The BATTLE Trial: A Bold Step toward Improving the Efficiency of Biomarker-Based Drug Development

Eric H. Rubin, Keaven M. Anderson, and Christine K. Gause

Summary: Successful completion of the Biomarker-integrated Approaches of Targeted Therapy for Lung Cancer Elimination (BATTLE) trial, reported in this issue of Cancer Discovery, is an important advance in the effort to improve clinical trial approaches to the simultaneous development of new therapeutics with matching diagnostic tests so that patients most likely to benefit from these therapies can be identified. Cancer Discovery; 1(1); 17–20. © 2011 AACR.

Commentary on Kim et al., p. 44 (1).

The Problem That BATTLE Was Designed to Solve

Advances in basic cancer research have led to a widely used discovery and development approach for drugs designed to inhibit specific cancer pathways. However, clinical trial designs have not kept pace with basic research advances, and use of traditional, histology-based, “all-comers” phase I and II trial designs for these drugs has led typically to failure in phase III studies or demonstration of “success” based on statistically significant but clinically questionable benefit in an all-comers population. For example, Table 1 lists 9 drugs that failed in 13 phase III trials of unselected (i.e., in the absence of a diagnostic test that predicts tumor responsiveness to a drug) patients with non–small cell lung cancer (NSCLC) during or after the completion of the BATTLE study (1). Indeed, only 2 drugs that target signaling pathways have been approved by the FDA for the treatment of unselected NSCLC patients: erlotinib [small molecule inhibitor of the epidermal growth factor receptor (EGFR) tyrosine kinase] and bevacizumab [monoclonal antibody targeting vascular endothelial growth factor (VEGF)]. Clearly, the approach to developing these kinds of drugs for NSCLC and other cancers needs to change, or we will continue to waste precious clinical trial resources in futile studies.

Table 1. Signaling pathway–targeting compounds that failed in phase III trials involving unselected NSCLC patients

<table>
<thead>
<tr>
<th>Agent</th>
<th>Target</th>
<th>Trial design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bexarotene</td>
<td>RXR</td>
<td>Add-on to vinorelbine/cisplatin</td>
</tr>
<tr>
<td>Sorafenib</td>
<td>Multikinase</td>
<td>Add-on to carboplatin/paclitaxel</td>
</tr>
<tr>
<td>Vandetanib</td>
<td>Multikinase</td>
<td>Add-on to gemcitabine/cisplatin</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Add-on to erlotinib</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Add-on to docetaxel*</td>
</tr>
<tr>
<td>Bevacizumab</td>
<td>VEGF</td>
<td>Add-on to carboplatin/paclitaxel</td>
</tr>
<tr>
<td>Cediranib</td>
<td>VEGFR</td>
<td>Add-on to carboplatin/paclitaxel</td>
</tr>
<tr>
<td>Figitumumab</td>
<td>IGF1R</td>
<td>Add-on to carboplatin/paclitaxel</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Add-on to erlotinib</td>
</tr>
<tr>
<td>Lonaafarnib</td>
<td>Farnesyltransferase</td>
<td>Add-on to carboplatin/paclitaxel</td>
</tr>
<tr>
<td>PFS3512676</td>
<td>TLR9</td>
<td>Add-on to carboplatin/paclitaxel</td>
</tr>
<tr>
<td>Vadimezan</td>
<td>Tumor vasculature</td>
<td>Add-on to carboplatin/paclitaxel</td>
</tr>
</tbody>
</table>

