Predictive Value of Phase I Trials for Safety in Later Trials and Final Approved Dose: Analysis of 61 Approved Cancer Drugs

Denis L. Jardim, Kenneth R. Hess, Patricia LoRusso, Razelle Kurzrock, and David S. Hong

Abstract

Phase I trials use a small number of patients to define a maximum tolerated dose (MTD) and the safety of new agents. We compared data from phase I and registration trials to determine whether early trials predicted later safety and final dose. We searched the U.S. Food and Drug Administration (FDA) website for drugs approved in nonpediatric cancers (January 1990–October 2012). The recommended phase II dose (RP2D) and toxicities from phase I were compared with doses and safety in later trials. In 62 of 85 (73%) matched trials, the dose from the later trial was within 20% of the RP2D. In a multivariable analysis, phase I trials of targeted agents were less predictive of the final approved dose (OR, 0.2 for adopting ± 20% of the RP2D for targeted vs. other classes; P = 0.025). Of the 530 clinically relevant toxicities in later trials, 70% (n = 374) were described in phase I. A significant relationship (P = 0.0032) between increasing the number of patients in phase I (up to 60) and the ability to describe future clinically relevant toxicities was observed. Among 28,505 patients in later trials, the death rate that was related to drug was 1.41%. In conclusion, dosing based on phase I trials was associated with a low toxicity-related death rate in later trials. The ability to predict relevant toxicities correlates with the number of patients on the initial phase I trial. The final dose approved was within 20% of the RP2D in 73% of assessed trials. Clin Cancer Res; 20(2); 281–8. ©2013 AACR.

Introduction

Phase I clinical trials are a critical first step in cancer drug development. A primary objective is defining a maximum tolerated dose (MTD) and the safety of new agents. We compared data from phase I and registration trials to determine whether early trials predicted later safety and final dose. We searched the U.S. Food and Drug Administration (FDA) website for drugs approved in nonpediatric cancers (January 1990–October 2012). The recommended phase II dose (RP2D) and toxicities from phase I were compared with doses and safety in later trials. In 62 of 85 (73%) matched trials, the dose from the later trial was within 20% of the RP2D. In a multivariable analysis, phase I trials of targeted agents were less predictive of the final approved dose (OR, 0.2 for adopting ± 20% of the RP2D for targeted vs. other classes; P = 0.025). Of the 530 clinically relevant toxicities in later trials, 70% (n = 374) were described in phase I. A significant relationship (P = 0.0032) between increasing the number of patients in phase I (up to 60) and the ability to describe future clinically relevant toxicities was observed. Among 28,505 patients in later trials, the death rate that was related to drug was 1.41%. In conclusion, dosing based on phase I trials was associated with a low toxicity-related death rate in later trials. The ability to predict relevant toxicities correlates with the number of patients on the initial phase I trial. The final dose approved was within 20% of the RP2D in 73% of assessed trials. Clin Cancer Res; 20(2); 281–8. ©2013 AACR.

Materials and Methods

Search strategy

Anticancer agents newly approved between January 1990 and October 2012 were identified on the FDA website (7).
Agents approved for the treatment of solid and hematologic malignancies were selected for further analysis. Agents approved for pediatric cancer, supportive care, locoregional treatment, and agents whose basic compound had already been approved before 1990 were excluded. Updated package inserts for each agent were reviewed to identify indications dose schemas and clinical trials leading to approval. An extensive search was concomitantly done through MEDLINE to identify phase I trials for each of the agents selected from the FDA database analysis. We searched for phase I trials of the single agent and for different approved combinations and schedules described in package inserts.

Matching clinical trials

Phase I trials of FDA-approved anticancer agents were matched to correlate doses and toxicities with those leading to approval of the respective agents. The criteria for matching trials were: the phase I trial enrolled patients with nonpediatric cancer and explored either monotherapy (as FDA-approved) or the same combination and schedule as described in the FDA package insert and later clinical trials of the respective agent; the phase I trial started before the later clinical trial; and the matched trials included similar patient populations of solid versus hematologic malignancies. A single phase I trial could have been matched with more than one later clinical trial as long as different tumor types and indications were involved and the studies were referenced in the FDA package insert. If more than one phase I trial met our criteria, the trial conducted in the United States was used as our analysis was restricted to the United States. FDA-approved drugs or combinations with no published trials meeting the above criteria were excluded from analysis.

