>
<tr>
<td>9,01</td>
<td>0.33</td>
<td>2.56e-07</td>
<td>4.60</td>
<td>0.67</td>
<td>5.98e-02</td>
</tr>
<tr>
<td>1,90</td>
<td>0.33</td>
<td>1.50e-01</td>
<td>4.06</td>
<td>0.67</td>
<td>5.25e-09</td>
</tr>
<tr>
<td>1,90</td>
<td>0.33</td>
<td>3.91e-12</td>
<td>0.64</td>
<td>0.67</td>
<td>2.34e-04</td>
</tr>
<tr>
<td>0,91</td>
<td>0.33</td>
<td>3.75e-02</td>
<td>0.46</td>
<td>0.67</td>
<td>1.04e-06</td>
</tr>
<tr>
<td>0,19</td>
<td>0.33</td>
<td>1.46e-11</td>
<td>5.50</td>
<td>0.69</td>
<td>1.91e-02</td>
</tr>
<tr>
<td>8,20</td>
<td>0.50</td>
<td>6.48e-05</td>
<td>5.05</td>
<td>0.69</td>
<td>2.52e-08</td>
</tr>
<tr>
<td>8,02</td>
<td>0.50</td>
<td>2.88e-07</td>
<td>0.55</td>
<td>0.69</td>
<td>1.87e-05</td>
</tr>
<tr>
<td>2,80</td>
<td>0.50</td>
<td>1.80e-01</td>
<td>7.21</td>
<td>0.80</td>
<td>1.30e-04</td>
</tr>
<tr>
<td>2,08</td>
<td>0.50</td>
<td>7.03e-11</td>
<td>2.71</td>
<td>0.80</td>
<td>9.61e-02</td>
</tr>
<tr>
<td>0,82</td>
<td>0.50</td>
<td>1.13e-02</td>
<td>2.17</td>
<td>0.80</td>
<td>8.44e-09</td>
</tr>
<tr>
<td>0,28</td>
<td>0.50</td>
<td>9.89e-10</td>
<td>1.27</td>
<td>0.80</td>
<td>3.16e-08</td>
</tr>
<tr>
<td>7,30</td>
<td>0.61</td>
<td>6.48e-04</td>
<td>7.12</td>
<td>0.80</td>
<td>8.64e-06</td>
</tr>
<tr>
<td>7,03</td>
<td>0.61</td>
<td>1.92e-07</td>
<td>1.72</td>
<td>0.80</td>
<td>2.40e-02</td>
</tr>
<tr>
<td>3,70</td>
<td>0.61</td>
<td>1.28e-01</td>
<td>6.31</td>
<td>0.90</td>
<td>1.13e-03</td>
</tr>
<tr>
<td>3,07</td>
<td>0.61</td>
<td>7.50e-10</td>
<td>3.61</td>
<td>0.90</td>
<td>5.98e-02</td>
</tr>
<tr>
<td>0,73</td>
<td>0.61</td>
<td>2.00e-03</td>
<td>6.13</td>
<td>0.90</td>
<td>5.04e-06</td>
</tr>
<tr>
<td>0,37</td>
<td>0.61</td>
<td>3.96e-08</td>
<td>3.16</td>
<td>0.90</td>
<td>7.88e-08</td>
</tr>
<tr>
<td>8,11</td>
<td>0.64</td>
<td>8.64e-06</td>
<td>1.63</td>
<td>0.90</td>
<td>3.74e-03</td>
</tr>
</tbody>
</table>

4 Sample size for translational trials

The question of appropriate size for a translational trial remains open. There appear to be three vehicles for addressing this question if we make decisions based on the empirical information: the progression or other properties of the exact distribution, mean squared error, or the separate bias and variance values. The latter two seem to be the least computationally intensive. There might arise cases where the bias and variance behave differently, so I will suggest a sample size approach using them, rather than the MSE.

