Analysis of responder-based endpoints: improving power through utilising continuous components

James Wason (✉ james.wason@newcastle.ac.uk)  
Newcastle University  https://orcid.org/0000-0002-4691-126X

Martina McMenamin  
MRC Biostatistics Unit

Susanna Dodd  
University of Liverpool

Research

**Keywords:** Augmented binary method; composite endpoint, efficiency, responder analysis, statistical analysis

**Posted Date:** March 19th, 2020

**DOI:** https://doi.org/10.21203/rs.2.11783/v3

**License:** ☝️ ☑️ This work is licensed under a Creative Commons Attribution 4.0 International License.  
Read Full License

**Version of Record:** A version of this preprint was published at Trials on May 25th, 2020. See the published version at https://doi.org/10.1186/s13063-020-04353-8.
Abstract

Background: Clinical trials and other studies commonly assess effectiveness of an intervention through use of responder-based endpoints. These classify patients based on whether they meet a number of criteria which often involve continuous variables categorised as being above or below a threshold. The proportion of patients who are responders is estimated and, where relevant, compared between groups. An alternative method called the augmented binary method keeps the definition of the endpoint the same but utilises information contained within the continuous component to increase the power considerably (equivalent to increasing the sample size by >30%). In this article we summarise the method and investigate the variety of clinical conditions that use endpoints to which it could be applied.

Methods: We reviewed a database of Core Outcome Sets (COS) that covered physiological and mortality trial endpoints recommended for collection in clinical trials of different disorders. We identified responder-based endpoints where the augmented binary method would be useful for increasing power.

Results: Out of the 287 COS reviewed, we identified 67 new clinical areas where endpoints were used that would be more efficiently analysed using the augmented binary method. Clinical areas that had particularly high numbers were rheumatology (11 clinical disorders identified), non-solid tumour oncology (10 identified), neurology (9 identified), and cardiovascular (8 identified)

Conclusions: The augmented binary method can potentially provide large benefits in a vast array of clinical areas. Further methodological development is needed to account for some types of endpoint.

Background

In clinical trials gathering evidence about the effectiveness of a medical intervention, it is necessary to specify a primary endpoint. An endpoint should represent how patients respond after being given the treatment; it should be expected that the distribution of the endpoint will be more favourable if a treatment is effective than if it is ineffective. In many disorders it is difficult to specify just one specific endpoint, as an intervention may have a variety of effects that cannot be adequately measured through one measurement. For this reason, it is common in many conditions to combine multiple distinct endpoints (which we will refer to as components) into a composite endpoint.

Composite endpoints have been recommended when there is large variability in the disease manifestation, e.g. complex multisystem diseases, allowing multiple equally relevant outcomes to be considered without the need to correct for multiplicity. They have also been advocated for rare diseases, where they might improve the power by increasing the number of events observed. On the other hand, they have been criticised for making trial results more difficult to interpret (1).

One specific type of composite endpoint is a composite responder endpoint, which divides patients into responders and non-responders on the basis of the set of components. Some of these components may be binary (present or absent), some may be continuous. In the case of continuous components, some
Dichotomisation is necessary, so that patients are responders only if the continuous component is above or below a specified threshold. In Table 1, we provide examples of some commonly used responder-based endpoints and their definitions. In some cases (such as tumour response in Table 1), a patient must meet all the criteria to be a responder; in other cases (such as Rheumatoid Arthritis in Table 1) a patient must meet a set number. Some responder endpoints are not composite and are just formed from a single dichotomised continuous endpoint.

Responder endpoints are appealing as they simplify several (potentially complex) pieces of information into one responder/non-responder variable. The proportion of patients who are responders serves as an easy to interpret measurement of the effectiveness of a treatment.

From a statistical point of view, however, this appealing simplicity comes at a non-appealing cost when one or more component is continuous. Dichotomising continuous variables loses information, a point which has been made several times (e.g. [2–4]). This means that if considering one continuous endpoint, it is substantially more efficient to analyse it as a continuous variable rather than dichotomise it and test as a binary variable. As a rule of thumb, the minimum cost of dichotomisation is requiring a 35% higher sample size for the same level of statistical precision(2).