Data extraction

The phase I trial data extracted included information about trial characteristics, toxicities and dosing endpoints. Information about doses of experimental agents and toxicities from matched later trials were also obtained. From these data, we enumerated the number and type of clinically relevant toxicities in later trials defined as: treatment-related toxicities leading to death, treatment delays and discontinuations, and toxicities among the three most frequent grade 3/4 laboratory and nonlaboratory toxicities with an overall incidence of at least 1%. Targeted agents were defined in our study as any agent with extra- or intracellular targets different than those of cytotoxic chemotherapy (8), excluding immunotherapy, antihormones, or conjugates. When the agent was approved in combination, the experimental agent was used for classification.

Toxicities were graded according to the criteria used in each trial. Similar toxicities were categorized under the same group as long as they were not exclusionary. All deaths reported by investigators as "possibly," "probably," or "definitely" related to treatment were considered toxicity-related deaths.

Statistical analysis

Toxicity correlations were summarized using descriptive statistics. We determined for each matched comparison the percentage of phase I DLTs present among the four most frequent grade 3 and 4 laboratory and nonlaboratory toxicities on matched later trials. This analysis was based on the following question: was the DLT represented in the four most frequent grade 3/4 adverse events in the matched later trial? We dichotomized the answer for this question as "yes," when at least 50% of DLTs were represented, or "no." We also described the absolute percentage of DLTs captured on later trials as the top four high-grade toxicities.

In addition, we determined how often clinically relevant toxicities from later trials occurred in respective phase I trials. When a clinically relevant toxicity was a composite adverse event, such as infection, at least one component should have been present in phase I for the toxicity to be considered. This analysis was also based on a question: Were clinically relevant toxicities in later trials described in the respective phase I trial? Again, the threshold for a "yes" answer was that at least 50% of toxicities were described in the phase I trial.

Rates were calculated by dividing total number of toxicity-related deaths by total number of patients assessed for toxicity. Comparing treatment-related toxicity death rates was performed with a truncated Poisson model for the positive counts and a binomial logistic regression model for zeros versus nonzero counts. The doses of experimental agents on trials leading to drug approval were compared with the RP2Ds and percentage relation of calculated RP2D.

Proportions from independent groups were compared using the $\chi^2$ test. Multiple logistic regression was used to adjust these comparisons for other possible confounding study factors. A running lines smoother was used to graphically depict how the probability of a binary variable changed according to values of an interval-scaled covariate. Concordance between binary variables measured for the same units was assessed using the $\kappa$ statistic. Analyses were performed using S+ 8.2 for Windows (TIBCO Software Inc).

Results

Search results

Between January 1990 and October 2012, we identified 91 FDA-approved drugs for nonpediatric cancer treatment. After applying previously described criteria, we included 61 approved drugs, 78 phase I trials with different approved schedules and/or combinations, and 88 later clinical trials, each referenced in respective FDA package inserts (Supplementary Fig. S1A). A total of 33,715 patients with cancer were included, 3,499 from phase I trials and 30,216 from later trials leading to regulatory drug approval. Trials included in the analysis are given in Supplementary Tables S1 and S2.
Overall, we performed 88 matched comparisons of phase I trials and their respective later clinical trial(s) to correlate drug-related toxicities and drug doses. The number of matched comparisons included in each subanalysis varied according to the availability of information required for each analysis. The characteristics of the agents and their supporting phase I trials are depicted in Supplementary Table S3.

**Comparison of RP2D and doses used in later trials**

For analysis, three matched trials were excluded because the optimal dose on phase I was not clearly defined. Overall, in 51 (60%) of 85 matched comparisons the dose of the agent adopted in the later trial was identical to the RP2D, but in 62 (73%) of them the dose was within 20% of the RP2D (Fig. 1A). In 73% of phase I trials, the RP2D was determined on the basis of a MTD defined by the protocol; when type of agent was examined 97% of cytotoxics used MTD for RP2D versus 58% of targeted agents (P = 0.0001). We found no difference in predicting future dose of the agent if the RP2D was based on MTD versus not (58% vs. 64%, respectively; P = 0.59).

