Supplementary materials

Table S1 The details of search strategy.

Table S2 Characteristics of included studies.

Table S3 The details of included responder analysis methods.

Table S4 The characteristics of included studies.

Table S5 Reporting quality of included studies.

Table S6 The details of simulation performance.

9 Table S1 Characteristics of included studies.

Study	Country	Methods investigated	No. of trials simulated	Clinical area	Sample size of trials	No. of simulation replicates	Outcomes	Thresholds	Criteria to assess the performance of simulation
Peacock 2012	UK	Standard binary method Distributional approach	4	Low birth weight in pregnant women with or without smoking	1254	1000	Birth weight	The percentage of babies with low birth weight (birth weight < 2500 g)	Bias, Coverage, Power
				Low birth weight in pregnant women with different age	587				
				Low birth weight in pregnant women with or without urinary tract infection	44379				
				Low birth weight in pregnant women with or without drug use	906				
Garofolo 2013	USA	Adjusted baseline severity analyses Unadjusted baseline severity analyses	1	Hyperglycemic acute ischemic stroke	1400	1000	“success” and “failure”	For unadjusted analyses, scores of 0 to 1 (or 0 to 2) were considered to be “successes” while scores greater than 1 (or 2) were considered to be “failures,” For adjusted analyses, subjects in a less severe prognosis group at baseline must achieve a better outcome to be considered a trial "success,”.	Type-I error, Power

Wason 2013	UK	Standard binary method Augmented binary method	1	Cancer patients with capecitabine	106	NR	Probability of treatment success	we define a patient as successful if no toxicity or death occurs, no new lesions develop and the tumor size shrinkage between baseline and the final observation is greater than 30%.	Coverage, Power Precision or variance Type-I error
Sauzet 2015	Germany	Skew-normal distribution method Normal distribution method	4	Proportion of low birth weight women with or without smoking	876	20000	Low birth weight	The percentage of babies with low birth weight (birth weight < 2500 g)	Bias, Precision or variance
				Proportions of high blood pressure in white or non-white patients	1896		High blood pressure	NR	
				Proportions of obesity primipari or multipari	1781		Obesity	The usual threshold of obesity is 30 kg/m^2.	
				Proportions of preterm primipari or multipari	1730		Preterm	Preterm means gestational age (GA) is under 37 completed weeks	
Jiang 2016	USA	Advanced nonasymptotic Bayesian (ANB) approach Simple asymptotic Bayesian (SAB) approach Simple nonasymptotic bayesian (SNB)	1	Aerobic exercise program for pain	145	1000	Pain	Difference in pain (DP) defined as the initial pain scores minus two times the pain scores at the end of the study	Coverage, Power, Precision or variance

		method Traditional beta-binomial (TBB) method							
Wason 2016	UK	Standard binary method Augmented binary method	1	Rheumatoid arthritis	918	5000	ACR DAS28	ACR score was greater than or equal to the relevant threshold (20, 50 and 70) at 24 weeks. DAS28 score at 24 weeks was below an absolute threshold (2.6 or 3.2)	Type-I error, Power, Coverage
Zhang 2016	USA	Standard binary method Model-based approach	1	Parkinson's disease	NR	10000	UPDRS	UPDRS scores at 1-, 2-, and 3-months post randomization are consistently below the baseline score.	Bias, Precision or variance

UPDRS, the Unified Parkinson's Disease Rating Scale; ACR, the American College of Rheumatology; DAS28, the Disease Activity Score in 28 joints; NR, Not reported.

Table S2 The details of search strategy.

