Evaluation of Statistical Designs in Phase I Expansion Cohorts: The Dana-Farber/Harvard Cancer Center Experience

Suzanne E. Dahlberg, Geoffrey I. Shapiro, Jeffrey W. Clark, Bruce E. Johnson

Background
Phase I trials have traditionally been designed to assess toxicity and establish phase II doses with dose-finding studies and expansion cohorts but are frequently exceeding the traditional sample size to further assess endpoints in specific patient subsets. The scientific objectives of phase I expansion cohorts and their evolving role in the current era of targeted therapies have yet to be systematically examined.

Methods
Adult therapeutic phase I trials opened within Dana-Farber/Harvard Cancer Center (DF/HCC) from 1988 to 2012 were identified for sample size details. Statistical designs and study objectives of those submitted in 2011 were reviewed for expansion cohort details.

Results
Five hundred twenty-two adult therapeutic phase I trials were identified during the 25 years. The average sample size of a phase I study has increased from 33.8 patients to 73.1 patients over that time. The proportion of trials with planned enrollment of 50 or fewer patients dropped from 93.0% during the time period 1988 to 1992 to 46.0% between 2008 and 2012; at the same time, the proportion of trials enrolling 51 to 100 patients and more than 100 patients increased from 5.3% and 1.8%, respectively, to 40.5% and 13.5% ($\chi^2$ test, two-sided $P < .001$). Sixteen of the 60 trials (26.7%) in 2011 enrolled patients to three or more subcohorts in the expansion phase. Sixty percent of studies provided no statistical justification of the sample size, although 91.7% of trials stated response as an objective.

Conclusions
Our data suggest that phase I studies have dramatically changed in size and scientific scope within the last decade. Additional studies addressing the implications of this trend on research processes, ethical concerns, and resource burden are needed.

The goal of a phase I clinical trial is to assess the safety of a new agent or a combination including an investigational agent, determine the maximum tolerated dose (MTD) and recommended phase II dose, and evaluate the side-effect profile (1–21). The statistical design of such studies most commonly follows the Fibonacci 3 + 3 dose escalation scheme to determine the MTD, although other approaches, including Bayesian designs, have also been implemented (1–20). Six to fifteen patients are traditionally enrolled to a small single expansion cohort for further evaluation of toxicities, making for a final sample size of roughly 30 to 50 patients per phase I trial to evaluate three to six different dose levels (7,20). These trials have historically not been designed to assess efficacy but are conducted with the intention of therapeutic benefit (20–22). Emerging targeted therapies directed against activated pathways have increased the number of trials that restrict enrollment to patients with different cancers with specific biomarkers. Therefore, the clinical trials process is evolving such that phase I studies more frequently include large expansion cohorts or multiple expansion cohorts to study cohorts with high accrual in the first-in-human setting (20,23). Although statistical and clinical literature has been written about traditional phase I trial design and optimal approaches for determination of the MTD, little focus has been placed on the frequency, scientific objectives, and statistical justification of phase I expansion cohorts or their evolving role in the current era of targeted therapies (24,25).

The primary intention of the expansion cohort has evolved into an opportunity to garner efficacy data for particular cancer types and to enroll patients with different tumor sites with similar genetic aberrations that may serve as effective targets of the experimental therapy (20,23–26). Trials with large expansion cohorts do not always provide statistical justification for their sample sizes, even though the total number of patients to be enrolled may approximate or even exceed the number of patients required for evaluation of the agent in a stand-alone phase II trial testing a prespecified hypothesis. The variability in the approach to the design of these cohorts is the topic of much discussion and debate within scientific and institutional review committees because of the lack of established guidelines and because of the perceived inefficiencies of...
enrolling to cohorts. The expansion cohorts have been addressed using published studies, but to our knowledge, the degree to which large phase I trials are becoming more common has not yet been quantified or addressed at the protocol-specific or design level for ongoing studies within a specific institution (25).

To address these issues, the sample size of adult therapeutic phase I trials activated at Dana-Farber/Harvard Cancer Center (DF/HCC), including Dana-Farber Cancer Institute, Massachusetts General Hospital, and Beth Israel Deaconess Medical Center, over the last 25 years was studied, and details of the statistical operating characteristics of phase I expansion cohort designs within trials submitted over the course of 1 recent year were examined to assess the trends of the studies within our institution and to inform recommendations addressing these trends with respect to the research process, ethical concerns, and resource burden that they bear.

