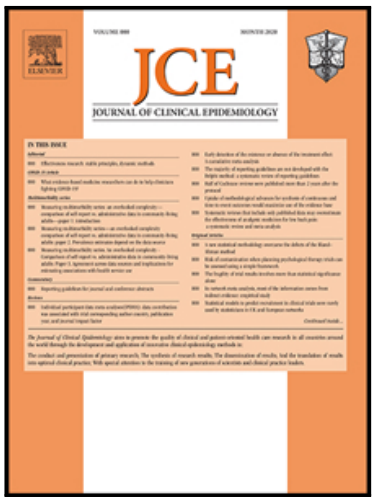


The Net Benefit of a treatment should take the correlation between benefits and harms into account

Marc Buyse , Everardo D. Saad , Julien Peron , Jean-Christophe Chiem , Mickaël De Backer , Eva Cantagallo , Oriana Ciani

PII: S0895-4356(21)00094-9
DOI: https://doi.org/10.1016/j.jclinepi.2021.03.018
Reference: JCE 10461

To appear in: *Journal of Clinical Epidemiology*

Accepted date: 18 March 2021

Please cite this article as: Marc Buyse , Everardo D. Saad , Julien Peron , Jean-Christophe Chiem , Mickaël De Backer , Eva Cantagallo , Oriana Ciani , The Net Benefit of a treatment should take the correlation between benefits and harms into account, *Journal of Clinical Epidemiology* (2021), doi: https://doi.org/10.1016/j.jclinepi.2021.03.018

Highlights

- The assessment of benefits and harms from experimental treatments often ignores the association between outcomes
- The method of generalized pairwise comparisons (GPC) takes into account the association between endpoints
- A Net Benefit computed using GPC leads to very different conclusions about the benefit/risk of treatment than when only marginal benefits are used
- When data from randomized clinical trials are available, the benefit/risk assessment should use GPC rather than marginal treatment effects

The Net Benefit of a treatment should take the correlation between benefits and harms into account

Marc Buyse, ScD [1,2], Everardo D. Saad, MD [3], Julien Peron, MD, PhD [4,5],
Jean-Christophe Chiem, PhD [3], Mickaël De Backer, PhD [6],
Eva Cantagallo, MSc [7] and Oriana Ciani, PhD [8,9]

[1] International Drug Development Institute, San Francisco, CA, USA; [2] Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BioStat), Hasselt University, Diepenbeek, Belgium; [3] International Drug Development Institute, Louvain-la-Neuve, Belgium; [4] Hospices Civils de Lyon, departments of Oncology and Biostatistics, Pierre-Benite, France; [5] University of Lyon 1, CNRS UMR 5558, Biometry and Evolutive Biology Laboratory, Biostatistics-Health Team, Villeurbanne, France; [6] Institut de statistique, biostatistique et sciences actuarielles, Université Catholique de Louvain, Louvain-la-Neuve,

Belgium; [7] European Organisation for Research and Treatment of Cancer (EORTC), Brussels, Belgium; [8] CERGAS - Università Commerciale L. Bocconi, Milan, Italy; [9] University of Exeter Medical School, Evidence Synthesis & Modelling for Health Improvement, Exeter, UK

Address for correspondence:

Marc Buyse, ScD

IDDI

Av. Provinciale 30

1340 – Louvain-la-Neuve, Belgium

marc.buyse@iddi.com

SOURCES OF SUPPORT

This research was partially funded by the regions of Wallonia (BioWin Consortium Agreement No 7979) and Brussels-Capital (Innoviris). EC works as a Fellow at EORTC and was supported by a grant from the EORTC Cancer Research Fund.

KEYWORDS

Generalized pairwise comparisons, prioritized outcomes, benefit/harm, Net Benefit

RUNNING TITLE

Prioritized outcomes to assess Net Benefit

WORD COUNT: 4,626

WHAT IS NEW

A Net Benefit computed using the method of generalized pairwise comparisons and taking into account the association between two binary endpoints leads to very different conclusions about the benefit/risk of treatment than when only marginal benefits are used. When data from randomized clinical trials are available, the benefit/risk assessment should use generalized pairwise comparisons rather than marginal treatment effects.

ABSTRACT

Objective: The assessment of benefits and harms from experimental treatments often ignores the association between outcomes. Generalized pairwise comparisons (GPC) can be used to assess the Net Benefit of treatment in a randomized trial accounting for that association.

Study Design and Settings: We use GPC to analyze a fictitious trial of treatment versus control, with a binary efficacy outcome (response) and a binary toxicity outcome, as well as data from two actual randomized trials in oncology. In all cases, we compute the Net Benefit for scenarios with different orders of priority between response and toxicity, and a range of odds ratios (ORs) for the association between outcomes.

Results: The GPC Net Benefit was quite different from the benefit/harm computed using marginal treatment effects on response and toxicity. In the fictitious trial using response as first priority, treatment had an unfavorable Net Benefit if OR<1, but favorable if OR>1. With OR=1, the Net Benefit was 0. Results changed drastically using toxicity as first priority.

Conclusion: Even in a simple situation, marginal treatment effects can be misleading. In contrast, GPC assesses the Net Benefit as a function of the treatment effects on each outcome, the association between outcomes, and individual patient priorities.

INTRODUCTION

The assessment of benefits and harms from experimental treatments usually comprises qualitative and quantitative considerations, and relies on sources of data as varied as hospital records, pharmacovigilance databases, randomized and non-randomized clinical trials, etc.[1] A review of methods for benefit/harm assessment in systematic reviews identified four stages: a review of the reported benefits and harms of an intervention; quantitative assessments of the intervention's benefits as compared with harms; decision-making at the population level; and decision-making at the individual level.[2] In this paper we will focus on the quantitative assessment of benefit/harm (also called benefit/risk). The quantitative assessment of benefit/risk is typically based on aggregate data (i.e., summary statistics) for the outcomes observed with the intervention as compared with standard of care or competing interventions. In particular, efficacy and safety are usually analyzed separately, possibly using different data sources, and the results of these analyses are combined quantitatively into a benefit/harm (also called benefit/risk) assessment. Many methods have been proposed and used to perform this quantitative assessment, but it has long been recognized that a limitation of approaches based on aggregate data is that the association between the different outcomes of interest is ignored.[3]

Randomized clinical trials provide the most reliable evidence about (some of the) benefits and harms of experimental interventions, if only because confounding is eliminated by design and the risk of selection and accidental bias thereby reduced as much as possible. However, even in randomized trials, marginal analyses that purport to estimate the effect of an intervention on efficacy and toxicity outcomes independently of each other do not

reflect the association between these outcomes, which is crucial in interpreting the benefit/risk of the intervention.[4] In a hypothetical trial of an experimental treatment having better efficacy but also higher toxicity than the standard of care, it is crucial to know if patients deriving benefit are also those having harm, or if, conversely, patients are suffering toxicity without any gain in efficacy. For benefit/risk analyses of randomized clinical trials to account for the association between benefits and harms of an intervention, individual patient data must be available.[3]Oncology offers some striking situations in which understanding the association between efficacy and toxicity is needed for a meaningful benefit/harm assessment. Three prototypical situations may exist. In the first situation, there is independence between the benefits and harms, *i.e.*, patients who have toxicity may or may not respond to treatment. For instance, DNA-intercalating agents such as anthracyclines may induce late cardiac toxicities that are more likely to occur in patients who are frail or who have cardiac risk factors before receiving the treatment, independently of whether these drugs also produce an antitumor effect in these patients while they are treated.[5] In the second situation, there is a positive association between response and toxicity, *i.e.*, patients who have toxicity are also more likely to respond. For instance, inhibitors of the epidermal growth factor receptor (EGFR) pathway induce severe skin rash that is associated with response while on treatment.[6,7] In the third situation, there is a negative association between response and toxicity, *i.e.*, patients who have toxicity may be less likely to respond. For instance, patients with enzyme deficiencies may experience excessive toxicity that leads to stopping fluoropyrimidine- and irinotecan-based therapies, hence patients having toxicity are unlikely to derive benefit from treatment.[8,9] Thus, the

proper assessment of the benefit/risk of a treatment requires knowledge of the association between benefits and harms.

