

SOFTWARE

Open Access



# Increasing power in the analysis of responder endpoints in rheumatology: a software tutorial

Martina McMenamin<sup>1,2\*</sup> , Michael J. Grayling<sup>3</sup>, Anna Berglind<sup>4</sup> and James M. S. Wason<sup>1,3</sup>

## Abstract

**Background:** Composite responder endpoints feature frequently in rheumatology due to the multifaceted nature of many of these conditions. Current analysis methods used to analyse these endpoints discard much of the data used to classify patients as responders and are therefore highly inefficient, resulting in low power. We highlight a novel augmented methodology that uses more of the information available to improve the precision of reported treatment effects. Since these methods are more challenging to implement, we developed free, user-friendly software available in a web-based interface and as R packages. The software consists of two programs: one that supports the analysis of responder endpoints; the second that facilitates sample size estimation. We demonstrate the use of the software to conduct the analysis with both the augmented and standard analysis method using the MUSE study, a phase IIb trial in patients with systemic lupus erythematosus.

**Results:** The software outputs similar point estimates with smaller confidence intervals for the odds ratio, risk ratio and risk difference estimators using the augmented approach. The sample size required in each arm for a future trial using the novel approach based on the MUSE data is 50 versus 135 for the standard method, translating to a reduction in required sample size of approximately 63%.

**Conclusions:** We encourage trialists to use the software demonstrated to implement the augmented methodology in future studies to improve efficiency.

**Keywords:** Composite responder endpoint, Augmented binary method, Systemic lupus erythematosus, Shiny, Software

## Background

Composite endpoints combine a number of individual outcomes in order to assess the effectiveness or efficacy of a treatment. They are typically used in situations where it is difficult to identify a single relevant endpoint to sufficiently capture the change in disease status incited by the treatment, however they may be employed for multiple purposes [1–3]. A subset of these endpoints, known as composite responder endpoints,

are commonly used in studies of rheumatic conditions [4–6]. These endpoints allocate patients as either ‘responders’ or ‘non-responders’ based on whether they cross predefined thresholds in the individual continuous outcomes or respond in individual binary outcomes, and are typically treated as a single binary endpoint. A review of core outcome sets across a range of disease areas identified 13 conditions within the domain of rheumatology where an endpoint assuming this structure is recommended to be reported as a primary or secondary endpoint [7]. Table 1 details a typical composite endpoint in each of these 13 conditions

\*Correspondence: mmcm@hku.hk

<sup>1</sup> MRC Biostatistics Unit, University of Cambridge, Cambridge, UK  
Full list of author information is available at the end of the article



© The Author(s) 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

**Table 1** List of rheumatic conditions where composite responder endpoints containing at least one continuous component are used

Condition	Endpoint	Response definition
Acute Gout	Proportion of patients who responded*	1. sUA level of < 6.0 mg
Ankylosing spondylitis	ASAS20 response	1. 20% improvement and $\geq 10$ units of change (on a 0–100 scale) in each of 3 domains 2. No worsening of a similar amount in the fourth domain (Components are physical function, pain, inflammation and patient's global assessment)
I idiopathic arthritis-associated uveitis	Best corrected visual acuity above threshold and no light perception	1. Best-corrected visual acuity, thresholds $\leq 20/50, \leq 20/200$ 2. No light perception 3. Estimate contribution of amblyopia, yes/no
Juvenile arthritis	Response	1. Improvement by 30% in at least 3 of: a. MD global assessment; b. parent or patient global assessment c. functional ability; d. number of joints with active arthritis; e. number of joints with limited range of motion; f. erythrocyte sedimentation rate
Juvenile dermatomyositis	Responder index	1. $\geq 4$ point reduction from baseline in safety of estrogen in lupus national assessment (SELENA) systemic lupus erythematosus disease activity index (SLEDAI) score 2. No worsening (increase of < 0.30 points from baseline) in physician's global assessment (PGA) 3. No new British Isles Lupus Assessment Group of SLE clinics (BILAG) A organ domain score or 2 new BILAG B organ domain scores compared with baseline
Prevention of fracture in high-risk populations	Response	1. Bone mineral density increase 2. Occurrence of new vertebral fractures
Proliferative and membranous lupus renal disease	Urinary protein levels within normal range*	1. Between 6 and 8.3 g per deciliter (g/dL)
Rheumatoid arthritis	ACR20 response	1. $\geq 20\%$ improvement in ACR score 2. Can be combined with additional requirements e.g. no additional medication
Sarcopenia prevention	Occurrence of sarcopenia	Heterogeneity in precise definition, but severe sarcopenia defined by all of the following: 1. Low muscle strength (assessed with chair stand test or grip strength) 2. Low muscle quantity/quality 3. Low physical performance as assessed with gait speed test or short physical performance battery
Sjogren's syndrome	Response	1. > 30% reduction in analog scales evaluating dryness, pain and fatigue
Systemic lupus erythematosus	SRI responder index	1. SLEDAI change e.g. $\leq -4$ 2. PGA change e.g. < 0.3 3. No Grade A or more than one Grade B in BILAG
Systemic sclerosis	SCP in normal range, no renal crisis	E.g. 1. < 3.0 mg/dl not drug related 2. No renal crisis
Vasculitis disorders	Response/partial improvement*	1. 50% improvement in disease activity score