Embase <1974 to 2023 October 09>
#1 ("Responder* analysis" or "Responder* analyses" or "Response* rate*" or "Dichotomi*" or "dichotomous" or "Distribution* approach*" or "Minimal clinically important difference*" or "Minimal important difference*" or "Minimum clinically important difference*" or "Minimum important difference*").ti. **6,365**
#2 (Dichotomi* or dichotomous or Binary or Categor*).ti. **34,114**
#3 Continuous.ti. **94,997**
#4 2 and 3 **394**
#5 1 or 4 **6,692**
Ovid MEDLINE(R) ALL <1946 to October 09, 2023>
#1 ("Responder* analysis" or "Responder* analyses" or "Response* rate*" or "Dichotomi*" or "dichotomous" or "Distribution* approach*" or "Minimal clinically important difference*" or "Minimal important difference*" or "Minimum clinically important difference*" or "Minimum important difference*").ti. **4,394**
#2 (Dichotomi* or dichotomous or Binary or Categor*).ti. **32,428**
#3 Continuous.ti. 81,139
#4 2 and 3 **371**
#5 1 or 4 **4,698**
PUBMED
#1 (((((((("Responder* analysis"[Title]) OR ("Responder* analyses"[Title])) OR ("Response* rate*"[Title])) OR ("Dichotomi*"[Title])) OR ("dichotomous"[Title])) OR ("Distribution* approach*"[Title])) OR ("Minimal clinically important difference*"[Title])) OR ("Minimal important difference*"[Title])) OR ("Minimum clinically important difference*"[Title])) OR ("Minimum important difference*"[Title]) **3,326**
#2 (((Dichotomi*[Title]) OR (dichotomous[Title])) OR (Binary[Title])) OR (Categor*[Title]) **32,406**
#3 Continuous[Title] **81,102**
#4 ((((Dichotomi*[Title]) OR (dichotomous[Title])) OR (Binary[Title])) OR (Categor*[Title])) AND (Continuous[Title]) **371**
#5 (((((Dichotomi*[Title]) OR (dichotomous[Title])) OR (Binary[Title])) OR (Categor*[Title])) AND (Continuous[Title])) OR ((((((((("Responder* analysis"[Title]) OR ("Responder* analyses"[Title])) OR ("Response* rate*"[Title])) OR ("Dichotomi*"[Title])) OR ("dichotomous"[Title])) OR ("Distribution* approach*"[Title])) OR ("Minimal clinically important difference*"[Title])) OR ("Minimal important difference*"[Title])) OR ("Minimum clinically important difference*"[Title])) OR ("Minimum important difference*"[Title])) **3,631**
Web of Science Core Collection
#1 "TI=(""Responder* analysis"") OR TI=(""Responder* analyses"") OR TI=(""Response* rate*"") OR TI=(""Dichotomi*"") OR TI=(""dichotomous"") OR TI=(""Distribution* approach*"") OR TI=(""Minimal clinically important difference*"") OR TI=(""Minimal important difference*"") OR TI=(""Minimum clinically important difference*"") OR TI=(""Minimum important difference*"") **9187**
#2 "TI=(""binary"") OR TI=(""categor*"") OR TI=(""Dichotomi*"") OR TI=(""dichotomous"") **167384**
#3 "TI=(""continuous"") **207479**

#4	"#3 AND #2 **994**
#5	"#4 OR #1 **10080**
#6	"#4 OR #1 and 2.1 Synthesis or 9.92 Statistical Methods or 9.162 Numerical Methods or 9.28 Pure Maths or 6.277 Asian Studies or 7.57 Modelling & Simulation or 7.63 Mechanics (Citation Topics Meso) **707**

Table S3 The details of included responder analysis methods.

Method	Definition	Components Considered	Statistical Model Used	Adjustment for Covariates
Standard binary method	Treats the overall composite endpoint as a binary outcome (responder vs. non-responder).	- Binary outcome (responder vs. non-responder) - Withdrawal or rescue therapy (treated as non-responders)	Logistic regression model	- Baseline - Treatment arm
Augmented binary method	Combines continuous and binary components to estimate the probability of response between treatment arms.	- Continuous outcome score - Withdrawal or rescue therapy (treated as non-responders)	- Continuous GEE model for outcome scores - Two logistic regression models for withdrawal / rescue therapy	- Baseline - Treatment arm - Outcomes - Additional covariates if needed
Distributional method	Uses the underlying continuous data distribution to estimate responder proportions without dichotomization.	- Mean difference in continuous outcome - Proportion below a predefined threshold (cut-point)	- Assumes normal distribution for both groups - Uses sample means and standard deviation to estimate proportions below a cut-point - CI for difference in means translated to CI for difference in proportions	- Baseline - Treatment arm
Model-based method	Estimate probability of response using a model fitted to original continuous/discrete data.	One or more continuous/discrete variables	Logistic regression, GLM, or other parametric models	- Baseline - Treatment arm - Additional covariates if needed

Table S4 Reporting quality of included studies.

