A gentle statistical introduction to Generalized Pairwise Comparisons (GPC) and the Net Treatment Benefit (NTB)

Marc Buyse, ScD, Everardo D. Saad, MD, Mickaël De Backer, PhD,

Samuel Salvaggio, PhD, Sarah Kosta, PhD, Emilie Barré, MSc,

Vaiva Deltuvaite-Thomas, PhD, Johan Verbeeck, PhD

May 2023

Pairwise comparisons

Assume we wish to compare a group of patients receiving some new intervention (called "treatment") with a group of patients receiving standard of care (called "control"). However, instead of comparing group-level data such as means, proportions or hazard rates, we use generalized pairwise comparisons (GPC), which consists in comparing each patient under treatment with each and every patient under control by forming all possible pairs of patients, taking one from each group (Buyse 2010). We start by considering the simple case of one outcome (or endpoint) of interest. Denote this outcome by X for patients in the treatment group and by Y for patients in the control group. A pairwise score (u_{ij}) is defined as follows for the i^{th} patient in the treatment group and the j^{th} patient in the control group:

$$u_{ij} = \begin{cases} +1 \text{ if } X_i \succ Y_j \\ -1 \text{ if } X_i \prec Y_j \\ \quad 0 \text{ otherwise} \end{cases}$$

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

The concepts of superiority and inferiority depend on the type of variable considered. They can easily be defined for binary variables (for instance, treatment success or

response versus failure or non-response), ordered categorical variables (for instance, increasing grades of toxicity), continuous variables (for instance, level of circulating tumor DNA, with a smaller value being preferable), or time-to-event variables (for instance, overall survival time, with larger values being preferable). Thus, the pairwise score u_{ij} is computed for each and every pair, and it is equal to 1 if the pair favors treatment ("win"), to -1 if the pair favors control ("loss"), and to 0 if the pair favors neither treatment nor control ("tie") (Pocock *et al.* 2012).

Multiple prioritized outcomes

In many cases, we wish to compare several outcomes between two groups. As a general rule, each outcome is subject to one statistical analysis. The GPC method allows the comparison of multiple outcomes in the same formal analysis, as long as an order of priority can be chosen for the different outcomes. For simplicity, assume that there are two outcomes of interest, e prioritized because an improvement in one of these outcomes is preferred to an improvement in the other. The outcomes are again denoted X and Y in the treatment and the control groups, respectively, with priorities denoted by subscripts 1 or 2 (1 being of higher priority than 2). The pairwise score is defined as follows for the i^{th} patient in the experimental group and the j^{th} patient in the control group:

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

where the symbols " $\succ$ " and " $\prec$ " again denote superiority and inferiority, respectively, and $\asymp$ denotes neutrality on the scale of the variables. One advantage of GPC is that the method can include any number of prioritized outcomes of any type. The method was first proposed by Finkelstein and Schoenfeld in an AIDS clinical trial, with death as the outcome of first priority, and CD4 cell counts as the second priority outcome for patients

still alive at the time of analysis (Finkelstein and Schoenfeld 1999). Another advantage of using prioritized outcomes is that the correlation between the outcomes is properly taken into account (Buyse *et al.* 2021). Last but not least, for times to events, the time to most important outcome can be considered instead of the time to first outcome. This makes GPC different from conventional analyses with composite endpoints, which consider the outcome occurring first in a given patient, regardless of its relative importance to other components of the composite endpoint. As such, the approach used in GPC and similar methods is more clinically relevant and has received much attention lately, for example, in cardiovascular trials (Pocock *et al.* 2012).

Thresholds of clinical relevance

For outcomes that are ordered and for continuous outcomes, a threshold of clinical relevance can be specified so that a difference smaller than the threshold is ignored in pairwise comparisons (Buyse 2010). For example, a difference of less than 3 months in survival might be considered too small to be of clinical relevance. In this case, the definition of the pairwise score is denoted as follows for the i^{th} patient in the treatment group and the j^{th} patient in the control group:

$$u_{ij} = \begin{cases} +1 \text{ if } X_i - Y_j > \tau \\ -1 \text{ if } X_i - Y_j < -\tau \\ 0 \text{ otherwise} \end{cases}$$

where " τ " denotes the threshold of clinical relevance.

