Efficient modelling of complex responder endpoints to improve trial power

James Wason

www.newcastle-biostatistics.com/staff/james_wason

SUPPORTED BY

NIHR | National Institute for Health and Care Research

Responder endpoints

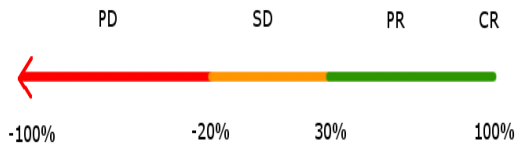
- In many conditions, especially ones with heterogeneous symptoms, specifying a single primary endpoint is difficult.
- One approach to this is combining multiple distinct endpoints (components) into a single composite endpoint.
- One type of composite divides patients into responders and non-responders; responders have to meet specified criteria on the components.

Responder endpoints - examples

1. Oncology: 30% decrease in target tumour lesions length and no new lesions.

2. Rheumatology: 20% decrease in number of tender and swollen joints and three other continuous components.

3. Type II Diabetes: 1. Glycated haemoglobin A_{1c} concentration ≤ 6.5; Fasting glucose concentration ≤ 5.6 mmol/L; No non-study pharmacological treatment given



MEASUREMENTS OF RESPONSE

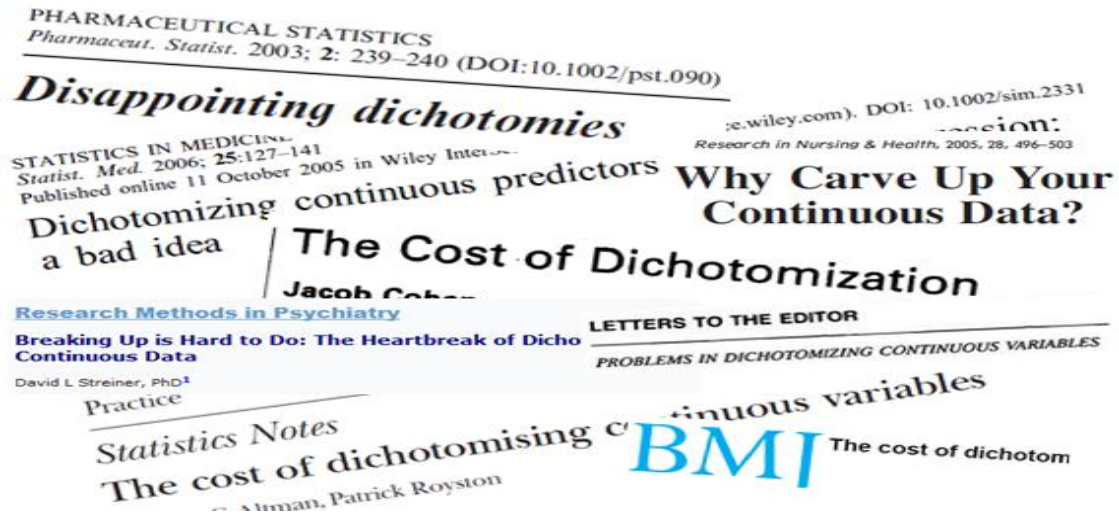
ACR 20 Response :

A decrease of at least 20% in both the number of tender and swollen joints along with 20% reduction in at least three of the following:

Patient's assessment of disease status
Patient's assessment of pain
Patient's assessment of physical function
Physician's assessment of disease status
C-reactive protein level

Responder endpoints

- These endpoints all involve continuous components being dichotomised.
- Usual analysis would treat response/non-response as binary outcome.
- Dichotomisation loses substantial statistical efficiency but difficult to avoid if combining continuous and binary components.



