

Subgroup Analyses in Oncology Trials: Regulatory Considerations and Case Examples



Anup K. Amatya¹, Mallorie H. Fiero¹, Erik W. Bloomquist¹, Arup K. Sinha¹, Steven J. Lemery^{2,3}, Harpreet Singh^{2,3}, Amna Ibrahim^{2,3}, Martha Donoghue^{2,3}, Lola A. Fashoyin-Aje^{2,3}, R. Angelo de Claro^{2,3}, Nicole J. Gormley^{2,3}, Laleh Amiri-Kordestani^{2,3}, Rajeshwari Sridhara³, Marc R. Theoret^{2,3}, Paul G. Kluetz^{2,3}, Richard Pazdur^{2,3}, Julia A. Beaver^{2,3}, and Shenghui Tang¹

ABSTRACT

Subgroup analyses are assessments of treatment effects based on certain patient characteristics out of the total study population and are important for interpretation of pivotal oncology trials. However, appropriate use of subgroup analyses for regulatory decision-making and product labeling is challenging. Typically, drugs approved by the FDA are indicated for use in the total patient population studied; however, there are examples of restriction to a subgroup of patients despite positive study results in the entire study population and also extension of an indication to the entire study population despite positive results appearing primarily in one or more subgroups. In this article, we summarize key issues related to subgroup analyses in the benefit–risk assessment of cancer drugs and provide case examples to illustrate approaches that the FDA

Oncology Center of Excellence has taken when considering the appropriate patient population for cancer drug approval. In general, if a subgroup is of interest, the subgroup analysis should be hypothesis-driven and have adequate sample size to demonstrate evidence of a treatment effect. In addition to statistical efficacy considerations, the decision on what subgroups to include in labeling relies on the pathophysiology of the disease, mechanistic justification, safety data, and external information available. The oncology drug review takes the totality of the data into consideration during the decision-making process to ensure the indication granted and product labeling appropriately reflect the scientific evidence to support patient population for whom the drug is safe and effective.

Introduction

Despite the use of patient inclusion and exclusion criteria, the total patient population comprises clinically relevant distinct subgroups identifiable by demographics, clinical histories, and molecular and disease characteristics. Treatment effect assessments within such subgroups of the patient population, known as subgroup analyses, are important for interpretation of information derived from pivotal oncology trials. A consistent benefit–risk assessment across subgroups provides greater assurance that the treatment benefit observed applies to the entire patient population studied. However, pronounced treatment effects within distinct subgroups may be obscured in the analysis of the total patient population. For example, a subgroup of patients with tumors harboring a specific biomarker may have a more favorable outcome to treatment with a drug targeted against that biomarker compared with those with tumors lacking this biomarker of interest. In such cases, analyses restricted to all randomized patients [i.e., the intention-to-treat

(ITT) population] may not adequately characterize the true extent of treatment effects in these subgroups.

Additionally, multiple efficacy endpoints add another layer of complexity to the benefit–risk assessment in a subgroup. In a typical pivotal oncology trial where patients are randomized to an investigational or control arm, efficacy endpoints are generally related to tumor progression and survival. For example, progression-free survival (PFS) is a common primary efficacy endpoint in clinical trials evaluating treatment for metastatic disease in cancer patients. Overall survival (OS) is another commonly used endpoint generally in the metastatic setting for the evaluation of both efficacy and safety. An anticancer therapy that prolongs PFS is not considered safe and effective if the therapy results in a detrimental effect on OS. This can also apply in certain situations to a subgroup analysis in that a therapy is not considered beneficial in a subgroup where treatment results in a credible detriment to OS, regardless of advantages observed in other endpoints.

In a regulatory setting, decisions regarding the scope of the indicated population for approval require an overall benefit–risk assessment from multiple perspectives, including careful examination of important subgroups. Regulatory agencies are tasked with ensuring that the product labeling provides adequate safety and efficacy data for the intended population. Most indications are granted based on the broad population enrolled in the trial and typically subgroup analyses are supportive in these approvals. Occasional recent oncologic drug approvals in the United States have either restricted an indication to a subgroup of patients despite positive study results in entire study population or have allowed an indication for the entire study population, despite positive results primarily in one or more subgroups. Such a decision to either narrow or expand an indication is not straightforward and may transcend statistical principles either due to their limitations described in section 3, clinical relevance, or because of the unmet medical need in life-threatening disease settings.