\* Denotes a single dichotomized continuous variable

along with the response criteria. The endpoints range from single dichotomised measures such as that used in acute gout and vasculitis disorders, to combinations of continuous and discrete outcomes, such as those used in juvenile arthritis, systemic sclerosis and ankylosing spondylitis. As the review considered only outcomes listed within core outcome sets, it is likely a

conservative estimate of the number of rheumatology diseases using these measures.

Employing composite endpoints as the primary outcome measure in a study has many advantages. Proponents of composite endpoints believe that they are appropriate as they estimate the net clinical benefit of an intervention by accounting for the multiple factors of interest in a given disease [8–10]. This is especially

important in complex, multisystem, chronic diseases typical in rheumatology, to ensure that while a patient may have improved overall on one scale, that a flare in a different organ domain is not introduced on another scale. Furthermore, in the case of diseases with large variation in symptoms, employing a composite endpoint will avoid an arbitrary choice of a single outcome [11, 12]. However, many problems with the application of composite endpoints have been raised in the literature. In practice, composites may be inconsistently defined and provide opportunities for post-hoc changes [13]. Composite endpoints may also be driven by less important or subjective components, meaning that a promising treatment effect may not translate to benefit for patients. Moreover, they have the tendency to become very complicated and therefore difficult for physicians and patients to understand.

Additional criticisms arise from the analysis of these endpoints. The endpoints are typically treated as binary measures based on whether or not the patient responded, meaning the analysis is straight-forward to implement. However, for composites containing continuous outcome measures, this is at the expense of losing large amounts of information contained in those components [14]. In the context of phase II oncology responder endpoints, Wason and Seaman [15] proposed a novel technique to address these issues, using a more complex model to retain information on how close patients were to the response thresholds in the continuous measures. This has since been developed to include different types of endpoints [16–18] and for application in rare diseases [19]. It has also been successfully applied retrospectively in trials, including in rheumatoid arthritis [20] where the efficiency gains translated to a reduction in required sample size of at least 30%, and systemic lupus erythematosus (SLE), where the resulting required sample size was reduced by 60% [16].

One limitation of these methods is that they are more difficult to implement. Therefore, in this paper we demonstrate the use of free, user-friendly online software for conducting analyses of composite responder endpoints using the augmented approach. We illustrate this using the MUSE trial (NCT01438489) [21], which assessed the efficacy of anifrolumab in patients with SLE. Furthermore, we show how a second software tool may be used to establish the required sample size for a future study in SLE.

In what follows we give a brief description of the methods. In Sect. 2 we summarise the MUSE trial data and demonstrate the capability of the software; in Sect. 3 we describe the software output from the application and in Sect. 4 we discuss the implications for practice.

### Standard binary approach

We refer to the analysis method routinely applied to composite responder endpoints as the binary approach. This consists of collapsing the outcome information to form a binary response variable based on whether or not the patients meet the overall response criteria. This response variable is analysed using an appropriate binary analysis method, such as logistic regression. The treatment effect can then be reported in terms of odds ratios, risk ratios or risk differences along with confidence intervals and  $p$  values.

### Augmented approach

The augmented approach involves using a more sophisticated model that jointly models data from each of the components using a latent variable framework. The information contained in the continuous components is retained and used to weight patients differently in the analysis, based on how close their readings were to the response threshold. The probability of response in each arm is subsequently obtained which can then be used to form treatment effect estimates in terms of odds ratios, risk ratios or risk differences, as in the standard binary case. The increased efficiency compared to the binary approach is due to making inference on the probability of response without discarding any of the continuous data. In datasets where many patients' continuous readings are close to the dichotomisation threshold, this may have a substantial impact on the precision of the estimate and hence on the conclusions reached. More technical detail on the specification and assumptions of the models used in the augmented approach for a range of outcome types is provided elsewhere [15–20].

