The Role of Non-Randomized Trials for the Evaluation of Oncology Drugs

Deborah Armstrong, Professor of Oncology, Johns-Hopkins Kimmel Cancer Center
Gideon Blumenthal, Clinical Team Leader, FDA
Mace Rothenberg, Senior Vice President, Clinical Development & Medical Affairs &
Chief Medical Officer, Pfizer, Inc.
Richard Simon, Chief, Biometric Research Branch, National Cancer Institute
Josh Sommer, Executive Director, Chordoma Foundation
Lisa LaVange, Director, Office of Biostatistics, FDA
Richard Pazdur, Director, Office of Hematology and Oncology Products, FDA

Introduction

Randomized, controlled clinical trials (RCTs) provide the most reliable information about the effects of therapeutic interventions. RCTs minimize many sources of potential bias, enable a more precise, thorough, and reliable characterization of efficacy and safety than do other types of clinical research, and are usually needed for FDA drug approval. Further, because extension of overall survival provides the most objective and direct measure of clinical benefit, RCTs measuring overall survival are considered the gold-standard in oncology, although demonstration of a survival benefit is not always required for approval. Single-arm trials are usually considered appropriate only in the context of accelerated approval because assessment of the effect of the drug on overall survival (OS), progression-free survival (PFS), and on adverse events generally cannot be clearly distinguished from the pace and complications of the underlying disease. It is difficult to assure that external control groups are comparable with regard to important prognostic factors or comparable in the use of follow-up procedures for measuring PFS. In contrast to PFS or OS, the overall response rate of a drug can be measured in a single arm trial, as absent the therapy, it is extremely rare for a tumor to spontaneously regress. Most cancer therapies historically have provided only modest benefit with significant toxicity, making it difficult to interpret the results of efficacy studies in the absence of a concurrent control group. Although the FDA approved oncology drugs on the basis of tumor response in the 1970s and early 1980s, it began calling for approvals to be supported by improvements in survival or patient symptoms in the mid-1980s based on advice from the Oncologic Drugs Advisory Committee (1). It may be time now to re-evaluate this position once again.

Recent advances in biomedical science and technology have resulted in the development of several highly effective, molecularly targeted anti-cancer drugs as well as the ability to identify those subsets of patients most likely to respond to these drugs. The high anti-tumor activities of these drugs were recognized very early in clinical development. This has prompted many to question whether traditional RCTs are always the appropriate approach for demonstrating clinical benefit to support initial FDA approval (2-4). Despite the scientific advantages of randomized trials measuring overall survival, there are several commonly recognized barriers to their conduct (5). For drugs which target rare cancers or low-frequency, molecularly defined subsets of more common cancers, it may be prohibitive to screen and enroll enough patients within a reasonable period of time to conduct a large randomized trial. It may also be ethically

and logistically challenging to enroll patients on a placebo-controlled or active-controlled (in cases where the chemotherapy control is highly toxic or marginally effective) trial, when patients and physicians are aware of the potential for benefit with the experimental agent. In this situation, patients may enroll in clinical trials in the hopes of gaining access to the investigational agent, even if that access may only come in the form of crossover to the experimental arm after progressing on the control arm. In unblinded randomized studies, dropout of patients assigned to the control arm before treatment can be administered can threaten the integrity of the study and reliability of its results. A high rate of patient crossover within or external to the study could confound analysis of overall survival, resulting in an underestimation of the magnitude of clinical benefit. High crossover may also confound analysis of progression-free survival if there is a strong bias to take patients off the control arm in an unblinded study. Indeed, for a new drug being developed with very strong biological rationale in a biomarker-selected population of patients, it may not be possible to conduct a randomized clinical trial at all and particularly one without a crossover of the control arm to the new drug, and overall survival data may therefore not be obtainable. Situations in which randomized trials may not be feasible or ethical include:

1. A new drug that demonstrates unprecedented effects on overall response rate in a setting of high unmet need with no effective therapies: in this situation, equipoise is lost and a randomized trial is not appropriate.
2. An already approved molecularly targeted agent is being tested in a rare tumor histology expressing the appropriate biomarker: in this situation, a randomized trial may simply not be possible due to the very low numbers of patients. In this scenario, the drug is also supported by the existing safety database developed in previous settings.