¹US Food and Drug Administration, Center for Drug Evaluation and Research, Office of Biostatistics, Silver Spring, Maryland. ²US Food and Drug Administration, Center for Drug Evaluation and Research, Office of New Drugs, Office of Oncologic Diseases, Silver Spring, Maryland. ³US Food and Drug Administration, Oncology Center of Excellence, Silver Spring, Maryland.

Corresponding Author: Anup K. Amatya, Office of Biostatistics, Division V, Center for Drug Evaluation and Research, US Food and Drug Administration, 10903 New Hampshire Avenue, Silver Spring, MD 20993. E-mail: anup.amatya@fda.hhs.gov

Clin Cancer Res 2021;27:5753–6

doi: 10.1158/1078-0432.CCR-20-4912

©2021 American Association for Cancer Research

The Regulatory Agency Guidance on Subgroup Analysis

A number of issues relevant to subgroup analyses are discussed in various FDA guidance for industry documents. Additionally, the regulation 21 CFR 314.50 states that effectiveness data must be presented by gender, age, and racial subgroups (1). A brief discussion of subgroups is included in section 5.7 of the International Council for Harmonization Efficacy guideline (ICH E9; ref. 2). As outlined in ICH E9, sponsors should prespecify the subgroups they plan to analyse to investigate the potential heterogeneity of treatment effects across patient subgroups. It allows sponsors to deliberate among experts to identify cogent subgroups and formally test treatment effect in those subgroups or descriptively assess consistency of treatment effect across subgroups that are considered prognostic or predictive.

Principles for pursuing an indication in a molecular subset of a disease are outlined in the FDA guidance for industry on developing targeted therapies for low-frequency molecular subsets of a disease (3). This guidance states that when clinical trials, designed based on enrollment of patients with a set of biologically appropriate variants, are successful and other conditions for approval are met, the FDA will generally approve the drug for all patients who meet the prespecified inclusion criteria for the trial, regardless of the extent to which patients with various molecular alterations were represented in the clinical trial. A variety of strategies to enrich the study population by selecting a subset of patients in which the potential effect of a drug can be more readily demonstrated are discussed in the FDA guidance on enrichment strategies (4). This guidance includes an in-depth discussion on trial designs in a variety of settings including adequate evaluation of the marker-negative subgroup when pursuing a predictive enrichment strategy. The guidance states that the interpretation of the study finding would have to consider the magnitude of the effect in the marker-negative group. The recommendations and strategies for adequate representation of older adults in cancer clinical trials to enable benefit-risk evaluation in this population are outlined in the Inclusion of Older Adults in Cancer Clinical Trials guidance (5). It includes an example of a subgroup analysis approach using hierarchical testing in a randomized controlled trial that enrolls younger and older adults and stratifies by age.

Statistical Issues in Subgroup Analyses

Although it is important to identify subgroups that respond differently to a treatment, a valid statistical analysis of subgroups in a pivotal trial to inform the appropriate patient population for treatment use is challenging. Subgroups in pivotal trials that primarily focus on the total patient population usually lack adequate sample size (power) for rigorous statistical inference. This problem is often more pronounced in oncology drug trials where the primary efficacy analysis for all randomized patients is generally a time-to-event analysis. Statistical power for such analyses is dependent on the anticipated number of events at the time of analysis rather than the actual sample sizes. Furthermore, lack of control for multiplicity may result in false-positive conclusions regarding subgroup effects. Nonetheless, several authors have discussed statistical approaches that may inform a decision to include or exclude subgroups in the population for treatment use (6–9).

For regulatory purposes, it is helpful to distinguish the intent of subgroup analyses prespecified in the study protocol as either

inferential, supportive, or exploratory (7, 8). Such categorization guides the statistical analysis plan and provides context when interpreting and evaluating credibility of the findings based on the subgroups.

Inferential subgroup analyses are prespecified with adequate power and alpha control, with an intent to establish efficacy in a subgroup of interest. Supportive subgroup analyses are prespecified, but do not include prospective planning for testing within subgroups. These analyses are aimed at descriptively investigating consistency of treatment effect across subgroups after the treatment effect is shown to be significant in the total population. Exploratory subgroup analyses are generally included in the protocol to maximize the utility of the clinical trial data to generate hypotheses or to gain further insight into potentially predictive mechanistic elements of the treatment and biological characteristics of the disease.