NOTE: Shading indicates compounds that were studied in the BATTLE trial.
Abbreviations: RXR, retinoid X receptor; TLR9, Toll-like receptor 9.
*Met primary progression-free survival end point of a hazard ratio <0.80, no difference in overall survival.
Although this problem is well recognized and frequently discussed, in practice little has changed in the use of traditional phase I and II trial designs in cancer drug development. In fact, this lack of change has resulted from the difficulty of selecting a diagnostic test to identify responsive patient subgroups in early clinical trials, and even when a test has been selected, it has often been incorrect. A good example is the selection of EGFR protein expression to predict responsiveness to the EGFR-targeting antibody cetuximab. This choice was rational, based on preclinical studies of the antibody–receptor interaction (2), but subsequent studies indicated that EGFR protein expression, as assessed by immunohistochemical (IHC) analyses, is not a useful predictor for responsiveness to cetuximab in the clinic (3). Similarly, although intuitively appealing and supported by preclinical experiments (4), insulin-like growth factor 1 receptor (IGF1R) protein expression alone has not been useful in selecting patients who will benefit from treatment with IGF1R-targeting antibodies (5). These kinds of errors are, in some measure, due to the frequent depiction of cancer pathways as relatively simple, well-understood linear networks, when in reality our understanding of these pathways, and their perturbation by therapeutics, remains poor. In addition, agnostic, systems biology approaches to identifying “molecular signatures” for responsive patient subgroups have been hampered by the requirement for relatively large clinical datasets for signature validation, which are not available in early phase I and II trials of a new anticancer agent. Further, molecular signatures derived from preclinical studies have not yet proved reliable for predicting benefit in the clinic (6, 7). Thus, despite major advances in understanding cancer biology over the past 30 years, only 8 diagnostic tests used to select responsive patient subgroups are included in cancer drug labels (estrogen receptor IHC, HER-2 IHC/DNA hybridization assay, EGFR IHC, C-KIT IHC, BCR-ABL chromosome, PML-RAR chromosome, 5 del chromosome, and RAS mutation), with only 3 of these FDA approved (HER-2 IHC/DNA hybridization assay, EGFR IHC, and C-KIT IHC).

BATTLE TRIAL DESIGN AND WHAT WE CAN LEARN FROM A DRUG DEVELOPMENT PERSPECTIVE

The BATTLE trial investigators must be recognized and congratulated as bold innovators in their efforts to move beyond traditional clinical trial designs that are ineffective in simultaneously developing a new therapeutic and a matching diagnostic test. The investigators showed that treatment allocation based on results from multiple assays performed on computed tomography–guided biopsy specimens is feasible and associated with minimal safety risk. The study used an adaptive randomization design for this allocation, which was based on ongoing analyses of the rate of 8-week disease control obtained for 20 biomarker–treatment groups (4 treatments, with 5 biomarker groups, yields 20 combinations; the number of patients in some groups was small, as shown in Table 2 of ref. 1). The results indicate that 8 of the 20 biomarker–treatment matches met the predefined criterion for efficacy: a >80% probability of achieving a >30% 8-week disease control rate (DCR; ref. 1). Some matches that met the efficacy criterion are consistent with our current understanding of markers predictive for drug efficacy, such as KRAS/BRAF mutations predicting for response to sorafenib (8). However, other matches are more difficult to understand, including the finding that the highest DCR for erlotinib (40%) was in the VEGF/VEGFR-2 biomarker group. One might have expected the highest DCR for erlotinib to have been in the EGFR biomarker group (which had a 35% DCR that did not meet the efficacy criterion).

Another consideration is that issues with the BATTLE study design complicate interpretation of results. First, as noted by the investigators, partial exclusion of patients with previous erlotinib treatment confounded the adaptive randomization process because these patients could be randomized to only 2 of the 4 treatment arms. Second, technical qualifications of the assays used for biomarker grouping were not reported; thus, the choice of cutoffs for these assays (to determine whether a patient’s tumor was “positive” or “negative” for a given biomarker group) could be questioned. These issues make it difficult to conclude that predictive biomarkers have been identified for the treatments. In addition, although the authors assert that the study “validated pre-specified hypotheses regarding predictive biomarkers,” the precise biomarker hypotheses, as well as associated type I and type II statistical errors, are not clear. Thus, the study should be considered as generating a hypothesis rather than as confirming a particular biomarker hypothesis.

From a drug developer perspective, it is of interest to ask whether BATTLE results will influence development plans for the investigational drugs included in the study, or whether BATTLE results might have altered development plans if the results had been available before phase III studies were initiated for bexarotene, sorafenib, or vandetanib, all of which yielded negative results in an unselected NSCLC population (Table 1). With regard to future development, although it is not clear whether the manufacturers of bexarotene, sorafenib, or vandetanib will initiate confirmatory studies in biomarker-defined patient populations identified in BATTLE as potentially responsive to these drugs, 2 BATTLE-like studies have been initiated recently, with support from multiple pharmaceutical companies: BATTLE-FL (front line; NCT01263782) and BATTLE-2 (NCT01248247). BATTLE-FL involves lung cancer patients who are chemotherapy naïve for metastatic disease and includes the combination of pemetrexed and carboplatin as a “control group,” with other treatment arms involving addition of anti-VEGF (bevacizumab), anti-EGFR (cetuximab), or anti-IGF1R (cixutumumab) antibodies to the pemetrexed + carboplatin backbone. Similar to BATTLE, BATTLE-2 involves previously treated lung cancer patients and includes erlotinib and sorafenib treatment arms as well as 2 other investigational treatments: a combination of the AKT inhibitor MK-2206 with the MEK inhibitor AZD6244 and a combination of MK-2206 with erlotinib. As noted in the article by Kim and colleagues (1), this study involves an approach to biomarker selection and tumor classification that is different from the approach of the BATTLE trial. In the first half of the study, clinically validated biomarkers (such as KRAS mutation) will define biomarker groups to be used in adaptive treatment allocation. A limited set of additional prespecified biomarkers will be evaluated in tumor
biopsies. After analysis of results, biomarkers considered to be potentially predictive for response to each experimental treatment will be selected for use in treatment allocation decisions in the second half of the study.