Assuming that avoiding dichotomisation is desirable, it is not obvious how this is possible when the responder endpoint consists of a mix of continuous and binary components. One approach would be to use the approach of Lachenbruch(5) or Hu and Proschan(6) which use separate test statistics for each component, and form an overall test through appropriate weighting. However, this loses the clinical interpretability of the endpoint, and does not allow efficient estimation of the probability of response. Even in the case of a single continuous component, there may be compelling clinical reasons to keep a responder endpoint dichotomised (7): ease of interpretation to researchers and patients, wide acceptance as important, meaningful clinical diagnosis (e.g. diabetes or hypertension).

This motivates statistical methods that can be used to keep to what is clinically relevant by inferring the proportion of patients who are responders, but utilise information contained in continuous components to improve the efficiency. For the single-component responder, this idea dates back to the 90s, where Suissa and Blais (8,9) proposed methods for doing this for a single continuous component case. To our knowledge, this method rarely is applied in practice despite its advantages over analysing the endpoint as binary. More recently, an approach known as the augmented binary method has been developed that allows composite responder endpoints (that consist of at least one continuous component) to be analysed in a more efficient way, whilst maintaining the definition of the endpoint.

In this paper (and associated supplementary material) we first describe the augmented binary method, focusing on its advantages and drawbacks. The main novel contribution of the paper is a review that identifies new clinical areas where trial efficiency can be improved through use of the augmented binary method. Finally, we discuss some further developments to the method that are motivated by the review.

**The augmented binary method - intuition, benefits and drawbacks**
The augmented binary method extends previous work focused on a single dichotomised continuous endpoint (8,9) to composite responder endpoints with a mixture of continuous and binary endpoints. The original motivation was solid-tumour oncology (10,11), but subsequent papers have focused on developing the methodology for rheumatology (12) and rare diseases using composite endpoints (13).

For simplicity we focus on the case of a composite responder endpoint that combines a dichotomised continuous component with a binary component. For example, response in solid-tumour oncology consists of the sum of target lesion diameters shrinking by at least 30% from a baseline scan (dichotomised continuous) and no new tumour lesions appearing on a scan (binary). The traditional, binary analysis would work with the data on whether or not each patient is a responder or not. If a patient meets the criteria they are a responder, otherwise not. If analysing a randomised controlled trial (RCT), then one might test for a difference between arms in the proportion of patients who are responders with an established method that gives an effect size, confidence interval and p-value (e.g., logistic regression).

A detailed description of how to fit the method is provided in the supplementary material, including R code that can be used for the case of a composite responder endpoint formed from a single continuous and single binary component. The main intuition behind the method is to first fit a more sophisticated model to the data from the different components, and second to use this model to estimate a probability of response and test for a difference between arms. The second step can be thought of weighting the different patients as a proportion of a response with this proportion depending on how close the continuous component was to the threshold. This is demonstrated in Figure 1, where patients are measured on a continuous and binary component. The continuous measurement must be above 1 for the patient to be a responder, however patients must also meet additional binary criteria. The binary method treats the information as 0s and 1s whereas the augmented binary method uses a ‘response weight’ which is determined from the underlying model and is higher as the continuous component increases. The supplementary material contains a link to an R package that can be used to fit the model.

The benefit of the method is primarily the increased power. By better using the available information, the proportion of patients who respond (and therefore any differences between arms in a RCT) can be estimated more precisely. In more statistical language, the variance of the estimate is lower and the width of the confidence interval (CI) is narrower. Simulation studies presented by Wason and Seaman (10) found that the average gain in power was equivalent to increasing the sample size by at least 30%. The gain can be considerably higher depending on the scenario, predominantly depending on how well the dichotomisation point divides patients. This gain in power has been confirmed a re-analysis of the OSKIRA-1 trial (14) in rheumatoid arthritis, which showed the reduction in CI width was equivalent to an increase in sample size of >50% (12). It should be emphasised that this gain in power does not rely on additional data being collected – it just comes from using the existing data more efficiently.

There are some additional benefits of the approach. First, due to the underlying model being fitted, it better allows for missing data on different components (10) (it is generally not obvious how to handle missing data on a specific component of a composite outcome). This is especially true when there is the
possibility of some components having more missing data than others. Second, it may also help address issues of misclassification due to measurement error: if a patient is truly close to the responder threshold then a measurement error will have a potentially very large impact on the binary method, but a small impact on the augmented binary method.