Case Examples from the Recent FDA Approvals in Oncology

The first set of examples discussed in this section illustrates approvals in the ITT population despite decreased treatment effect in an important subgroup. The second set of examples illustrates the exception to the typical approach to approval in the total study population. These recent examples demonstrate clinical rationale and mechanistic justification driving interpretation of results when making regulatory decisions regarding treatment effects in subgroups.

Approval in the ITT Population

CHECKMATE-057 was the pivotal trial supporting approval of nivolumab for second-line treatment of patients with metastatic nonsquamous non-small cell lung cancer (mNSQ-NSCLC) who had experienced disease progression during or after one prior platinum doublet-based chemotherapy regimen (10). The trial demonstrated a statistically significant improvement in OS between nivolumab arm and docetaxel arm for the ITT population [hazard ratio (HR), 0.73 (95% CI, 0.60–0.89); $P < 0.002$] with 2.8 months prolongation in the median OS. There was no strong evidence of an effect on OS in the subgroup of mNSQ-NSCLC patients with PD-L1 $< 1\%$ [HR, 0.90 (95% CI, 0.66–1.24); ref. 11]. However, interpretation of these exploratory subgroup analyses was limited because of missing data (22%) on PD-L1 expression and the trial was not powered to detect an effect in these subgroups. Additionally, it was noted PD-L1 $< 1\%$ patients were not harmed by nivolumab as compared with the active control and roughly 10% of these patients had objective response with sustained duration of response (10); and this effect may be considered important to patients and prescribers. Specific details regarding these subgroups were included in product labeling to assist patients and providers with individual treatment decisions where the risks of immune-mediated toxicity may alter the risk-benefit assessment for tumors with low expression of PD-L1.

The approval of pembrolizumab for first-line treatment for advanced non-small-cell lung cancer was based on KEYNOTE-042. The trial demonstrated a 4.6-month improvement in the estimated median OS for patients on single-agent pembrolizumab compared with patients on carboplatin-based chemotherapy in the ITT population of patients who expressed PD-L1 [TPS $\geq 1\%$; HR, 0.81 (95% CI, 0.71–0.93); $P = 0.0018$]. In the subgroup of patients with TPS 1–49%, the estimated median OS was 13.4 months in the pembrolizumab arm and 12.1 months in the chemotherapy arm [HR, 0.92 (95% CI, 0.77–1.11); ref. 12]. Although the results of this exploratory

subgroup analysis raised uncertainty regarding the benefit of pembrolizumab in this subgroup, it was not possible to draw definitive conclusions, as the trial was not powered for the subgroup analysis. In addition, the comparator in KEYNOTE-042 was an active control, and there was no apparent detrimental effect on OS in this subpopulation of patients who were randomized to pembrolizumab as compared with chemotherapy.

The approval of margetuximab in combination with chemotherapy for the later line treatment of metastatic HER2-positive breast cancer was based on the SOPHIA trial. The trial met its primary PFS analysis endpoint in the ITT population [HR, 0.76 (95% CI, 0.59–0.98); $P = 0.03$]. An exploratory subgroup analysis by CD16A allotypes indicated possible differential effect of margetuximab. Although patients carrying the CD16A low-affinity F allele (F/F or F/V) appeared to have a greater benefit from margetuximab versus trastuzumab [HR, 0.68 (0.52–0.90)], PFS showed a trend in favor of trastuzumab in the subgroup of patients with the CD16A V/V genotype [HR, 1.78 (0.87, 3.62); ref. 13]. However, the FDA granted approval for the entire study population because there was no clear mechanistic basis for the heterogeneity of results by the CD16A allotypes; the sample size in the CD16A-V/V subgroup was small (69 of total $N = 536$); and some patient characteristics were not balanced between treatment arms within CD16A V/V genotype subgroup, which may cause bias in the subgroup results. Further reports from clinical trials that correlate CD16A F158V genotype with trial efficacy endpoints will be submitted as part of post-marketing commitment to further characterize the clinical benefit of margetuximab that may inform product labeling.