In the exploratory analysis, neither patient numbers nor number of dose levels of phase I trials or year of agent approval predicted the future dose of the agent (Supplementary Table S4). These results were reproduced when variables were analyzed as continuous covariates in logistic regression models (Supplementary Fig. S2). In the multivariate analysis, phase I trials with targeted agents performed poorly in relation to nontargeted agents in the odds...
of predicting a future dose within 20% of RP2D [OR, 0.2; 95% confidence interval (CI), 0.03–0.8; \( P = 0.025 \); Table 1].

### Relative frequency of DLTs in trials leading to agent approval

Of the 78 phase I trials involved in matched comparisons, 13 reported no DLTs (5 hormonal, 4 targeted, 3 radioimmune conjugates, and 1 cytotoxic); hence, 75 matched comparisons were included in this analysis. The median number of DLTs per phase I trial was 2 (0–11). Considering only trials presenting DLTs, the median number of DLTs per trial was 2.5 (1–8) for cytotoxic agents and 3 (1–11) for targeted agents. Overall, 53% (109 of 204) of DLTs described in all phase I trials were among the four most common grade 3/4 laboratory and nonlaboratory toxicities in later trials leading to regulatory agent approval. Twenty-one of 75 matched comparisons included (28%) did not meet the prespecified criteria of having at least 50% of DLTs in the respective phase I trial described as “top 4” grade 3 and 4 toxicities in later trials. Phase I trials with cytotoxic agents showed a trend to have DLTs represented in the four most frequent grade 3/4 adverse events of later trials than phase I trials with targeted agents \( (P = 0.012) \). Similar trends were observed for nonsolid versus solid tumors and single center versus multicenter phase I trials \( (P = 0.067 \text{ and } 0.16, \text{ respectively; Table 2}) \).

### Ability of phase I trials to describe later clinically relevant toxicities

A total of 4 of 88 matched comparisons were excluded from analysis because the registration trial did not describe toxicities that met our criteria for clinical relevance. We identified 530 clinically relevant toxicities in the later trials, of which 374 (70.6%) were at least cited in respective phase I trials and 126 (24%) described as DLTs. In 16 of 84 matched comparisons the phase I trial described less than 50% of clinically relevant toxicities. The factors that were significantly related to a better performance in describing these toxicities were the higher number of patients included in a phase I trial \( (P = 0.026) \) and the inclusion of diverse tumor types \( (P = 0.024; \text{ Table 2}) \).

In addition, the relationship between number of patients included in a phase I trial and probability of describing at least 50% of clinically relevant toxicities in later trials was significant \( (P = 0.0032, \text{ logistic regression with number of patients as a continuous variable; Fig. 1B}) \). Beyond 60 patients on the phase I trial, there was no significant improvement in the ability to describe clinically relevant toxicities in later trials.

As mentioned above, of the 78 phase I trials that were involved in matched comparisons, 13 described no DLTs, and in only 3 of the later trials of these agents were clinically relevant toxicities not described.

### Concordance between high-grade toxicities and mortality in early and later trials

We analyzed the concordance of reporting grade 3 and 4 toxicities as one of the top 3 toxicities in phase I and later trials matched for drugs. Of 28 toxicities occurring frequently enough for analysis, 20 had poor agreement \( (k < 0.4) \), 7 had fair to good agreement \( (k 0.4–0.75) \), and only the occurrence of “fever/chills” had excellent agreement \( (k > 0.75) \) in the context of being reported as the top grade 3 and 4 toxicities among both early and later trials (Table 3).

Among 3,499 participants in all phase I studies, 21 deaths (0.6%) were considered at least possibly related to treatment. In later trials with information about treatment-related deaths, 402 toxicity-related deaths were registered for 28,505 participants (1.41%). The treatment-related death rates were 1.01%, 1.44%, and 1.74% depending whether the trial adopted a dose less than 80%, 80 to 120%, or more than 120% of RP2D, respectively, \( (P = 0.0015) \).