There are also drawbacks. First, it is undoubtedly more complex to apply the method compared with standard binary approaches. Some code is available (see supplementary material) for applying the method in specific cases but a more generic implementation in different commonly used statistical programs is a high priority for the future. Second, the method makes more assumptions, for instance that the distribution of the continuous components is normal. This means that it is necessary to check this prior to analysing the data and use a suitable correction if assumptions are not met, such as applying a Box-Cox transformation (15) to ensure the continuous component is normally distributed. Third, if the number of components or number of timepoints at which the endpoint is measured is large, applying the method can require a large amount of computational time. This is generally not an issue for an analysis of a single trial; however assessing the performance of the method on a large number of computer simulations can become infeasible.

Up to now, the method has been applied to datasets in solid tumour oncology(10,11), rheumatoid arthritis(12) and systemic lupus erythematosus (SLE)(16). Based on personal experience of peer-reviewing clinical trial papers and discussion with a wider group of clinicians, we hypothesised that there might be a much greater number of diseases where the augmented binary method could be useful. We decided that a more systematic attempt to identify these clinical areas was warranted.

**Materials And Methods**

We made use of the COMET (Core Outcome Measures in Effectiveness Trials) database ([http://www.comet-initiative.org/resources](http://www.comet-initiative.org/resources)), which lists completed and ongoing projects in core outcome set (COS) development. COS represent the minimum that should be measured and reported in all clinical trials of a specific condition(17,18).

We reviewed physiological and mortality trial outcomes (categorised according to (19)) recorded within all core outcome sets (COS) in the COMET database that were published before 2016. The restriction to 2016 was so we could utilise the taxonomy of COS presented in Dodd et al (19), which considered COS up until then. This led to 287 COS papers to review, which were split amongst the three authors (JW MM and SD) to review. Each core outcome set paper was reviewed to determine if any responder (composite or categorised continuous) endpoints were recommended for reporting in all clinical trials within that condition. In some cases, a potentially relevant endpoint was not clearly described in the core outcome paper. In this case, we examined RCTs that had used the endpoint to determine whether it was a suitable responder endpoint.

**Results**
This process allowed us to identify 39 clinical areas (additional to solid tumour oncology, rheumatoid arthritis, and SLE) where the augmented binary method could be utilised to gain efficiency. An additional 28 clinical areas had used responder endpoints formed from a single categorised/dichotomised continuous variable. Table 2 breaks down the number by clinical classification. A full listing of these clinical areas is given in supplementary material. These are given by clinical classification in supplementary tables 1a-1m.

The clinical classifications that the method appears most useful to in terms of number of endpoints are rheumatology (11 found), non-solid tumour oncology (10 found), neurology (9 found), and cardiovascular (8 found).

We note that this was not a systematic review and represents a likely substantial underestimate of the number of clinical areas where suitable endpoints are used, as our review only covered clinical areas which were covered by a COS published by 2016. As an example, Table 1 mentions type II diabetes and shows diabetes remission would be a suitable endpoint; however, since there was no associated COS published by 2016, it does not appear in the identified clinical areas (although gestational diabetes does).

**Discussion**

In this paper we have highlighted and summarised previous statistical work on an efficient analysis approach called the augmented binary method, which can be used to improve analysis of composite responder outcomes. The method allows retention of clinically relevant endpoints whilst improving the power of analyses by an amount equivalent to a considerable increase in sample size. As well as describing previous work we have conducted a review of new clinical areas that the method could be used for. We have also provided a worked example of fitting the model using publicly available R code in the supplementary material.

Through our review of core outcome sets we have found a great deal of new disease areas where the augmented binary method could be applied to gain power. We acknowledge that many of the core outcome sets were developed prior to best-practice guidance(20) existed and therefore the quality of them may differ. We also do not know whether the augmented binary method is currently being used in any areas; however our intuition is that it is not due to the relatively low number of citations of the methodological papers describing the method. We hope that this paper may help improve the uptake of the method.

Although the results indicate the widespread utility of the method, there are several areas where further methodological research is required to fully realise the possible benefits.

There are several endpoints which are typically analysed using time-to-event methods. Many progression, remission and relapse endpoints are used and the time until such a negative event occurs is the quantity of interest. Although the augmented binary method is well developed for composite responder outcomes
that are analysed at a single timepoint or longitudinally, further work is needed to apply it to time-to-event outcomes.

In some cases, the composite responder outcomes are particularly complex with more than two components and with response being defined as meeting some, but not all, of the criteria. Recent work in this area (16) shows the potential efficiency gain is even larger in this case. In addition, the method, with some modification to the underlying latent variable model, could be applied in the case of a categorised responder endpoint with more than 2 levels.

