

COMMENTARY

What is the weight of expectation bias in oncology trials?

Cinzia Dello Russo^{1,2}  | Pierluigi Navarra¹ ¹Department of Healthcare Surveillance and Bioethics, Fondazione Policlinico Universitario A. Gemelli IRCCS, Università Cattolica Del Sacro Cuore, Rome, Italy²Department of Pharmacology & Therapeutics, Institute of Systems Molecular and Integrative Biology (ISMIB), University of Liverpool, Liverpool, UK**Correspondence**

Pierluigi Navarra, Chair of Pharmacology, Department of Healthcare Surveillance and Bioethics, Fondazione Policlinico Universitario A. Gemelli IRCCS, Università Cattolica Del Sacro Cuore, Rome, Italy.

Email: pierluigi.navarra@unicatt.it

In biomedical research, expectation bias, also referred to as Rosenthal effect, is the distorting effect on the results of an experiment caused by the expectation that the investigator, or the patient, has about the results themselves. The classical way to reduce the influence of this bias on the interpretation of results is through the blinding, or masking, of treatments, both terms referring to keeping the patients, the investigators, or the assessors unaware of the assigned treatment.

Talking of expectations, can we postulate that the expectation bias will be greater the higher the unmet medical need underlying the research is? And, for the above reason, is the expectation bias stronger in oncology studies? This postulate does not seem to apply to the patients, since expectation (acting in patients via the placebo effect) is rarely associated with tumour positive responses.¹ But, what about the investigators? If one looks at the recommendations of main regulatory bodies, the answer to this question will be that the investigator expectation is indeed considered an important confound in the interpretation of outcomes, thus requiring a blinded independent review to prevent the bias. In fact, in the current Food and Drug Administration (FDA) guideline on clinical trial endpoints for the approval of cancer drugs and biologics, the adoption of an independent blinded review is recommended for all the endpoints based on tumour assessments, namely disease-free survival, event-free survival, overall response rate (ORR), complete response, time to progression and progression-free survival (PFS), although for disease-free survival, event-free survival, time to progression and PFS, a decision should be taken on a case-by-case basis. Likewise, in the most recent European Medicines Agency (EMA) guideline, still available in draft, it is recommended that “if the study has to be conducted open label, this has implications with respect to choice of study endpoints, independent review, conduct of sensitivity analyses and other measures to be undertaken to limit potential bias related to the open-label nature of the trial”.

We could not say how much the recommendations of FDA and EMA guidelines were based on the analysis of clinical evidence, and how much on methodological reasoning. As we will see below, some evidence is available, but limited to PFS in phase-3 trials. For certain, it is now possible to gain a quantitative measure of the influence of investigator expectation over the results of oncology trials, since many papers have been published so far that report the assessment of 2 pivotal endpoints, PFS and ORR, carried out both at local level (local assessment, LA) and by a blinded independent central review (BICR), within the same trial. If the assessment is conducted at local level, the local investigator, that is, the oncologist, will be in contact with the radiologist, hence will be aware of the treatment assigned (even though they are not directly involved in the imaging assessment). However, they will be blinded to the outcomes of BICR assessment, and the masking of assigned treatment will be maintained. By looking at the putative differences in assessment between LAs and BICRs in an adequately sized sample of clinical trials, an estimation can be obtained of the weight of expectation bias in this setting, if any.

In the last 2 years, we have carried out extensive research on the topic of investigator expectation bias in oncology trials. We collected and analysed all phase-2 and phase-3 trials recorded in clinicaltrials.gov and EudraCT databases and reporting the results of PFS and/or ORR assessments carried out by both LAs and BICR within the same trial.

First, we focused on PFS in phase-3 trials, a topic that has been matter of debate for more than a decade. Initially, Dodd *et al.*, based on the analysis of 7 phase-3 trials showing no difference between the assessments of PFS carried out by LAs and BICR, raised the issue of BICR as an unnecessary, expensive and time-consuming procedure, which should not be used on a regular basis in confirmative phase-3 trials.² Soon after, a group of researchers from a consortium of pharmaceutical companies further expanded the initial observation by Dodd and colleagues, reporting the results of an analysis on

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2023 The Authors. *British Journal of Clinical Pharmacology* published by John Wiley & Sons Ltd on behalf of British Pharmacological Society.

27 phase-3 clinical trials where a dual estimation of PFS by LAs and BICR was carried out.^{3,4} These authors found a strong correlation ($R = .947$) between LA and BICR assessments of PFS and concluded that LA evaluation provides a reliable estimate of PFS, making the blinding of assessors unnecessary.^{3,4} Based on this evidence, a series of recommendation papers were published, suggesting the possibility to limit or even to abandon the blinded independent evaluation of PFS.⁵⁻⁷ A practical effect of this effort by the scientific community can be found in the FDA guideline, where PFS is 'not always recommended', whereas EMA guideline was not influenced by these proposals.

