Phase II trials play an essential role in the drug development pathway. This step is often where initial activity of a novel agent or a combination of therapeutics is gauged, and phase II trials inform the decision to take a novel agent into the more expensive phase III process. However, conclusions about agent activity as determined in the phase II setting can be highly dependent on trial design (1, 2).

In oncology, allocating development resources appropriately is a complex endeavor. There are more than 1,100 antineoplastic agents in preclinical development and nearly 500 antineoplastic agents undergoing phase I or phase II testing (3, 4). However, many oncology drugs fail to prove efficacy in phase III trials despite putative activity in earlier clinical investigation (5). The costs of drug development and the increasing number of investigational agents underscore the need to rigorously evaluate the efficiency and performance of the drug development process and the phase II trial in particular. Although there has been considerable attention paid to the statistical design of phase II trials in oncology (6–10), there are few studies evaluating trends in trial design and implementation across medical specialties.

We hypothesized that significant quantifiable differences exist between phase II trials conducted by oncologists and those conducted by other medical subspecialties. An analysis of these differences might illuminate the components of phase II design and implementation that best identify the agents that will eventually prove efficacious and cost-effective. We analyzed phase II trials published in a single year, 2002, and represented in the MEDLINE database. Our aim was to compare trial design, characteristics, and conclusions and to identify trends that distinguish the trials conducted in oncology from those conducted by other medical subspecialties.

**Materials and Methods**

**Identification and selection of studies.** We conducted a computerized literature search for all 2002 published phase II trials. We searched MEDLINE using the commercially available SilverPlatter search engine. The search strategy first pulled records that included the term "phase II" and then filtered for studies published in 2002.
“Clinical-Trial-Phase II.” This set was then limited to publication year 2002. The final set included 859 publications.

An algorithm was developed to determine which publications would be included in the analysis (Table 1). Publications were included if they (a) described an original trial explicitly labeled as phase I/II, phase II, or phase II/III; or (b) described a trial that seemed to be a first attempt to test activity of an experimental intervention or combination of interventions. Publications in languages other than English and trials involving nonhuman subjects were excluded. It is likely that our search may have missed some legitimate phase II trials because neither our search criteria nor the MEDLINE database are likely to be exhaustive. Additionally, our criteria included publications from journal supplements. As such, it includes some phase II trials that did not undergo formal peer review.

All abstracts of these 859 publications were reviewed by one of the authors (L. Michaelis). Once the included set was defined, the publications were classified according to predetermined variables. All data for classification were abstracted from the abstracts alone; full publications were not reviewed.

**Classification system.** The inclusion algorithm yielded 703 publications of phase I/II, II, or II/III trials. The following information was recorded for each publication: author; title; journal and citation; brief description of experimental intervention; medical or surgical subspecialty conducting the trial; number of subjects; trial design; and trial conclusions (Table 2). All classification decisions were made solely by the reader based only on the information in the published abstracts. All data were recorded, sorted, and tallied using Microsoft Excel spreadsheet software. Following the classification of data, we did statistical analysis using available $\chi^2$ computational software. There was no correction for multiple testing.

**Results**

**Fields of study.** Our search yielded 703 publications included in MEDLINE for the year 2002, which could be classified, according to our inclusion algorithm, as phase II trials. A total of 586/703 (83%) were trials conducted in the field of oncology. Non-oncology trials accounted for the remaining 117 publications conducted by 26 other medical and surgical subspecialties. Of these, more than a quarter (30/117, 26%) tested drugs or interventions in the subspecialty of infectious disease. The majority of the latter (16/30, 53%) were phase II trials of HIV interventions (Fig. 1).

Of the 586 publications on oncology phase II trials, 33% (191/586) enrolled subjects with primary breast, colorectal, or lung cancers (Fig. 2). Approximately half of all phase II publications (289/586 or 49%) in oncology were trials examining combinations of therapeutic agents, either a new agent added to commonly used regimens or a novel way of scheduling or dosing a combination of Food and Drug Administration (FDA)–approved agents. Approximately 38% (228/586) tested a single agent or modality as monotherapy. These 228 trials included the investigation of FDA-approved agents for new indications or the use of novel clinical entities as monotherapy. The remaining trials examined the timing or application of multimodality therapy, including radiation therapy or surgery in combination with chemotherapy.