Approval in Subgroups

PAOLA-1 was the pivotal trial supporting approval of olaparib in combination with bevacizumab for the first-line maintenance treatment of adult patients with advanced epithelial ovarian cancer whose cancer is associated with homologous recombination deficiency (HRD)-positive status. The study demonstrated a statistically significant improvement in the primary endpoint of PFS in the ITT population [HR, 0.59 (95% CI, 0.49–0.72); $P < 0.0001$; ref. 14]. During the review, based on exploratory subgroup analysis, it was concluded that the treatment effect in the ITT population was driven by the HRD-positive subgroup [HR, 0.33 (95% CI, 0.25–0.45); ref. 14]. In the HRD-negative subgroup (34% of ITT), no treatment effect was observed for PFS, and OS numerically favored the placebo + bevacizumab arm. Although the result was based on an exploratory subgroup analysis, the differential efficacy between the HRD-positive and HRD-negative populations was biologically plausible based on the mechanism of action of olaparib with strong preclinical and clinical data demonstrating improved outcome across multiple PARP inhibitors. PARP inhibitors are known to cause cell death from gross genetic disarray in the presence of HRD (15). FDA decided to limit the indication to the HRD-positive population due to a lack of evidence of benefit and some evidence for potential harm in HRD-negative population.

The FDA approval of eribulin for the treatment of patients with unresectable or metastatic liposarcoma was based on Study 309 that enrolled patients with advanced liposarcoma (32%) or leiomyosarcoma (68%). Study 309 demonstrated a statistically significant improvement in the primary endpoint of OS in the ITT population [HR, 0.75 (95% CI, 0.61–0.94); ref. 16]. In a prespecified exploratory subgroup analysis based on histologic subtype, the overall OS results appeared to be strongly driven by a large treatment effect in the smaller group of patients with liposarcoma

[HR, 0.51 (95% CI, 0.35–0.75); ref. 16] as compared with the estimated effect in the subgroup of patients with leiomyosarcoma [HR, 0.90 (95% CI, 0.69–1.18); ref. 16]. Based on the differential treatment effects also observed among the secondary endpoints and biological differences between the two sarcomas, it was concluded that substantial evidence of effectiveness was limited to the liposarcoma subgroup.

Atezolizumab in combination with paclitaxel protein-bound received accelerated approval from the FDA for the treatment of adult patients with unresectable locally advanced or metastatic triple-negative breast cancer whose tumors express PD-L1. The recommendation for the approval was based on efficacy and safety data from the IMpassion130 study. The study was designed to formally test the treatment effect in the ITT population and PD-L1⁺ subgroup by splitting alpha between the two populations. Although the PFS results were statistically significant for the ITT population [HR, 0.79 (95% CI, 0.68–0.92); PFS medians 7.0 vs. 5.5 months; ref. 17], the magnitude of difference in median PFS was relatively short, which suggested uncertainty in the clinical benefit of adding atezolizumab to paclitaxel protein-bound for the entire study population. Therefore, the indication for this approval was limited to the PD-L1⁺ population in which the PFS benefit was greater [PFS HR, 0.60 (0.48–0.77); medians 7.4 vs. 4.8 months; ref. 17], and there was a favorable trend in OS benefit [OS HR, 0.71 (95% CI, 0.54–0.93); ref. 17].

Discussion

Determining the appropriate population to include in a new indication for a drug can be challenging when there are cogent subgroups that appear to show a heterogeneous response to treatment in a clinical trial. An effective therapy should be made available to all patients who derive benefit from it. However, broad use of the therapy may not be appropriate when the benefit-risk assessment is not favorable in important subgroups. A rigorous statistical analysis of clinical trial data is required for the regulatory evaluation of overall efficacy of a new drug application. However, statistical analysis alone of late-stage clinical trial data by subgroups has a limited ability to inform identification of the appropriate patient population for the approved indication.

In oncologic drug review, the challenges and consequences of interpreting subgroup analyses are heightened because the cancers are life-threatening and regulatory decisions are often based on multiple endpoints (e.g., PFS and OS) in a single pivotal trial. Multiple endpoints exacerbate the multiplicity problem in subgroup analysis. A single pivotal trial does not allow the opportunity to replicate and establish analytical credibility of subgroup findings, particularly when such finding is not anticipated *a priori*. When a stronger treatment effect is anticipated in a cogent pathophysiologic subgroup, a trial should be prespecified to formally evaluate this subgroup and be powered to detect such an effect. The interpretation of the study finding in the total study population should consider the adequacy of data in the complementary subgroup in the context of the disease setting. Ultimately, incorporation of results of subgroup analyses into regulatory decision-making should take into account the pathophysiologic rationale, nonclinical and pharmacologic support, clinical experience with similar agents, other supportive endpoint results, benefit-risk evaluation in the subgroups, medical need of the therapy, and other practical considerations. If needed, post-marketing commitments or requirements may be requested to better define the full extent of a drug's effect. Such decisions are guided by a goal to ensure

promising and safe therapeutic options are available to all patients who benefit from the treatment and to provide useful information regarding clinically important and scientifically supported observed differences in treatment effects to promote optimal decision-making among patients with cancer and their health care providers.