## Implementation

### MUSE trial summary

To illustrate how the analysis can be conducted using the software, we focus on the MUSE trial [21]. The trial was a phase IIb, randomised, double-blind, placebo-controlled study investigating the efficacy and safety of anifrolumab in adults with moderate to severe SLE. Patients ( $n = 305$ ) were randomised (1:1:1) to receive anifrolumab (300 mg or 1000 mg) or placebo, in addition to standard therapy every 4 weeks for 48 weeks. The primary end point in the study was the percentage of patients achieving an SLE Responder Index (SRI) response at week 24, with sustained reduction of oral corticosteroids ( $< 10$  mg/day and less than or equal to the dose at week 1 from week 12 through 24), which is typically referred to as 'SRI + OCS'. As detailed in Table 1, SRI is comprised of a continuous Physician's Global Assessment (PGA) measure, a continuous SLE Disease Activity Index (SLEDAI) measure

and an ordinal British Isles Lupus Assessment Group (BILAG) measure [22].

Table 2 shows the decomposition of responders and non-responders in each of the components by treatment arm. In both the treatment and the control arm, almost all patients are responders in both the PGA and BILAG measures. This indicates that these components do not enrich the composite endpoint in this study and so it is the SLEDAI and taper measures that are responsible for driving response rates. Previous work has shown that we may expect smaller efficiency gains than if three or four components determined response [23]. For the purposes of this analysis we combine the ordinal and binary components to form a single indicator, as modelling the ordinal component directly vastly increases computing time for only a small increase in efficiency [16].

**Analysis**

The software to implement the analysis is a Shiny application, a Graphical User Interface (GUI) for programming language ‘R’ which can be accessed at <https://martinamcm.shinyapps.io/augbin/>. Underlying code and documentation are available as indicated on the homepage. In addition, an R package to implement the augmented and standard binary methods is available at [https://github.com/martinamcm/augbin\\_rheum](https://github.com/martinamcm/augbin_rheum) and can be installed through the ‘devtools’ library using the ‘install\_github’ function. We focus on demonstrating the use of the Shiny app in what follows, however similar functionality is offered through the package.

The user begins by selecting the analysis tab and uploading the csv file using the ‘Upload Files’ panel. A table displaying the uploaded data will be shown on the right-hand side (see Fig. 1). Note that the data displayed in Fig. 1 is not the real data due to patient confidentiality but that the real trial data is used in what follows. In order to conduct the analysis, the user must organise the columns in the dataset prior to uploading, so that

patient ID comes first, followed by treatment arm, the continuous outcomes, the binary outcome and the baseline measures for the continuous variables. Failure to upload a dataset in this format will cause problems at the analysis stage as variables are identified by column order rather than variable name.

The raw data can be visualised using boxplots, histograms, density plots or bar graphs in the ‘Raw Data Plots’ panel. The user must then select the structure of the composite endpoint, where this can be one or two continuous components and zero or one binary components. As the ordinal and binary outcomes have been combined, the SLE endpoint has two continuous and one binary component. Details of the model fitted can be viewed by selecting ‘Generate model’. Both of these steps are demonstrated in the Additional file 1.

The analysis is initiated in the ‘Analysis’ panel by selecting the response threshold for the continuous outcomes. In this example the SLEDAI threshold is -4 and the PGA threshold is 0.3, where patients with readings below these values are considered to be responders and are otherwise treated as non-responders in the analysis.