Measures of treatment effect

When the GPC method is used, the measure of treatment effect is the Net Treatment Benefit (NTB). The estimated NTB is the mean pairwise score, i.e., the sum of all pairwise

scores divided by the number of pairs formed between each patient from the treatment group and each patient from the control group:

$$NTB = \sum_{i=1}^{n_T} \sum_{j=1}^{n_C} u_{ij} \Big/ n_T . n_C$$

where NTB denotes the estimated Net Treatment Benefit, n_T denotes the number of patients in the treatment group and n_C the number of patients in the control group. For a single outcome, the NTB represents the net probability that a random patient has a better outcome in one treatment group than in the other. For a single outcome with no threshold of clinical relevance, the NTB represents the net probability that a random patient has a better outcome in one treatment group than in the other (Buyse *et al.* 2010). For a single outcome with a threshold of clinical relevance, the NTB represents the net probability that a random patient has an outcome better by at least the threshold of clinical relevance in one treatment group than in the other (Péron *et al.* 2016a). For multiple prioritized outcomes, the NTB represents the net probability that a random patient in the treatment group has a better outcome in one treatment group than in the other, either for the outcome of highest priority, or, conditional on no benefit for the outcome of highest priority, for the outcome of next highest priority, and so on. One key advantage of the NTB is that for prioritized outcomes, the contributions of the individual outcomes to the NTB are additive. While the NTB is easy to interpret (Péron et al. 2016a) and has desirable statistical properties (Verbeeck et al. 2021), other measures of treatment effect have been proposed when using GPC: the win ratio (Pocock et al. 2012) and the win odds (Brunner et al. 2021), with the latter being preferred in situations with large numbers of patients being censored for time-to-event endpoints.

Benefit/risk analyses

The GPC method with prioritized outcomes allows arbitrarily complex benefit/risk analyses to be conducted in a mathematically rigorous way. In contrast to group-level results for each outcome analyzed separately, it allows the analyst to estimate the NTB, which accounts for the correlation between the various outcomes (Buyse *et al.* 2021). Some of the outcomes considered may favor the experimental treatment while others may favor the control group. For example, a treatment that works against a certain tumor type may also produce more toxicity. This approach has been used to investigate the benefit/risk of three experimental treatments compared with gemcitabine for advanced pancreatic cancer: gemcitabine + erlotinib (Péron *et al.* 2015), FOLFORINOX (Péron *et al.* 2016b) and gemcitabine + nab-paclitaxel (Péron *et al.* 2019).

REFERENCES

1. Brunner, E, Vandemeulebroecke, M, Mütze, T. Win odds: An adaptation of the win ratio to include ties. Stat Med 40: 3367– 3384, 2021.
2. Buyse M. Generalized pairwise comparisons of prioritized outcomes in the two-sample problem. Stat Med 2010; 29: 3245-57.
3. Buyse M, Saad ED, Peron J, Chiem JC, De Backer M, Cantagallo E, Ciani O. The Net Benefit of a treatment should take the correlation between benefits and harms into account. J Clin Epidemiol 137:148-158, 2021.
4. Finkelstein, D., and D. Schoenfeld. 1999. Combining mortality and longitudinal measures in clinical trials. Stat Med 18:1341–1354.
5. Péron J, Roy P, Ding K, Parulekar W, Roche L, Buyse M. Benefit-risk assessment of adding erlotinib to gemcitabine for the treatment of advanced pancreatic cancer. Brit J Cancer 112:971-976, 2015.

6. Péron J, Roy P, Ozenne B, Roche L, Buyse M. The net chance of a longer survival as a patient-oriented measure of benefit in randomized clinical trials. JAMA Oncology 2:901-5, 2016a.
7. Péron J, Roy P, Conroy T, Desseigne F, Ychou M, Gourgou-Bourgade S, Stanbury T, Roche L, Ozenne B, Buyse M. An assessment of the benefit-risk balance of FOLFORINOX in metastatic pancreatic adenocarcinoma. Oncotarget 7:82953-60, 2016b.
8. Péron J, Giai J, Maucourt-Boulche D, Buyse M. The benefit-risk balance of nab-paclitaxel in metastatic pancreatic adenocarcinoma. Pancreas 48:275-80, 2019.
9. 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. Eur Heart J 2012; 33: 176-82.
10. Verbeeck J, Deltuvaite-Thomas V, Berckmoes B, , Burzykowski T, Aerts M, Thas O, Buyse M, Molenberghs G. Unbiasedness and efficiency of non-parametric and UMVUE estimators of the probabilistic index and related statistics. Stat Meth Med Res 30:747-68, 2021.