BMJ Open

Target and actual sample sizes for studies from two trial registries from 1999 to 2020: an observational study

Adrian Gerard Barnett [1], Paul Glasziou [2]

To cite: Barnett AG, Glasziou P. Target and actual sample sizes for studies from two trial registries from 1999 to 2020: an observational study. *BMJ Open* 2021;**11**:e053377. doi:10.1136/bmjopen-2021-053377

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2021-053377).

Received 12 May 2021
Accepted 23 November 2021

[1]Australian Centre for Health Services Innovation and Centre for Healthcare Transformation, School of Public Health & Social Work, Queensland University of Technology, Brisbane, Queensland, Australia
[2]CREBP, Bond University, Robina, Queensland, Australia

Correspondence to
Adrian Gerard Barnett;
a.barnett@qut.edu.au

ABSTRACT

Objectives To investigate differences between target and actual sample sizes, and what study characteristics were associated with sample sizes.

Design Observational study.

Setting The large trial registries of clinicaltrials.gov (starting in 1999) and ANZCTR (starting in 2005) through to 2021.

Participants Over 280 000 interventional studies excluding studies that were withheld, terminated for safety reasons or were expanded access.

Main outcome measures The actual and target sample sizes, and the within-study ratio of the actual to target sample size.

Results Most studies were small: the median actual sample sizes in the two databases were 60 and 52. There was a decrease over time in the target sample size of 9%–10% per 5 years, and a larger decrease of 18%–21% per 5 years for the actual sample size. The actual-to-target sample size ratio was 4.1% lower per 5 years, meaning more studies (on average) failed to hit their target sample size.

Conclusion Registered studies are more often under-recruited than over-recruited and worryingly both target and actual sample sizes appear to have decreased over time, as has the within-study gap between the target and actual sample size. Declining sample sizes and ongoing concerns about underpowered studies mean more research is needed into barriers and facilitators for improving recruitment and accessing data.

Strengths and limitations of this study

- ► All analyses were repeated using two trial registries.
- ► The registries had very large sample sizes with little missing data.
- ► The registry data are completed by researchers and have some data entry errors and poor reporting.
- ► There were changes over time in the types of studies of registered, so differences in sample size over time should be interpreted in light of these changes.

INTRODUCTION

Sample size is a key element of most research study designs. Researchers should aim to collect a large enough sample to answer their research question with a good statistical power, for example, recruiting a sufficient number of patients to demonstrate a hypothesised difference in efficacy between two treatments. However, researchers do not want to collect more data than necessary as this wastes time and resources.

The target sample size should be estimated at the study design stage. Researchers then collect data until that target is achieved or until they run out of time or money. This sounds straightforward, but in practice many studies struggle to recruit their target sample size and difficulties with recruitment are a common reason why trials end early.[1–3] Recruiting sufficient participants is crucial to a trial's validity, and in recognition of the difficulties around trial recruitment there is large and ongoing research effort aimed at increasing recruitment and retention.[4 5]

Inadequate sample sizes mean studies are underpowered and so true associations may be missed or estimated with large uncertainty. Theoretical work has shown how underpowered studies contribute to the ongoing problem of poor quality research.[6 7] Generally, larger sample sizes are needed to tackle the pervasive problem of studies with low power,[8] although small samples are often appropriate for pilot or feasibility studies.

Sample size calculations depend on a range of assumptions that should reflect current knowledge. The practical application of these assumptions has been criticised in terms of a general lack of understanding of uncertainty, and the approach of reverse-engineering assumptions to get a desired target sample size.[9 10]

In this paper, we examine sample sizes using two large trial registries containing information on health and medical studies. We examined the difference between the target and actual sample size, what study characteristics were associated with sample size, and if sample sizes have declined over time. The

aim is to contribute to the ongoing work on improving study designs and the quality of research.[11]

METHODS

Trial registries

Trial registries were introduced to counter the serious problem of unreported trials.[12] Trials cannot now be published in any high-profile medical journal without a prospective registration, hence there has been good uptake of trial registries, although they have not eliminated the problem of unreported trials or poorly reported trials.[13–16] For our purposes, the high uptake of registries provides a large and comprehensive data set to study sample sizes.

Trial registries contain details on the study characteristics, including the study design, disease(s), outcome(s), key dates and funding. Researchers are responsible for posting and updating their studies.

We downloaded data from two large trial registries:

- Australian New Zealand Clinical Trials Registry (ANZCTR) started in 2005.
- clinicaltrials.gov run by the US National Library of Medicine started in 1999 and publicly available in 2000.

ANZCTR was chosen because of the authors' familiarity with the region, and clinicaltrials.gov was chosen because it is the largest international registry. Both registries make their data available for research.

Ethics approval

All the data are publicly available and do not involve human participants, hence this study did not require ethics approval.

Inclusion and exclusion criteria

We included studies from interventional studies and did not include observational studies. This is because these two study types are unlikely to be comparable and there were many study characteristics (eg, blinding) that are not applicable for observational studies. Interventional studies are those where participants were prospectively assigned to one or more health-related interventions in order to study the intervention's effects.