**Use of control subjects.** Of the 703 phase II trials examined, only 143 (20%) abstracts mentioned any control subjects, which are subjects not exposed to the experimental intervention in either a prospective or historical design. Although the traditional phase II trials may implicitly invoke historical response rates, we opted against including such trials in the controlled classification. Among oncology trials, only 13% (77/586) of published abstracts mentioned control subjects, compared with 56% (66/117) of published abstracts from other specialties ($P < 0.001$). In 54 of the 66 (81%) non-oncology trials that included a control population, the controls were chosen prospectively and were randomly assigned to the nonintervention arm. In the oncology trials, the controls were much more likely to be either cohort controls or historical control populations ($44/77, 57%; P < 0.01$). Of the 703 phase II trials examined, 50 (7%) were classified as randomized, blinded, and controlled studies. Of those 50, 43 were conducted in specialties other than oncology. Otherwise stated, 50/117 (43%) of non-oncology phase II trials were done with a randomly selected control population that was blinded to the

---

**Table 1. Abstract inclusion algorithm**

<table>
<thead>
<tr>
<th>Question</th>
<th>Yes, include abstract</th>
<th>No, exclude abstract</th>
</tr>
</thead>
<tbody>
<tr>
<td>Was the abstract published during 2002?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Does the article describe an original trial in human subjects?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is the abstract written in English?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Does the article explicitly state that the trial is a phase II, phase I/II, or phase II/III trial?</td>
<td>Yes, include abstract</td>
<td>No, then</td>
</tr>
<tr>
<td>Does the abstract imply that the trial is the first test of activity of a given intervention previously only tested for safety?</td>
<td>Yes, include abstract</td>
<td>No, exclude abstract</td>
</tr>
</tbody>
</table>

---

**Table 2. Classification system: trial design and conclusions**

<table>
<thead>
<tr>
<th>Trial design</th>
<th>Does the abstract explicitly list any control population?</th>
<th>Yes, if yes, then is such a control described as</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No</td>
<td>• randomly assigned, placebo-controlled arm</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• cohort matched</td>
</tr>
<tr>
<td>Does the abstract state or imply that the assignment to intervention is random?</td>
<td>Yes or No</td>
<td></td>
</tr>
<tr>
<td>Does the abstract state or imply that either physicians or patients are blinded to the intervention?</td>
<td>No</td>
<td>• Double-blinded</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Single-blinded</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Not blinded/not stated</td>
</tr>
<tr>
<td>Does the abstract state or imply that the dose of the drug(s) or exposure to the intervention changed as individuals were enrolled?</td>
<td>Yes or No</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2. Classification system: trial design and conclusions**

<table>
<thead>
<tr>
<th>Conclusions</th>
<th>How does the abstract summarize trial findings?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Claims that the intervention is active, that the data support phase III trials of the intervention, or that the data warrant further study of the intervention</td>
</tr>
<tr>
<td></td>
<td>• Claims that the data do not support additional study, that the intervention is not active, or that the intervention is dangerous</td>
</tr>
<tr>
<td></td>
<td>• The abstract does not make any claims about activity of the intervention</td>
</tr>
</tbody>
</table>
intervention, but only 7 (1%) trials conducted by oncology specialists met those same criteria.

In addition, we found that oncology trials examining novel targeted agents were unlikely to use control subjects or even historical controls in their studies. Of the trials looking at monotherapy treatment regimens, 47/228 were trials on novel targeted agents. In this subset, 9/47 or 19% were controlled trials of any kind. Of the 101 trials where cytotoxic chemotherapy was tested as monotherapy, six trials, or 17%, used control agents. All but 3 of the 289 oncology trials looking at combination chemotherapy included cytotoxic therapies, rather than agents thought to work through cytostatic or targeted mechanisms. Of the three oncology trials testing combinations of targeted agents, one included control subjects, and one randomized subjects to varied dosing schedules.

**Trial conclusions.** We attempted to tabulate the overall conclusions of the abstracts. We did this by developing a sorting system that labeled each abstract according to its concluding claim. The following three labels were used:

1. The abstract claims that the intervention is active, that the data support phase III trials of the intervention, or that the data warrant further study of the intervention.
2. The abstract claims that the data do not support additional study, that the intervention is not active, or that the intervention is dangerous.
3. The abstract does not make any claims about activity of the intervention.