Study	Aims of the simulation	Dependence of samples	Starting seed	Random number generator	Failures occur during simulation	Software to perform simulations	Software to perform analysis	Justification for data generation	Scenarios and statistical methods evaluated	Number of simulations	Any justification for number of simulations	Criteria to evaluate the performance of statistical methods
Peacock 2012	1	1	0	0	0	1	1	1	1	1	0	1
Garofolo 2013	1	1	0	0	0	1	1	1	1	1	0	1
Wason 2013	1	1	0	0	0	0	0	1	1	0	0	1
Sauzet 2015	1	1	0	0	0	1	1	1	1	1	0	1
Jiang 2016	1	1	0	0	0	1	1	1	1	1	0	1
Wason 2016	1	1	0	0	0	1	1	1	1	1	0	1
Zhang 2016	1	1	0	0	0	1	1	1	1	1	0	1

Table S5 Reporting quality of included studies.

Study	Simulation performance					
	Bias	**Precision or variance**	**Accuracy**	**Type-I error**	**Power**	**Coverage**
Standard binary methods vs. other responder analysis methods						
Wason 2016				Not favor augmented binary method	Favor augmented binary method	Favor augmented binary method
Wason 2013 - Simulation data		Favor augmented binary method		Favor augmented binary method	Not favor augmented binary method	
Wason 2013 - Trial data		Favor augmented binary method				Favor augmented binary method
Peacock 2012 - Trial 1 data					Favor distributional approach	Favor distributional approach
Peacock 2012 - Trial 2 data	Favor distributional approach				Favor distributional approach	Favor distributional approach
Peacock 2012 - Trial 3 data	Favor distributional approach				Favor distributional approach	Favor distributional approach
Peacock 2012 - Trial 4 data	Not favor distributional approach					Favor distributional approach
Zhang 2016	Favor model-based approach	Favor model-based approach				
Responder analysis for different data distribution types						
Sauzet 2015 - normal distribution real data set	Not favor normal distribution method					Not favor normal distribution method
Sauzet 2015 - lognormal distribution real data set	Not favor normal distribution method					Not favor normal distribution method
Sauzet 2015 - Inverse transformation normal distribution real data set	Not favor normal distribution method					Not favor normal distribution method
Sauzet 2015 - left skewed normal distribution real	Not favor normal distribution method					Not favor normal distribution method

data set						
Sauzet 2015 – simulation for different sample size from 20-500, skewness from ±1-±20, Log standard deviation from 0.02-1	Not favor normal distribution method					
Responder analysis with or without adjusted covariates						
Garofolo 2013				Not favor adjusted baseline severity analyses	Not favor adjusted baseline severity analyses	
Different Bayesian responder analysis methods						
Jiang 2016 - Simulation data		Not favor SAB			Favor SAB	
Jiang 2016 - Trial data		Favor ANB			Favor ANB	Favor ANB

Table S6 The details of simulation performance.

Study	**Objective**	**Simulation performance**						**Author's conclusions on performance**
		Bias	**Precision or variance**	**Accuracy**	**Type-I error**	**Power**	**Coverage**	
Wason 2016	To demonstrate whether the method does not inflate the type I error rate and show that it substantially improves the power of analyses.				There is no evidence of type I error rate inflation for the augmented binary method.	The augmented binary method provides a large gain in power for the example data set and end point across the different sample sizes considered.	The augmented binary method has the reduction in CI width, and gives an estimate closer to its mean from across all the replicates.	By using our proposed augmented binary method, the response rates for ACR and DAS28 end points could be estimated to the same degree of precision (i.e. same CI width) as the standard binary approach but using a much smaller sample size. The method does not cause any inflation to the type I error rate
Peacock 2012 – Trial 1	To propose a dual approach that analyses continuous data using both means and proportions to replace dichotomisation alone and that may be useful in certain situations.					For a power of 80%, the sample size required to obtain a significant difference in distributional method is smaller than thresholds method	The distributional method gives a CI that is nearly a third of the width of that of thresholds method	In analyzing data, precision as well as power can be an issue and that, using the distributional approach, tighter CIs for differences in proportions, relative risks and odds ratios can be computed where the data are intrinsically continuous.
Peacock 2012 – Trial 2		Bias for the estimated difference in				The estimated power is	The thresholds method calculated CI is	

		proportion using distributional method (0.18) is smaller than thresholds method (0.32)				increased from 62% (thresholds method) to almost 100% (distributional method).	on average twice as wide as the distributional one. The coverage of the 95% distributional CI as the percentage of cases that did contain the 'true' difference in proportion 0.062 was 94%.	
Peacock 2012 – Trial 3		The bias for distributional estimated odds ratio is 0.042 compared with a mean bias of 0.048 for thresholds method				The estimated power is increased from 20% (thresholds method) to almost 79% (distributional method).	The thresholds method calculated CI is on average twice as wide as the distributional one. The coverage of the 95% CI was exactly 95% for distributional method.	
Peacock 2012 – Trial 4		The bias for distributional estimated odds ratio is 0.138 compared with a mean bias of					The thresholds method calculated CI is on average twice as wide as the	