### Discussion

Dosing based on phase I trials was associated with a safe toxicity-mortality rate in later trials that led to FDA agent approval.

---

**Table 1.** Multiple logistic regression analysis correlating phase I trial characteristics and chances of establishing a later dose within 20% of RP2D \( (N = 82) \)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Contrast</th>
<th>OR (95% CI)</th>
<th>( P )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Targeted agents</td>
<td>Yes vs. no</td>
<td>0.2 (0.03–0.8)</td>
<td>0.025</td>
</tr>
<tr>
<td>Exclusive solid tumors</td>
<td>Yes vs. no</td>
<td>0.2 (0.03–1.2)</td>
<td>0.08</td>
</tr>
<tr>
<td>Combination treatment</td>
<td>Yes vs. not</td>
<td>4.2 (0.7–27)</td>
<td>0.13</td>
</tr>
<tr>
<td>Phase I population</td>
<td>Specific vs. varied</td>
<td>0.3 (0.1–1.6)</td>
<td>0.16</td>
</tr>
<tr>
<td>PO Route</td>
<td>Yes vs. no</td>
<td>3.3 (0.5–20)</td>
<td>0.20</td>
</tr>
<tr>
<td>Number of dose levels</td>
<td>&gt;6 vs. ( \leq 6^{a} )</td>
<td>0.3 (0.05–1.9)</td>
<td>0.22</td>
</tr>
<tr>
<td>Dose type</td>
<td>Variable vs. flat</td>
<td>0.3 (0.03–2.4)</td>
<td>0.24</td>
</tr>
<tr>
<td>Number of patients</td>
<td>&gt;36 vs. ( \leq 36^{a} )</td>
<td>2.0 (0.4–11)</td>
<td>0.42</td>
</tr>
<tr>
<td>Number of sites of phase I</td>
<td>Multiple vs. single</td>
<td>1.1 (0.3–3.6)</td>
<td>0.90</td>
</tr>
<tr>
<td>Approval</td>
<td>( &gt;2004 ) vs. ( \leq 2004^{a} )</td>
<td>0.9 (0.3–3.3)</td>
<td>0.92</td>
</tr>
</tbody>
</table>

**NOTE:** Values with statistical significance are in bold.

\( ^{a} \text{Cutoff points for analysis were established on the basis of the median values of the distribution for the variables analyzed.} \)
approval. Most-later trials (73%) adopted a dose within 20\% of the RP2D found in the phase I trial. In addition, a substantial proportion (70\%) of clinically relevant toxicities found in the registration trials were described in earlier trials.

Previous studies showed that identifying the RP2D is a heterogeneous process (9). For traditional cytotoxic agents higher doses and increased exposure to treatment correlate with better results (10) with dose determination driven by toxicities. In contrast, targeted agents produce broader activity at different dose levels (11) and parameters used to delineate the RP2D can include pharmacodynamic and pharmacokinetic information (12), although most targeted drugs in early trials are subject to traditional endpoints to determine a MTD (13). Importantly, as MTD is determined within a short window of time (3–4 weeks) in a small number of patients, chronically used targeted agents will likely result in toxicities that are not detected during the MTD window (2). Indeed, after adjusting for confounding effects in the multivariate analysis, our data suggested that phase I trials with targeted agents are less predictive of the final dose used in trials leading to FDA approval of cancer drugs compared with other classes of agents.

Importantly, though including a limited number of patients and multiple dose levels of drugs, early trials seem to achieve their key objective assuring safety during later steps of drug development. Indeed, among approximately