We have focused on how the method can improve power of trials. An alternative way to use this improved power would be to reduce the sample size needed for a target power level. A barrier to widespread use of the method in this way is sample size estimation. Methods for conducting sample size estimation for trials using responder endpoints analysed using latent variable models are developed in McMenamin et al (21). Barriers to this include the requirement for pilot data to inform the values used for the required parameters. Further work on ensuring these methods can be used in practice, such as software that can be used in a wide variety of situations, is a priority for future research.

Conclusion

In this paper we have shown that responder composite outcomes are used as primary clinical trial endpoints in many diseases. Analysing data from these trials using the augmented binary approach would improve power equivalent to increasing the sample size by at least 35%. Further methods research is needed to improve time-to-event analyses using these outcomes as events.

Declarations

Ethics approval and consent to participate: not applicable

Consent for publication: not applicable

Availability of data and materials: The datasets used are available from the authors of (19) upon request. R code for using the augmented binary method is provided in supplementary material.

Competing interests: not applicable

Funding: MMM and JW are supported by funding from the MRC (grant code MC_UU_00002/6). JW is also supported by Cancer Research UK (C48553/A1811). None of the funding bodies had a role in the design of the study, analysis, interpretation of data or writing the manuscript.

Authors contributions: All authors contributed to the design of the manuscript and interpretation of the data. JW developed the first draft, MMM and SD critically revised the manuscript and approved the final version. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.
Acknowledgements: not applicable

Abbreviations
CI – confidence interval; COMET - Core Outcome Measures in Effectiveness Trials; COS – core outcome set; RCT – randomized controlled trial; SLE - systemic lupus erythematosus

References
1. Ross S. Composite outcomes in randomized clinical trials: arguments for and against. Am J Obstet Gynecol. 2007 Feb 1;196(2):119.e1-119.e6.
2. Altman DG, Royston P. The cost of dichotomising continuous variables. BMJ. 2006;332:1080.
3. Senn S. Disappointing dichotomies. Pharm Stat. 2003 Oct;2(4):239–40.
4. Owen S V., Froman RD. Why carve up your continuous data? Res Nurs Health. 2005 Dec;28(6):496–503.
5. Lachenbruch PA. Analysis of data with excess zeros. Stat Methods Med Res. 2002;11:297–302.
6. Hu Z, Proschan M. Two-part test of vaccine effect. Stat Med. 2015 May 20;34(11):1904–11.
7. DeCoster J, Iselin AR, Gallucci M. A conceptual and empirical examination of justifications for dichotomization. Psychol Methods. 2009;14:349–66.
8. Suissa S. Binary methods for continuous outcomes: a parametric alternative. J Clin Epidemiol. 1991;44:241–8.
9. Suissa S, Blais L. Binary regression with continuous outcomes. Stat Med. 1995;14:247–55.
10. Wason JMS, Seaman SR. Using continuous data on tumour measurements to improve inference in phase II cancer studies. Stat Med. 2013;32(26):4639–50.
11. Lin C-J, Wason JMS. Improving phase II oncology trials using best observed RECIST response as an endpoint by modelling continuous tumour measurements. Stat Med. 2017 Dec 20;36(29):4616–26.
12. Wason JMS, Jenkins M. Improving the power of clinical trials of rheumatoid arthritis by using data on continuous scales when analysing response rates: an application of the augmented binary method. Rheumatology. 2016;55(10):1796–802.
13. McMenamin M, Berglind A, Wason JMS. Improving the analysis of composite endpoints in rare disease trials. Orphanet J Rare Dis. 2018;13(1):81.
14. Weinblatt ME, Genovese MC, Ho M, Hollis S, Rosiak-Jedrychowicz K, Kavanaugh A, et al. Effects of fostamatinib, an oral spleen tyrosine kinase inhibitor, in rheumatoid arthritis patients with an inadequate response to methotrexate: Results from a phase iii, multicenter, randomized, double-blind, placebo-controlled, parallel-group study. Arthritis Rheumatol. 2014 Dec 1;66(12):3255–64.
15. Box GEP, Cox DR. An Analysis of Transformations. Vol. 26, Source: Journal of the Royal Statistical Society. Series B (Methodological). 1964.
16. McMenamin M, Barrett JK, Berglind A, Wason JMS. Employing latent variable models to improve efficiency in composite endpoint analysis. 2019 Feb 19;arXiv:1902.07037.