In this paper, we use fictitious and actual examples in oncology to illustrate how the recently proposed method of generalized pairwise comparisons (GPC) of two or more prioritized outcomes can be employed to assess the net effect of treatment when individual patient data are available from randomized trials.[10] When using GPC, we refer to this net effect as the "Net Benefit", which can be favorable or unfavorable depending on whether benefits or harms predominate in a given analysis (always comparing an experimental treatment *vs* control). Our goal is to contrast a benefit/harm analysis conducted using the Net Benefit against a traditional assessment based on the "marginal net benefit", *i.e.*, differences between benefits and harms computed using the average (marginal) effects of treatment on each outcome (and independently of the effect on other outcomes). The Net Benefit we refer to is not to be mistaken for the intervention's "average net health benefit" as defined by Stinnett and Mullahy.[11] In essence, their "average net health benefit" determines the effectiveness of an intervention with a minimum health effect that society would demand in return for its investment. Except for the fact that cost is a component of the "average net health benefit", the latter may be considered conceptually similar to the "marginal net benefit".

We first consider a fictitious trial in which patients are randomized to receive an experimental treatment or control. We show that even in a simple situation with only two binary outcomes, marginal treatment effects on response and toxicity can be grossly

misleading, whilst the Net Benefit provides a relevant assessment of the benefit/harm relationship. We then apply the method to two actual clinical trials in oncology where an assessment of the relationship between efficacy and safety endpoints is key to interpreting the differences between the randomized arms.

METHODS

Generalized pairwise comparisons of prioritized outcomes

The method of GPC has been described in detail elsewhere and is summarized in Appendix 1.[10] Essentially, each individual in the intervention group is compared against each individual in the control group and the resulting pair is assigned a score of +1, -1 or 0 depending on whether the pair is a win, a loss or a tie. The Net Benefit is then the average of scores for all possible pairs. In the simple situation of a single binary efficacy outcome ("response") and a single binary safety outcome ("toxicity"), the score is assigned either by prioritizing response first and toxicity second, or vice-versa by prioritizing toxicity over response (see Appendix 1).

Fictitious clinical trial

Assume that the control treatment has a response probability of 0.2 and no toxicity (a toxicity probability of 0), while the experimental treatment has a higher response probability of 0.5, but at the cost of a toxicity probability of 0.6 (Table 1). At the patient level, the association between response and toxicity can be characterized generally by the three prototypical situations depicted in Tables 1(A), 1(B) and 1(C). These tables display the 2x2 cross-tabulation of response by toxicity for the experimental treatment (left-hand

side of each table) and control (right-hand side). The margins of these 2x2 tables are the response and toxicity probabilities used above to calculate the marginal net benefit (in this case, unfavorable). The control treatment has no toxicity, so the cross-tabulation of response by toxicity reduces to the probability of response for this treatment. In contrast, the experimental treatment has some toxicity, so the cross-tabulation of response by toxicity can show three different situations for this treatment. Table 1(A) shows a situation in which response and toxicity are independent. In this situation, the odds of having a response are independent of the odds of having a toxicity; this corresponds to a response *vs.* toxicity odds ratio equal to 1. Table 1(B) shows a situation in which toxicity is positively associated with response (patients who have a toxicity tend to have a response). The most extreme positive association compatible with the margins of the table occurs when all patients who do not have a response also do not have a toxicity; this corresponds to a response *vs.* toxicity odds ratio equal to infinity. Table 1(C) shows a situation in which toxicity is negatively associated with response (patients who have no toxicity tend to have a response). The most extreme negative association compatible with the margins of the table occurs when all patients who do not have a response have a toxicity; this corresponds to a response *vs.* toxicity odds ratio equal to 0. If only the margins are known, the cell probabilities of the 2x2 tables for these three situations (OR = 1, ∞ or 0) shown in Table 1 can be calculated for any chosen OR using the formula of Appendix 2 in the Supplementary Material.

Table 1 here

A clinical trial in resectable colorectal cancer

An actual example for the need of a benefit/harm analysis is provided by a recent meta-analysis of trials comparing 3 months *vs.* 6 months of oxaliplatin-based adjuvant chemotherapy for patients with stage III colon cancer.[12] The longer duration of 6 months was considered the standard of care in many countries, but this treatment leads to significant toxicity, in particular, an oxaliplatin-induced cumulative, dose-dependent peripheral sensory neuropathy (PSN), a clinically serious issue that has a clear impact on quality of life. PSN is particularly bothersome when it is graded 3 or 4 according to the National Cancer Institute Common Terminology Criteria for Adverse Events,[13] which corresponds to sensory loss or paresthesia that interfere with activities of daily living or function. Here we consider one of the phase III trials included in the meta-analysis, the International Duration Evaluation of Adjuvant (IDEA) France (ClinicalTrials.gov Identifier: NCT00958737). In this trial, 25% of the patients randomized to 6 months of therapy (the control arm) experienced PSN of grade 3 or 4, compared with only 8% of the patients randomized to 3 months of therapy. This difference in the incidence of severe PSN is to be contrasted with a better disease-free survival (DFS) for patients randomized to 6 months of therapy than for those randomized to 3 months of therapy (hazard ratio of 1.24 against the shorter duration, $P = 0.011$).[14]

Table 2 shows the cross-classification of response (no DFS event) by toxicity (grade 3 or 4 PSN) in the same situations as for the fictitious example: no association (Table 2(A)), strong positive association between response and toxicity (Table 2(B)), and strong negative association (Table 2(C)). The published results of IDEA France do not provide this cross-

classification,[14] and therefore we will explore the full range of possibilities, assuming the same odds ratio for the long and short treatment duration.

Table 2 here

A clinical trial in advanced lung cancer

A second actual example for the need of a benefit/harm analysis is provided by a recent trial comparing combination chemotherapy (cisplatin plus pemetrexed) with the targeted agent afatinib in patients with advanced non-small-cell lung cancer (ClinicalTrials.gov Identifier: NCT00949650). Afatinib is an EGFR tyrosine kinase inhibitor with anti-tumoral activity in patients harboring activating EGFR mutations; unfortunately, it also induces acute toxicities, especially in Asian patients.[15] Here we consider the cohort of 83 Japanese patients randomized in study LUX-Lung 3.[16] In this cohort, the proportion of patients who were progression-free at 6 months favored afatinib over combination chemotherapy (Figure 1), but almost half of the patients receiving this drug had to undergo dose reductions because of severe skin rash or diarrhea versus none of the patients receiving combination chemotherapy (Figure 2). Using the detailed published data of Figures 1 and 2, we could reconstruct a dataset with individual patient data with treatment indicator, PFS status at 6 months and the occurrence of severe skin rash and/or diarrhea.[16]

Figures 1 and 2 here

Software

A ShinyApp is available at https://benefit.shinyapps.io/Net_Benefit_App/ to perform the calculations shown in this paper, or to explore other scenarios for a binary efficacy outcome

and a binary safety outcome. Software to implement generalized pairwise comparisons is available in the R package BuyseTest, which can be freely downloaded from GitHub and CRAN.

RESULTS

Fictitious clinical trial – Marginal net benefit

If only aggregate data were available for the fictitious trial described above (as opposed to patient-level data), the higher response probability with the experimental treatment than with control (difference in marginal response probabilities of 0.5 – 0.2 = 0.3) would point to a benefit of the experimental treatment in terms of its efficacy, while the toxicity observed with the experimental treatment (difference in marginal toxicity probabilities of 0.6 – 0 = 0.6) would point to harm of the experimental treatment in terms of its safety. Under a simple quantitative decision framework that takes these two outcomes into account without any priority assigned to them, the marginal net benefit (the difference between the marginal benefit and the marginal harm from the experimental treatment and control, all calculated from the margins of Table 1) would thus be equal to -0.3 (= 0.3 – 0.6) against the experimental treatment. A naïve quantitative benefit/harm assessment might thus conclude unfavorably against the experimental treatment, given that the marginal net benefit is negative.

Fictitious clinical trial – Net Benefit

Even though patient-level data are not available, the Net Benefit can be estimated using GPCs of prioritized outcomes that are based on the odds ratios for each of the three

prototypical situations shown in Table 1 and also for all intermediate levels of association between response and toxicity. In Figures 3 and 4, the x-axis represents the odds ratios between response and toxicity, while the y-axis represents the Net Benefit of the experimental treatment as compared with control.