The FDA has recently created the Breakthrough Therapy designation, which seeks to expedite the clinical development of drugs that are intended to treat serious and life-threatening diseases and for which preliminary clinical data indicate that the drug may provide a substantial improvement over available therapies. While this program has contributed to the expedited development of several recently approved oncology drugs, it remains unclear what the appropriate development path is in the absence of a traditional randomized trial. In this proposal, we will describe how single-arm trials can be used to support approval.

Can single-arm trials support traditional approval?

One potential approach when large effects are seen in early clinical studies or when very rare populations are being studied is to confirm clinical benefit in a single-arm trial. This approach is not without precedent. The FDA has granted traditional approval on the basis of substantial, durable responses in some situations (Table 1). A notable example is the 2006 approval of imatinib in a series of very rare, life-threatening malignancies that express imatinib-sensitive tyrosine kinases after imatinib previously received approval for treatment of CML and GIST. Due to the rarity of these diseases, these approvals were based partially on response rates observed in a phase 2, single-arm trial that enrolled patients with multiple tumor types, as well as on published case reports (6). Another notable example is the 2012 approval of vismodegib for metastatic and locally advanced basal cell carcinoma (BCC), a rare disease that can lead to severe disfigurement and morbidity and for which no systemic treatments were available at the time (7). Although vismodegib had been tested in a fairly small sample of patients, its mechanism of action was well understood as preclinical studies had demonstrated that it inhibits the Hedgehog signaling pathway, known to play an important role in BCC pathogenesis. The basis for the vismodegib

approval included the strong mechanistic rationale, and the fact that it produced durable objective tumor responses for disfiguring and symptomatic cutaneous tumors thought to reflect direct clinical benefit. Therefore, the drug received traditional, rather than accelerated, approval.

Although the vismodegib example is somewhat unique in that tumor responses in cutaneous malignancies are thought to confer direct clinical benefit, it helps to illustrate a set of standards that could be used to determine whether a single-arm trial is robust enough to support traditional approval: 1) the drug mechanism of action is supported by strong scientific rationale and/or preclinical data; 2) the drug is intended for a well-defined patient population; 3) the drug produces substantial, durable tumor responses that clearly exceed those offered by any existing available therapies; and 4) the benefits outweigh the risks. These standards are similar to criteria that have been described in previous proposals advocating for alternative paths to FDA approval (3, 4). However, there are still many reservations about the quality of evidence that can be produced in single-arm trials. For example, response rate, the primary endpoint typically used in single-arm trials, is classically considered a surrogate for long-term clinical benefit that must be validated in randomized confirmatory trials. Another challenge is identifying a comparative dataset to use for historical controls, particularly when the investigational drug is being studied in a molecularly defined or small patient population. Finally, unless a drug produces a unique toxicity, a single-arm trial makes it difficult to discern whether adverse events are due to the drug, disease, aging, or other factors, especially in the case of cardiac and pulmonary adverse events. We will discuss how these limitations can be mitigated in the remainder of the proposal.