		0.071 for thresholds method					distributional one. The coverage of the 95% CI was 92% for distributional method.	
Jiang 2016 - simulations	The purpose of this article is to develop a Bayesian distributional methodology that not only retains statistical precision but also reflects the true uncertainty.		The SAB method with correction tends to overestimate the true variation when cutoff points are far away from zero no matter what the sample size is.			The SAB approach always deflates the true posterior uncertainty even when sample size is very large.		The SAB approach does improve the precision over usual methods for dichotomized data, but it does not reflect the true uncertainty especially when the threshold is in the tails of distribution.
Jiang 2016 – trial data			The posterior variance of log(or) for ANB approach (i.e., 0.161) is approximately 40% smaller than the one based on the TBB approach (i.e., 0.262).			the TBB approach fails to detect the difference between two groups because its posterior credible interval includes 1. With the use of the ANB method, however, the	The 95% CI of TBB is nearly twice as wide as 95% CI of ANB	

						difference is significant although it is small		
Zhang 2016	To elaborate on advantages of model based method and illustrate them with a series of simulation studies mimicking a study of Parkinson's disease, which involves longitudinal continuous data in the definition of a responder.	The empirical estimator is known to be unbiased in any finite sample, while the model-based estimators may have a finite-sample bias that decreases to 0 with increasing sample size. We can reduce the sample size while maintaining the same precision by using the model-based estimator instead of the empirical estimator.	For parameter estimation, the model-based approach does require real knowledge and careful modeling of the distribution of X.					The model-based approach can be more efficient, and more effective for dealing with missing data, than the usual approach based on dichotomization. For parameter estimation, the model-based approach generally requires correct specification of the distribution of X.
Wason 2013 - simulation	We demonstrate these increases in precision and power using simulated data. We also apply augmented binary method to real data from a phase II cancer trial and		The augmented binary approach improves the precision of the estimated probability of success.		The deviation from 0.05 is generally greater for the logistic regression than for the augmented binary approach	The augmented binary approach is not with the highest power.		The augmented binary approach can be used to increase the power to detect a difference between the probabilities of success under two different treatments in a comparative trial.
Wason 2013 - trial			The augmented binary method				The augmented binary method	

	show that it results in a considerably narrower confidence interval for the probability of tumour response.		can change the estimate of the success probability considerably in some cases				gives reductions in the width of the CI in all cases.	
Garofolo 2013	While the bias decreases with increasing sample size, the adjusted estimates of the treatment effect parameter are consistently less biased than the unadjusted estimates.				The type I error rates for both the unadjusted and categorically-adjusted methods are within the 95% confidence limits for all the sample sizes, hovering close to the nominal 5% level.	There is no drastic difference in the unadjusted and categorically-adjusted methods with respect to power in these varying treatment effect scenarios.		Under various treatment effect settings, the operating characteristics of the unadjusted and adjusted analyses do not substantially differ. Power and type I error are preserved for both the unadjusted and adjusted analyses.
Sauzet 2015 – normal distribution real data set	To investigate if the distributional method remains reliable in the case of deviations from normality and propose a generalisation of the distributional method to allow for skewness in distributions using the skew-normal	Normal distribution method is similar with skew-normal distribution method					Normal distribution method is similar with skew-normal distribution method	The distributional method with its applicability for common skewed data allows researchers to provide both continuous and dichotomised estimates without losing information or precision. This will have the effect of providing a practical
Sauzet 2015 – lognormal distribution real data set		Normal distribution method is similar with skew-normal distribution method					Normal distribution method is similar with skew-normal distribution	

	distribution.						method	understanding of the difference in means in terms of proportions.
Sauzet 2015 – Inverse transformation normal distribution real data set		Normal distribution method is similar with skew-normal distribution method					Normal distribution method is similar with skew-normal distribution method	
Sauzet 2015 – left skewed normal distribution real data set		Normal distribution method is similar with skew-normal distribution method					Normal distribution method is similar with skew-normal distribution method	
Sauzet 2015 – simulation for different sample size from 20-500, skewness from ±1-±20, Log standard deviation from 0.02-1		For data almost normal the skew normal method did not perform well unless the sample size was large enough.						