Figure 3 here

Figure 3 shows the Net Benefit of the experimental treatment when achievement of a response has first priority, and avoidance of toxicity second priority. Table S2 of Appendix 3 in the Supplementary Material shows the calculations of the Net Benefit in the case of independence (OR = 1 as in Table 1(A)). Figure 3 shows the contribution of each outcome to the Net Benefit; the contribution of the first priority outcome is the increased response rate of 0.3 with the experimental treatment (straight blue line in Figure 3). This contribution is independent of the association between the two outcomes, because in pairwise comparisons, the difference between the proportions of wins and losses is always 0.3 in favor of the experimental treatment. The red line shows the contribution of the second priority outcome, toxicity, for pairwise comparisons that are ties for the first priority outcome. The contribution of the second priority outcome always favors control, which has no toxicity, but its magnitude depends on the association between outcomes. The black line is the Net Benefit which, in this example, favors the experimental treatment if the association between response and toxicity is positive (OR>1), but favors control if the association between response and toxicity is negative (OR<1). Thus, with achievement of a response having first priority, the experimental treatment would be preferable to control if and only if response to the experimental treatment was obtained at the price of toxicity for responding patients (OR>1). This comes from the fact that the second outcome comes into

play when there is a tie for the first outcome; hence the best-case scenario is when the patients who do best for response also do worst for toxicity. If, in contrast, lack of response to the experimental treatment was associated with toxicity, then the control treatment would be preferable (OR<1). If there was no association between response and toxicity (OR=1), then the Net Benefit would be equal to 0 in this scenario with response as the first priority. In other words, the conclusion about the Net Benefit of the experimental treatment would change depending on the association between response and toxicity, even if response remains as the first priority.

Figure 4 here

The above observations drastically change if the order of priorities for outcomes is reversed. Figure 4 shows the Net Benefit of the experimental treatment when avoidance of toxicity has first priority. In this example, the high rate of toxicity due to the experimental treatment is never compensated by the higher response rate obtained with the experimental treatment, so the experimental treatment has an unfavorable Net Benefit regardless of the association between response and toxicity. However, the magnitude of the Net Benefit again depends on the association between response and toxicity, and ranges from -0.28 (in the case of a strong negative association, OR≈0) to -0.68 (in the case of a strong positive association, OR≈∞). Obviously, the patterns observed in this example are not general properties of the Net Benefit; as the (more realistic) example in colorectal cancer will confirm, the Net Benefit is a function of the effects of each treatment on both outcomes, the association between these effects, and the priority of each outcome.

Resectable colorectal cancer trial

At the time of analysis in the IDEA France trial, the number of patients still alive and disease-free (response outcome) was 688 (of 1002) in the group receiving 3 months of treatment *vs.* 744 (of 1008) in the group receiving 6 months of treatment, which reflected a significant benefit of the longer treatment duration: the marginal benefit of the shorter treatment in terms of response was equal to 688/1002 - 744/1008 = - 0.051 (Table 2). In contrast, the number of patients who did not experience a grade 3 or 4 PSN (toxicity outcome) was 923 (of 1002) in the group receiving 3 months of treatment *vs.* 753 (of 1008) in the group receiving 6 months of treatment, which favored the shorter treatment duration: the marginal benefit in terms of toxicity was equal to 923/1002 - 753/1008 = 0.174 (Table 2).[14] The marginal net benefit thus calculated is therefore equal to -0.051 + 0.174 = 0.123 in favor of the shorter treatment duration.

Table 2 shows the cross-tabulation of the response outcome and the toxicity outcome. The published results of the trial only provide marginal numbers of events, and therefore an assumption must be made, as in the fictitious trial, about the magnitude of the odds ratio between response and toxicity. Once again, we assume odds ratios of 0, 1 or infinity for both treatment arms. In reality, the odds ratios between response and toxicity could differ for different treatment durations, but the results shown here do not crucially depend on this assumption.

Figure 5 here

Figure 5 shows the Net Benefit of the shorter treatment duration when achievement of a response has first priority, and avoidance of toxicity second priority. Despite the lower proportion of patients alive and disease-free in the shorter treatment duration (a harm of -

0.051), the Net Benefit is always positive regardless of the association between response and toxicity; the Net Benefit ranges from 0.007 if all patients with grade 3 or 4 PSN also had a DFS event (Table 2 (C)) to 0.06 when no patient with grade 3 or 4 PSN also had a DFS event (Table 2 (B)). If DFS events were independent of grade 3 or 4 PSN, the Net Benefit would be equal to 0.05. Hence, regardless of the assumption made about the response *vs.* toxicity odds ratio, the Net Benefit in favor of the shorter duration is considerably smaller than the marginal net benefit (equal to 0.123).

Figure 6 here

Figure 6 shows the Net Benefit of the shorter treatment duration when avoidance of toxicity has first priority. Due to the higher proportion of patients who did not experience grade 3 or 4 PSN in the shorter treatment duration (a benefit of 0.174), the Net Benefit is always positive regardless of the association between response and toxicity; the Net Benefit ranges from 0.007 if all patients with grade 3 or 4 PSN also had a DFS event (Table 2 (C)) to 0.181 when no patient with grade 3 or 4 PSN also had a DFS event (Table 2 (B)). Once again, the magnitude of the Net Benefit depends on the association between response and toxicity. Note that if patient-level data were available, a single association between response and toxicity would be calculated, and therefore a single Net Benefit would be calculated once the order of priorities between response and toxicity had been decided. Such an analysis, based on individual patient data, is currently in progress.

Advanced lung cancer trial Figure 1 shows that the proportion of patients alive and progression-free at 6 months was higher for patients treated with afatinib (81%) than for those treated with combination chemotherapy (45%). The proportion of patients having an

efficacy benefit was 0.36 (0.81-0.45) in favor of afatinib. In contrast, Figure 2 shows that 24 (44%) of the 54 patients treated with afatinib suffered severe skin rash and/or diarrhea, versus none treated by combination chemotherapy. Hence the proportion of patients harmed by these toxicities was 0.44 (0.44 - 0). The marginal benefit/harm is equal to 0.36-0.44 = -0.08, which suggests a net harm of afatinib for these three outcomes. Net Benefit analyses using GPC for these three outcomes are shown in Table 3, for each outcome separately, and with all three of them in two different orders of priority. The treatment difference is statistically significant for each of these three outcomes (Table 3). If being alive and progression-free at 6 months is the outcome of first priority, the Net Benefit of 0.16 still favors afatinib despite the severe skin rash and diarrhea (though this Net Benefit is no longer statistically significant, *P*=0.20, Table 3). If not experiencing these toxicities is preferred, then the negative Net Benefit of -0.26 indicates a net harm of afatinib as compared with combination chemotherapy. Of note, neither of these analyses of prioritized outcomes is close to the marginal net benefit of -0.08, which is a mathematical construct that does not account for the correlation structure of the benefit and harms considered.

DISCUSSION

The examples provided in this paper demonstrate that the independent calculation of the net effect from a treatment as the difference between a marginal benefit and a marginal harm is potentially misleading and should not generally be used. If individual patient data are available for the comparison of an experimental treatment with standard of care, GPC can be used to estimate the Net Benefit, a mathematically sound metric that takes into account the association between the outcomes considered. In our hypothetical trial, for

example, we have assumed the common situation in which the control treatment is poorly effective but caused no toxicity, while the experimental treatment is highly effective but caused substantial toxicity. In this situation, the association between response and toxicity is only relevant for the experimental treatment. The situation of independence (Table 1(A)) is expected when the experimental treatment has beneficial effects that are independent of its harmful effects. The situation of positive association (Table 1(B)) is expected when beneficial and harmful effects of the experimental treatment are on the same mechanistic pathway, or when the treatment is likely to be given for a longer time period if it is efficient for a given patient. The situation of negative association (Table 1(B)) may be the least plausible biologically, as it corresponds to a treatment that causes toxicity only in patients who do not respond to it. It might happen typically when the occurrence of a toxicity impairs the delivery of the treatment and of any other effective treatment, or when the toxicity directly alters the efficacy endpoint (for example when the toxicity is fatal and the efficacy endpoint is overall survival).

