Revista Española
de Educación Médica

Getting the best out of single case data.

Aprovechar al máximo los datos de casos únicos.

Jimmie Leppink [1,*]

[1,*] Leppink Analytics, Ondara; leppinkanalytics@gmail.com, https://orcid.org/0000-0002-8713-1374

Recibido: 28/5/25; Aceptado: 2/6/25; Publicado: 2/6/25

Given its potential for personalized healthcare and education, *single case designs* (SCDs) (1-4) – in which the same cases are measured repeatedly on the same outcome(s) of interest – has enjoyed increasing interest over time. In a publication in Volume 3 of this journal (5), three key statistical methods for SCDs are compared and contrasted: *randomization tests* (4), the Bayesian percentage of all non-overlapping data or *PAND-B* (5), and a *time series regression* (TSR) method (5). Although PAND-B is the only of these methods that allows for the incorporation of knowledge from theory or previous research and does not require frequently violated assumptions that invalidate TSR (5), recent work (6) demonstrates that PAND-B does not fully account for order information in ordinal or quantitative outcomes (e.g., '3' is not only different from '1' or '2'; '3' is also higher than '2', and '2' is higher than '1') and returns invalid estimates when the two conditions compared have unequal sample sizes. To address these shortcomings while retaining the benefits of being useful for all levels of measurement including nominal (e.g., different shoe colors or qualitatively different emotions that cannot undisputedly be categorized on a dimension like 'better' or 'worse') and incorporating knowledge from theory or previous research, a *generalized relative effect* (GRE) statistic was developed (6), which is based on Brunner-Munzel's relative effect (RE) statistic (7). Although the GRE introductory paper provides a step-by-step tutorial for how to obtain GRE point and interval estimates with basic R (8) code and interpret the outcomes, that paper focuses on two larger samples of participants. Therefore, the remainder of this paper uses the data from the aforementioned publication in Volume 3 of this journal (5) to provide an example of GRE in a single context.

In this example (4- 5), there are 14 dizziness ratings (higher is dizzier) from the same person, 7 from a control (placebo) condition [5, 4, 5, 5, 5, 4, 6] and 7 from a treatment condition [6, 7, 6, 7, 4, 6, 7]. If we can assume the dizziness ratings as an outcome of at least ordinal level of measurement, GRE follows the computational procedure of RE, which is visualized in Table 1, and adds a prior distribution to account for knowledge from theory or previous research. In Table 1, the rating difference of treatment minus control is calculated for each pairwise comparison. As displayed in the lower part of the table, a negative difference (in red), which indicates higher ratings in the control condition, occurs only 5 times, there are 5 comparisons which result in a '0' difference or tie, and the other 39 comparisons result in higher ratings for the treatment. Next, the ties are equally divided over 'control' (hence 5 + 2.5 = 7.5) and 'treatment' (hence 39 + 2.5 = 41.5). These numbers are then rescaled back from the n_C x n_E (i.e., the sample sizes of both conditions) comparisons to the original sample size of $n_C + n_E = N$. This yields about 11.857 in favor of treatment and about 2.143 in favor of control. These numbers are then combined with a Binomial prior distribution – in the absence of prior knowledge, 1 in favor of either condition – to obtain a Binomial posterior distribution: B(11.857, 2.143) + B(1, 1) = B(12.857, 3.143). This is a distribution with a posterior median (point estimate) of about 0.816 and a 95% credible interval of [0.584; 0.952]. Although these estimates are similar to those of PAND-B (5), they are more accurate because the order information in the data is accounted for more precisely (6). The 95% credible interval excludes 0.5, which is the value under the null hypotheses of no difference between conditions, and since higher ratings indicate more dizziness, the treatment seems to increase dizziness.