Alternative approaches

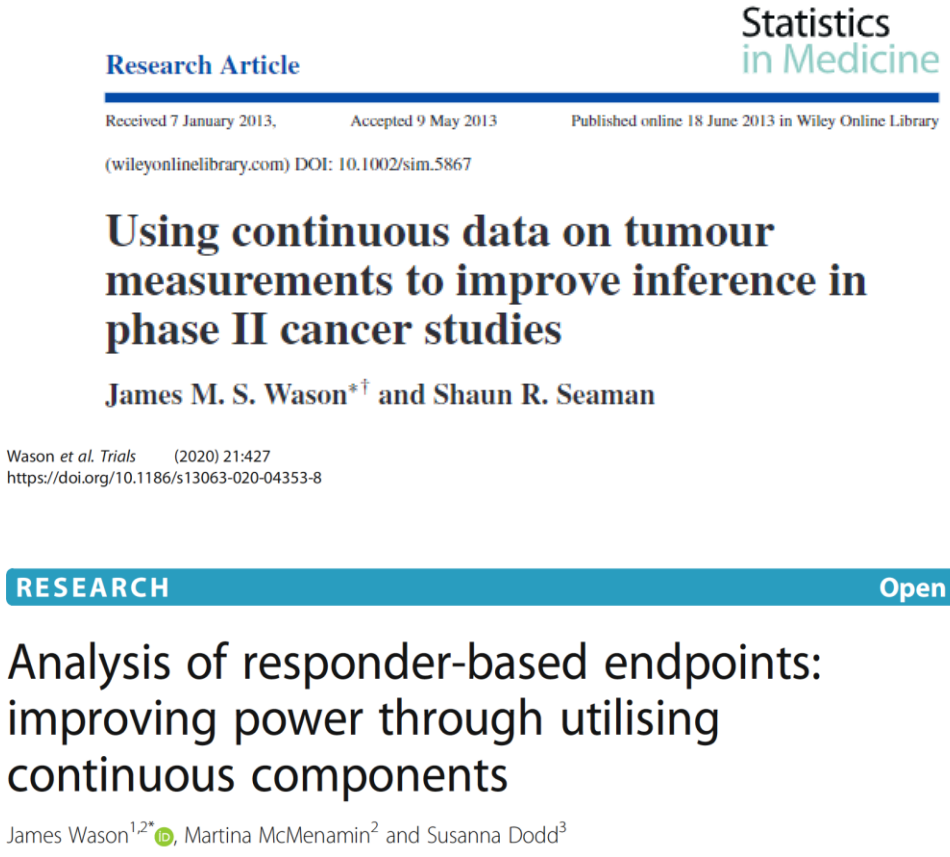
1. Pick a single (continuous) primary outcome and analyse that;
2. If response has >2 levels then using ordinal approaches can reclaim some lost efficiency;
3. Approaches for analysing composite endpoints (e.g. Win ratio[1]) could be used.

- However, assuming there's a good reason to use responder outcomes:
- How can we make inference about proportion of responders with maximum efficiency?

[1]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 Jan 1;33(2):176-82.

Augmented binary method

- Methods I've previously proposed (with others) retain the responder endpoint but use the continuous component to improve precision.



Research Article

Statistics in Medicine

Received 7 January 2013, Accepted 9 May 2013 Published online 18 June 2013 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.5867

Using continuous data on tumour measurements to improve inference in phase II cancer studies