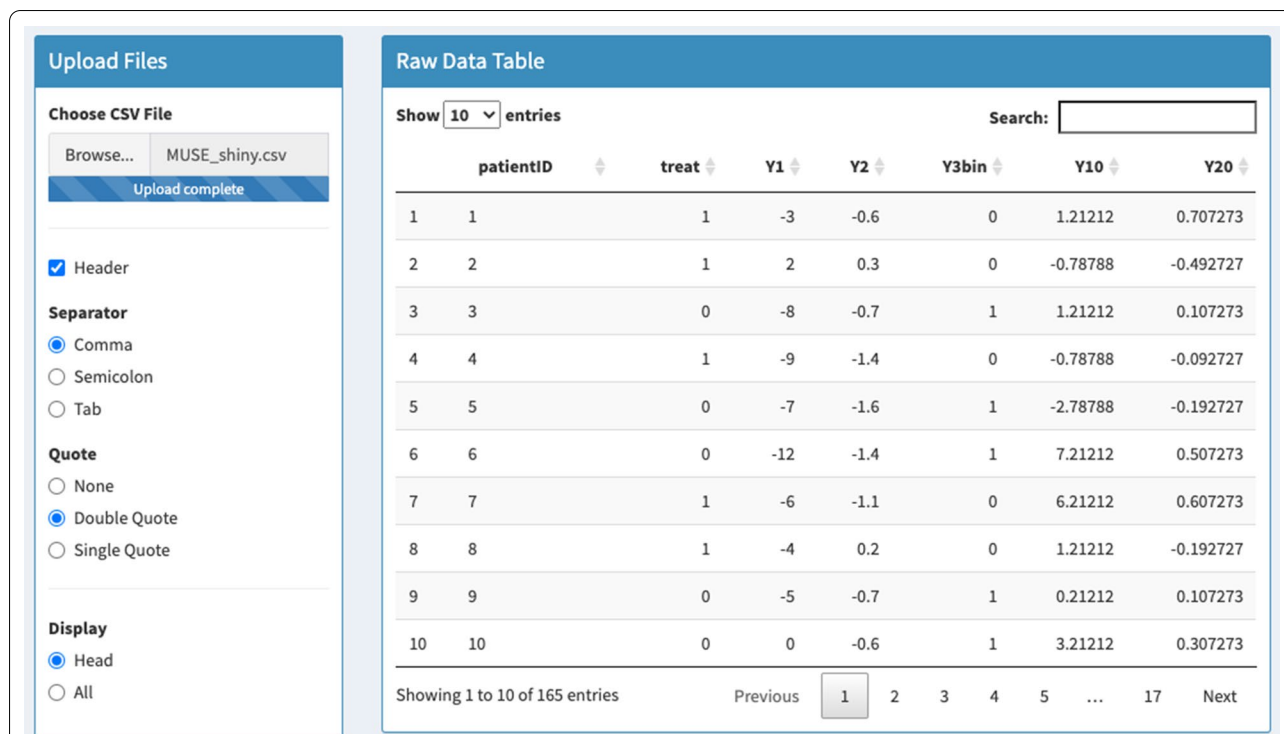
**Sample size determination**

A critical aspect of planning a future study using the augmented approach is how to determine the sample size, in order to avail of the efficiency gains. The ‘Multi-SampSize’ Shiny application allows users to determine sample sizes required through using preliminary data to inform the estimates. This may be in the form of pilot trial data, trial data from earlier phase studies or another source. The sample size estimation app can be accessed at <https://martinamcm.shinyapps.io/multsampsize/>, where detailed documentation is also included. Alternatively, the methods are implemented as an R package which can be installed using the ‘install\_github’ function, as detailed at [http://github.com/martinamcm/mult\\_sampsize](http://github.com/martinamcm/mult_sampsize). In order to demonstrate the capabilities of the app, we can

**Table 2** Observed response rates in each of the SRI + OCS components in the anifrolumab 300 mg arm and placebo arm of the MUSE trial

Components	Response criteria	Treatment arm	
		Anifrolumab 300 mg	Placebo
SLEDAI	Improvement of at least 4 points (change from baseline $\leq -4$ )	58/89	41/76
PGA	No flare/worsening of disease as measured by PGA (change from baseline $<0.3$ )	87/89	75/76
BILAG	No flare/worsening of disease as measured by BILAG (no new Grade A or more than one Grade B compared to baseline)	86/89	72/76
OCS	Sustained reduction in oral corticosteroids	53/95	37/87
Overall SRI + OCS response	Must responds in all four components	34/95	18/87

SLE index is comprised of a continuous SLEDAI outcome, continuous PGA outcome, ordinal BILAG outcome and binary OCS measure



**Fig. 1** MUSE trial data is uploaded in the left-hand panel where the user can indicate preferences such as whether the file includes column headers and whether to display some or all of the data. The raw data is viewed in the right-hand panel where users may also search for particular subjects

assume that we wish to design a future trial in SLE which will use the augmented approach as the primary analysis method. This will be directly informed by estimates from the MUSE study.