### Table 2. Correlation of relevant safety information from phase I trials and later trials

<table>
<thead>
<tr>
<th>Trial/agent characteristic</th>
<th>Was DLT represented in the four most frequent grade 3/4 adverse events of later trials?</th>
<th>Were the clinically significant toxicities on later trials described on the respective phase I trial?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>“yes” (%)</td>
</tr>
<tr>
<td>Overall</td>
<td>75</td>
<td>54 (72)</td>
</tr>
<tr>
<td>Route^d</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drug class</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cytotoxic</td>
<td>36</td>
<td>29 (81)</td>
</tr>
<tr>
<td>Targeted</td>
<td>33</td>
<td>21 (64)</td>
</tr>
<tr>
<td>Other</td>
<td>6</td>
<td>4 (67)</td>
</tr>
<tr>
<td>Monotherapy or not</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single agent</td>
<td>62</td>
<td>45 (73)</td>
</tr>
<tr>
<td>Combination</td>
<td>13</td>
<td>9 (69)</td>
</tr>
<tr>
<td>Number of patients</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11–36</td>
<td>37</td>
<td>28 (76)</td>
</tr>
<tr>
<td>37–153</td>
<td>38</td>
<td>26 (68)</td>
</tr>
<tr>
<td>No. of Sites</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single center</td>
<td>27</td>
<td>22 (81)</td>
</tr>
<tr>
<td>Multicenter</td>
<td>44</td>
<td>29 (66)</td>
</tr>
<tr>
<td>Exclusive solid tumors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>18</td>
<td>16 (89)</td>
</tr>
<tr>
<td>Yes</td>
<td>57</td>
<td>38 (67)</td>
</tr>
<tr>
<td>Dose type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flat</td>
<td>28</td>
<td>19 (68)</td>
</tr>
<tr>
<td>Variable</td>
<td>47</td>
<td>35 (74)</td>
</tr>
<tr>
<td>Year drug approval</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1990–2004</td>
<td>34</td>
<td>26 (76)</td>
</tr>
<tr>
<td>2005–2012</td>
<td>41</td>
<td>28 (68)</td>
</tr>
<tr>
<td>Phase I population</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Varied tumors</td>
<td>60</td>
<td>43 (72)</td>
</tr>
<tr>
<td>Tumor-specific</td>
<td>15</td>
<td>11 (73)</td>
</tr>
</tbody>
</table>

NOTE: Values with statistical significance are in bold.

^aAt least 50\% of DLTs should have been among the 4 most frequent laboratory and nonlaboratory adverse events of later trial for a “yes.”

^bAt least 50\% of clinically relevant toxicities should have been described on a phase I trial for a “yes.”

^cAfter omitting “other.”

^dOne agent administered by subcutaneous route was not included in this analysis.
29,000 patients enrolled in later trials, only 1.41% succumbed to toxicities at least possibly related to treatment. The treatment-related mortality presented a significant relationship with the percentage of RP2D adopted in later trials, with higher doses associated with higher mortality. These data should be interpreted cautiously as safety and survival endpoints might not be correlated (14). Previous reviews reported a global toxicity mortality rate range between 0.4% and 0.59% for oncologic agents in early trials (15–17). In agreement, we report a rate of 0.6% for deaths possibly due to toxicity of agents in phase I trials of drugs subsequently approved by the FDA.

There are also selection differences in phase I versus later trials; the former usually recruit patients having more advanced cancers and varied tumor types. Drug toxicities might differ amongst these populations. Nevertheless, there was high predictive ability for phase I trials to identify clinically relevant toxicities. Attribution of toxicities to experimental agents is also not straightforward in phase I trials and errors are frequent in early trials (18). Interestingly, toxicities easily attributable to experimental drugs, such as hand–foot syndrome, neurotoxicity, peripheral neuropathy, neutropenia, and mucositis, were most concordant when comparing top 3 grade 3 and 4 toxicities between early and later trials.

Our DLT analysis showed approximately 50% of DLTs not the most frequent high-grade events in later clinical trials. Previous studies comparing DLTs in phase I trials of cytotoxic versus targeted agents showed that DLTs are more frequently identified and are more prone to be hematologic toxicities in phase I trials of cytotoxic agents (4, 19). Our findings also suggest that DLTs in phase I trials of cytotoxic agents are more likely to be among frequently occurring high-grade toxicities during later stages of drug development compared with targeted agents.