James M. S. Wason*† and Shaun R. Seaman

Wason *et al. Trials* (2020) 21:427
https://doi.org/10.1186/s13063-020-04353-8

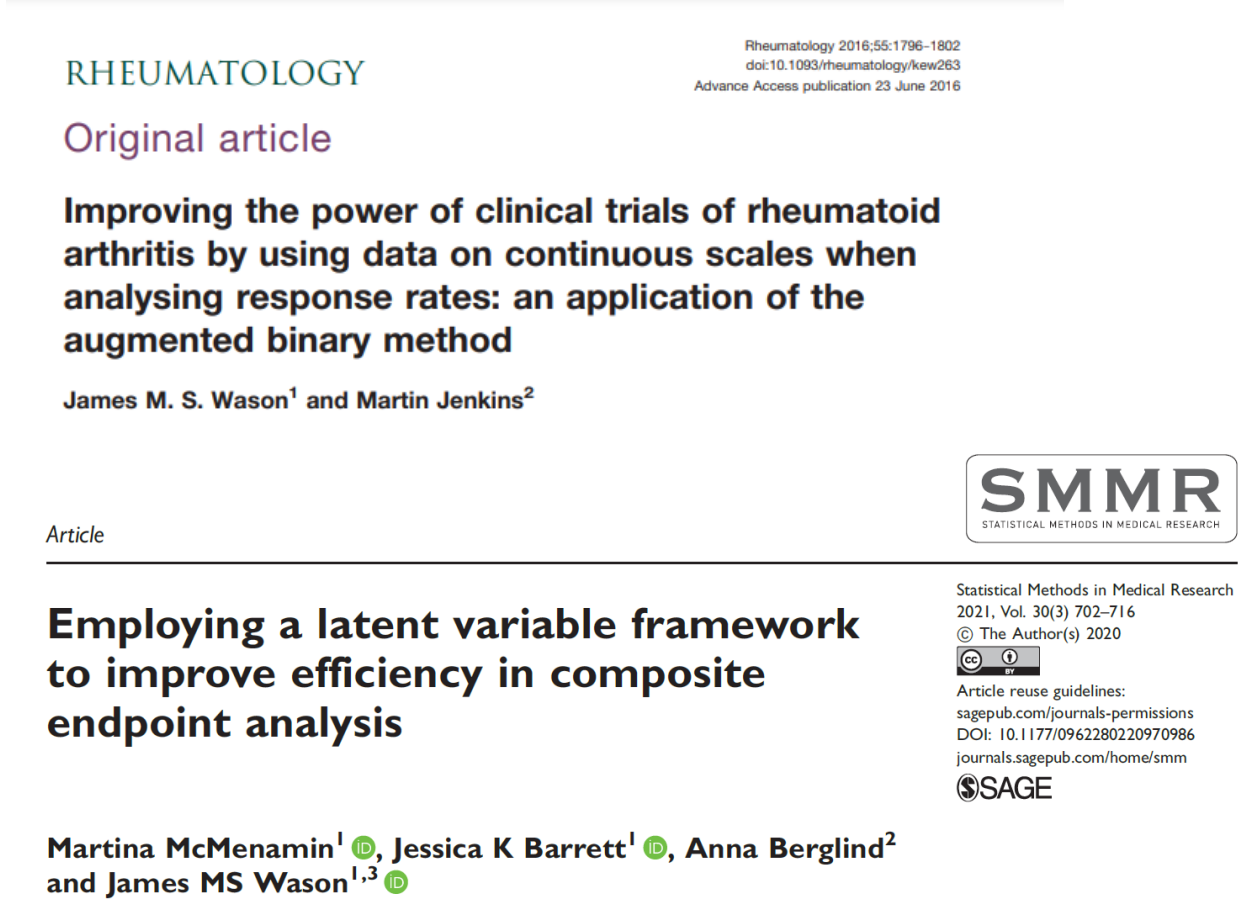
Trials

RESEARCH Open Access

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

James Wason[1,2*], Martina McMenamin[2] and Susanna Dodd[3]

Check for updates



RHEUMATOLOGY

Rheumatology 2016;55:1796–1802
doi:10.1093/rheumatology/kew263
Advance Access publication 23 June 2016

Original article

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

James M. S. Wason[1] and Martin Jenkins[2]

Article

SMMR STATISTICAL METHODS IN MEDICAL RESEARCH

Employing a latent variable framework to improve efficiency in composite endpoint analysis

Statistical Methods in Medical Research
2021, Vol. 30(3) 702–716
© The Author(s) 2020

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0962280220970986
journals.sagepub.com/home/smm

SAGE

Martina McMenamin[1], Jessica K Barrett[1], Anna Berglind[2] and James MS Wason[1,3]

Example: E-CLAD UK trial

- Randomised trial comparing extracorporeal photopheresis (ECP) vs standard of care for treatment of chronic lung allograft dysfunction (CLAD) following lung transplant.
- Primary endpoint is 'lung function stabilisation' which consists of the following criteria:

1. Lower than 10% loss in FEV1* from baseline to 24 weeks
2. Lower than 10% loss in FVC† from baseline to 24 weeks
3. No rapid decline confirmed being due to progressive CLAD at 12 weeks.