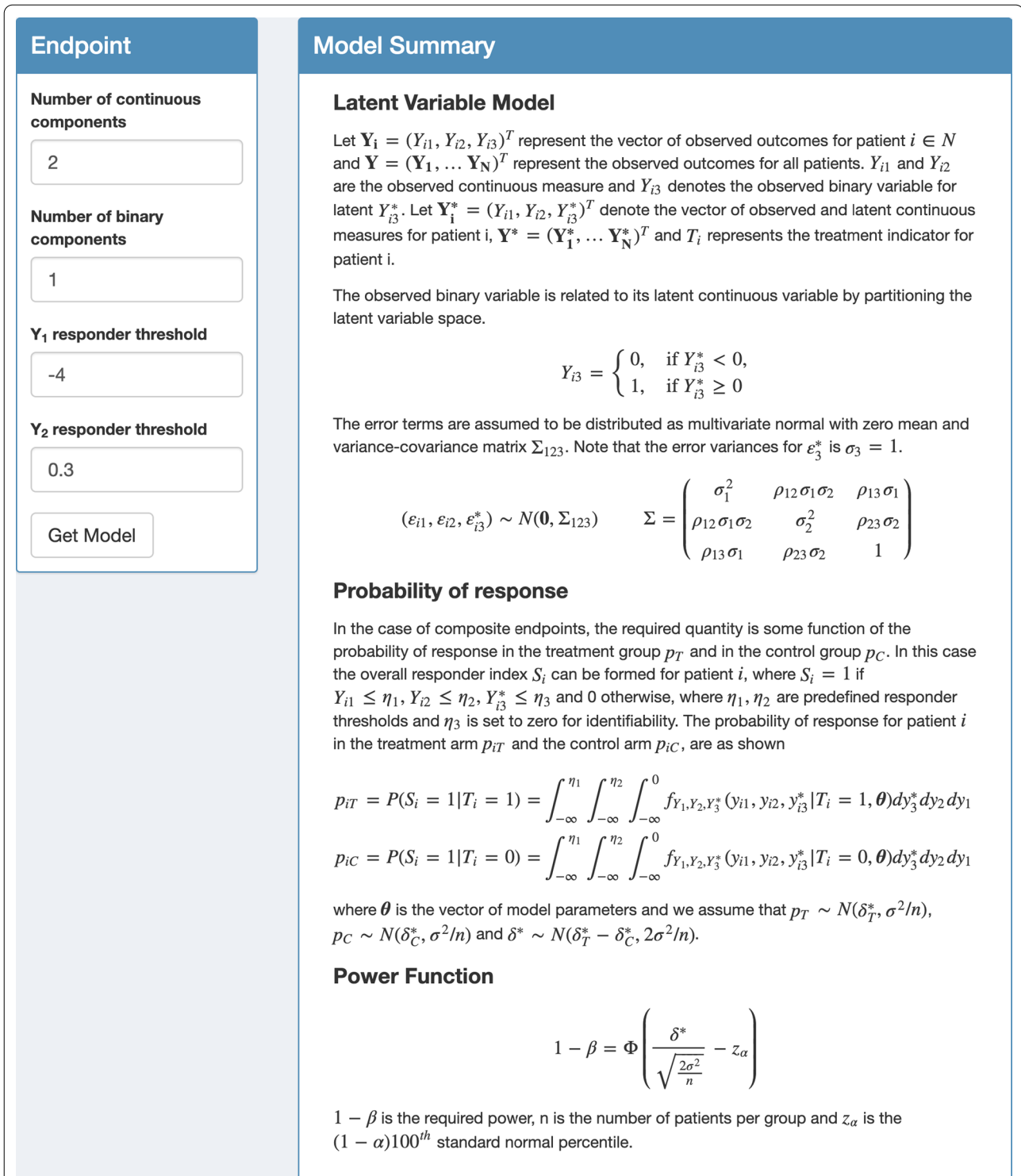
The user should select the ‘Sample Size’ tab and choose the ‘Composite’ option to proceed. Note that the app also accommodates co-primary and multiple primary endpoints, which also feature in rheumatology [23]. The user must select the number of continuous and binary components and the corresponding response thresholds, as before. Selecting ‘Get Model’ displays the relevant model assumed along with the power function used to determine the sample size (Fig. 2).

The pilot data can be uploaded using the ‘Parameter Estimates’ panel, where the columns must be ordered as before. Further guidance is available at <https://github.com/martinamcm/MultSampSize>. Clicking ‘Obtain Estimates’ runs the analysis and provides estimates for the probability of response in each arm, the risk difference and its variance.