We found that 668/703 (95%) abstracts made a concluding claim about the efficacy of their intervention. In 520 of those 668 abstracts (78%), the claim was that the experimental intervention warranted further development or study. Approximately 76% (422/558) of the trials conducted by oncologists made that claim, whereas 98/110 (89%) of trials conducted by other specialties made similar claims about their results. Overall, published phase II trials seem to overwhelmingly advocate additional study of tested interventions. That said, oncology trials were significantly less likely to conclude that the intervention was safe and active and/or worthy of additional study (76% versus 89%; $P < 0.01$).

**Discussion**

Our data show that phase II trials in oncology are much more prevalent in the medical literature than are phase II trials in other medical and surgical subspecialties and, importantly, very different in design. The difference in the number of published phase II trials is one of this analysis’ most striking findings and accounts for the fact that the comparator set in our analysis has many fewer trials. A number of factors may contribute to this finding. First, it is possible that oncologists are conducting phase II trials at the same frequency as other specialties, but that oncology journals are more likely to publish these uncontrolled phase II trials than are journals that focus on other fields. Second, the large number of published phase II oncology trials may reflect the current state of drug development;

![Phase II abstracts: Other Medical Subspecialties](https://www.aacrjournals.org/cancerres/2007/13(8)/2402/PhaseIIabstracts-Otherspecialties.png)
if there are many more novel clinical entities and untested interventions to investigate in the treatment of malignant diseases, there would be more phase II trials done and published. The skewing in the number of published trials might also be due to the current paradigm of cancer treatment, which largely relies on combination chemotherapy. Therefore, new entities undergo phase II testing as monotherapy and also as additions to combination regimens. The number of combinations of interventions potentially leads to a greater number of phase II trials. One issue that we are not able to answer with this study is the impact of this high number of published phase II trials. Does this speed drug development in the specialty? Does it mean that more phase III trials are undertaken, and if so, is that beneficial or harmful?

Our study also found that, overall, relatively few phase II trials used any explicit control subjects in their design, and that phase II trials conducted with oncology subjects were significantly less likely to use controls than were other medical and surgical subspecialties. Is there a fundamental reason why phase II trials should not include controls? Not according to the FDA, which, in their monograph General Considerations for the Clinical Evaluation of Drugs, state that the majority of investigational drugs should be evaluated using placebo and/or active drug controls (11). In describing the requirements for phase II trials, the U.S. Code of Federal Regulations spells out the following criteria: “Phase 2 studies are typically well controlled, closely monitored, and conducted in a relatively small number of patients, usually involving no more than several hundred subjects.”

As noted earlier, traditionally, phase II trials in the practice of oncology have been designed as single-arm trials with historical control data incorporated into the thresholds for response rate. In one commonly used design, the null hypothesis is that the drug’s effect is less than some minimum response rate (12). Thus, failure to reject the null hypothesis suggests that the drug is inactive and should not be tested further for that indication. However, rejection of the null hypothesis does not prove the alternative hypothesis (a response rate much greater than the threshold used for the null hypothesis). Not surprisingly, the positive predictive value is low, which may be due to the lack of concurrent controls (8, 13). It is reasonable for trials of cytotoxics, as single agents, to be conducted as uncontrolled trials, assuming the response rate is the endpoint. However, we found that even when novel agents, thought to act by targeted mechanisms, were tested as monotherapy, the great majority of trials, 81%, did not control for the intervention.

Other medical specialties, according to our data, are not as narrow in their designs. For example, the FDA guidelines for

![Fig. 2. Classification of oncology abstracts by subspecialty.](image-url)
drugs developed for the treatment of rheumatoid arthritis spell out that: "Once a reasonably safe range of doses has been established, randomized, parallel-arm dose comparison trials are ordinarily recommended... Active controlled designs that specify an arm with a well-characterized, known therapy can also be very useful" (14).

