

clinical benefit of anti-cancer therapies based on comparable parameters while upholding accountability for reasonableness. Both platforms represent very important tools to ensure the appropriate use of limited resources in delivering cost-effective and more affordable cancer care globally.

Inarguably, data weighted and reported in these evolving scoring systems will help to identify those therapies with proven clinical benefit and guide clinical decision-making and provide physicians, patients and their families with independent, balanced information.

Guided by evidence-based data reported through the ASCO and ESMO frameworks, decision makers will need to balance efficacy and reimbursement of anti-cancer treatments with regional socio-economic realities. In parallel, professional societies and organizations, including ASCO and ESMO, have a shared responsibility to help map these necessary directions.

In our dedicated efforts to improve oncology practice and promote quality and high-value cancer care for all our patients, we can and will do better.

M. Bertagnolli<sup>1,2†</sup> & J. Tabernero<sup>3\*‡</sup>

<sup>1</sup>Division of Surgical Oncology, Brigham and Women's Hospital/Dana Farber Cancer Institute, Boston; <sup>2</sup>Richard E. Wilson Professor of Surgery at Harvard Medical School, Boston, USA; <sup>3</sup>Medical Oncology Department, Vall d'Hebron University Hospital, Vall d'Hebron Institute of Oncology (VHIO), Barcelona, Spain  
(\*E-mail: jtabernero@vhio.net)  
†ASCO President 2018-2019.  
‡ESMO President 2018-2019.

## Funding

None declared.

## Disclosure

MB: Research funding to the Alliance for Clinical Trials in Oncology Foundation: AbbVie, Agenus, Astellas, AstraZeneca, Baxalta, Bayer HealthCare, Breast Cancer Research Foundation, Bristol-Myers Squibb, Celgene, Complion, Czarnowski, Derse, Inc., Eisai, Exelixis, Genentech, GHI, Gilead Sciences, GSK Total, Incyte Corporation, Janssen, Jazz Pharma, Leidos, Lexicon Pharma, Lilly, Maltrex, Merck, Millennium, Novartis, Pfizer,

Pharmacyclics, Robert Wood, Johnson Foundation, Roche/Genentech, Sagerock Advisors, Sanofi, STO, Taiho Oncology, Takeda, Tesaro and Teva. JT: Personal financial interest in form of scientific consultancy role for Array Biopharma, AstraZeneca, Bayer, BeiGene, Boehringer Ingelheim, Chugai, Genentech, Inc., Genmab A/S, Halozyme, Imugene Limited, Inflection Biosciences Limited, Ipsen, Kura Oncology, Lilly, MSD, Menarini, Merck Serono, Merrimack, Merus, Molecular Partners, Novartis, Peptomyc, Pfizer, Pharmacyclics, ProteoDesign SL, Rafael Pharmaceuticals, F. Hoffmann-La Roche Ltd, Sanofi, SeaGen, Seattle Genetics, Servier, Symphogen, Taiho, VCN Biosciences, Biocartis, Foundation Medicine, HaliDX SAS and Roche Diagnostics. Institutional financial interest in form of financial support for clinical trials or contracted research for Agendia BV, Amgen SA, Debiopharm International SA, Janssen-Cilag SA, Mologen AG, Novartis Farmacéutica SA, Pharma Mar, Roche Farma SA, Laboratorios Servier SL and Symphogen A/S.