References

1. Food and Drug Administration. Health human services, code of federal regulation, 21CFR314.50, 2013. Available from: <https://www.fda.gov/media/133660/download>.
2. International Conference on Harmonization International Conference on Harmonization (ICH) Guidance, E9 statistical principles for clinical trials (ICH E9 guidance), 1998. Available from: https://database.ich.org/sites/default/files/E9_Guideline.pdf.
3. Food and Drug Administration. Developing targeted therapies in low-frequency molecular subsets of a disease guidance for industry. Available from: <https://www.fda.gov/media/117173/download>.
4. Food and Drug Administration. Enrichment strategies for clinical trials to support determination of effectiveness of human drugs and biological products. Available from: <https://www.fda.gov/media/121320/download>.
5. Food and Drug Administration. Inclusion of older adults in cancer clinical trials guidance for industry. Available from: <https://www.fda.gov/media/135804/download>.
6. Dmitrienko A, Muysers C, Fritsch A, Lipkovich I. General guidance on exploratory and confirmatory subgroup analysis in late-stage clinical trials. *J Biopharm Stat*. 2016;26:71–98.
7. Alosch M, Huque MF, Bretz F, D'Agostino RB Sr. Tutorial on statistical considerations on subgroup analysis in confirmatory clinical trials. *Stat Med* 2017;36:1334–60.
8. Dane A, Spencer A, Rosenkranz G, Lipkovich I, Parke T. PSI/EFSPi Working Group on Subgroup Analysis. Subgroup analysis and interpretation for phase 3 confirmatory trials: white paper of the EFSPi/PSi working group on subgroup analysis. *Pharm Stat* 2019;18:126–39.
9. Jones HE, Ohlssen DI, Neuenschwander B, Racine A, Branson M. Bayesian models for subgroup analysis in clinical trials. *Clin Trials* 2011;8:129–43.
10. Kazandjian D, Suzman DL, Blumenthal G, Mushti S, He K, Libeg M, Keegan P, Pazdur R. FDA approval summary: nivolumab for the treatment of metastatic non-small cell lung cancer with progression on or after platinum-based chemotherapy. *Oncologist*. 2016;21:634–42.
11. Drugs@FDA: FDA-Approved Drugs. OPDIVO (nivolumab), supplement. 5 label and letter 10/09/2015. Available from: https://www.accessdata.fda.gov/drugsatfda_docs/label/2015/125554s005lbl.pdf.
12. Drugs@FDA: FDA-Approved Drugs. KEYTRUDA (pembrolizumab), supplement. 47 label and letter 04/11/2019; 2019. Available from: https://www.accessdata.fda.gov/drugsatfda_docs/label/2019/125514s047lbl.pdf.
13. Drugs@FDA: FDA-Approved Drugs. MARGENZA (margetuximab) FDA Application Review Files. Available from: https://www.accessdata.fda.gov/drugatfda_docs/nda/2020/761150Orig1s000MultidisciplineR.pdf.
14. Arora S, Balasubramanian S, Zhang H, Berman T, Narayan P, Suzman D, et al. FDA approval summary: olaparib monotherapy or in combination with bevacizumab for the maintenance treatment of patients with advanced ovarian cancer. *Oncologist*. 2021;26:e164–e172.
15. Ledermann JA. PARP inhibitors in ovarian cancer. *Ann Oncol* 2016;27:i40–i44.
16. Osgood CL, Chuk MK, Theoret MR, Huang L, He K, Her L, Keegan P, Pazdur R. FDA approval summary: eribulin for patients with unresectable or metastatic liposarcoma who have received a prior anthracycline-containing regimen. *Clin Cancer Res*. 2017;23:6384–9.
17. Narayan P, Wahby S, Gao JJ, Amiri-Kordestani L, Ibrahim A, Bloomquist E, et al. FDA approval summary: atezolizumab plus paclitaxel protein-bound for the treatment of patients with advanced or metastatic TNBC whose tumors express PD-L. *Clin Cancer Res*. 2020;26:2284–89.

Authors' Disclosures

No disclosures were reported.

Received December 29, 2020; revised April 15, 2021; accepted June 10, 2021; published first June 11, 2021.