Translating tumor response into clinical benefit

Objective criteria for measuring tumor response rates were first published in 1960 and have been used in many cancer clinical trials since that time to assess therapeutic efficacy (8, 9). In many hematologic malignancies, such as Hodgkin's lymphoma and pediatric leukemia, durable complete responses have historically been used to support new drug approval. This considerable experience allows many single-arm trials to rely on historical controls. For a particular disease, it may be possible to establish response rate thresholds which predict long-term clinical outcomes in order to validate response rate in that disease. One way this can be accomplished is through meta-analyses evaluating the relationship of response rates to long-term clinical outcomes (10). A classic example is an analysis of 26 RCTs in ovarian cancer published between 1975 and 1989 which found that large improvements in response rates were needed for clinically meaningful improvements in survival (11). In a recent trial-level and patient-level meta-analysis, 14 active-controlled RCTs in metastatic non-small cell lung cancer (NSCLC) that had been submitted to the FDA between 2003 and 2013 were evaluated (12). This analysis found that a treatment effect on ORR was strongly correlated with a treatment effect on PFS (R^2=0.89). The analysis did not indicate a strong correlation between improvement in ORR and improvement in OS, but this may have been confounded by high crossover and long post-progression survival in the studies of targeted therapies in molecularly-enriched patient populations. The analysis also found that drugs which are tested in a molecularly defined subset of metastatic NSCLC, such as crizotinib in patients with ALK-rearranged NSCLC or erlotinib and afatinib in patients with EGFR mutation positive NSCLC, that produce a large effect on overall response rate, are likely to have a large effect on progression-free survival. Another analysis from academic investigators suggested that studies enriched for patients with molecular drug targets were associated with higher therapeutic benefit as compared with unselected populations (13). One key aspect of the meta-analysis in NSCLC described here relative to older studies is the ability to include

studies of molecularly targeted drugs that elicited high response rates, which improved the ability to establish a correlation between response and longer term outcomes and to correlate treatment effects with respect to the two outcomes. Earlier meta-analyses were hindered by the lack of highly active agents and consequently a lack of studies with high response rates to include in these analyses. With the advent of molecularly targeted drugs that often produce large response rates, it may now be possible to reconsider the role of response rates and identify settings in which response rates of a sufficient magnitude and duration may be predictive of long-term clinical benefit.

Beyond the use of response rate as a surrogate measure, it is a real-world tool used daily in the clinic for ongoing assessment of patients. It is widely accepted by both patients and physicians that significant and prolonged reduction of tumor burden is clinically meaningful and could potentially be considered for the purposes of regulatory approval to constitute direct clinical benefit in some cases. Objective response is usually one of the earliest signals of efficacy to emerge in a clinical trial and use of response rate as an endpoint could substantially shorten the time to final analysis compared to event-driven endpoints such as PFS or OS which may occur over a more protracted period of time. The clinical relevance of response rates may be dependent on the nature of the disease, the location of tumors, and the symptoms associated with a specific tumor. For many cancer types, significant reduction of tumor burden is likely to be accompanied by symptomatic improvement. Further investigation and validation of disease-specific patient reported outcomes in randomized trials may assist in aligning radiographic responses with symptomatic benefit, lending support to the use of response rate as an endpoint sufficient for regulatory approval in a specific setting. Whether a tumor response is clinically meaningful or not also depends on the depth, duration, and type of response. While RECIST criteria are helpful because they allow comparison to historical controls, these criteria classify responses into either complete or partial responses and do not fully capture quantitative differences in partial responses, and methods that characterize responses as continuous variables, such as waterfall plots that measure the greatest depth of response, may be more representative (14, 15). The RECIST criteria also do not capture qualitative differences in tumor location. Improved radiographic techniques, alternative data collection and representation, and validated patient-reported outcomes may be able to better characterize tumor response and associated reductions in symptomatic burden in the future.

The appropriateness of response rate and duration as an endpoint may also be dependent upon the nature and magnitude of benefit offered by any existing therapies. For example, if a new drug that produces large response rates is being tested in a disease setting where an existing therapy offers a significant survival advantage, a RCT is needed to ensure that the survival advantage is not lost. However, if the drug produces large response rates in a disease setting with no available therapies or where the available therapy provides only a marginal efficacy advantage with high toxicity, a RCT may not be needed.