**Results Analysis**

Figure 3 shows the probability of response in each arm of the MUSE trial using both the augmented approach and the standard binary approach. The probability of response

in the control group is similar using both methods, estimated as 0.240 using the augmented method and 0.256 using the binary method. The corresponding estimated probability of response in the treatment group is 0.351 compared to 0.395. We can expect some small differences in these quantities as the augmented model is accounting for how close a patient is to the responder threshold, whereas the binary approach only accounts for whether a patient was a responder or non-responder. The log-odds ratio, log-risk ratio and risk difference treatment effects are shown for each method along with the 95% confidence intervals. The point estimates are similar for both methods however the augmented approach reports each of the treatment effects more precisely. Given that the augmented binary method reports the treatment effect with a reduction in confidence interval width of approximately 37%, one would need to increase the sample size by 270%  $((1/0.37) \times 100)$  in this example to gain a similar level of precision whilst using the binary analysis method. The ‘Goodness-of-fit’ panel indicates how well the latent variable model fits the data, where a ‘good fit’ is assumed if the residuals follow the chi-squared distribution shown by the red line. The goodness-of-fit plots for the MUSE dataset are shown in the Additional file 1.



**Fig. 2** 'MultSampSize' app with sample size calculator for co-primary, multiple primary and composite endpoints. The interface for the composite endpoint is shown where the number of continuous and binary components and response thresholds for the continuous measures are selected in the 'Endpoint' panel. 'Get Model' generates the model summary of the latent variable model and the power function



**Sample size determination**

The ‘Sample Size Estimation’ panel displays the power curve and highlights the number of patients needed per arm to attain a desired power and alpha level, which can be set by the user. These values are also provided assuming the standard binary method was used as shown in Fig. 4. Assuming a one-sided test with alpha level 0.05, a target power of 80% and using the observed risk difference from the standard approach in the MUSE trial, dictates a required sample size of 50 individuals per arm for a future SLE study, compared with 135 patients per arm that would be required using the standard method. The power curve for both the augmented and binary approaches is shown in Fig. 4. Note that the values shown in the ‘Parameter Estimates’ panel can be used as inputs for the R package to determine the sample size with more flexibility, as the user can modify the estimates provided by the data.

**Discussion**

In this paper we highlight novel methods to address inefficiencies in the analysis of composite responder endpoints commonly used in rheumatology, which use more sophisticated models to retain the information provided by continuous components. As this approach is more difficult to implement, we developed user-friendly, free to use software. This software conducts

the analysis using both methods and offers sample size determination for future studies using the augmented technique as the primary analysis method. We demonstrated the functionality of the apps using the MUSE trial dataset, a phase II study in patients with SLE using a composite comprised of two continuous and one binary outcomes. The analysis took approximately 5 min to complete, where the gains in efficiency resulted in 37% reduction in confidence interval width for the log-odds estimate, equating to approximately 60% reduction in required sample size. This means that 60% fewer patients could be recruited by using the augmented analysis method without requiring any additional data to be collected. Using the MUSE trial to inform a future study in SLE indicated that the novel approach would require 50 per arm versus 135 required for the standard approach to detect the risk difference of 0.14 estimated by the binary method.

As the structure of the composite endpoints vary substantially and may be quite complex, the efficiency gains offered by this technique also depends on many factors. In particular, the number of continuous and binary components, response probabilities in each arm, the responder thresholds and correlation between components. Both Shiny applications therefore report the results for the standard and novel approaches. In the case of sample size estimation, the investigator has



**Fig. 4** The MUSE trial dataset is uploaded in the 'Parameter Estimates' panel, where the probability of response in each arm, treatment effect and variance is shown for both the augmented and binary approaches. The power curve for a future study based on MUSE trial results is shown in the 'Sample Size Estimation' panel

the option to recruit the number of patients dictated using the binary approach and benefit from the additional power instead.

The methods underpinning the apps allow for any number of continuous, ordinal and binary components to be included in the composite endpoint however the app currently only implements this for up to two continuous components along with up to one binary component. Each additional continuous endpoint may add a substantial amount of efficiency and so future work will involve updating the software to allow for more continuous components to facilitate endpoints such as those used in juvenile arthritis. In its current form the software may

still be used for such endpoints however the additional components will have to be combined as a binary indicator, retaining the most informative continuous outcomes. However, it is important to note that responder endpoints with a more complex structure exist within rheumatology that cannot yet be accommodated by the software. In particular a common endpoint in osteoarthritis is the OARSI/OMERACT responder criterion [24, 25]. A 'low responder' would require 20% or more change in KOOS pain subscale score or an absolute change of 3 units if the baseline score is 15 or less, and a global change in pain of 'slightly better' or 'much better'. A 'high responder' must achieve 50% or more change in KOOS pain subscale



score or absolute change of 3 units if baseline score is 15 or less. The underlying methodology would require some extension in order to be applicable to this more complex endpoint however, this would be possible to do. With extension to the methodology, an important future aim for this software would be to include facilitating analysis of these more complex rheumatic endpoints that allow response thresholds in individual components to depend on values of other components within the endpoint. We welcome suggestions from users for future development on the Github pages of the respective apps.