Methods

Data Collection
The adult therapeutic phase I trials opened within DF/HCC were identified in our institutional clinical trials database for the years 1988 to 2012 because the planned sample sizes were available in the database. Information about the total planned study sample sizes was extracted from this database. To collect details on the study objectives and statistical designs for the expansion cohorts, we reviewed phase I trials submitted in the year 2011, the most recent year for which all protocols were available for review when this study was initiated while allowing for sufficient time for protocol amendments. For this analysis, all trials must have been identified as phase I in the database; pediatric trials, banking protocols, expanded access trials, and studies with sample sizes (n < 15) too small to warrant identification as a phase I trial were not included in this study. We also excluded phase I/II trials. Each protocol submitted in 2011 was reviewed for details on stated objectives, total expansion cohort sample size, number of expansion cohorts, type of expansion cohort (clinical/per-protocol or genetic) and form of statistical justification. The DF/HCC Institutional Review Board determined that this study did not require approval because the data do not include any identifiable information and did not require human subjects. The clinical trial identifiers for the protocols reviewed are listed in Supplementary Table 1 (available online).

Statistical Analysis
The variables were summarized using descriptive statistics and exact binomial confidence intervals. Two-sided P values were computed using the χ² test and the Kruskal–Wallis test. A P value of less than .05 was considered statistically significant.

Table 1. Summary statistics for planned sample sizes of adult phase I trials from 1988 to 2012 in 5-year intervals

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary statistics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total number of studies</td>
<td>201</td>
<td>305</td>
<td>552</td>
<td>774</td>
<td>864</td>
</tr>
<tr>
<td>No. of phase I trials</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Planned sample size</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25th percentile</td>
<td>25</td>
<td>26.8</td>
<td>28</td>
<td>24.3</td>
<td>36</td>
</tr>
<tr>
<td>Median</td>
<td>30</td>
<td>35</td>
<td>34</td>
<td>33.5</td>
<td>56</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>33.8 (15.2)</td>
<td>37.9 (17.2)</td>
<td>37.9 (16.9)</td>
<td>46.0 (44.1)</td>
<td>73.1 (85.6)</td>
</tr>
<tr>
<td>75th percentile</td>
<td>39</td>
<td>42.5</td>
<td>40</td>
<td>50</td>
<td>80</td>
</tr>
</tbody>
</table>

Results

Between 1988 and 2012, 522 adult therapeutic phase I trials were identified among a total of 2696 adult therapeutic protocols (19.4%) submitted during that period. The total number of therapeutic studies has more than quadrupled during the past 25 years. Although the number of phase I trials has grown at a similar pace, there is a marked trend in the proportion of phase I studies in recent years, increasing to 24.9% of the total number of protocols over the last 5-year period. The size of phase I trials has also grown during that time. During the first 5 years (1988–1992), the average proposed sample size of a phase I therapeutic study opened within DF/HCC was 33.8 patients. As shown in Table 1, these numbers have increased over time to an average proposed sample size of 73.1 patients in the most recent 5 years (2008–2012). Figure 1 displays the median sample size by year, which has statistically significantly increased over time from 25 patients to 55 patients in the most recent year studied (P < .001). Figure 1 also displays a dramatic increase in median sample size starting in 2005 (27–29).

When we categorized total planned sample sizes as shown in Figure 2, we saw that the proportion of trials that planned to enroll fewer than 50 patients dropped from 93.0% during the time period 1988 to 1992 to 46.0% between 2008 and 2012; concurrently, there was a corresponding increase in the proportion of trials enrolling 51 to 100 patients and more than 100 patients from 5.3% and 1.8%, respectively, to 40.5% and 13.5%, respectively, over the 25-year timeframe of evaluation (P < .001).