Within this framework, we carried out a first study looking at possible discrepancies between the assessment of PFS carried out by LAs and BICRs, respectively, in the setting of phase-3 trials.⁸ Usually, these trials involve a comparison between the experimental treatment and a control therapy, so that the results are expressed as the hazard ratio (HR) of PFS curves. On a sample of 28 randomized controlled trials, we calculated a *discrepancy index*, defined as the ratio of HRs assessed by LAs and BICRs, respectively. With a null hypothesis of no-difference between the 2 estimates, the expected discrepancy index is = 1; in this study, we obtained an average discrepancy index of 0.98 (95% confidence interval: 0.927-1.032), thereby confirming—although via a different methodological approach—the previous findings by Amit *et al.*⁴ It is noteworthy that the samples of the 2 studies were largely independent between each other, with only three trials in common. In conclusion, the evidence from ourselves and other groups is altogether consistent with the notion that there is no significant difference between the assessment of PFS carried out in blinded or in open-label conditions, indicating that the expectation of investigators does not influence the measure of PFS in the setting of phase-3 trials.

We then moved to investigate putative differences in the assessment of ORR between LAs and BICR in the context of phase-2 trials.⁹ Overall response rate is often chosen as the primary endpoint in the early phase of clinical development, seeking for preliminary evidence of efficacy. In this second study, 20 trials including the assessment of ORR by both LAs and BICR were selected, but the total number of comparisons analysed was 33, as many trials had >1 comparison carried out in the same study. Since the protocols involved single groups of patients, HRs were not available, and the discrepancy index was calculated as the ratio between each LA-assessed ORR and the corresponding ORR assessed by BICR.⁹ Again, the null hypothesis of no-difference was associated to an expected discrepancy index = 1, with values >1 indicating a *more optimistic* evaluation by LAs and vice versa for values <1. At variance with the study on PFS, here we found an average discrepancy index of 1.175 (95% confidence interval: 1.083-1.264). In 18 of 33 comparisons, a dual assessment of ORR and PFS was carried out; in this subgroup, the average discrepancy index for PFS was 1.092 (95% confidence interval: 0.96-1.22), and no significant correlation was found between ORR and PFS.⁹ In conclusion, a difference of +17.5% in favour of unblinded investigators strongly suggested that, at least in the setting of uncontrolled phase-2 studies, the expectation bias can influence in a significant manner the results.

The above conclusion raised the question of why investigators' expectation should influence the assessment of ORR but not that of PFS, both endpoints being obtained through the same methodological approach, that is, the evaluation of imaging and patients' categorization using validated criteria. To address this point, we set a third study aimed at analysing differences in the assessment of ORR and PFS that could account for the discrepancies between the results of the 2 previous studies. In this study we also completed the analysis of our database, looking at the putative discrepancies in assessment between LAs and BICRs in the settings of PFS in phase-2 trials and ORR in phase-3 controlled trials.¹⁰ We confirmed in phase-2 trials the previous observation about the lack of significant difference between the assessments of PFS carried out by LAs or BICR, which led us to conclude that the expectation of investigators does not influence the measure of PFS regardless of whether it is carried out within phase-2 or phase-3 trials. We also found that LAs tend to overestimate ORR compared with BICR in phase-3 trials as well. However, such overestimate was observed equally in experimental and control groups; therefore, expressing the results as HRs, the 2 errors tend to compensate each other, and the comparison between the assessments of LAs and BICR shows no significant difference.¹⁰

This last finding is highly relevant in the context of our working hypothesis since it stands against a possible role of expectation bias in influencing the assessment of ORR by local investigators. In fact, in this case we would expect the overestimate to be significantly higher in experimental groups compared with controls. Instead, the present results rather suggest that a methodology bias, inherent to the measurement of ORR, may be responsible for the discrepancy in assessment between ORR and PFS. In the same paper, we attempted to analyse this phenomenon, indicating as possible determinants: (i) a limited number of measurements for ORR, compared with repeated measures with PFS; (ii) the time-to-response, which is a variable for ORR, whereas PFS is always measured after a response is established; (iii) the type of treatment, with small molecules in general inducing faster responses than immunotherapies; (iv) moreover, some protocols may assess ORR at fixed times, while other may consider the *best response* to calculate ORR.¹⁰

In conclusion, after completing the analysis of the whole dataset of phase-2 and phase-3 trials reporting the assessments of PFS and ORR by both LAs and BICR, we may now attempt to answer to the question entitling this paper: (i) concerning PFS, it is now well established that the expectation bias of investigators does not influence the results of assessment, either in phase-2 or in phase-3 studies; (ii) likewise, no differences are observed between the assessments of ORR carried out by LAs and BICRs, provided that trials have a control group; (iii) a strong concern remains on the assessment of ORR in single-group nonrandomized phase-2 trials, since a significant difference in the assessment between LAs and BICR is recorded in this setting. This conclusion is of special interest, if one considers that an increasing number of oncology drugs nowadays is granted a conditional approval based on the evidence from nonrandomized phase-2 studies.