## Conclusions

Novel methods to analyse composite responder endpoints can now be easily applied. We encourage trialists to use the software demonstrated in future studies in rheumatology to improve efficiency and reduce biases arising from measurement error.

## Availability and requirements

Project name: AugBin. Project home page(s): Shiny apps: <https://github.com/martinamcm/AugBin>, <https://github.com/martinamcm/MultSampSize>. R packages: [https://github.com/martinamcm/augbin\\_rheum](https://github.com/martinamcm/augbin_rheum), [https://github.com/martinamcm/mult\\_sampsize](https://github.com/martinamcm/mult_sampsize). Operating system(s): Any. Programming language: R/Shiny. Other requirements: None. License: None. Any restrictions to use by non-academics: None.

## Abbreviations

SLE: Systemic Lupus Erythematosus; SRI: SLE Responder Index; PGA: Physician's Global Assessment; SLEDAI: Systemic Lupus Erythematosus Disease Activity Index; BILAG: British Isles Lupus Assessment Group; GUI: Graphical User Interface.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s41927-021-00224-0>.

**Additional file 1.** Further app functionality and computation detail for the AugBin and MultSampSize Shiny applications.

## Acknowledgements

We thank the reviewers for their feedback which helped to improve the article and software.

## Authors' contributions

M.M.M. wrote the R code underlying the software, A.B. provided trial dataset and data specific interpretation for the application included, M.G. and J.W. tested and revised the software. All authors contributed to writing and reviewing drafts of the manuscript. All authors read and approved the final manuscript.

## Funding

MMM is funded by the NIHR Cambridge Biomedical Research Centre and JW is supported by funding from the Medical Research Council (MRC), grant code

MC\_UU\_00002/6. None of the funding bodies had a role in the design of the study, analysis, interpretation of data or writing the manuscript.

## Availability of data and materials

Simulated example datasets and underlying R code for both web applications is available at <https://github.com/martinamcm>.

## Declarations

### Ethics approval and consent to participate

The authors were given consent to use the MUSE trial dataset in the application.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no conflicts of interest.

### Author details

<sup>1</sup>MRC Biostatistics Unit, University of Cambridge, Cambridge, UK. <sup>2</sup>WHO Collaborating Centre for Infectious Disease Epidemiology and Control, School of Public Health, The University of Hong Kong, Hong Kong, Hong Kong, Special Administrative Region, China. <sup>3</sup>Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, UK. <sup>4</sup>Late Respiratory & Immunology, Biometrics, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden.

Received: 3 January 2021 Accepted: 16 August 2021

Published online: 07 December 2021

## References

- Ross S. Composite outcomes in randomized clinical trials: arguments for and against. *Am J Obstet Gynecol*. 2007;196(2):1991–6. <https://doi.org/10.1016/j.ajog.2006.10.903>.
- Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C. Composite outcomes in randomized trials: greater precision but with greater uncertainty? *JAMA*. 2003;289:2554–9. <https://doi.org/10.1001/jama.289.19.2554>.
- Montori V, Permyer-Miralda G, Ferreira-Gonzalez I, Busse J, Pacheco-Huergo V, Bryant D, et al. Validity of composite endpoints in clinical trials. *BMJ*. 2005;330:594–6. <https://doi.org/10.1136/bmj.330.7491.594>.
- Cornec D, Devauchelle-Pensec V, Mariette X, Jousse-Joulin S, Berthelot JM, Perdriger A, et al. Development of the Sjögren's Syndrome Responder Index, a data-driven composite endpoint for assessing treatment efficacy. *Rheumatology*. 2015;54(9):1699–708. <https://doi.org/10.1093/rheumatology/kev114>.
- Ibrahim F, Tom BD, Scott DL, Prevost AT. A systematic review of randomised controlled trials in rheumatoid arthritis: the reporting and handling of missing data in composite outcomes. *Trials*. 2016;17(1):272. <https://doi.org/10.1186/s13063-016-1402-5>.
- Helliwell PS, Kavanaugh A. Comparison of composite measures of disease activity in psoriatic arthritis using data from an interventional study with golimumab. *Arthritis Care Res (Hoboken)*. 2014;66(5):749–56. <https://doi.org/10.1002/acr.22204>.
- Wason J, McMenamin M, Dodd S. Analysis of responder-based endpoints: improving power through utilising continuous components. *Trials*. 2020;21:427. <https://doi.org/10.1186/s13063-020-04353-8>.
- Wittkop L, Smith C, Fox Z, et al. Methodological issues in the use of composite endpoints in clinical trials: examples from the HIV field. *Clin Trials*. 2010;7(1):19–35. <https://doi.org/10.1177/1740774509356117>.
- 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(2):176–82. <https://doi.org/10.1093/eurheartj/ehr352>.
- Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?" *Lancet*. 2005;365(9453):82–93. [https://doi.org/10.1016/S0140-6736\(04\)17670-8](https://doi.org/10.1016/S0140-6736(04)17670-8).