## References

1. Cherny NI, de Vries EGE, Dafni U et al. Comparative assessment of clinical benefit using the ESMO-Magnitude of Clinical Benefit Scale Version 1.1 (ESMO-MCBS v1.1) and the ASCO Value Framework Net Health Benefit Score (ASCO-NHB v2). *J Clin Oncol* 2019; 37(4): 336–349.
2. Global Burden of Disease Cancer Collaboration, Fitzmaurice C, Akinyemiju TF et al. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 29 cancer groups, 1990 to 2016. A systematic analysis for the global burden of disease study. *JAMA Oncol* 2018; 4(11): 1553–1568.
3. IMS Health. MIDAS, December 2015; Market Prognosis, March 2016; IMS Institute for Healthcare Informatics, May 2016.
4. Saluja R, Arciero VS, Cheng S et al. Examining trends in cost and clinical benefit of novel anticancer drugs over time. *J Oncol Pract* 2018; 14(5): e280–e294.
5. IMS Institute for Healthcare Informatics. Global Oncology Trend Report: a review of 2015 and outlook to 2020.
6. Cherny NI, Dafni U, Bogaerts J et al. ESMO-Magnitude of Clinical Benefit Scale version 1.1. *Ann Oncol* 2017; 28(10): 2340–2366.
7. Schnipper LE, Davidson NE, Wollins DS et al. Updating the American Society of Clinical Oncology Value Framework: revisions and reflections in response to comments received. *J Clin Oncol* 2016; 34(24): 2925–2934.

doi:10.1093/annonc/mdz057  
Published online 18 February 2019

## On the need to adjust for multiplicity in confirmatory clinical trials with master protocols

Recent advances in tumor biology and targeted therapies have led to clinical trials assessing treatment efficacy in multiple patient populations or subgroups using a *master protocol* [1]. These can lead to efficiency gains by testing several statistical hypotheses in the same clinical trial.

Although much of the development of novel design approaches has been in exploratory phase II trials that may not include a control group, there is growing interest in such methods

in confirmatory randomized controlled trials. These might be phase III trials with subgroup analyses or phase II/III trials combining exploratory and confirmatory elements. In these settings, multiple hypothesis tests can lead to statistical error rate inflation. A recent Food and Drug Administration draft guidance [1] notes the risk of 'potential overinterpretation of findings', but offers no clear suggestions as to when statistical correction for multiplicity should be implemented.

Before considering multiplicity adjustment, we first describe some recently used innovative designs. Our aim is not to give a comprehensive review, but to illustrate key relevant concepts. Based on the principle that correction is necessary when there are multiple opportunities to make a (regulatory) claim of efficacy

for a particular treatment in any population or subgroup [2, 3], we propose when corrections are needed if the innovative designs are used in confirmatory trials.

Interest in stratified medicine has led to *subgroup selection*, or *treatment interaction*, trials [4] comparing one or more experimental treatments with a control treatment in a population comprising two or more pre-defined subgroups, for example based on biomarker levels. Such trials assess treatment efficacy in the whole population and the subgroups.

Examples of these trials include the Gefitinib versus docetaxel in previously treated non-small-cell lung cancer (INTEREST) trial [5] that assessed non-inferiority in the overall population and superiority in high epidermal growth factor receptor (EGFR)-gene-copy patients, and the Biomarker-integrated Approaches of Targeted Therapy for Lung Cancer Elimination 1 (BATTLE 1) trial [6], a phase II trial comparing four non-small-cell lung cancer treatments in five biomarker-defined subgroups to determine the best treatment for each subgroup.

Subgroup selection trials may also include interim analyses to decide whether recruitment should continue from the whole population or be restricted to one or more subgroups where the experimental treatment appears most effective, as in an *adaptive enrichment design* [7–9].

When experimental treatment options depend on one or more biomarkers within a single cancer type, several treatments may be evaluated in parallel in an *umbrella trial* [1]. If each experimental treatment is evaluated in a single subgroup, the parallel sub-studies are essentially independent trials, each considering one treatment in its own distinct patient group, conducted together for administrative or operational reasons such as saving time, reducing costs or facilitating recruitment.

The sub-studies, which are often randomized, could be two-arm comparisons, each comparing one experimental treatment to a control, or multi-arm comparisons of several drugs within each sub-study. An example of the former is FOCUS4: a molecularly stratified trial programme in colorectal cancer [10] which included randomized sub-studies of a different experimental treatment in each of five biomarker-defined subgroups, with early stopping for futility. The Biomarker-Driven Protocol for Accelerating Development of Therapies for Squamous Cell Lung Cancer (Lung-Map) trial [11] has a similar design, with different treatments assessed in parallel in marker-specific subgroups.