GPC generally provides a different assessment of benefits and harms than that provided by marginal probabilities. As our examples illustrate, this difference can be both quantitative and qualitative. The quantitative component of the difference is illustrated in Figures 3 to 6, which show that the Net Benefit depends on the association between outcomes. Other statistical models have been proposed to take this association into account,[17-19] and have proven particularly useful when benefit/harm is assessed on a large number of outcomes.[20,21] The qualitative component represents the ability to choose priorities explicitly and in a personalized manner. This is particularly relevant for informing decision-

making at the individual level.[2] Some patients will prioritize response first, whilst others will prefer to avoid toxicity. The numerical examples of Table 1 and Figure 3 show that the decision to take the experimental treatment may depend on these patient preferences. In oncology, for instance, while some patients will prefer to avoid substantial toxicities from aggressive late-line therapies when the expected gain in efficacy is minimal (say, an extension of survival by a few weeks only), other patients may have a marked preference for being treated.[22,23] Similar variation has been reported or is likely to exist in other medical specialties.[24] Benefit/harm analyses, to be patient-relevant, need to explore a range of potential patient preferences and a range of clinical situations for which different baseline risks may call for different benefit/risk assessments.[17,25] Patient-dependent priorities in GPC are in line with the objective of the Food and Drug Administration's patient-focused drug development initiative.[26] As such, GPC may help pave the way to patient-centric clinical research, and further to personalized medicine.[1]

Although the priorities chosen for the various outcomes may be highly individual, they can also be obtained from a consensus of an expert panel for decision-making at the population level.[2] This approach may be used, for example, when regulatory or health-technology-assessment authorities need to make decisions about approval and reimbursement of an experimental treatment. Quantitative models that estimate the Net Benefit as shown in this paper not only take the association between the various outcomes considered into account, they can also be fitted under different scenarios that reflect patient preferences or agreed-upon priorities. Approval or reimbursement of an experimental treatment could be granted when the consensual set of preferences is associated with a favorable Net Benefit. Research

is ongoing to characterize the Net Benefit under multiple (possibly many) scenarios of outcome priorities, an approach that could be used to inform the approval or reimbursement of an experimental treatment when a sufficient proportion of the scenarios considered to be plausible conclude that the Net Benefit is favorable.

In practice, benefit/harm will typically be assessed on many outcomes, not just one binary efficacy outcome and one binary safety outcome as in our simplified examples. Moreover, benefit(s) and harm(s) are usually assessed via outcomes based on different types of variables. These outcomes will often be measured on continuous scales (including times to event), with associated thresholds of clinical relevance, such as an improvement of at least 6 months in survival time.[27] GPC easily handle any type and any number of variables, provided patient-level data are available.[10] The analysis of the trial in resectable colorectal cancer (Table 2) and in advanced lung cancer (Figures 1 and 2) are shown here for illustration; in practice, time-to-event outcomes such as DFS in early disease and PFS in advanced disease would not be analyzed simply by counting the number of events by treatment arm because such analyses do not take individual times to event into account, and depend crucially on the duration of follow-up. A more meaningful way to analyze DFS or PFS, using GPC, is to assess the DFS or PFS Net Benefit by at least *m* months,[28] *m* being a threshold of minimal clinically relevant difference, corresponding to the preference of one stakeholder. The approach can then be extended by analyzing two prioritized outcomes, one capturing the survival Net Benefit by at least *m* months, the other severe toxicities. A threshold can also be chosen for toxicities, if these are graded on an ordered scale.[13] For illustration, these analyses were performed on three randomized trials for patients with

metastatic pancreatic cancer, and showed very different benefit/harm profiles of erlotinib plus gemcitabine,[28] FOLFIRINOX,[29] and nab-paclitaxel plus gemcitabine[30] as compared with gemcitabine alone or plus placebo. The possibility to choose thresholds of clinical relevance for efficacy and safety outcomes may make GPC particularly attractive for personalized medicine.

GPC analyses are conceptually simple and easily explained, but they can only be performed when patient-level data are available from randomized clinical trials that measured all important efficacy and safety outcomes. This is a limitation they share with other methods for benefit/risk assessment that account for the association between the outcomes of interest.[3,17-19,31,32] There has been a recent push to making data from randomized trials available for further analyses, so there is good hope that in the future, benefit/harm analyses will be possible that are at once mathematically correct and patient-relevant.[33-35]

Author contributions

Marc Buyse: Conceptualization; Data curation; Formal analysis; Writing - original draft; Writing - review & editing

Everardo D. Saad: Conceptualization; Writing - original draft; Writing - review & editing

Julien Peron: Conceptualization; Methodology; Writing - review & editing

Jean-Christophe Chiem: Conceptualization; Data curation; Writing - review & editing

Mickaël De Backer: Conceptualization; Methodology; Writing - review & editing

Eva Cantagallo: Conceptualization; Methodology; Writing - review & editing

Oriana Ciani: Conceptualization; Writing - review & editing

Conflicts of Interest

Marc Buyse declares no conflict of interest in relation to this study.

Everardo D. Saad declares no conflict of interest in relation to this study.

Julien Peron declares no conflict of interest in relation to this study.

Jean-Christophe Chiem declares no conflict of interest in relation to this study.

Mickaël De Backer declares no conflict of interest in relation to this study.

Eva Cantagallo declares no conflict of interest in relation to this study.

Oriana Ciani declares no conflict of interest in relation to this study.

REFERENCES

1. European Medicines Agency (EMA). Benefit-risk methodology project. Work package 2 report: Applicability of current tools and processes for regulatory benefit-risk assessment. EMA/549682/2010 - Revision 1, 2010. Available at https://www.ema.europa.eu/en/documents/report/benefit-risk-methodology-project-work-package-2-report-applicability-current-tools-processes_en.pdf (accessed on 29/08/2019).
2. Boyd CM, Singh S, Varadhan R, Weiss CO, Sharma R, Bass EB, Puhan MA. Methods for Benefit and Harm Assessment in Systematic Reviews. Methods Research Report. (Prepared by the Johns Hopkins University Evidence-based Practice Center under contract No. 290-2007-10061-I). AHRQ Publication No. 12(13)-EHC150-EF. Rockville, MD: Agency for Healthcare Research and Quality; November 2012.
3. Puhan MA, Singh S, Weiss CO, Varadhan R, Boyd CM. A framework for organizing and selecting quantitative approaches for benefit-harm assessment. BMC Med Res Methodol 2012;12:173.
4. Evans SR, Follmann D. Using Outcomes to Analyze Patients Rather than Patients to Analyze Outcomes: A Step toward Pragmatism in Benefit:risk Evaluation. Stat Biopharm Res 2016;8:386-93.
5. Lotrionte M, Biondi-Zoccai G, Abbate A, et al. Review and meta-analysis of incidence and clinical predictors of anthracycline cardiotoxicity. Am J Cardiol 2013;112:1980-4.
6. Abdel-Rahman O, Fouad M. Correlation of cetuximab-induced skin rash and outcomes of solid tumor patients treated with cetuximab: a systematic review and meta-analysis. Crit Rev Oncol Hematol 2015;93:127-35.
7. Liu HB, Wu Y, Lv TF, et al. Skin rash could predict the response to EGFR tyrosine kinase inhibitor and the prognosis for patients with non-small cell lung cancer: a systematic review and meta-analysis. PLoS One 2013;8:e55128.
8. Ezzeldin H, Diasio R. Dihydropyrimidine dehydrogenase deficiency, a pharmacogenetic syndrome associated with potentially life-threatening toxicity following 5-fluorouracil administration. Clin Colorectal Cancer 2004;4:181-9.
9. Glimelius B, Garmo H, Berglund A, et al. Prediction of irinotecan and 5-fluorouracil toxicity and response in patients with advanced colorectal cancer. Pharmacogenomics J 2011;11:61-71.
10. Buyse M. Generalized pairwise comparisons of prioritized outcomes in the two-sample problem. Stat Med 2010;29:3245-57.