The total sample size of phase I studies does not directly address the issue of the sample sizes of the expansion cohorts. The 60 phase I clinical trials submitted to DF/HCC in 2011 were reviewed to abstract details on the statistical design and objectives of the expansion cohorts. The median planned sample size of the expansion phase, defined as the sum total of all patients enrolled at the recommended phase II dose across all cohorts, was 27 total patients, and the mean was 58.4 patients; these patients were enrolled to a median of one cohort per trial (mean = 2.23), but 16 of the trials (26.7%) enrolled patients to three or more cohorts as part of the total expansion phase. As shown in Table 2, the median planned sample size of the expansion phase of a study increases with the number of planned expansion cohorts, from 14.5 (mean = 20) patients among studies with a single expansion cohort to 60 (mean = 144.5) patients among studies with at least three cohorts.

When we examined the types of cohorts to which patients were being registered, we found six protocols (10.0%) enrolled to expansion cohorts defined by genetic characteristics not used as eligibility or selection criteria, 17 studies (28.3%) enrolled patients to cohorts defined by genetic aberrations required for protocol...
eligibility, 32 trials (53.3%) registered patients to more traditional clinical cohorts defined, for example, by type of cancer, and five studies (8.3%) enrolled patients to both genetically and clinically defined cohorts.

The statistical operating characteristics of these expansion cohorts varied. Eight protocols (13.3%) provided formal power calculations justifying the sample sizes in the context of a formal hypothesis test with prespecified type I and type II error rates; 16 studies (26.7%) included some type of probability or feasibility statement, and the largest number of trials (n = 36; 60.0%) provided no statistical justification of the sample size and committed only to descriptions of the endpoints to be calculated, despite an average sample size of 30.7 patients among them and objectives stated to include response and progression-free survival among 32 (88.9%) and 16 (44.4%) of them, respectively.

All of the trials (100.0%) listed toxicity as an objective of the expansion cohort, and all but one study (n = 59; 98.3%) included pharmacokinetic or correlative studies; 55 trials (91.7%) stated
that response was a study objective. Progression-free survival was a planned aim of 34 trials (56.7%), whereas overall survival was stated as an endpoint in 17 studies (28.3%). It should be clarified that these protocols generally defined determination of the MTD to be the primary endpoint.

### Discussion

Motivated by the observation that phase I trials are more commonly used to provide early efficacy data and to identify specific target patient populations, our study quantifies the degree to which adult therapeutic phase I studies have changed in sample size over the last 25 years and summarizes the statistical designs and scientific objectives of 60 adult therapeutic phase I trials submitted to DF/HCC in 2011. Our data demonstrate the dramatic increase in the number and size of phase I clinical trials at this cancer center over the last 25 years, as well as the variability or paucity of statistical rigor that accompanies them despite increased interest in efficacy endpoints. When compared with previous reports that reviewed publications, the distinct advantages of our study include a more exhaustive review over time at the protocol- and statistical-design–specific levels within a large National Cancer Institute–funded cancer center, which is essential in light of the fact that many publications of phase I studies fail to report on all expansion cohorts, and especially in light of recent results indicating that nearly one-third of trials remain unpublished (24,25,30).

We recognize the limitation that our study reports on a single institutional experience; however, it is unlikely that phase I trial designs differ considerably from those at other institutions given the breadth of our phase I program, variety of study sponsors, and frequency of multicenter trials. Our results reflect the planned sample sizes of these trials, not the final accrual goals, which are best obtained by reviewing the publications of these trials. Studies activated in 2011 may not yet be closed to accrual and may have been subject to amendments after our review, which in our experience frequently result in increases to phase I sample sizes. In contrast, at any point in time our institution could have declined to participate in amended versions of these studies. Our desire to keep protocol information de-identifiable limits our ability to report on certain aspects of our studies, including maximum sample sizes within any time period.

Given our observation that 60.0% of the phase I studies provided no statistical justification of the expansion cohort sample sizes, the clear trend toward increased sample size over time, and the types of endpoints stated in these protocols, the likely goal is to gather information about efficacy early in the drug development process, in addition to information about toxicity. We maintain, however, that use of the term “phase I” to describe the phase of development does not preclude the need to justify the sample size in the protocol. Establishing stopping rules and formal criteria for termination once the recommended phase II dose has been identified has been discussed previously and should be revisited in the context of this changing paradigm; discussions of the ethical implications of this approach to phase I study conduct should follow suit.