### ALTERNATIVE APPROACHES TO CODEVELOPMENT OF A NEW THERAPEUTIC WITH A MATCHING PREDICTIVE DIAGNOSTIC TEST

Although it is not difficult to point out flaws in the BATTLE trial and to question the significance of results in terms of subsequent development of drugs and biomarkers included in the trial, it is more problematic to suggest an alternative, more efficient approach to codevelopment of new therapeutics with matching predictive biomarkers.

Two major innovations of the BATTLE trial are its operational and statistical approaches. From an operational perspective, BATTLE successfully pioneered an ambitious goal of incorporating 4 different treatment arms (requiring cooperation of 4 different pharmaceutical companies) and 5 different biomarker classifiers within a single study, with treatment allocation based on results of a diagnostic biopsy. Considerable operational efficiency is gained by the lack of “screen failures” versus a traditional approach to selecting only biomarker-positive patients for separate phase II studies. For example, using a traditional single-arm phase II Simon 2-stage approach with null and alternative hypotheses that match those of BATTLE—the null hypothesis is a DCR of 30%, and the alternative hypothesis is a DCR of 50%, with 20% type I error rate and 80% power—with treatment of only biomarker-positive patients in each phase II study (using the data in Supplementary Table S1 of ref. 1 to calculate the frequency of biomarker-positive patients), up to 299 patients would need to be enrolled and undergo biopsy (to obtain 20 biomarker-positive patients for each treatment), compared with the 200 patients required for BATTLE (ref. 9; Table 2). Even with an assumption that all separate phase II studies would stop early, 225 patients (on average) would be needed (Table 2).

Furthermore, among the 4 treatment arms in the BATTLE trial, erlotinib was the only one FDA approved for lung cancer before initiation of the trial. As described previously (10), if the erlotinib arm is viewed as a control group, then additional operational efficiency is gained by inclusion of 3 experimental treatments with 1 control treatment in a single study, as opposed to the traditional approach of separate randomized phase II trials, each comparing one experimental treatment to erlotinib.

The second major innovation of the BATTLE trial is the statistical approach, using adaptive rather than equal randomization. Adaptive randomization allowed selection of the best treatment arm for each enrolled patient based on accumulating knowledge of the DCR for each of the 20 biomarker–treatment matches evaluated in the study. Simulations performed by the BATTLE statisticians indicate that, even with a requirement for equal randomization among approximately the first 90 of 200 patients (to avoid early skewing of the adaptive randomization process), the adaptive randomization approach provides higher expected overall DCRs than does an equal randomization approach (9). The adaptive approach is attractive to both patients and physicians because it epitomizes the idea of “personalized medicine.” Notably, a similar approach is used in the Investigation of Serial Studies to Predict Your Therapeutic Response with Imaging and Molecular Analysis 2 (I-SPY 2) trial, which involves patients with newly diagnosed breast cancer who are eligible for neoadjuvant treatment with a taxane (11).

A caveat regarding the use of adaptive randomization is that the differences in expected DCRs for adaptive versus equal randomization in the BATTLE trial simulations were relatively small (9), and that Korn and Freidlin (12) have reported similar simulations for 2-arm studies, in which the advantages of adaptive versus equal randomization (or alternate fixed randomization, such as 2:1) in the probability of disease control were found to be quite small and of questionable advantage from a trial design perspective.

An alternative approach to BATTLE, using fixed randomization with a similar number of patients, can be used to...