- Target sample size of 90 patients to be enrolled.

*Forced Expiratory Volume in 1 second

† Forced Vital Capacity

How does it work?

- Let R_i be response indicator for participant *i* and T_i be treatment indicator.
- Of interest to make inference on R_i, e.g.

$$P(R_i = 1 \,|T_i = 1) \; - P(R_i = 1 \,|T_i = 0)$$

- R_i defined by:

1. Y_{i1} percentage reduction in FEV1
2. Y_{i2} percentage reduction in FVC
3. Y_{i3} a binary indicator as to whether there was rapid decline due to progressive CLAD.

- R_i = 1 if: $Y_{i1} \leq 10\%, Y_{i2} \leq 10\%, Y_{i3} = 0$

How does it work?

- Instead of making inference on R_i, the augmented binary method involves two steps:

1) Fit a suitable joint model to Y_{i1}, Y_{i2}, Y_{i3}

2a) Use the parameter estimates from this model to estimate the quantity of interest

$$P(R_i = 1 \,|T_i = 1) \; - P(R_i = 1 \,|T_i = 0)$$

2b) Use the covariance of the parameter estimates and the delta method to estimate the standard error of the quantity of interest and get a confidence interval + Wald test.

How does it work?

- Latent variable model can be used for step 1: allows modelling correlation between continuous and binary components.

$$Y_{i1} = \mu_1 + \alpha_1 X_{i1} + \beta_1 T_i + \varepsilon_{i1}$$
$$Y_{i2} = \mu_2 + \alpha_2 X_{i2} + \beta_2 T_i + \varepsilon_{i2}$$
$$Y_{i3}^* = \mu_3 + \beta_3 T_i + \varepsilon_{i3}$$

- Y_{i3} = 1 if $Y_{i3}^* > 0$; $Y_{i3} = 0$ if $Y_{i3}^* \leq 0$ (latent variable)
- X_{i1} and X_{i2} are baseline measurements of FEV1 and FVC.
- $(\varepsilon_{i1}, \varepsilon_{i2}, \varepsilon_{i3})$ distributed as $N(0, \Sigma)$,

How does it work?

- Step 2a involves a multidimensional integration over a multivariate normal distribution – computationally efficient to evaluate, e.g. with pmvnorm in R.

$$p_{ij} = P(R_i = 1 | T_i = j) = \int_{-\infty}^{10} \int_{-\infty}^{10} \int_{-\infty}^{0} f_{Y_1,Y_2,Y_3}(y_1, y_2, y_3 | T_i = j, \boldsymbol{\theta})\, dy_1 dy_2 dy_3,$$

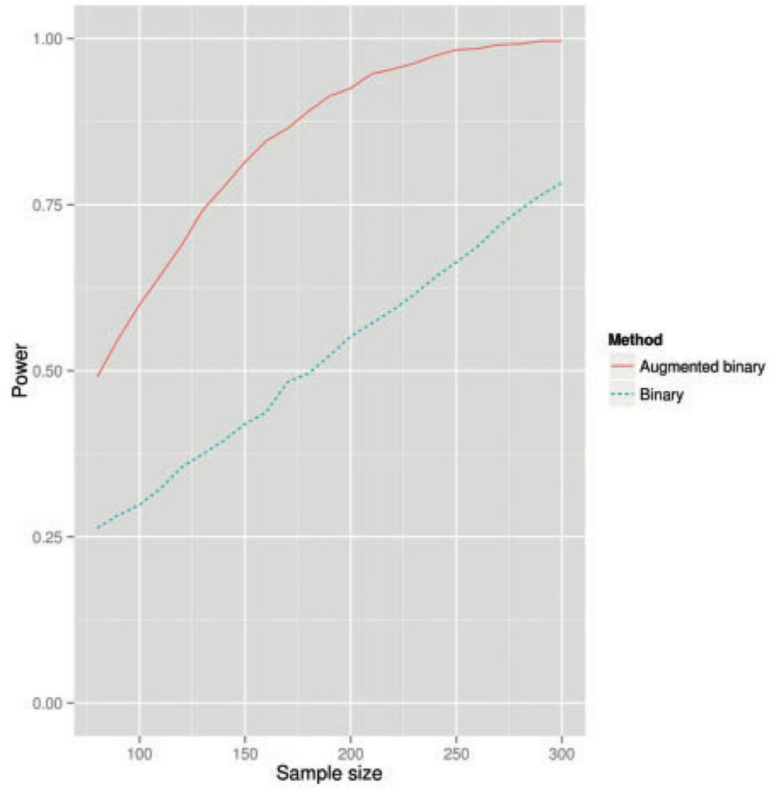
where $\boldsymbol{\theta}$ is the vector of latent variable model parameters.