The development of therapies targeting a particular mutation that may be present in multiple tumor types has led to the *basket trial* design.

Some authors [1, 12] use this term for a trial in which patients with a common mutation are recruited from populations with different tumor types. The effect of the experimental treatment can then be assessed both in the whole recruited group and in the individual tumor types. Often such trials are small and non-randomized, with all patients receiving the experimental treatment.

An example of such a trial is the imatinib study B2225 [13]. This open-label phase II trial investigated the effect of imatinib in patients expressing imatinib-sensitive tyrosine kinases in 40 different life-threatening malignancies, including solid tumors and hematologic malignancies.

The term *basket trial* has also been used to describe a more complex trial in which recruited patients have different tumor

types and different mutations [14]. Patients are then assigned to experimental treatments according to their mutation. Acknowledging that this type of design extends the simpler definition to include more than one mutation type, such a trial has also been called a *matrix trial*, although this term has also been used more generally, including both basket and umbrella trials as special cases.

An example of a matrix trial is the National lung matrix trial [15] in which seven drugs are evaluated in 18 targeted molecular cohorts, with drugs evaluated in one to five cohorts.

Multiplicity issues may arise when multiple hypothesis tests are conducted, particularly in confirmatory trials. Conventionally, adjustments are made when multiple confirmatory hypothesis tests are conducted in a trial [16], but not when they are conducted in separate trials. This position has arisen mainly in simpler settings than those now being developed, and the use of innovative trial designs has led to reconsideration of multiplicity issues, with little current consensus [2, 3, 17].

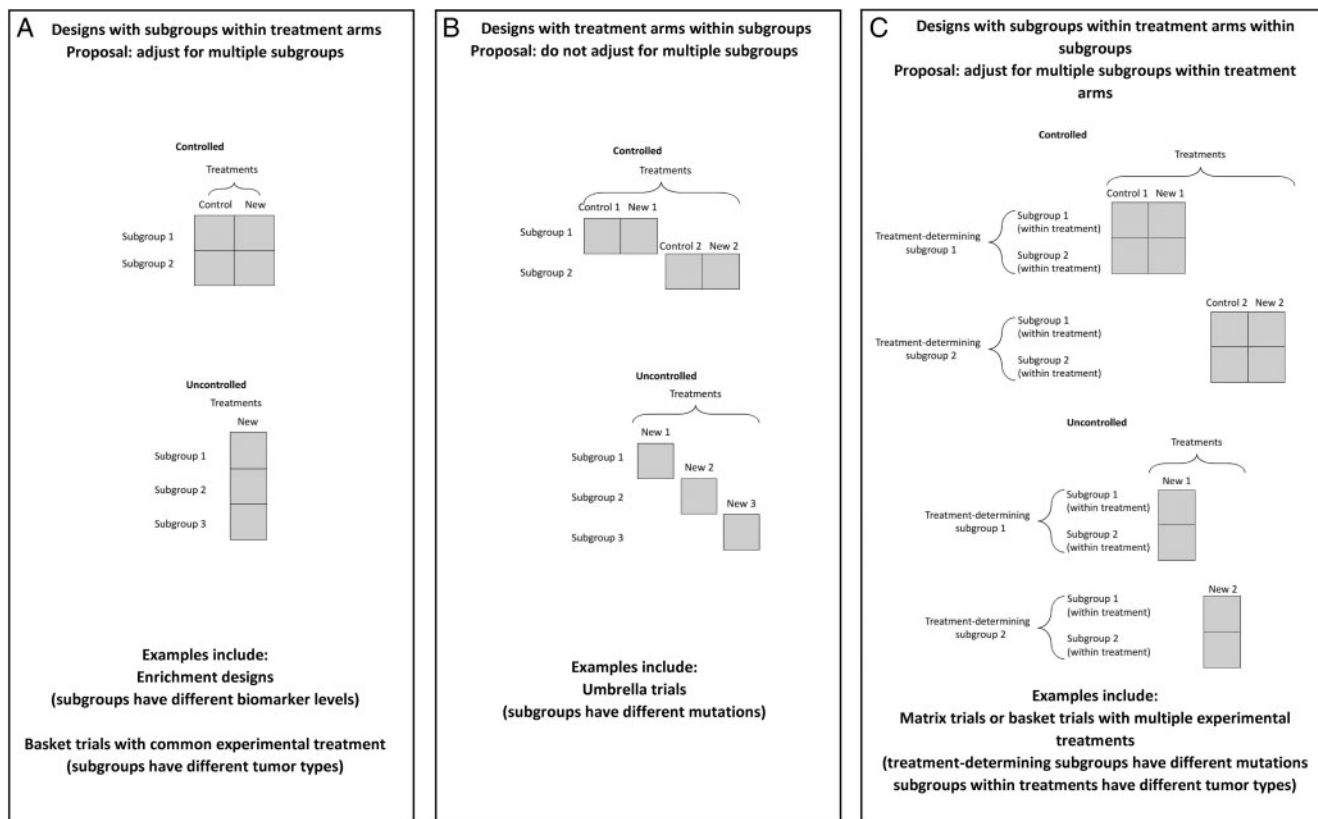
In the confirmatory setting, a basic principle is that correction for multiplicity is needed if testing of multiple hypotheses, whether through testing in multiple subgroups or at interim analyses, leads to multiple opportunities to inform a single claim of effectiveness for a drug [2].