11. Carneiro AV. Composite outcomes in clinical trials: uses and problems. *Rev Port Cardiol*. 2003;22(10):1253–63.
12. Lauer MS, Topol EJ. Clinical trials—multiple treatments, multiple end points, and multiple lessons. *JAMA*. 2003;289(19):2575–7. <https://doi.org/10.1001/jama.289.19.2575>.
13. Cordoba G, Schwartz L, Woloshin S, Bae H, Gøtzsche PC. Definition, reporting, and interpretation of composite outcomes in clinical trials: systematic review. *BMJ*. 2010;341: c3920. <https://doi.org/10.1136/bmj.c3920>.
14. Senn S. Disappointing dichotomies. *Pharm Stat*. 2003;2:239–40. <https://doi.org/10.1002/pst.090>.
15. Wason J, Seaman SR. Using continuous data on tumour measurements to improve inference in phase II cancer studies. *Stat Med*. 2013;32(26):4639–50. <https://doi.org/10.1002/sim.5867>.
16. McMenamin M, Barrett JK, Berglind A, Wason JM. Employing a latent variable framework to improve efficiency in composite endpoint analysis. *Stat Methods Med Res*. 2021;30(3):702–16. <https://doi.org/10.1177/0962280220970986>.
17. Lin CJ, Wason JMS. Improving phase II oncology trials using best observed RECIST response as an endpoint by modelling continuous tumour measurements. *Stat Med*. 2017;36(29):4616–26. <https://doi.org/10.1002/sim.7453>.
18. Wason JM, Seaman SR. A latent variable model for improving inference in trials assessing the effect of dose on toxicity and composite efficacy endpoints. *SMMR*. 2020;29(1):230–42. <https://doi.org/10.1177/0962280219831038>.
19. McMenamin M, Berglind A, Wason J. Improving the analysis of composite endpoints in rare disease trials. *Orphanet J Rare Dis*. 2018;13:81. <https://doi.org/10.1186/s13023-018-0819-1>.
20. Wason JM, 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. <https://doi.org/10.1093/rheumatology/kew263>.
21. Furie R, Khamashta M, Merrill J, Werth V, Kalunian K, Brohawn P, et al. Anifrolumab, an anti-interferon alpha receptor monoclonal antibody, in 23 moderate-to-severe systemic lupus erythematosus. *Arthritis Rheumatol*. 2017;69:376–86. <https://doi.org/10.1002/art.39962>.
22. Luijten K, Tekstra J, Bijlsma J, Bijl M. The Systemic Lupus Erythematosus Responder Index (SRI); a new SLE disease activity assessment. *Autoimmun Rev*. 2012;11(5):326–9. <https://doi.org/10.1016/j.autrev.2011.06.011>.
23. 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; [arXiv:1912.05258](https://arxiv.org/abs/1912.05258).
24. Pham T, van der Heijde D, Altman RD, et al. OMERACT-OARSI initiative: osteoarthritis research society international set of responder criteria for osteoarthritis clinical trials revisited. *Osteoarthr Cartil*. 2004;12(5):389–99. <https://doi.org/10.1016/j.joca.2004.02.001>.
25. Dougados M, LeClaire P, Van der Heijde D, Bloch DA, Bellamy N, Altman RD. Response criteria for clinical trials on osteoarthritis of the knee and hip: a report of the Osteoarthritis Research Society International Standing Committee for Clinical Trials response criteria initiative. *Osteoarthr Cartil*. 2000;8(6):395–403. <https://doi.org/10.1053/joca.2000.0361>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