Table 1. Example of GRE computation.

		control						
		5	4	5	5	5	4	6
treat	6	1	2	1	1	1	2	0
	7	2	3	2	2	2	3	1
	6	1	2	1	1	1	2	0
	7	2	3	2	2	2	3	1
	4	-1	0	-1	-1	-1	0	-2
	6	1	2	1	1	1	2	0
	7	2	3	2	2	2	3	1
	difference	-2	-1	0	1	2	3	
	frequency	1	4	5	15	18	6	
	if at least ordinal		control 7.5		treat 41.5			
	if nominal		same 5		different 44			

However, if we were in a situation that the outcome variable was only of nominal level of measurement (e.g., '4' and '5' as different ratings but the latter not necessarily indicating more dizziness), we can slightly adapt the RE computational procedure to obtain valid GRE point and interval estimates: all pairwise comparisons resulting in different ratings would simply be coded as 'different' (here: 44) while all comparisons resulting in no difference would be coded as 'same' (here: 5). Subsequently, we apply the same rescaling procedure as for outcomes that are of at least ordinal level of measurement, and this time obtain about 12.571 in favor of treatment and about 1.429 in favor of control. These numbers are then combined with a Binomial prior distribution – in the absence of prior knowledge, 1 in favor of either condition – to obtain a Binomial posterior distribution: B(12.571, 1.429) + B(1, 1) = B(13.571, 2.429). This is a distribution with a posterior median (point estimate) of about 0.863 and a 95% credible interval of [0.643; 0.973]. The 95% credible interval excludes 0.5, which is the value under the null hypotheses of a 50%-50% same-different distribution.

Although GRE suffers from the same statistical power issue as other methods when samples are small (e.g., assuming at least ordinal level of measurement, GRE with a default B(1, 1) prior, GRE and RE are very similar in terms of statistical power), it does provide point and interval effect size estimates for all levels of measurement, like PAND-B it can account for prior knowledge yet contrary to PAND-B accounts more precisely for order information and is also valid for unequal samples. Researchers should therefore consider using GRE instead of PAND-B when analyzing single case data, and the aforementioned tutorial (6) provides a step-by-step guide for how to compute GRE in basic R code.

Finally, when using GRE, statistical power challenges – which are common for single case designs – can be reduced in one or several of the following ways, depending on the possibilities in a given context at hand: increasing the number of measurements of the time series (1), using relevant theory or previous research to justify informative priors (e.g., B(5, 1) or B(1, 5) instead of B(1, 1)) or one-sided instead of two-sided testing when only differences in a specific direction are to be expected (6), and/or combining findings from

single studies into meta-analyses (3) to acquire more precise estimates of differences of interest.

Funding: There was no funding.

Declaration of conflict of interest: The author declares that he has no conflict of interest.

Author contributions: There is only one author, who wrote the first draft and revised the final version.

References

1. McDonald S, Vieira R, Johnston DW. Analysing N-of-1 observational data in health psychology and behavioural medicine: a 10-step tutorial for beginners. *Health Psych Behav Med.* **2020**, 8(1), 32-54. https://doi.org/10.1080/21642850.2019.1711096
2. Parker RI, Brossart DF. Evaluating single-case research data: a comparison of seven statistical methods. *Behav Therap.* **2003**, 34, 189-211. https://doi.org/10.1016/S0005-7894(03)80013-8
3. Tanious R, Onghena P. A systematic review of applied single-case research published between 2016 and 2018: study designs, randomization, data aspects, and data analysis. *Behav Res Meth.* **2021**, 53, 1371-84. https://doi.org/10.3758/s13428-020-01502-4
4. Onghena P. One by one: the design and analysis of replicated randomized single-case experiments. In: Van de Schoot R & Miocevic M (Eds.), *Small sample size solutions: a guide for applied researchers and practitioners* (chapter 6). New York: Routledge; **2020**. https://www.taylorfrancis.com/chapters/oa-edit/10.4324/9780429273872-8/one-one-patrick-onghena
5. Leppink J. Diseños experimentales de caso único y sus estadísticas [Single case experimental designs and their statistics]. *Rev Esp Educ Med*, **2022**, *3*(3), 1-3. https://doi.org/10.6018/edumed.528911
6. Leppink J. El efecto relative generalizado (ERG): su introducción y un tutorial en R [The generalized relative effect (GRE): its introduction and a tutorial in R. *CERES HEMJ*. **2025**, 3(2), e79. https://doi.org/10.62234/ceresv3n2-002
7. Karch JD. Psychologists should use Brunner-Munzel's instead of Mann-Whitney's U test as the default nonparametric procedure. *Adv Methods Pract Psychol Sci*. **2021**, 4(2). https://doi.org/10.1177/2515245921999602
8. R Core Team. R: A language and environment for statistical computing (version 4.4.1) [Computer software]. Retrieved (May 24, 2025) from: https://www.r-project.org