11. Stinnett AA, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. Med Decis Making 1998;18:S68-80.
12. Grothey A, Sobrero AF, Shields AF, et al. Duration of Adjuvant Chemotherapy for Stage III Colon Cancer. N Engl J Med 2018;378:1177-88.
13. U.S. Department of Health and Human Services. National Cancer Institute. Cancer Therapy Evaluation Program. Common Terminology Criteria for Adverse Events (CTCAE). Version 5.0. Available at https://ctep.cancer.gov/protocoldevelopment/electronic_applications/docs/CTCAE_v5_Quick_Reference_5x7.pdf (Accessed 15/01/2020).
14. Andre T, Vernerey D, Mineur L, et al. Three Versus 6 Months of Oxaliplatin-Based Adjuvant Chemotherapy for Patients With Stage III Colon Cancer: Disease-Free Survival Results From a Randomized, Open-Label, International Duration Evaluation of Adjuvant (IDEA) France, Phase III Trial. J Clin Oncol 2018;36:1469-77.
15. Kozuki T. Skin problems and EGFR-tyrosine kinase inhibitor. Jpn J Clin Oncol 2016;46:291-8.
16. Kato T, Yoshioka H, Okamoto I, et al. Afatinib versus cisplatin plus pemetrexed in Japanese patients with advanced non-small cell lung cancer harboring activating EGFR mutations: Subgroup analysis of LUX-Lung 3. Cancer Sci 2015;106:1202-11.
17. Cai T, Tian L, Wong PH, Wei LJ. Analysis of randomized comparative clinical trial data for personalized treatment selections. Biostatistics 2011;12:270-82.
18. Claggett B, Tian L, Castagno D, Wei LJ. Treatment selections using risk-benefit profiles based on data from comparative randomized clinical trials with multiple endpoints. Biostatistics 2015;16:60-72.
19. Henderson NC, Varadhan R. Bayesian bivariate subgroup analysis for risk–benefit evaluation. Health Serv Outcomes Res Methodol 2018;18:244-64.
20. Kappos L, Radue EW, O'Connor P, et al. A placebo-controlled trial of oral fingolimod in relapsing multiple sclerosis. N Engl J Med 2010;362:387-401.
21. Spanu A, Aschmann HE, Kesselring J, Puhan MA. Benefit-harm balance of fingolimod in patients with MS: A modelling study based on FREEDOMS. Mult Scler Relat Disord 2020;46:102464.
22. McQuellon RP, Muss HB, Hoffman SL, Russell G, Craven B, Yellen SB. Patient preferences for treatment of metastatic breast cancer: a study of women with early-stage breast cancer. J Clin Oncol 1995;13:858-68.
23. Fu AZ, Graves KD, Jensen RE, Marshall JL, Formoso M, Potosky AL. Patient preference and decision-making for initiating metastatic colorectal cancer medical treatment. J Cancer Res Clin Oncol 2016;142:699-706.
24. Bewtra M, Reed SD, Johnson FR, et al. Variation Among Patients With Crohn's Disease in Benefit vs Risk Preferences and Remission Time Equivalents. Clin Gastroenterol Hepatol 2019.
25. Yu T, Fain K, Boyd CM, et al. Benefits and harms of roflumilast in moderate to severe COPD. Thorax 2014;69:616-22.
26. Food and Drug Administration (FDA). Benefit-risk assessment in drug regulatory decision-making. Draft PDUFA VI Implementation Plan (FY 2008-2012). Available at https://www.fda.gov/media/112570/download (Accessed 29/08/2019).
27. Peron J, Roy P, Ozenne B, Roche L, Buyse M. The Net Chance of a Longer Survival as a Patient-Oriented Measure of Treatment Benefit in Randomized Clinical Trials. JAMA oncology 2016;2:901-5.
28. Peron J, Roy P, Ding K, Parulekar WR, Roche L, Buyse M. Assessing the benefit-risk of new treatments using generalised pairwise comparisons: the case of erlotinib in pancreatic cancer. British journal of cancer 2015;112:971-6.
29. Peron J, Roy P, Conroy T, et al. An assessment of the benefit-risk balance of FOLFIRINOX in metastatic pancreatic adenocarcinoma. Oncotarget 2016;7:82953-60.
30. Peron J, Giai J, Maucort-Boulch D, Buyse M. The Benefit-Risk Balance of Nab-Paclitaxel in Metastatic Pancreatic Adenocarcinoma. Pancreas 2019;48:275-80.

31. Boers M, Brooks P, Fries JF, Simon LS, Strand V, Tugwell P. A first step to assess harm and benefit in clinical trials in one scale. J Clin Epidemiol 2010;63:627-32.
32. Chuang-Stein C. A new proposal for benefit-less-risk analysis in clinical trials. Control Clin Trials 1994;15:30-43.
33. Rockhold F, Bromley E, Wagner E, Buyse M. Open Science: The open clinical trials data journey. Clin Trials (in press) 2019.
34. Briggs JP, Palevsky PM. Clinical Trial Data Sharing: The Time Is Now. J Am Soc Nephrol 2019;30:1556-8.
35. Taichman DB, Sahni P, Pinborg A, et al. Data Sharing Statements for Clinical Trials: A Requirement of the International Committee of Medical Journal Editors. Ann Intern Med 2017;167:63-5.

Table 1. Fictitious trial: probabilities of response and toxicity by randomized treatment group, assuming no toxicity in the control arm (N/A, not applicable).

(A). Assuming independence between response and toxicity (odds ratio = 1).

	Experimental treatment			**Control**		
	No Toxicity	**Toxicity**	**Total**	**No Toxicity**	**Toxicity**	**Total**
No response	0.2	0.3	**0.5**	0.8	0	**0.8**
Response	0.2	0.3	**0.5**	0.2	0	**0.2**
Total	**0.4**	**0.6**	**1**	**1**	**0**	**1**
	Odds ratio = 1			N/A		

(B). Assuming a positive association between response and toxicity (odds ratio = ∞).

	Experimental treatment			**Control**		
	No Toxicity	**Toxicity**	**Total**	**No Toxicity**	**Toxicity**	**Total**
No response	0.4	0.1	**0.5**	0.8	0	**0.8**
Response	0	0.5	**0.5**	0.2	0	**0.2**
Total	**0.4**	**0.6**	**1**	**1**	**0**	**1**
	Odds ratio = ∞			N/A		

(C). Assuming a negative association between response and toxicity (odds ratio = 0).

	Experimental treatment			**Control**		
	No Toxicity	**Toxicity**	**Total**	**No Toxicity**	**Toxicity**	**Total**
No response	0	0.5	**0.5**	0.8	0	**0.8**

Response	0.4	0.1	**0.5**	0.2	0	**0.2**
Total	**0.4**	**0.6**	**1**	**1**	**0**	**1**
	Odds ratio = 0			N/A		

Table 2. French IDEA trial: number of "responses" (no disease-free survival [DFS] event) and "toxicities" (grade 3 or 4 peripheral sensory neuropathy [PSN]) by randomized treatment group. Marginal numbers were abstracted from the original publication.[14]

(A). Assuming independence between response and toxicity (odds ratio ≈ 1).

	3 months			6 months		
	No Gr 3-4 PSN	**Gr 3-4 PSN**	**Total**	**No Gr 3-4 PSN**	**Gr 3-4 PSN**	**Total**
DFS event	289	25	**314**	197	67	**264**
No DFS event	634	54	**688**	556	188	**744**
Total	**923**	**79**	**1002**	**753**	**255**	**1008**
	Odds ratio ≈ 1			Odds ratio ≈ 1		

(B). Assuming a positive association between response and toxicity (odds ratio = ∞).

	3 months			6 months		
	No Gr 3-4 PSN	**Gr 3-4 PSN**	**Total**	**No Gr 3-4 PSN**	**Gr 3-4 PSN**	**Total**
DFS event	314	0	**314**	264	0	**264**
No DFS event	609	79	**688**	489	255	**744**
Total	**923**	**79**	**1002**	**753**	**255**	**1008**
	Odds ratio = ∞			Odds ratio = ∞		

(C). Assuming a negative association between response and toxicity (odds ratio = 0).