- Step 2b uses the delta method and $\mathrm{Cov}(\widehat{\boldsymbol{\theta}})$ to get variance of between-group difference in response probability.

Augmented binary method

- The method has been tested on oncology[1], rheumatoid arthritis[2] and SLE datasets[3].
- Compared to analysing as binary variable, using the method is equivalent to increasing the sample size by >30%.
- …without type I error rate inflation or extra data being required.
- Power gain achieved varies between an effective increase in sample size between 30% and 1000%[3]!

FIG. 2 Power of standard and augmented binary methods for different sample sizes



[1]Wason J. and Seaman S. (2013) Statistics in Medicine, 32: 4639-4650
[2]Wason J, Jenkins M. (2016) Rheumatology 55:1796-1802
[3]McMenamin et al (2021) Statistical Methods in Medical Research 30(3) 702–716

Case studies

- In Newcastle we are now running four trials that are utilising the augmented binary method prospectively as the primary analysis.
- Three are CTIMPs and one is an ATIMP, all regulated by MHRA*.

*none of these trials are to be used in product filing and so the amount of regulatory scrutiny is much lower in such trials

E-CLAD

Participants	People with Chronic Lung Allograft Dysfunction following lung transplant
Intervention	Extracorporeal photopheresis (ECP)
Comparator	Standard of care
Outcome	Lung function stabilisation at 24 weeks: 1) $\leq 10\%$ decrease in FEV1; 2) $\leq 10\%$ decrease in FVC; 2) No progressive CLAD at 12 weeks
Augmented binary effect	Trial was powered for 80% power using Binary endpoint (n=90); we argued that power using augmented approach would be >90% as per funder requirement.

OPERA

Participants	People with Primary Biliary Cholangitis with non-response risk score > 50%*
Intervention	Obeticholic acid
Comparator	Ursodeoxycholic acid (standard first-line therapy)
Outcome	Biochemical remission, defined as normalisation of liver biochemistry (ALP and bilirubin) at 6 months.
Augmented binary effect	Used registry data to simulate efficiency gain achieved, allowing reduction of sample size from 156 to 106.

*this is a risk score that was derived to predict the probability of not responding to first-line therapy

RIGD 2

Participants	Young people (12-24 years old) with newly diagnosed Graves Disease
Intervention	Standard antithyroid drug treatment (ATD) + 1 dose of 500mg rituximab
Comparator	Standard ATD + placebo
Outcome	Remission at 36 months (not requiring ATD, receiving thyroidectomy), non-supressed TSH concentration and FT3 below laboratory reference range.
Augmented binary effect	Used data from previous (single-arm) phase II to investigate efficiency increase: allowed reduction in sample size from 180 to 124

Methormab

METH^or^MAB

Newcastle University

Participants	People with moderate psoriasis (PASI score 5-10)
Intervention	Adalimumab
Comparator	Methotrexate
Outcome	Reduction of 75% in PASI score at 16 weeks with no topical steroid use
Augmented binary effect	Funding panel was concerned by optimistic effect size so we said that we would switch to using augmented binary method as primary analysis to allow us adequate power to detect smaller differences.