COMPETING INTERESTS

Cinzia Dello Russo and Pierluigi Navarra declare no conflict of interest.

ACKNOWLEDGEMENTS

Open Access Funding provided by Universita Cattolica del Sacro Cuore within the CRUI-CARE Agreement.

CONTRIBUTORS

PN and CDR conceived the paper. PN drafted the manuscript. CDR critically reviewed the manuscript.

FUNDING INFORMATION

This work received no financial support.

ORCID

Cinzia Dello Russo  <https://orcid.org/0000-0002-2538-3832>

Pierluigi Navarra  <https://orcid.org/0000-0002-4424-650X>

REFERENCES

1. Chvetzoff G, Tannock IF. Placebo effects in oncology. *J Natl Cancer Inst.* 2003;95(1):19-29. doi:[10.1093/jnci/95.1.19](https://doi.org/10.1093/jnci/95.1.19)
2. Dodd LE, Korn EL, Freidlin B, et al. Blinded independent central review of progression-free survival in phase III clinical trials: important design element or unnecessary expense? *J Clin Oncol.* 2008;26(22):3791-3796. doi:[10.1200/JCO.2008.16.1711](https://doi.org/10.1200/JCO.2008.16.1711)
3. Amit O, Bushnell W, Dodd L, Roach N, Sargent D. Blinded independent central review of the progression-free survival endpoint. *Oncologist.* 2010;15(5):492-495. doi:[10.1634/theoncologist.2009-0261](https://doi.org/10.1634/theoncologist.2009-0261)
4. Amit O, Mannino F, Stone AM, et al. Blinded independent central review of progression in cancer clinical trials: results from a meta-

analysis. *Eur J Cancer.* 2011;47(12):1772-1778. doi:[10.1016/j.ejca.2011.02.013](https://doi.org/10.1016/j.ejca.2011.02.013)

5. Dancey JE, Dodd LE, Ford R, et al. Recommendations for the assessment of progression in randomised cancer treatment trials. *Eur J Cancer.* 2009;45(2):281-289. doi:[10.1016/j.ejca.2008.10.042](https://doi.org/10.1016/j.ejca.2008.10.042)
6. Stone AM, Bushnell W, Denne J, et al. Research outcomes and recommendations for the assessment of progression in cancer clinical trials from a PhRMA working group. *Eur J Cancer.* 2011;47(12):1763-1771. doi:[10.1016/j.ejca.2011.02.011](https://doi.org/10.1016/j.ejca.2011.02.011)
7. Pignatti F, Hemmings R, Jonsson B. Is it time to abandon complete blinded independent central radiological evaluation of progression in registration trials? *Eur J Cancer.* 2011;47(12):1759-1762. doi:[10.1016/j.ejca.2011.05.009](https://doi.org/10.1016/j.ejca.2011.05.009)
8. Dello Russo C, Cappoli N, Navarra P. A comparison between the assessments of progression-free survival by local investigators versus blinded independent central reviews in phase III oncology trials. *Eur J Clin Pharmacol.* 2020;76(8):1083-1092. doi:[10.1007/s00228-020-02895-z](https://doi.org/10.1007/s00228-020-02895-z)
9. Dello Russo C, Cappoli N, Pilunni D, Navarra P. Local investigators significantly overestimate overall response rates compared to blinded independent central reviews in phase 2 oncology trials. *J Clin Pharmacol.* 2021;61(6):810-819. doi:[10.1002/jcph.1790](https://doi.org/10.1002/jcph.1790)
10. Dello Russo C, Navarra P. Local investigators significantly overestimate overall response rates compared to blinded independent central reviews in uncontrolled oncology trials: a comprehensive review of the literature. *Front Pharmacol.* 2022;13:858354. doi:[10.3389/fphar.2022.858354](https://doi.org/10.3389/fphar.2022.858354)

How to cite this article: Dello Russo C, Navarra P. What is the weight of expectation bias in oncology trials? *Br J Clin Pharmacol.* 2023;1-3. doi:[10.1111/bcp.15680](https://doi.org/10.1111/bcp.15680)