	3 months			6 months		
	No Gr 3-4 PSN	**Gr 3-4 PSN**	**Total**	**No Gr 3-4 PSN**	**Gr 3-4 PSN**	**Total**
DFS event	235	79	**314**	9	255	**264**
No DFS event	688	0	**688**	744	0	**744**
Total	**923**	**79**	**1002**	**753**	**255**	**1008**
	Odds ratio = 0			Odds ratio = 0		

Table 3. LUX-Lung 3 cohort of Japanese patients[16] (PFS6 = indicator for being alive and progression-free at 6 months, δ = contribution to Net Benefit, Δ = Net Benefit, *P* = P-value, unadjusted for multiplicity)

	Proportion of Wins	**Proportion of Losses**	**Proportion Neutral**	**δ**	**Δ**	***P***
Single outcomes						
PFS6	0.45	0.08	0.47	0.37	0.37	0.002
Skin rash	0	0.28	0.72	-0.28	-0.22	<0.0001
Diarrhea	0	0.22	0.78	-0.22	-0.28	0.005
Prioritized outcomes						
1) PFS6	0.45	0.08	0.47	0.37	0.37	0.002
2) Skin rash	0	0.13	0.34	-0.13	0.24	0.05
3) Diarrhea	0	0.08	0.26	-0.08	0.16	0.21
1) Skin rash	0	0.28	0.72	-0.28	-0.28	<0.0001
2) Diarrhea	0	0.16	0.56	-0.16	-0.44	<0.0001
3) PFS6	0.24	0.06	0.26	0.18	-0.26	0.04

Figure 1. Progression-free survival Kaplan-Meier curves for afatinib (yellow) versus combination chemotherapy (cisplatin + pemetrexed) for Japanese patients randomized in trial LUX-Lung 3.[16] (reproduced with permission of the authors, under the terms of the Creative Commons Attribution-NonCommercial License)

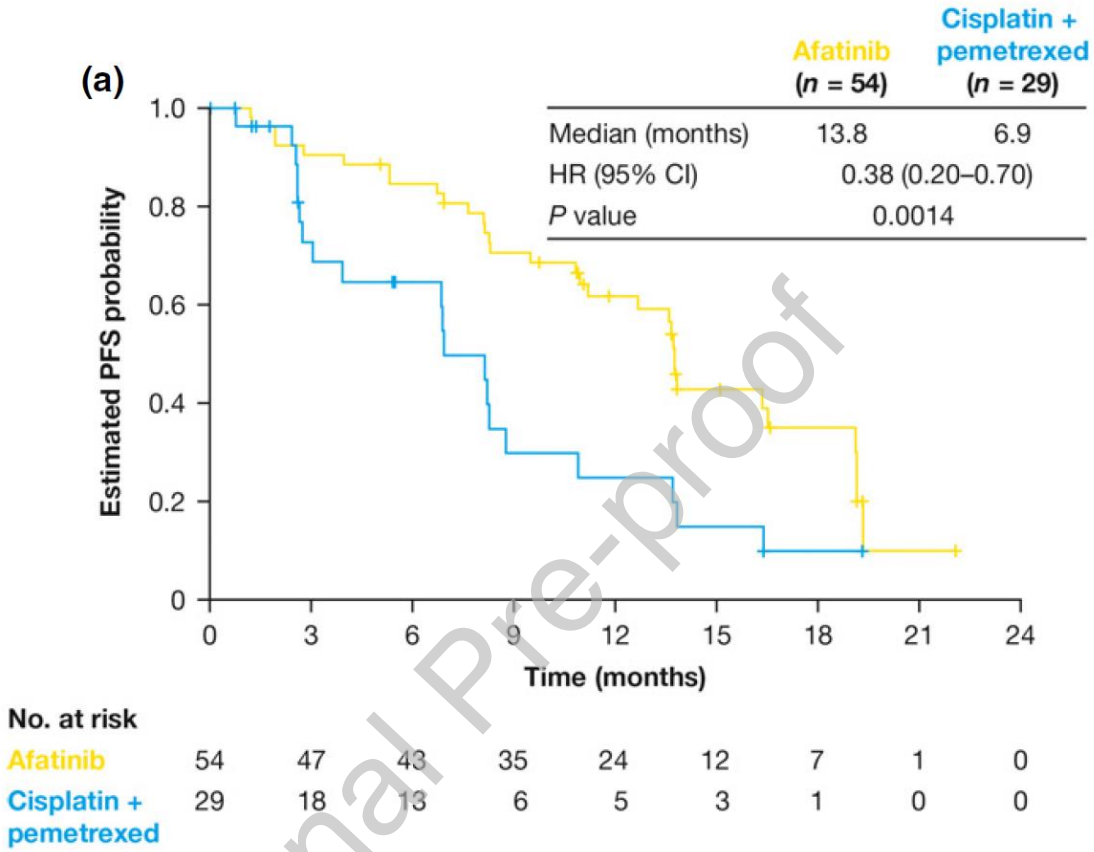


Figure 2. Treatment duration, afatinib dosage, and reason for dosage reduction (skin rash and/or diarrhea for Japanese patients randomized in trial LUX-Lung 3.[16] (reproduced with permission of the authors, under the terms of the Creative Commons Attribution-NonCommercial License)

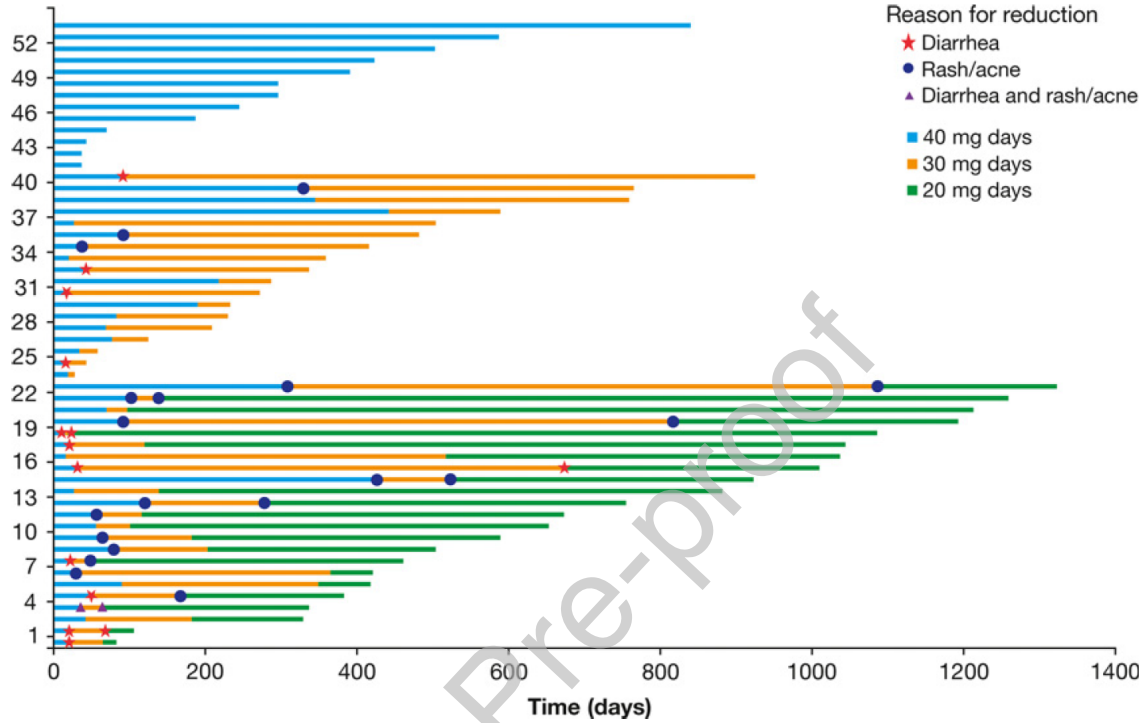


Figure 3. Fictitious trial: Net Benefit as a function of the response *vs.* toxicity odds ratio in the experimental arm, assuming no toxicity in the control arm, and with achievement of response as first priority outcome. Black line, Net Benefit; blue line, contribution of response; red line, contribution of toxicity. Dotted line, second priority outcome.