Software implementation

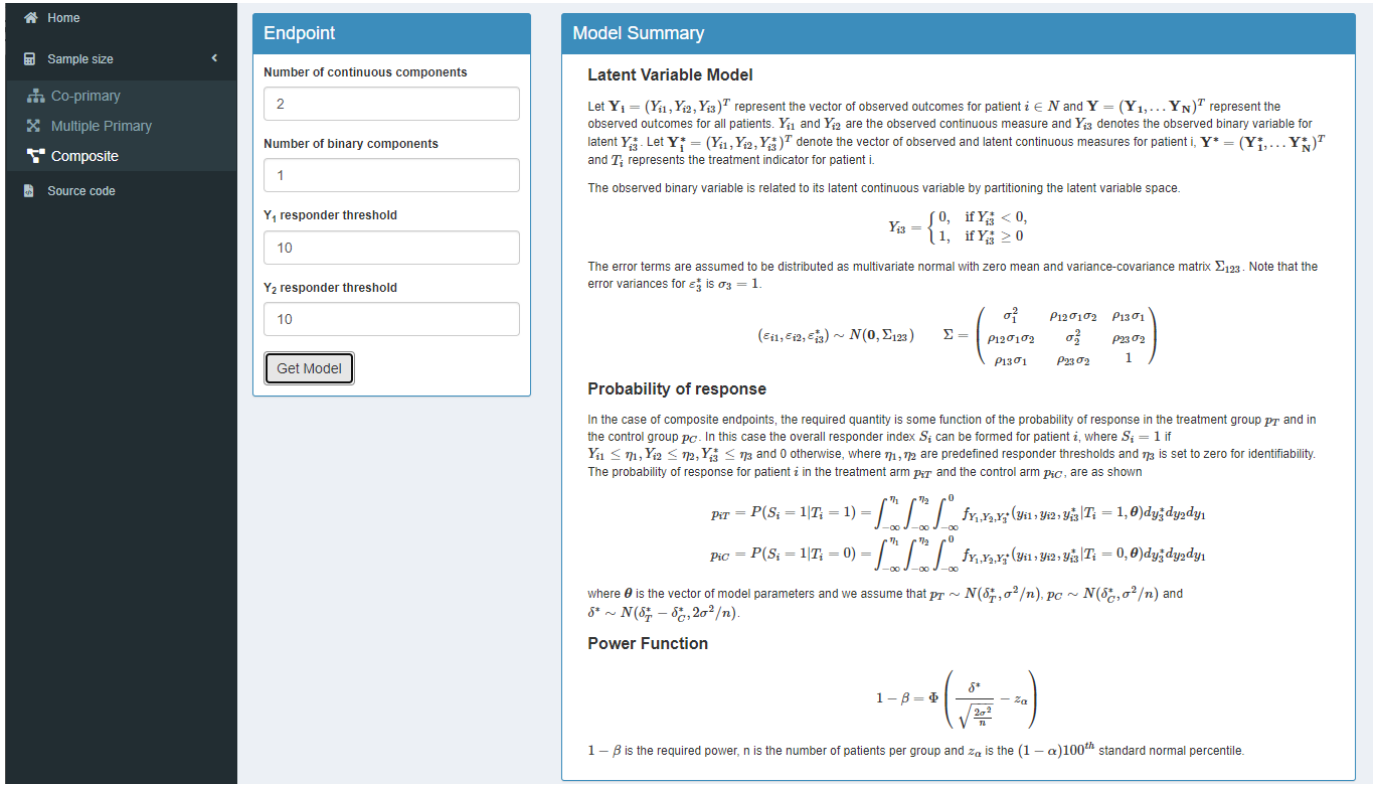
McMenamin *et al. BMC Rheumatol* (2021) 5:54
https://doi.org/10.1186/s41927-021-00224-0

BMC Rheumatology

SOFTWARE Open Access

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

Martina McMenamin[1,2*], Michael J. Grayling[3], Anna Berglind[4] and James M. S. Wason[1,3]



https://martinamcm.shinyapps.io/multsampsize/

Future work

- Several diseases use time-to-event outcomes where the event includes a continuous component, e.g.:

1. Multiple Sclerosis: time to progression on EDSS
2. Lupus: time to flare (although depends on definition of flare)
3. Rheumatoid arthritis: time to sustained remission.

- Further work on extending methods to analyse these types of endpoints would be valuable.
- Use of sample size re-estimation approaches when no pilot data is available to choose $\hat{\sigma}^2$.