---

**Table 2. Alternative approach to BATTLE**

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Biomarker eligibility requirement</th>
<th>Estimated proportion of eligible patients (%)</th>
<th>Sample size needed without early stopping</th>
<th>Expected sample size with early stopping</th>
</tr>
</thead>
<tbody>
<tr>
<td>Erlotinib</td>
<td>EGFR mutation</td>
<td>15</td>
<td>133</td>
<td>100</td>
</tr>
<tr>
<td>Vandetanib</td>
<td>VEGFR-2 overexpression (IHC score &gt;100)</td>
<td>40</td>
<td>50</td>
<td>38</td>
</tr>
<tr>
<td>Erlotinib + bexarotene</td>
<td>RXRα overexpression (nuclear IHC score &gt;30)</td>
<td>80</td>
<td>25</td>
<td>19</td>
</tr>
<tr>
<td>Sorafenib</td>
<td>KRAS or BRAF mutation</td>
<td>22</td>
<td>91</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Total = 299</td>
<td>Total = 225</td>
</tr>
</tbody>
</table>

**NOTE:** This alternative approach uses 4 separate Simon 2-stage phase II trials that treat only biomarker-positive patients.
determine whether codevelopment of an experimental agent with a matching diagnostic test should move forward. For example, we consider a 240-patient trial with equal randomization comparing 3 experimental groups with a common control, with 5 distinct biomarker subpopulations equally distributed across the treatment groups (yielding 20 biomarker–treatment matches, as in BATTLE). A set of decision rules may be constructed on the basis of observed P values, comparing each experimental treatment with the control. If the test in the all-comers population is not significant at a prespecified threshold (e.g., \( P > 0.2 \)), then further development may not be considered; conversely, further development of an experimental treatment in all comers might be considered if the test is significant (e.g., \( P < 0.05 \)). An observed trend toward statistical significance (e.g., \( 0.05 < P < 0.20 \)) could trigger a comparison of the experimental treatment with the control in each biomarker subpopulation, with additional development in a specific subpopulation considered if \( P < 0.05 \). Using this example and assuming an DCR of 30% in all but one biomarker subpopulation for a given experimental treatment, we see that an underlying 80% DCR would be required to have a >50% probability of moving an experimental treatment forward, either in a biomarker subpopulation or in an all-comers population. Although an 80% DCR may seem high, precedence for this kind of efficacy exists in well-matched biomarker–treatment combinations (such as \( \text{BRAF} \) mutation and the \( \text{BRAF} \) inhibitor PLX4032) and is arguably an appropriate expectation for further development of new therapeutics with matching diagnostic tests.

Statistical designs different from those of BATTLE and I-SPY 2 have been proposed for randomized trials that include predictive biomarker hypotheses (13, 14). These designs do not require a prespecified biomarker test used for treatment assignment, which avoids the screen failure problem described above. The cross-validated adaptive signature design (14) is particularly attractive because it can test multiple biomarkers in a large trial with many clinical end points in the true population of interest. Statistical validity of selected subgroups is characterized by evaluating the signature selection procedure among multiple (e.g., 10) nonoverlapping groups, forming a 90% sample for each complement to the 10% subsample for training and using the 10% sample to test how well the predicted model works. Combining variability across these samples gives a composite predictive value of the signature selection procedure. Although signatures in the 10 subsets evaluated will vary, they presumably will predict patient outcomes and treatment benefit similarly well; that is, they are all different, but similar, “versions of the truth.” Not having to prespecify a set of biomarkers to test before enrollment has the major advantage of accommodating how little we often know at the beginning of a pivotal trial about which subgroups may benefit from a novel treatment regimen.

In summary, successful completion of the BATTLE study is an important milestone in the war against cancer. Biomarker-based approaches like those of BATTLE and I-SPY 2, as well as fixed randomization alternatives, represent major advances in trial design that should accelerate identification of predictive biomarkers for novel therapeutics.

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

REFERENCES
The BATTLE Trial: A Bold Step toward Improving the Efficiency of Biomarker-Based Drug Development

Eric H. Rubin, Keaven M. Anderson and Christine K. Gause


Updated version  Access the most recent version of this article at:
doi:10.1158/2159-8274.CD-11-0036

Cited articles  This article cites 14 articles, 8 of which you can access for free at:
http://cancerdiscovery.aacrjournals.org/content/1/1/17.full.html#ref-list-1

Citing articles  This article has been cited by 2 HighWire-hosted articles. Access the articles at:
/content/1/1/17.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.