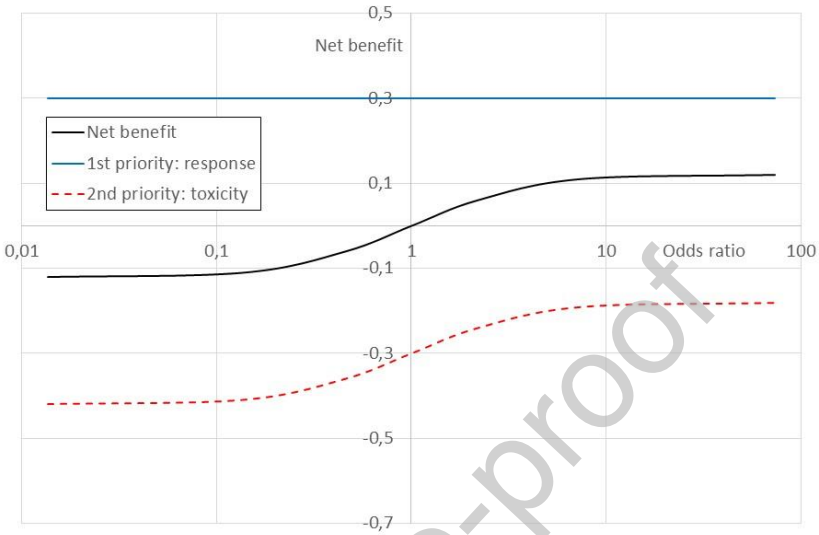


Figure 4. Fictitious trial: Net Benefit as a function of the response *vs.* toxicity odds ratio in the experimental arm, assuming no toxicity in the control arm, and with avoidance of toxicity as first priority outcome. Black line, Net Benefit; blue line, contribution of response; red line, contribution of toxicity. Dotted line, second priority outcome.

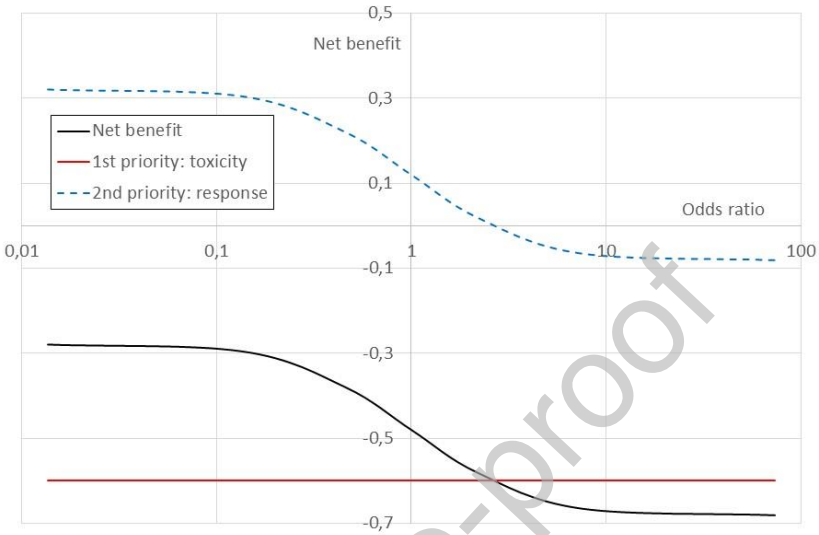


Figure 5. French IDEA trial: Net Benefit as a function of the response *vs.* toxicity odds ratio, assumed identical in both treatment arms, and with achievement of response (no DFS event) as first priority outcome. The Net Benefit is depicted considering 3-month duration as the experimental treatment. Black line, Net Benefit; blue line, contribution of response; red line, contribution of toxicity. Dotted line, second priority outcome.

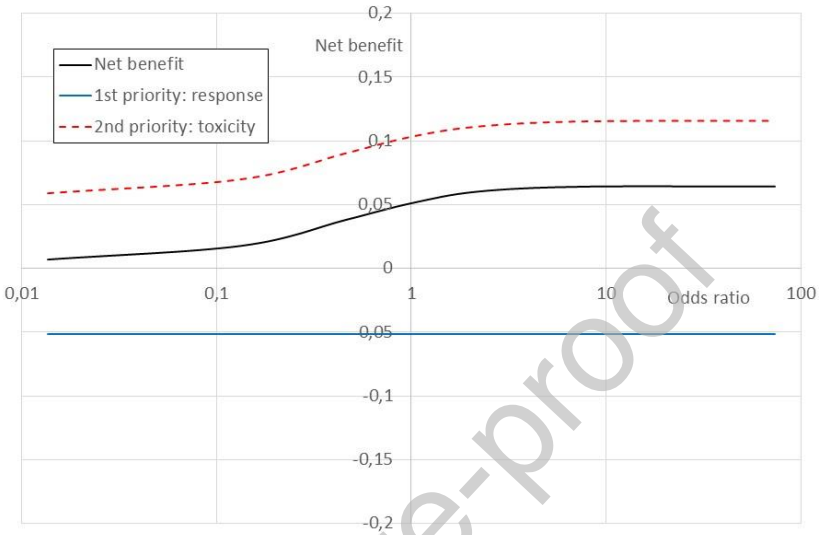
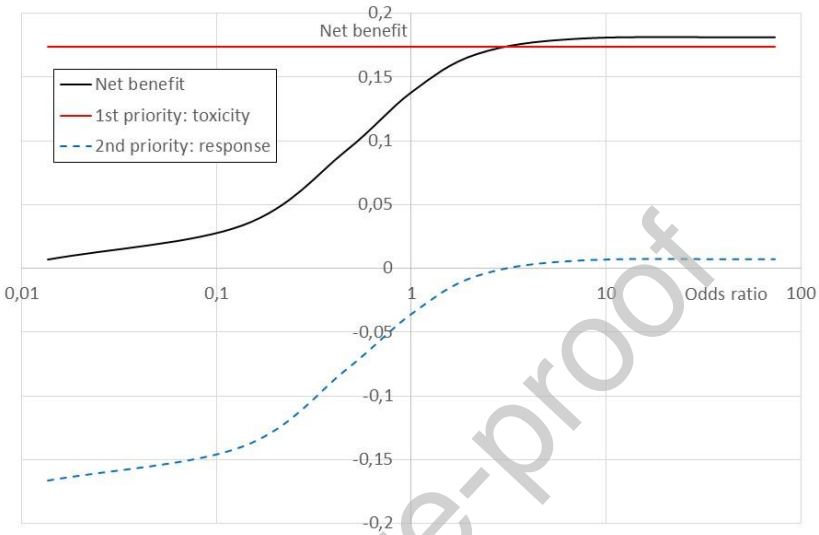


Figure 6. French IDEA trial: Net Benefit as a function of the response *vs.* . toxicity odds ratio, assumed identical in both treatment arms, and with avoidance of toxicity (no grade 3 or 4 PSN) as first priority outcome. The Net Benefit is depicted considering 3-month duration as the experimental treatment. Black line, Net Benefit; blue line, contribution of response; red line, contribution of toxicity. Dotted line, second priority outcome.



SUPPLEMENTARY MATERIAL

Appendix 1

Summary of the methods used in this paper

1. Generalized pairwise comparisons of prioritized outcomes and Net Benefit

Assume interest focuses on two outcomes, one capturing benefit (called "response" in the examples of the paper) and the other harm (called "toxicity" in the examples of the paper). These outcomes will be denoted X and Y in the treatment and the control groups, respectively. Assume further that these outcomes can be prioritized: their priority will be denoted by subscript 1 or 2. Hence the two variables $\{X_1, Y_1\}$ denote the outcome of first priority (*e.g.*, response) in the experimental and control groups, and $\{X_2, Y_2\}$ denote the outcome of second priority (*e.g.*, toxicity) in the experimental and the control groups, respectively. GPC are carried out by forming all possible pairs of patients, taking one patient from each group [1]. A pairwise score is defined as follows for the i^{th} patient in the experimental group and the j^{th} patient in the control group:

$$u_{ij} = \begin{cases} +1 \text{ if } X_{1,i} \succ Y_{1,j} \text{ or } (X_{1,i} = Y_{1,j} \text{ and } X_{2,i} \succ Y_{2,j}) \\ -1 \text{ if } X_{1,i} \prec Y_{1,j} \text{ or } (X_{1,i} = Y_{1,j} \text{ and } X_{2,i} \prec Y_{2,j}) \\ 0 \text{ otherwise} \end{cases}$$

where the symbols " $\succ$ " and " $\prec$ " denote superiority and inferiority, respectively.