Setting standards for single arm trial design

For the types of treatments discussed here, the outcome is expected to be clearly superior to that obtainable with currently available treatments for the disease. The outcome with available treatments should be documented, however, using historical control data for one or more large series of patients. Sponsors should develop a protocol specifying how historical control series will be selected and how data from historical controls will be analyzed. The patients in these series should be comparable with regard to important prognostic factors to the patients on the single arm clinical trial. Establishing this comparability

will either require individual patient data for the control patients or deriving the control series from clinical trials with well characterized eligibility and evaluation criteria. In addition to prognostic comparability, there should be comparability of response assessment or progression assessment if either response rate or progression- free survival is used as the primary endpoint. Having multiple control series provides the opportunity to document that the inter-study variation in outcomes for patients with available treatments is small or explained by known prognostic factors. One way in which comparable historical control data can be obtained is through new initiatives such as Project Data Sphere (16). Project Data Sphere aggregates patient-level information from multiple clinical trials and is intended to serve as a platform through which independent researchers can share and analyze historical patient-level data from phase III clinical trials. These types of initiatives may facilitate matching of study patients to historical controls for known prognostic characteristics, thereby creating a context for interpretation of clinical trial results from single arm studies.

A single arm study can be sized to have adequate statistical power for demonstrating that the better outcomes for the new treatment are unlikely to have occurred by chance alone. This type of analysis should take into account the size of the control series (17) being used to provide historical or external control data. If individual patient data are available for the control series, the statistical significance test comparing outcomes for patients on the single arm trial to outcomes for the control series can be adjusted for prognostic factors if there are strong prognostic factors to take into account. In many cases the patients in the single arm study will be selected for treatment based on a biomarker such as having a mutation in a gene that is a target of the treatment. In order to establish that the improved outcome of the patients in the single arm study is not solely due to a prognostic benefit of the biomarker, some information is needed about the potential prognostic effect of the marker used to select patients for treatment. In some cases it may be possible to measure the marker in archived tissue from a sufficient number of cases within the control series to estimate the potential prognostic effect of the marker. For example, a small retrospective case-controlled study stained archived non-small cell lung cancer tissue to identify ALK-positive and ALK-negative historical controls and evaluate survival outcomes in these patients (18). This study found that ALK translocation is not a favorable prognostic factor for survival in NSCLC. In situations where there is a prognostic effect, the sample size for the single arm study can be expanded to provide adequate statistical power for determining whether outcomes with the new treatment is greater than outcome for available treatments by more than the plausible prognostic effect of the marker.

Additional studies that can contribute to the evidence base for a new drug

For drugs that receive approval on the basis of single-arm trials, one frequently used approach to expand the safety database and obtain comparative safety information is to perform randomized trials in other stages of the disease or in different disease settings. This is commonly done for drugs that have received accelerated approval, where post-approval trials to confirm clinical benefit are often conducted in a less advanced stage of disease than the initial approval (19). One example where it may be extremely difficult to collect adequate safety data is the development of BRAF inhibitors in NSCLC. Because *BRAF* mutations are so rare (1-4%) in this disease setting (20), investigators will likely need to rely on safety information developed through studies in melanoma and other disease settings. Large data sets, such as those being developed through initiatives such as Project Data Sphere, may provide another potential way to obtain comparative safety information. Yet another potential way to expand the safety database may be

to perform randomized dose comparison trials when there is reason to believe that a lower dose might provide equal efficacy (21).

Conclusions

It is not always possible to perform a randomized, controlled clinical trial. While the RCT is ideal, rigorous single-arm trials can be used to support approval in certain situations. Improved science and technology have led to the development of new therapies for which the mechanism of action is well understood, the patient population can be well-defined, and tumor responses clearly exceed those offered by existing therapy in terms of both magnitude and duration. Response rate thresholds which are predictive of long-term outcomes or correlate with clinically meaningful symptomatic improvements can be defined for different disease settings, which may advance the use of response rate as an endpoint for approval. A key issue that must be resolved for single-arm trials to support regulatory approval is the identification of appropriate historical controls. The term "historical control" is often used to refer to different things, such as the response rate assumed to occur in the absence of any treatment, based on what is known about the natural history of the disease, or a response rate that has been observed in trials of similar patients receiving alternative treatments. Guidelines on methodology for prospective selection and analysis of historical control data are needed ensure appropriate use of historical comparator groups in evaluating results from a single-arm study. An FDA Guidance or Best Practices document should provide such guidelines as well as describe how adequate safety information can be developed and monitored in the post-market setting. The development of these guidelines will help to facilitate the use of single-arm trials that can both produce strong evidence as well as enable effective drugs to reach patients in need quickly.