In considering multiplicity issues in trials testing hypotheses based on different subgroups, as in the examples above, it is helpful to differentiate between trials on the basis of the role of the subgroups. We will consider three cases illustrated in Figure 1 and discussed in detail in the next three paragraphs.

We use the term *subgroups within treatment arms* to describe a setting in which the population is divided into subgroups, but in which the subgroup classification does not determine the experimental treatment options. In a non-randomized setting, all patients receive the same experimental treatment. In a randomized setting, all patients are randomized between the same two or more treatments, usually including a control (Figure 1A). The trial then assesses the effect of the same experimental treatment(s) in a number of subgroups. In this setting, since patients in different subgroups can receive the same experimental treatment(s), testing might also be conducted with subgroups combined as well as in individual subgroups.

This situation may be contrasted with one in which we have *treatment arms within subgroups*, when the population is divided into subgroups with distinct experimental treatment options available to each. In a non-randomized trial, the treatment is determined by the subgroup classification, and in a randomized trial, the treatments available for randomization depend on the subgroup so that randomization can only be undertaken for a patient once their subgroup is known (Figure 1B). The trial then assesses the effects of a number of experimental treatments, each in its corresponding subgroup.

An extension of the two settings above occurs when a trial includes subgroups of both types; in this case, subgroups are defined in two ways, with one of these restricting experimental treatment options, giving *subgroups within treatment arms within subgroups* (Figure 1C). An example is a situation in which patients are classified by two biomarkers, biomarker 1 and biomarker 2, with the experimental treatment depending on biomarker 1 but not on biomarker 2. Patients with a positive biomarker 1 status might then be allocated to treatment A, or randomized between A



**Figure 1.** Designs for clinical trials with master protocols: (A) with subgroups within treatment arms; (B) with treatment arms within subgroups; (C) with subgroups within treatment arms within subgroups.

and a control treatment C, and biomarker 1 negative patients be allocated to treatment B or randomized between B and C, in each case irrespective of biomarker 2 status. The trial then assesses treatment A in two subgroups of biomarker 1-positive patients who are either positive or negative for biomarker 2 and treatment B in two subgroups of biomarker 1-negative patients who are either positive or negative for biomarker 2.

Our aim is to correct for multiplicity when multiple hypothesis testing increases the chance of erroneously declaring a given ineffective treatment to be effective in all or some part of the population, but not for multiple hypothesis testing for different treatments when the use of each is restricted to a single subgroup. Using the definitions introduced above, we therefore propose that multiplicity corrections should be used in the setting with *subgroups within treatment arms*, or for the *subgroups within treatment arms within subgroups*, including the case of testing in combinations of subgroups, but not with *treatment arms within subgroups*.