17. Gargon E, Gurung B, Medley N, Altman DG, Blazeby JM, Clarke M, et al. Choosing Important Health Outcomes for Comparative Effectiveness Research: A Systematic Review. Scherer RW, editor. PLoS One. 2014 Jun 16;9(6):e99111.

18. Gargon E, Gorst SL, Harman NL, Smith V, Matvienko-Sikar K, Williamson PR. Choosing important health outcomes for comparative effectiveness research: 4th annual update to a systematic review of core outcome sets for research. Gillies K, editor. PLoS One. 2018 Dec 28;13(12):e0209869.

19. Dodd S, Clarke M, Becker L, Mavergames C, Fish R, Williamson PR. A taxonomy has been developed for outcomes in medical research to help improve knowledge discovery. J Clin Epidemiol. 2018 Apr;96:84–92.

20. Kirkham JJ, Davis K, Altman DG, Blazeby JM, Clarke M, Tunis S, et al. Core Outcome Set-STAndards for Development: The COS-STAD recommendations. PLOS Med. 2017 Nov 16;14(11):e1002447.

21. McMenamin M, Barrett JK, Berglind A, Wason JMS. Sample Size Estimation using a Latent Variable Model for Mixed Outcome Co-Primary, Multiple Primary and Composite Endpoints. 2019 Dec 11;ArXiv:1912.05258.

### Tables

**Table 1 - examples of responder endpoints used in different areas of medicine; italicised components denote continuous dichotomisations.**

| Clinical area       | Endpoint          | Components and definition                                                                 |
|---------------------|-------------------|-------------------------------------------------------------------------------------------|
| Oncology            | Tumour response   | 1. *Sum of longest diameter of target tumour lesions* >=30% shrinkage from baseline        |
|                     |                   | 2. No new tumour lesions                                                                  |
| Rheumatology        | ACR20             | 1. *Swollen joint count* >=20% improvement                                                |
|                     |                   | 2. *Tender joint count* >=20% improvement                                                  |
|                     |                   | 3. 20% improvement in at least three of:                                                   |
|                     |                   | 4. a. *patient assessment*                                                                  |
|                     |                   | b. *physician assessment*                                                                  |
|                     |                   | c. *pain scale*                                                                            |
|                     |                   | d. *disability/functional questionnaire*                                                    |
|                     |                   | e. *acute phase reactant (ESR or CRP)*                                                     |
|                     |                   | 5. No rescue therapy given                                                                 |
| Type II diabetes    | Diabetes remission| 1. *Glycated haemoglobin A$_{1c}$ concentration* <=6.5%                                     |
|                     |                   | 2. *Fasting glucose concentration* <=5.6 mmol/L                                             |
|                     |                   | 3. No non-study pharmacological treatment given                                              |
Table 2 - number of new clinical areas identified by classification; full list provided in supplementary material; if a condition had both composite and non-composite responder endpoints identified, they were only included in the composite column. *excludes solid-tumour oncology (as the utility of the method had previously been highlighted there)

| Classification                  | Number of conditions with suitable composite responder endpoints | Number of conditions with single-variable responder endpoints |
|---------------------------------|------------------------------------------------------------------|---------------------------------------------------------------|
| Bleeding and Transfusion        | 2                                                                | 1                                                             |
| Cancer*                         | 6                                                                | 4                                                             |
| Cardiovascular and circulation  | 5                                                                | 3                                                             |
| Dentistry and vision            | 2                                                                | 1                                                             |
| Gastroenterology                | 3                                                                | 1                                                             |
| Infectious diseases             | 3                                                                | 0                                                             |
| Lungs and airways               | 0                                                                | 2                                                             |
| Mental health and addiction     | 3                                                                | 1                                                             |
| Neurology                       | 2                                                                | 7                                                             |
| Orthopaedics and trauma         | 1                                                                | 3                                                             |
| Renal and urology               | 2                                                                | 1                                                             |
| Rheumatology                    | 8                                                                | 3                                                             |
| Unclassified                    | 2                                                                | 1                                                             |
| Total                           | 39                                                               | 28                                                            |

**Figures**
Figure 1

Illustration of how (hypothetical) response information from patients is weighted by the two different methods. Non-responders are made up of those in whom the continuous component is below 1 and those who do not respond according to another binary criterion. Underlying the augmented binary method is a joint model that is fitted to the continuous and binary data and yields fitted ‘response weights’ for each patient, which can then be compared between arms.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Reportingchecklist.docx
- Supplementarymaterialfinalrevised.docx