Table 1:

Traditional Approvals Based on Response Rate in Single-arm Trials (2002-2013)

Drug	Year	Approval Type	Indication(s)	N	ORR (95% CI)	mDOR (months)
Tositumomab	2003	NME	Relapsed, CD-20+, follicular, Non-Hodgkin Lymphoma (NHL)	40	68	16
Imatinib	2006	Supplement	Dermatofibrosarcoma protuberans (DFSP)	18	83	6.2
	2006	Supplement	Myelodysplastic syndrome (MDS/MPD)	31	84	4.6+ → 15+
	2006	Supplement	Adult aggressive systemic mastocytosis (ASM)	28	61	1 → 30
	2006	Supplement	Hypereosinophilic syndrome/ chronic eosinophiliic leukemia (HES/CEL)	176	74	1.5+ → 44
Bortezomib	2006	Supplement	Relapsed mantle cell lymphoma (MCL)	155	31	9.3
Cetuximab*	2006	Supplement	Recurrent squamous cell carcinoma of the head and neck (SCCHN)	103	12.6	5.8
Vorinostat	2006	NME	Recurrent cutaneous manifestations of cutaneous T-cell lymphoma (CTCL)	74	30	5.6
Dasatinib*	2006	NME	2nd-line Ph+ acute lymphoblastic leukemia (ALL)	36	MaHR - 42	4.8
Ixabepilone*	2007	Supplement	Refractory metastatic breast cancer	126	12.4	6
Bendamustine	2008	Supplement	Indolent B-cell Non-Hodgkin Lymphoma (NHL)	100	74	9.2
Romidepsin	2009	NME	2nd-line cutaneous T-cell lymphoma (CTCL)	167	34.5	13
Vismodegib	2012	NME	Metastatic basal cell carcinoma	33	30	7.6
	2012	NME	Locally advanced basal cell carcinoma	63	43	7.6
Bosutinib	2012	NME	2nd-line Ph+ chronic myelocytic leukemia (CML)	503	McyR - 53	18
Lenalidomide	2013	NME	Relapsed mantle cell lymphoma (MCL)	134	26	16.6
Denosumab	2013	Supplement	Giant cell tumor of bone	187	25	>8

Abbreviations: N, number of patients tested; ORR, overall response rate; CI, confidence interval; mDOR, median duration of response; NME: new molecular entity; Ph, Philadelphia chromosome; MaHR, major hematologic response; McyR, major cytogenetic response.

 * Cetuximab, dasatinib, and ixabepilone, approvals were supplemented by concurrent approvals in closely related settings or in combination regimens based on randomized trials

Adapted from Blumenthal, New Drugs in Oncology Seminar, ASCO 2014

References

(1)   Johnson, J.R., Williams, G. & Pazdur, R. End points and United States Food and Drug Administration approval of oncology drugs. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **21**, 1404-11 (2003).

(2)   Stewart, D.J. & Kurzrock, R. Cancer: the road to Amiens. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **27**, 328-33 (2009).

(3)   Sharma, M.R. & Schilsky, R.L. Role of randomized phase III trials in an era of effective targeted therapies. *Nature reviews Clinical oncology* **9**, 208-14 (2012).

(4)   Fleming, T. *et al. Development Paths for New Drugs with Large Treatment Effects Seen Early. Conference on Clinical Cancer Research.* <http://www.focr.org/sites/default/files/Panel4FINAL11411.pdf> (2011).

(5)   Sargent, D.J. & Hayes, D.F. Assessing the measure of a new drug: is survival the only thing that matters? *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **26**, 1922-3 (2008).