Is it ethically difficult for oncology trials to include controls? It has been suggested that, because physicians always have a treatment preference and their primary duty is to their patient, all randomized trials are unethical (15). In life-threatening diseases such as cancer, this may be even more problematic for some. In addition, it can be particularly difficult to randomize oncology patients to placebo, particularly in the United States. That said, many medical conditions (including, for example, HIV infection, sepsis, critical care illnesses, and acute coronary syndromes) conform to these same restrictions. Notably, controlling for a new intervention does necessitate a placebo arm; rather, it merely implies that one group will not be exposed to the experimental intervention. The use of randomized discontinuation, dose ranging, or active controls (i.e., randomization to a standard treatment arm) have all been proposed as ethically acceptable control methods for phase II trials (16, 17). In addition, the use of crossover designs can also reduce the potential exposure of subjects to inactive therapies (18). At a minimum, the use of explicitly described historical controls may allow for improved identification of bias and limitations of generalizability.

As one author writing on the development of pharmaceuticals for congestive heart failure wrote, "Even the most accomplished physicians cannot predict the natural course of patients and thus cannot interpret clinical responses in the absence of a control group" (19). Other alternative methods for controlling the effect of an intervention have been widely proposed and advocated for adoption in oncology trials and other specialties (20, 21). It is beyond the scope of this article to review the variety of designs available for phase II trials. However, our data do show remarkable monotony in the designs chosen for oncology trials as opposed to those chosen in other specialties.

Our study also found that non-oncology trials were more likely to identify interventions that they claim warranted advancement in the development process or additional investigation. Explanations for this discrepancy may well be due to publication bias, and that journals in medical specialties other than oncology are not as likely to publish phase II trials without findings of significant efficacy. Understanding whether publication bias accounts for these findings would require that we investigate the total number of trials conducted rather than just those that are published in the literature. Our survey does not allow for such a determination.

It is clear that a meta-analysis such as ours is too blunt an instrument to tease out many details about trends in phase II design. Among the obvious limitations is that it represents only 1 year’s worth of abstracts, and more time points may be necessary to define whether the identified trends are durable. In addition, there is limitation in using only the abstracts as the source data because many abstracts may not suitably record the details of the trial and the trial design. There may be cases where classification decisions were made based on information in the abstracts that would have differed if the full published study had been examined. However, we believe that any bias in classification decision making was spread out among oncology and non-oncology trials. Overall, we felt that the benefit of surveying such a wide number of trials merited the technique. It is, of course, beyond the scope of an analysis like this one to prove that randomized trials are the most efficient or valuable method of testing drugs at this early phase of development. We do, however, think that this analysis illustrates the remarkable homogeneity of phase II testing strategies in oncology, as opposed to other medical and surgical subspecialties.

Phase II trials serve as a crucial filter point in the pathway of bringing a drug to market. If the trial design is inadequate, three major errors can happen: patients can be needlessly exposed to risk, resources can be wasted by initiating phase III trials on drugs that were poorly evaluated in phase II testing, and new interventions that may actually have utility can be considered failures. We would argue that the ideal phase II trial is a randomized, blinded, and controlled design, whether or not the intervention is expected to work via a cytotoxic or a cytostatic mechanism. We acknowledge that such an ideal may be difficult to achieve; however, it is still a worthy goal. There are at least several examples of well-designed phase II trials that have led to early and efficient phase III approval even in the absence of meeting traditional objective response rate outcomes. For example, a randomized discontinuation phase II trial of the kinase inhibitor sorafenib found only a 4% objective response rate, but nonetheless showed a failure-free survival advantage after randomization, which was later confirmed in a phase III trial (8, 22). The pivotal phase II trial of bevacizumab and 5-fluorouracil for metastatic colorectal cancer used a randomized, placebo-controlled dose-ranging design, and results garnered from that study led the way to successful phase III studies (23).

We see this survey as a starting point for investigations into whether our current paradigm of trial design in oncology is the optimal one for a specialty that needs to accommodate an increasing number of therapies with greater and greater heterogeneity of mechanisms. Finding ways to make the current process more efficient, while also ensuring that good drugs are not missed and that patients are protected, remains the paramount goal. Given the massive societal costs of pharmaceuticals (particularly for oncology), more efficient use of resources for drug development is critical.

References
11. U.S. Department of Health and Human Services, Food and Drug Administration. General considerations...


Phase II Trials Published in 2002: A Cross-Specialty Comparison Showing Significant Design Differences between Oncology Trials and Other Medical Specialties

Laura C. Michaelis and Mark J. Ratain