Because of the historically low probability of response on phase I trials and the higher probability of experiencing toxicity at the MTD, expansion cohorts should also be accompanied by sample size justification in the form of a power calculation or probability statement, if for no reason other than clarifying the response rate or other endpoint of scientific interest enough to warrant further studies. In contrast to the observed mean sample size of 73.1 patients in the most recent 5 years of our study, we underscore the idea that large sample sizes are not needed to obtain a hint of efficacy in the phase I setting, where patients tend to be refractory to other available therapies. For example, with an expansion cohort of 20 patients, if the novel agent being tested has a true but unknown response rate of 10.0% in a patient population, there is a 0.88 chance that at least one response will be observed; this probability increases to 0.99 if the response rate is 20.0%. Similar calculations can be made for trials in which the primary goal is screening patients for cancer harboring an unknown mutation that responds particularly well to therapy. Small sample sizes can also be used to conduct formal hypothesis tests using familiar design methodology for one sample tests.

Justification of sample sizes becomes exceedingly complicated when there is interest in enrolling patients to multiple expansion cohorts and enrollment criteria for each group are defined by cancer type and/or detection of a particular genetic aberration. In our study, 46.7% of trials in 2011 planned more than one expansion cohort; 26.7% of all trials were designed with at least three planned expansion cohorts. This approach reduces the number of protocols needed to evaluate safety in each cohort by distinct studies and streamlines the contracting and review processes, but it is also not without its limitations. The statistical designs may consist of any number of cohorts, but in the absence of clearly defined operating characteristics, patients are arguably enrolled to a very large study that does not serve a clear rationale beyond toxicity evaluation. There usually are no reasons to believe that the toxicities observed at the MTD across separate cohorts should differ, however, statements regarding global oversight of toxicity and efficacy across cohorts are typically absent from these protocols. This counters the intent of minimizing the number of patients exposed to ineffective or toxic doses. Large trials also tend to be multi-institutional studies, which reduce the ability to recognize patterns in patient outcomes (20,23). The rules for presentation and publication of phase I trials with multiple expansion cohorts are generally ambiguous, often allowing for incomplete reporting of all expansion cohorts or permitting separate cohorts to be reported as independent standalone studies (37–41).

Factors contributing to our observation, particularly the dramatic change in phase I trials since 2005, include the discovery of genetic drivers of cancer and the corresponding successful
development of therapies targeting those aberrations (42–45). Advances in technology and the subsequent reduction in costs have almost certainly influenced this process (46,47). Although these are positive trends, these trials should be designed in the best possible way to address the therapeutic objectives, as well as the safety and best interest of each individual who participates in the trial.

In conclusion, our data suggest that phase I studies have changed in size and scientific scope over the last 25 years. We recommend that phase I studies justify their planned sample sizes so that any predefined expectation of benefit is stated a priori and provide unambiguous rules determining study completion and success. The appropriate stakeholders—the National Institutes of Health, the Food and Drug Administration, cooperative groups, and pharmaceutical industry—should join in clearly defining the appropriate parameters for what we have called phase I trials, set expectations for statistical justifications of proposed sample sizes, and consider an appropriate designation that goes beyond the designation of phase I. Institutional review boards should set expectations for an appropriate notation that goes beyond the designation of phase I. Institutional review boards should set expectations for an appropriate designation that goes beyond the designation of phase I. Institutional review boards should set expectations for an appropriate designation that goes beyond the designation of phase I.

References

29. Davies RB. Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika. 1987;74(1):33–43.


**Funding**

This study was supported by the Dana Foundation at Dana-Farber Cancer Institute.

**Notes**

The study sponsor had no role in the design of the study; the collection, analysis, and interpretation of the data; the writing of the manuscript; and the decision to submit the manuscript.

**Affiliations of authors:** Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA (SED); Department of Medical Oncology and Department of Medicine, Dana-Farber Cancer Institute and Brigham and Women’s Hospital, Boston, MA (GIS, BEJ); Department of Hematology and Oncology (JWC) and Department of Medicine (JWC), Massachusetts General Hospital, Boston, MA.