Our rationale is that with *subgroups within treatment arms*, the same treatment is evaluated in a number of subgroups, giving multiple opportunities to claim this treatment is effective. This contrasts with the *treatment arms within subgroups* setting, where each experimental treatment is considered for only one subgroup so that there are not multiple opportunities for an effectiveness claim for any particular treatment.

The following paragraphs illustrate the application of this proposed multiple testing strategy in the designs introduced previously.

In subgroup selection, treatment interaction or adaptive enrichment designs the aim is to assess the experimental treatment(s) in the whole population and/or in one or more subgroups. As treatment options are not determined by subgroup classification, we have *subgroups within treatment arms*. In a confirmatory trial setting, our proposal is therefore to adjust for multiplicity arising from hypothesis testing in multiple subgroups in such a trial, for example using a Holm–Bonferroni correction [18], or a method allowing for overlapping subgroups [19].

In umbrella trials, we have *treatment arms within subgroups*. Our proposal would therefore be to not adjust for multiplicity arising from the testing of multiple hypotheses in different subgroups. This proposal is consistent with the view that the sub-studies, though run under a single protocol, are essentially independent trials, as each subgroup receives a different experimental treatment so that the efficacy of each treatment is assessed only once.

In the simpler type of basket trial [12] one experimental treatment is evaluated in patients with different tumor types. We thus have *subgroups within treatment arms*. In a confirmatory trial, we would therefore propose adjustment for multiplicity arising from hypothesis testing in the different tumor types to protect the overall chance of erroneously claiming effectiveness of the experimental treatment.

In matrix trials or the more complex type of basket trial with distinct treatments tested in different mutations [14], we have *subgroups* (e.g. tumor types) *within treatment arms within*

*subgroups* (e.g. mutations). In confirmatory trials, we would therefore propose adjustment for the multiple hypothesis tests of the effectiveness of the same treatment in different tumor types, but not for testing in different mutations since mutation determines treatment.

Although we have focused on specific trial settings, the principles are generalizable. For example, the same design might be used if subgroups are specified not by a molecular biomarker but by another measurable patient characteristic that determines treatment. Similarly, the same design might be used in a trial not just with multiple tumor types, but also with subgroups defined by some other factor in which we are trying to ascertain subgroups for which a treatment is effective, requiring multiplicity to be handled in the same way. Consideration of the same concepts could also enable application of the novel trial designs in other therapeutic areas.

We have focused on multiplicity arising from multiple subgroups. Trials, including platform or multi-arm multi-stage trials, may also include multiple treatments, possibly in addition to multiple subgroups. Multiplicity issues in such settings are discussed elsewhere [3, 17]. The need for multiplicity correction depends on the relatedness of the hypotheses being tested. If different doses of the same drug are tested, multiplicity corrections will be more appropriate than when multiple arms are unrelated treatments, possibly with different modes of action, whether or not they are compared with a common control group. Error rates other than familywise error rates, including false discovery rates and family multiple error rates, could also be considered in such settings [3].

Multiplicity arising from testing a single treatment in a number of different indications could also be considered, as could methods for testing a treatment strategy using data from multiple subgroups each receiving a different experimental treatment [20].

N. Stallard<sup>1\*</sup>, S. Todd<sup>2</sup>, D. Parashar<sup>1,3,4</sup>, P. K. Kimani<sup>1</sup> & L. A. Renfro<sup>5</sup>

<sup>1</sup>Statistics and Epidemiology, Warwick Medical School, University of Warwick, Coventry; <sup>2</sup>Department of Mathematics and Statistics, University of Reading, Reading; <sup>3</sup>The Alan Turing Institute, London; <sup>4</sup>Warwick Cancer Research Centre, University of Warwick, Coventry, UK; <sup>5</sup>Division of Biostatistics, University of Southern California, Los Angeles, USA

(\*E-mail: n.stallard@warwick.ac.uk)

## Acknowledgements

The authors are grateful to three reviewers for their helpful comments.