The concepts of superiority and inferiority depend on the type of variable considered, but they can easily be defined for binary variables, for numerical variables (*e.g.*, a larger value for a continuous outcome) of for arbitrarily complex criteria including thresholds of clinical relevance (*e.g.*, a value larger by a certain quantity for a continuous outcome).In general,

u_{ij} can capture the overall treatment effect on any number of prioritized outcomes of any type [2]. As such, this approach permits an overall assessment of all identified benefits and harms from the experimental treatment using direct patient comparisons, rather than marginal treatment effects on the various outcomes. Importantly, outcomes may be categorical, numerical, times to events, or any patient-relevant outcomes, even though we focus in this paper on the simplified situation of two binary outcomes. The pairwise score u_{ij} is equal to 1 for pairs that favor the experimental treatment ("wins"), to -1 for pairs that favor control ("losses"), and to 0 for pairs that favor neither the experimental treatment nor control ("ties") [3].

The Net Benefit is estimated as the sum of the pairwise scores over all the pairs that can be formed between one patient from the experimental group and one patient from the control group:

$$\hat{\Delta} = \sum_{i=1}^{n_E} \sum_{j=1}^{n_C} u_{ij} \Big/ n_E . n_C$$

where $\hat{\Delta}$ denotes the estimated Net Benefit, n_E denotes the number of patients in the experimental group and n_C the number of patients in the control group. The Net Benefit Δ represents the difference between the probability that a pair of patients taken randomly, one from each treatment group, favors the experimental treatment, minus the probability that the pair favors the control. The next section shows that Δ depends on the correlation between the two binary outcomes.

References:

1. Buyse M. Generalized Pairwise comparisons. ***In****: Wiley StatsRef: Statistics Reference Online, https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118445112.stat08224, 2019.*

2. Buyse M. Prioritized multiple outcomes. ***In****: Wiley StatsRef: Statistics Reference Online, https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118445112.stat08158, 2019.*

3. Pocock SJ, Ariti CA, Collier TJ, Wang D. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. European Heart Journal 2012;33:176-82.

2. Dependency of the Net Benefit on the association between prioritized outcomes

Consider again two binary outcomes, with $\{X_1, Y_1\}$ the variables denoting the outcome of first priority respectively in the experimental group and control group, and $\{X_2, Y_2\}$ the variables denoting the outcome of second priority respectively in the experimental group and the control group. We have assumed without loss of generality that X_1 and Y_1 capture "response" (with X_1 =1 and Y_1= 1 denoting achievement of a response, and X_1 = 0 and Y_1 = 0 denoting lack of response), while X_2 and Y_2 capture "toxicity" (with X_2 = 0 and Y_2 = 0 denoting absence of toxicity, and X_2 = 1 and Y_2 = 1 denoting presence of toxicity). Table S1 schematically depicts a pairwise comparison.

Table S1. Pairwise comparison if response takes first priority, and toxicity second priority.

Response	**Toxicity**	**Pairwise score**
Patient on experimental arm has response, patient on control arm does not	Not considered	+1 (win)
Patient on control arm has response, patient on experimental arm does not	Not considered	-1 (loss)
Patients on experimental and control arms both have response, or	Patient on control arm has toxicity, patient on experimental arm does	+1 (win)

both do not	not	
	Patient on experimental arm has toxicity, patient on control arm does not	-1 (loss)
	Patients on experimental and control arms both have toxicity, or both do not	0 (tie)

Let $p_{1T} = P\{X_1 = 1\}$ and $p_{1C} = P\{Y_1 = 1\}$, where $P\{\cdot\}$ represents the probability. The Net Benefit for these two prioritized outcomes is:

$$\begin{aligned}\Delta = {} & p_{1T} - p_{1C} + P\{X_1 = 1, Y_1 = 1\} \\ & \times [P\{X_2 = 1, Y_2 = 0 | X_1 = 1, Y_1 = 1\} - P\{X_2 = 0, Y_2 = 1 | X_1 = 1, Y_1 = 1\}] \\ & + P\{X_1 = 0, Y_1 = 0\} \\ & \times [P\{X_2 = 1, Y_2 = 0 | X_1 = 0, Y_1 = 0\} - P\{X_2 = 0, Y_2 = 1 | X_1 = 0, Y_1 = 0\}]\end{aligned}$$

The last two terms of this expression for Δ depend on the association between the two outcomes. If the two outcomes are perfectly correlated (whether positively or negatively), the conditional probabilities are all equal to zero and the Net Benefit reduces to $\Delta = p_{1T} - p_{1C}$, in other words the second priority outcome does not play any role. If, in contrast, the two outcomes are independent, then the conditional probabilities reduce to unconditional probabilities, and the expression for Δ simplifies to

$$\Delta = p_{1T} - p_{1C} + (p_{2T} - p_{2C}) \times (1 + 2p_{1T}p_{1C} - p_{1T} - p_{1C})$$

where $p_{2T} = P\{X_2 = 0\}$ and $p_{2C} = P\{Y_2 = 0\}$. The notation for p_{2T} and p_{2C} is chosen for consistency with the situations discussed in the paper, for which an outcome of 0 is favorable for the second priority outcome. This expression shows that, if treatment has a positive effect on the first outcome ($p_{1T} - p_{1C} > 0$), Δ is larger or smaller than $p_{1T} - p_{1C}$ depending on whether the treatment effect on the second outcome, $p_{2T} - p_{2C}$, is positive or negative, respectively.

Appendix 2

Calculation of cell probabilities in 2x2 tables, given the margins and odds ratio

Suppose the margins of the 2x2 Table S2 cross-classifying response and toxicity are known, as well as the odds ratio characterizing the association between response and toxicity.

Table S2. Cross-classification of response and toxicity

	Toxicity	**No toxicity**	**Total**
No response	$P_{NR,T}$	$P_{NR,NT}$	P_{NR}
Response	$P_{R,T}$	$P_{R,NT}$	P_R
Total	P_T	P_{NT}	1
	Odds ratio = OR		

The cell probabilities in this table can then be calculated. $P_{R,NT}$, for instance, is given by the following formulas (Molenberghs and Lesaffre 1994):

$$P_{R,NT} = \begin{cases} \dfrac{1 + (P_R + P_{NT})(OR - 1) - F(P_R, P_{NT}, OR)}{2(OR - 1)} & if\ OR \neq 1 \\ P_R P_{NT} & if\ OR = 1 \end{cases}$$

where the function F is defined by

$$F(P_R, P_{NT}, OR) = \sqrt{[1 + (P_R + P_{NT})(OR - 1)]^2 + 4OR(1 - OR)P_R P_{NT}}$$

for $0 \leq P_R, P_{NT} \leq 1$ and $0 \leq OR \leq \infty$.

Reference:

Molenberghs G, Lesaffre E. Marginal modelling of correlated ordinal data using a multivariate Plackett distribution. Journal of the American Statistical Association, 89: 633-44, 1994.

Appendix 3

Example of the calculation of the Net Benefit

Supplementary Table S3. Hypothetical trial: Calculation of the Net Benefit for the case of independence between response and toxicity (odds ratio = 1 in experimental arm), when response is first priority outcome. Rows 1. to 16. represent all possible combinations of outcomes for each treatment arm, with probabilities in each arm calculated from the data of Table 2(A) of the article. For both response and toxicity, 0 and 1 indicate absence and presence of the outcome, respectively.

	Experimental arm			Control arm			
Win/Loss	**Response**	**Toxicity**	**Probability**	**Response**	**Toxicity**	**Probability**	**Joint Probability**
1. Win	1	0	0,2	1	1	0	0
2. Win	1	0	0,2	0	0	0,8	0,16
3. Win	1	0	0,2	0	1	0	0
4. Win	1	1	0,3	0	0	0,8	0,24
5. Win	1	1	0,3	0	1	0	0
6. Win	0	0	0,2	0	1	0	0
7. Tie	1	0	0,2	1	0	0,2	0,04
8. Tie	1	1	0,3	1	1	0	0
9. Tie	0	0	0,2	0	0	0,8	0,16
10. Tie	0	1	0,3	0	1	0	0
11. Loss	1	1	0,3	1	0	0,2	0,06
12. Loss	0	0	0,2	1	0	0,2	0,04
13. Loss	0	0	0,2	1	1	0	0
14. Loss	0	1	0,3	1	0	0,2	0,06
15. Loss	0	1	0,3	1	1	0	0
16. Loss	0	1	0,3	0	0	0,8	0,24
Probability of a win (sum of joint probabilities rows 1. to 6.)							0.4

Probability of a loss (sum of joint probabilities rows 11. to 16.)	0.4
Net Benefit (= probability of a win - probability of a loss)	0.0