Using Equations (5) and (6), the relationship between $\text{Var}(H^*)$, $\text{Bias}(H^*)$, and sample size can be examined quantitatively. Examples are shown in Figure 3 where the variance and bias of the empirical entropy are plotted for several representative true multinomial responses of length 5, and sample sizes ranging from 5 to 35. Figures 4 and 5 show analogous results for multinomial responses of length 3 and 2 respectively. MSE versus sample size is shown in Figure 6.

Qualitatively and quantitatively similar behavior results in all these circumstances. Small sample sizes yield relatively large bias and high variance in the empirical entropy. However, modest increases in sample size yield large proportionate reductions in both bias and variance. Large increases in sample size, as expected, can reduce the bias and variance to low absolute levels. But most of the benefit in relative reductions can be achieved by

![Figure 2](image)

Figure 2 Exact cumulative distributions for empirical entropy as sample size increases from 5 (left) to 35 (right) by 5. The true probabilities underlying all distributions are (0.46, 0.26, 0.15, 0.80, 0.05).
Table 4 Absolute errors ($\times 10^5$) between simulated means and theoretical means, and simulated variances and theoretical variances for selected binomial outcomes. All results are based on 10 000 replications

<table>
<thead>
<tr>
<th>$p$</th>
<th>$n$</th>
<th>$\Delta$Mean</th>
<th>$\Delta$Var</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5</td>
<td>-139</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>-4</td>
<td>-8</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>-2</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>14</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>-22</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>-17</td>
<td>-1</td>
</tr>
</tbody>
</table>

Table 5 Absolute errors ($\times 10^5$) between simulated means and theoretical means, and simulated variances and theoretical variances for selected multinomial outcomes of length 3. All results are based on 10 000 replications

<table>
<thead>
<tr>
<th>$p$</th>
<th>$n$</th>
<th>$\Delta$Mean</th>
<th>$\Delta$Var</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5</td>
<td>-605</td>
<td>121</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>104</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>-15</td>
<td>-9</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>-23</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>-33</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>-37</td>
<td>-4</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>-3</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 6 Absolute errors ($\times 10^5$) between simulated means and theoretical means, and simulated variances and theoretical variances for selected multinomial outcomes of length 5. All results are based on 10 000 replications

<table>
<thead>
<tr>
<th>$p$</th>
<th>$n$</th>
<th>$\Delta$Mean</th>
<th>$\Delta$Var</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5</td>
<td>-39</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>-144</td>
<td>-17</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>-13</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>96</td>
<td>-9</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>-174</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>-63</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>55</td>
<td>0</td>
</tr>
</tbody>
</table>

*Corresponds to Table 3.

sample sizes in the range of 10–20. In addition, higher sample sizes demonstrate diminishing returns with regard to both $\text{Var}(H^*)$ and $\text{Bias}(H^*)$ reduction.

These observations suggest that a useful strategy for selecting the size of a translational trial might be the following. During the design stage, investigators provide their best guess as to the true multinomial
Figure 3 Relationship between bias, variance, and sample size for empirically calculated entropy. The true probabilities used to generate each curve are: $A = (0.2, 0.2, 0.2, 0.2, 0.2)$, $B = (0.46, 0.26, 0.15, 0.08, 0.05)$, $C = (0.01, 0.15, 0.80, 0.03, 0.01)$, $D = (0.50, 0.30, 0.10, 0.50, 0.05)$.

Figure 4 Relationship between bias, variance, and sample size for empirically calculated entropy. The true probabilities used to generate each curve are: $A = (0.33, 0.34, 0.33)$, $B = (0.15, 0.35, 0.50)$, $C = (0.20, 0.75, 0.05)$, $D = (0.10, 0.25, 0.65)$.

Figure 5 Relationship between bias, variance, and sample size for empirically calculated entropy. The true probabilities used to generate each curve are: $A = (0.5, 0.5)$, $B = (0.1, 0.9)$, $C = (0.25, 0.75)$, $D = (0.6, 0.4)$.