In addition, we described a trend for single institutional phase I trials to perform better in describing DLTs that are

### Table 3. Concordance in top 3 grade 3/4 toxicities between earlier and later trials

<table>
<thead>
<tr>
<th>Toxicity</th>
<th>Ne</th>
<th>I, not L</th>
<th>L, not I</th>
<th>Both</th>
<th>( \kappa ) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excellent agreement ( \kappa \geq 0.75 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fever/chills</td>
<td>82</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>0.79 (0.50–1.00)</td>
</tr>
<tr>
<td>Fair to good agreement ( \kappa 0.4–0.75 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diarrhea</td>
<td>52</td>
<td>10</td>
<td>11</td>
<td>15</td>
<td>0.42 (0.21–0.63)</td>
</tr>
<tr>
<td>Hand-foot syndrome</td>
<td>77</td>
<td>0</td>
<td>6</td>
<td>5</td>
<td>0.59 (0.31–0.88)</td>
</tr>
<tr>
<td>Hyperbilirubinemia</td>
<td>78</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>0.42 (0.07–0.77)</td>
</tr>
<tr>
<td>Mucositis/stomatitis</td>
<td>73</td>
<td>7</td>
<td>2</td>
<td>6</td>
<td>0.52 (0.25–0.79)</td>
</tr>
<tr>
<td>Neurotoxicity</td>
<td>80</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>0.52 (0.15–0.88)</td>
</tr>
<tr>
<td>Neutropenia</td>
<td>28</td>
<td>4</td>
<td>14</td>
<td>42</td>
<td>0.59 (0.42–0.75)</td>
</tr>
<tr>
<td>Peripheral neuropathy</td>
<td>81</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>0.41 (0.00–0.83)</td>
</tr>
<tr>
<td>Poor agreement ( \kappa \leq 0.4 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abdominal pain</td>
<td>81</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>-0.03 (-0.07–0.00)</td>
</tr>
<tr>
<td>Anemia</td>
<td>40</td>
<td>10</td>
<td>28</td>
<td>10</td>
<td>0.07 (-0.12–0.26)</td>
</tr>
<tr>
<td>Anorexia</td>
<td>81</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>-0.04 (-0.07–0.01)</td>
</tr>
<tr>
<td>Dyspnea</td>
<td>80</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>0.18 (-0.19–0.55)</td>
</tr>
<tr>
<td>Fatigue/asthenia</td>
<td>28</td>
<td>8</td>
<td>28</td>
<td>24</td>
<td>0.22 (0.04–0.40)</td>
</tr>
<tr>
<td>Febrile neutropenia</td>
<td>75</td>
<td>4</td>
<td>6</td>
<td>3</td>
<td>0.31 (-0.01–0.63)</td>
</tr>
<tr>
<td>Hepatotoxicity</td>
<td>70</td>
<td>7</td>
<td>7</td>
<td>4</td>
<td>0.27 (-0.01–0.55)</td>
</tr>
<tr>
<td>Hyponatremia</td>
<td>83</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0.31 (-0.18–0.80)</td>
</tr>
<tr>
<td>Hypophosphatemia</td>
<td>80</td>
<td>2</td>
<td>6</td>
<td>0</td>
<td>-0.04 (-0.07–0.00)</td>
</tr>
<tr>
<td>Infection</td>
<td>73</td>
<td>0</td>
<td>13</td>
<td>2</td>
<td>0.20 (-0.04–0.44)</td>
</tr>
<tr>
<td>Leukopenia</td>
<td>67</td>
<td>7</td>
<td>9</td>
<td>5</td>
<td>0.30 (0.02–0.54)</td>
</tr>
<tr>
<td>Lymphopenia</td>
<td>73</td>
<td>6</td>
<td>9</td>
<td>0</td>
<td>-0.09 (-0.14–0.04)</td>
</tr>
<tr>
<td>Musculoskeletal</td>
<td>77</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>0.38 (0.05–0.72)</td>
</tr>
<tr>
<td>Nausea</td>
<td>61</td>
<td>14</td>
<td>9</td>
<td>4</td>
<td>0.10 (-0.12–0.33)</td>
</tr>
<tr>
<td>Pain</td>
<td>79</td>
<td>2</td>
<td>7</td>
<td>0</td>
<td>-0.04 (-0.08–0.00)</td>
</tr>
<tr>
<td>Rash</td>
<td>75</td>
<td>3</td>
<td>7</td>
<td>3</td>
<td>0.32 (0.00–0.63)</td>
</tr>
<tr>
<td>Skin toxicity</td>
<td>81</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>Thrombocytopenia</td>
<td>29</td>
<td>13</td>
<td>16</td>
<td>30</td>
<td>0.34 (0.15–0.54)</td>
</tr>
<tr>
<td>Vomiting</td>
<td>75</td>
<td>3</td>
<td>7</td>
<td>3</td>
<td>0.32 (0.00–0.63)</td>
</tr>
</tbody>
</table>