## Funding

This work was supported by UK Medical Research Council Grant/Award Number MR/N028309/1.

## Disclosure

The authors have declared no conflicts of interest.

## References

- Food and Drug Administration. Master Protocols: Efficient Clinical Trial Design Strategies to Expedite Development of Oncology Drugs and Biologics. Draft Guidance for Industry. Rockville, MD: Food and Drug Administration 2018.
- Benda N, Brandt A. Regulatory issues with multiplicity in drug approval: principles and controversies in a changing landscape. *J Biopharm Stat* 2018; 28(1): 3–9.
- Howard DR, Brown JM, Todd S, Gregory WM. Recommendations on multiple adjustment in multi-arm trials with a shared control group. *Stat Methods Med Res* 2018; 27(5): 1513–1530.
- Simon R. Biomarker based clinical trial design. *Chin Clin Oncol* 2014; 3(3): 39.
- Kim ES, Hirsh V, Mok T et al. Gefitinib versus docetaxel in previously treated non-small-cell lung cancer (INTEREST): a randomized phase III trial. *Lancet* 2008; 372(9652): 1809–1818.
- Liu A, Lee JJ. An overview of the design and conduct of the BATTLE trials. *Chin Clin Oncol* 2015; 4(3): 33.
- Simon N, Simon R. Adaptive enrichment designs for clinical trials. *Biostatistics* 2013; 14(4): 613–625.
- Stallard N, Hamborg T, Parsons N, Friede T. Adaptive designs for confirmatory clinical trials with subgroup selection. *J Biopharm Stat* 2014; 24(1): 168–187.
- Parashar D, Bowden J, Starr C et al. An optimal stratified Simon two-stage design. *Pharm Stat* 2016; 15(4): 333–340.
- Kaplan R, Maughan T, Crook A et al. Evaluating many treatments and biomarkers in oncology: a new design. *J Clin Oncol* 2013; 31(36): 4562–4570.
- Herbst RS, Gandara DR, Hirsch FR et al. Lung Master Protocol (Lung-MAP)—a biomarker-driven protocol for accelerating development of therapies for squamous cell lung cancer: SWOG S1400. *Clin Cancer Res* 2015; 21(7): 1514–1524.
- Woodcock J, LaVange LM. Master protocols to study multiple therapies, multiple diseases, or both. *N Engl J Med* 2017; 377(1): 62–70.
- Heinrich MC, Joensuu H, Demetri GD 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. *Clin Cancer Res* 2008; 14(9): 2717–2725.
- Renfro LA, Sargent DJ. Statistical controversies in clinical research: basket trials, umbrella trials, and other master protocols: a review and examples. *Ann Oncol* 2017; 28: 34–43.
- Middleton G, Crack LR, Popat S et al. The national lung matrix trial: translating the biology of stratification in advanced non-small-cell lung cancer. *Ann Oncol* 2015; 26(12): 2464–2469.
- Committee for Proprietary Medicinal Products (CPMP). Points to Consider on Multiplicity Issues in Clinical Trials. London: European Medicines Agency 2002.
- Dmitrienko A, Millen B, Lipkovich I. Multiplicity considerations in subgroup analysis. *Stat Med* 2017; 36(28): 4446–4454.
- Bretz F, Hothorn T, Westfall P. *Multiple Comparisons Using R*. Boca Raton, FL: CRC Press 2011.
- Sugitani T, Posch M, Bretz F, Koenig F. Flexible alpha allocation strategies for confirmatory adaptive enrichment clinical trials with a prespecified subgroup. *Stat Med* 2018; 37(24): 3387–3402.
- Sun H, Bretz F, Gerke O, Vach F. Comparing a stratified treatment strategy with standard treatment in a randomized clinical trial. *Stat Med* 2016; 35(29): 5325–5337.

doi:10.1093/annonc/mdz038  
Published online 31 January 2019