Probabilities of response for categories constructed for the outcome measurement. A variance–bias diagram can be constructed for the empirical entropy under the hypothesized probabilities. A sample size is selected such that it provides relative reductions in $\text{Var}[H^*]$ and $\text{Bias}[H^*]$ appropriate to the question at hand. If evidence emerges during the trial that the originally specified response probabilities are substantially incorrect, the process can be repeated with revised probabilities and the sample size projection updated. However, one might not want to use the observed probabilities continuously during the experiment as the true probabilities, because they are highly variable early in the trial.

4.1 Examples

4.1.1 Binary outcome

As an example, suppose we have a translational trial with a binary outcome (beneficial effect or not). Also, suppose investigators hypothesize that the true effect of the treatment is such that 25% of subjects will show a beneficial direction and magnitude of response, and 75% will not. This circumstance corresponds to Curve C in Figure 5. We can see therefore that a sample size of 15 subjects, for example, produces a very substantial reduction in bias and variance of $H^*$ or $\Delta H^*$. In other words, taking the observed frequency of success after a trial in 15 subjects, we know that we will have resolved a substantial fraction of the overall uncertainty regarding the success rate.

This situation appears to represent a slight statistical paradox. The estimate of the success rate from a sample of 20 subjects is not highly accurate in conventional terms. In fact, the width of the 95% confidence interval for the true success rate is greater than ±20% in such a sample. The reduction in variance and bias of the empirical entropy must be interpreted in light of 1) considerable overall uncertainty, and 2) the biological paradigm. It is not the information required for a comparison such as $\hat{p}$ versus $p_0$. Given the minimal amount of structure imposed on the problem, our sample will yield a
substantial fraction of the information available and should be useful for designing the next experiment.

4.1.2 Ordinal outcome

As a second example, consider a circumstance with more uncertainty in the outcome. Suppose a categorical outcome is classified into one of five categories: strong benefit, weak benefit, no change, slightly negative, or harmful. If we expect that the true response probabilities are \( (0.01, 0.15, 0.80, 0.03, 0.01) \) (see Table 1), the bias versus variance behavior can be seen in Curve C, Figure 3. A sample size in the range of 10–20 appears to provide a substantial reduction in the bias and variance of empirical entropy.

4.1.3 Maximum entropy

Investigators might not always be confident specifying all the multinomial outcome probabilities needed for determining the bias and variance of empirical entropy. An alternative would be to hypothesize a mean outcome, leaving open the question of the best choice for the remainder of the distribution. It is sensible to choose the outcome probabilities so as to maximize the entropy of the distribution, conditional on the specified mean (maximum entropy principle). This can be done analytically. If the mean is chosen to be consistent with a uniform distribution, the maximum entropy distribution will, in fact, be the uniform. More generally, maximum entropy for a specified mean will be found in the well-known Gibbs distribution. The Gibbs distribution is specified by

\[
p_i = \frac{e^{\mu_i}}{\sum_{j=1}^{m} e^{\mu_j}}, \quad j = 1, \ldots, m
\]

and

\[
\sum_{j=1}^{m} p_j x_j = \mu
\]

where \( \mu \) is the mean value and \( \xi \) and the \( p_j \)’s are determined jointly by the system of equations. This provides a sample size approach with somewhat weaker assumptions.

Suppose that outcome values are \( x = 1, 2, 3, 4, \) or \( 5 \), corresponding to multinomial classifications appropriate to the study, and the mean outcome value is hypothesized to be 2. The resulting Gibbs distribution is \( p = (0.46, 0.26, 0.15, 0.08, 0.05) \) (with \( \xi = -0.566 \)). The ordering of the probabilities is irrelevant. A bias versus variance curve for this case is shown in Figure 3 (Curve B), indicating that an appropriate sample size might be in the range of 10 to 15.