(6)   Heinrich, M.C. *et al.* Phase II, open-label study evaluating the activity of imatinib in treating life-threatening malignancies known to be associated with imatinib-sensitive tyrosine kinases. *Clinical cancer research : an official journal of the American Association for Cancer Research* **14**, 2717-25 (2008).

(7)   Axelson, M. *et al.* U.S. Food and Drug Administration approval: vismodegib for recurrent, locally advanced, or metastatic basal cell carcinoma. *Clinical cancer research : an official journal of the American Association for Cancer Research* **19**, 2289-93 (2013).

(8)   Oxnard, G.R. *et al.* When progressive disease does not mean treatment failure: reconsidering the criteria for progression. *Journal of the National Cancer Institute* **104**, 1534-41 (2012).

(9)   Zubrod CG, S.M., Frei Iii E et al. Appraisal of methods for the study of chemotherapy of cancer in man: comparative therapeutic trial of nitrogen mustard and triethylene thiophosphoramide. *J Chronic Dis* **11**, 7-33 (1960).

(10) Sherrill, B., Kaye, J.A., Sandin, R., Cappelleri, J.C. & Chen, C. Review of meta-analyses evaluating surrogate endpoints for overall survival in oncology. *OncoTargets and therapy* **5**, 287-96 (2012).

(11) Torri, V., Simon, R., Russek-Cohen, E., Midthune, D. & Friedman, M. Statistical model to determine the relationship of response and survival in patients with advanced ovarian cancer treated with chemotherapy. *Journal of the National Cancer Institute* **84**, 407-14 (1992).

(12) Blumenthal, G. *et al. Overall response rate (ORR) as a potential surrogate for progression-free survival (PFS): A meta-analysis of metastatic non-small cell lung cancer (mNSCLC) trials submitted to the U.S. Food and Drug Administration (FDA). ASCO Annual Meeting.* <http://meetinglibrary.asco.org/content/124927-144> (2014).

(13) Janku, F., Berry, D.A., Gong, J., Parsons, H.A., Stewart, D.J. & Kurzrock, R. Outcomes of phase II clinical trials with single-agent therapies in advanced/metastatic non-small cell lung cancer published between 2000 and 2009. *Clinical cancer research : an official journal of the American Association for Cancer Research* **18**, 6356-63 (2012).

(14) Schwartz, L. *Getting the Right Drug to the Right Patient Faster. ASCO Annual Meeting.* <http://meetinglibrary.asco.org/content/94975> (2014).

(15) Jain, R.K. *et al.* Change in tumor size by RECIST correlates linearly with overall survival in phase I oncology studies. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **30**, 2684-90 (2012).

(16) <https://www.projectdatasphere.org/projectdatasphere/html/home>.

(17) Korn, E.L. & Freidlin, B. Conditional power calculations for clinical trials with historical controls. *Statistics in medicine* **25**, 2922-31 (2006).

(18) Shaw, A.T. *et al.* Effect of crizotinib on overall survival in patients with advanced non-small-cell lung cancer harbouring ALK gene rearrangement: a retrospective analysis. *The lancet oncology* **12**, 1004-12 (2011).

(19) Johnson, J.R., Ning, Y.M., Farrell, A., Justice, R., Keegan, P. & Pazdur, R. Accelerated approval of oncology products: the food and drug administration experience. *Journal of the National Cancer Institute* **103**, 636-44 (2011).

(20) Lovly, C., Horn, L. & Pao, W. *My Cancer Genome: BRAF Mutations in Non-Small Cell Lung Cancer (NSCLC).* <http://www.mycancergenome.org/content/disease/lung-cancer/braf> (2013).

(21) Minasian, L., Rosen, O., Auclair, D., Rahman, A., Pazdur, R. & Schilsky, R.L. Optimizing Dosing of Oncology Drugs. *Clinical pharmacology and therapeutics* **96**, 572-9 (2014).