Abbreviations: I not L, occurrences only on phase I trial but not on later trial; L not I, occurrences only on later trial but not on phase I trial; Ne, neither trials.
re relevant to the future profile of drugs. Previous authors identified difficulties in recognizing relevant drug-induced side effects when multiple investigators are involved, due to unfamiliarity with drugs and number of patients diluted among multiple institutions (20). These data suggest that multiinstitutional phase I trials do not perform better in describing the safety of phase I agents.

Not surprisingly, the ability to predict clinically relevant toxicities in later trials correlated strongly with the number of patients in the phase I trial (up to ∼60 patients). Therefore, limited expansion of the number of patients on phase I trials is likely warranted if the objective of predicting toxicities is deemed important. Our data suggest, however, that no gain is acquired by including more than 60 patients when the objective is to delineate important toxicities that will occur in future clinical trials. There may, however, be other reasons for expanding these trials, such as better assessing responses in subsets of disease before proceeding with drug development (21).

There are several limitations to our study. For instance, our findings were generated using cancer drugs subsequently approved by the FDA. Obtaining similar correlations for drugs that were submitted to but not approved by the FDA might delineate how the lack of correlation between early and later trials could impact the ultimate success of drug approval. However, many of these failed studies are never published, so analysis and comparisons are problematic. Second, our data are based on reported toxicities and so it is not possible to know to what extent our findings are influenced by deficiencies in data reporting. Third, correlating phase I trials with later trials was only possible after applying specific criteria, as outlined in the Materials and Methods. Given these restrictive criteria, a few important drugs had to be excluded and the impact of their absence from our analysis is not known. Finally, as most phase I trials used the traditional 3 + 3 design, we cannot extend our findings to model-based designs.

In conclusion, early trials performed well in terms of dose prediction and description of safety profile of cancer drugs. Indeed, dosing based on phase I trials was associated with a low toxicity–mortality death rate on later trials. The ability to predict clinically relevant toxicities correlated with the number of patients on the initial phase I study.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Authors’ Contributions
Conception and design: D.L. Jardim, D.S. Hong
Development of methodology: D.L. Jardim, D.S. Hong
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): D.L. Jardim, P. LoRusso
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): D.L. Jardim, K.R. Hess, P. LoRusso, R. Kurzrock, D.S. Hong
Writing, review, and/or revision of the manuscript: D.L. Jardim, K.R. Hess, P. LoRusso, R. Kurzrock, D.S. Hong
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): D.L. Jardim
Study supervision: D.L. Jardim, P. LoRusso, D.S. Hong

Acknowledgments
The authors thank Joann Aaron, Department of Investigational Cancer Therapeutics, for editorial support.

Received August 10, 2013; revised October 18, 2013; accepted October 28, 2013. Published OnlineFirst November 4, 2013.

References
18. Mukherjee SD, Coombs ME, Levine M, Cosby J, Kowaleski B, Arnold A. A qualitative study evaluating causality attribution for serious


Predictive Value of Phase I Trials for Safety in Later Trials and Final Approved Dose: Analysis of 61 Approved Cancer Drugs

Denis L. Jardim, Kenneth R. Hess, Patricia LoRusso, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-13-2103

Supplementary Material
Access the most recent supplemental material at:
http://clincancerres.aacrjournals.org/content/suppl/2013/11/05/1078-0432.CCR-13-2103.DC1

Cited articles
This article cites 20 articles, 12 of which you can access for free at:
http://clincancerres.aacrjournals.org/content/20/2/281.full.html#ref-list-1

Citing articles
This article has been cited by 6 HighWire-hosted articles. Access the articles at:
/content/20/2/281.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.