5 Discussion

My purpose here is to add rigor to translational designs, not to justify small clinical trials generally. Small trials are not necessarily translational. For example, the recent Institute of Medicine monograph on small clinical trials [18] did not mention translational research at all. The important distinctions are the setting, purpose, nature of the outcome, and how the studies are designed, conducted, and interpreted. Small comparative trials using clinical outcomes are almost always the product of resource limitations and are rarely convincing. Translational clinical trials need no special justification – they have been and will continue to be performed broadly.

It is a well-known fact that when precision of estimation is low (and the prior chance of success is small), most of the positive (defined in terms of estimation or testing of a parameter) findings from clinical trials are false positives. A translational trial is no exception if taken out of context as an estimation design. However, it is the nature of such trials to be corroborated by both replication and parallel biological evidence, improving the overall properties of the discovery and early development process.

Investigators are convinced that they learn important information from such studies and the long run evidence supports the truth of this view. Medical science has developed effective treatments for many diseases, in part by discarding failures. The
early development process for most treatments did not proceed through large clinical trials, but was
guided by small targeted studies. These small
experiments are therefore proven to be informative
when properly used. This is the nature of transla-
tional trials.

Having said this, it is important to be aware of
the potential deficiencies of these designs. The most
obvious limitations are the lack of proven clinical
validity for the outcome and imprecise parameter
estimates. Not so obvious is the fact that the
translational trial paradigm partially confounds
three things: 1) the correctness of the disease
paradigm, 2) the selection of a relevant biological
outcome, and 3) the action of the therapy. Errors in
any of these components can masquerade as either a
positive or negative treatment effect. However, the
correctness of the disease paradigm and the
selection of a relevant outcome can be strongly
supported using evidence from earlier studies.

While entropy, information, or overall uncer-
tainty appears to offer a satisfactory framework for
viewing translational trials, it is not as serviceable
for larger studies with definitive clinical outcomes.
From comparative studies, for example, much more
is needed than overall reduction in uncertainty.
Such trials are appropriately designed for precision
in parameter estimates, relative risk estimates, and
absolute risks of side effects. Such trials also have
more sources of variability and are not likely to be
replicated. These characteristics appear to disqualify
overall entropy as the appropriate design objective
for definitive trials.

Even in the setting of translational research, we
do not have a common experiential basis to
understand quantitative measures of overall uncer-
tainty. Clinical researchers do not routinely think in
quantitative terms about their translational trial
designs. But their instincts to restrict studies to the
absolute minimum size to meet the objectives at
hand results in an operational optimization of
information. Formalizing this in terms of entropy
appears to be a potentially useful tool for improving
the methodology of these important trials.

References
1. Eggermont A, Newell H. Translational research in
clinical trials: the only way forward. Eur J Cancer
2. Minna JD, Gazdar AF. Translational research comes of
3. Crowley WF, Jr, Thier, SO. The continuing dilemma in
clinical investigation and the future of American health
care: a system-wide problem requiring collaborative
4. Rustgi AK. Translational research: what is it? Gastroen-
5. Saljo N. Translational study in cancer research. Intern Med
2002; 41: 770–73.
6. Sugarman J, McKenna WG. Ethical hurdles for
10. Applebaum D. Probability and information. Cambridge:
11. Kullback S. Information theory and statistics. New York:
13. Kullback S, Liebler RA. On information and suffi-
14. Bolland TW. The use of the expected-net-gain chart to
illustrate various aspects of sampling and sample design.
15. Lindley DV. On a measure of the information
provided by an experiment. Annals Math Stat 1956; 27:
986–1005.
16. Sebastiani P, Wynn HP. Maximum entropy sampling
Champaign, IL: Wolfram Media/Cambridge University
18. Evans CH Jr, Istead ST, eds. Small Clinical Trials: Issues
and Challenges. Washington, DC: National Academy Press,