We excluded a small number of retrospectively registered trials from ANZCTR before the registry started in 2005, and a small number that were missing the date the study was submitted to ANZCTR (details below).

We excluded studies from clinicaltrials.gov that had a status of 'withheld' because the available data for these studies were limited. We excluded studies that were terminated for safety reasons as they may have achieved their objective using a smaller sample size than planned. We excluded expanded access studies because we were not certain that these were comparable to interventional studies. We excluded studies where the type of sample size was not stated, as we had to know if the sample size was the target or actual. We excluded two studies that used a dummy sample size, for example, '9 999 999'. To avoid double-counting, we excluded clinicaltrials.gov studies if they had an ANZCTR number. We preferenced data from ANZCTR as it had more detailed information on sample size. The exclusions are shown in online supplemental figure 1.

We included all the available studies that met our inclusion/exclusion criteria and did not use a sample size calculation or formal hypothesis testing.

Data for both registries were downloaded on 1 February 2021 in XML format and then read into R (V.4.0.3).[17] Updated sample size data for clinicaltrials.gov were downloaded on 5 March 2021. All the code to replicate the data extraction and analyses, and data are openly available on GitHub (https://github.com/agbarnett/registries).[18] Results are reported using the Strengthening the Reporting of Observational Studies in Epidemiology guidelines for observational studies.[19]

Statistical methods

Models of sample size

Both registries had two measures of sample size: the target and actual. We used multiple regression to estimate what study characteristics were associated with the target and actual sample size. See online supplemental table 1 for the list of available study characteristics which differed by registry. For the models of actual sample size, we included the study status (eg, 'completed') as an independent variable, but we did not include study status for the models of target sample size as study status occurred after the target sample size and so any association could not be causal.

The clinicaltrials.gov database does not include a variable for whether studies are longitudinal. Hence, we searched each study's description for 'longitudinal' in order to extract this study design variable. We also searched for 'adaptive' or 'platform' trial to examine whether these study designs impacted sample sizes.[20]

Sample size had a strong positive skew with a small number of very large studies. To improve model fit and reduce the influence of a few very large studies, we log-transformed sample size (base e). We therefore present the effects of the study characteristics as the percent change in the geometric mean instead of the absolute difference in sample size.

Some study characteristics had a strong positive skew with a small proportion of very large numbers, for example, the number of primary outcomes (median 1, maximum 214 for clinicaltrials.gov). To reduce the potential for a few large studies to overly-influence the results, we log-transformed these variables using base 2, hence the parameters are the percent change in the sample size when the variable is doubled.

Most of the variables we used were mandatory, meaning researchers had to complete them and hence there was little item-missing data. The most amount of missing data was 2% for study purpose. For non-mandatory categorical variables with missing data, we included 'Missing' as its own category. Our reasoning was that investigators likely did not complete a question if they felt it was not relevant to their study and hence 'Missing' should be akin to 'Not

Table 1 Selected characteristics of the included studies from the two trial databases

Categorical variables, n (%)			
Variable	**Categories**	**ANZCTR**	**clinicaltrials.gov**
Study status	Active, not recruiting	1042 (6)	13 370 (5)
	Completed	7175 (41)	149 721 (55)
	Enrolling by invitation	–	2070 (1)
	Not yet recruiting	4259 (24)	13 659 (5)
	Recruiting	4072 (23)	39 237 (14)
	Stopped early	581 (3)	–
	Suspended	78 (0)	1393 (1)
	Terminated	–	16 813 (6)
	Unknown status	–	27 878 (10)
	Withdrawn	303 (2)	8019 (3)
Gender	All	14 882 (85)	231 998 (85)
	Female	1698 (10)	26 451 (10)
	Male	926 (5)	13 710 (5)
	Missing	4 (0)	1 (0)
Continuous variables, median (IQR)			
Variable		**ANZCTR**	**clinicaltrials.gov**
Year submitted		2014 (2011–2017)	2015 (2010–2018)
Number of primary outcomes		1 (1–2)	1 (1–2)
Number of secondary outcomes		3 (1–6)	2 (1–5)
Target sample size		66 (31–159)	70 (35–176)
Actual sample size		60 (26–140)	52 (22–139)

applicable.' This avoided excluding studies with small amounts of missing data. Details on the item-missing data are in online supplemental appendix 1.

We used the elastic net method to select the key variables from the larger subset of all variables.[21] We used 10-fold cross-validation to select the ideal penalty and hence which variables were included in the final model. We used a parsimonious model by choosing the penalty within one SE of the minimum cross-validated mean square error.

We checked the variance inflation factor of the final models to detect collinearity using a threshold of five. We checked the residuals of the final model to verify they were unimodal, approximately symmetric, and with no large outliers.

Target versus actual sample size

We calculated the sample size ratio of the actual divided by target and created a histogram of the ratio. We described

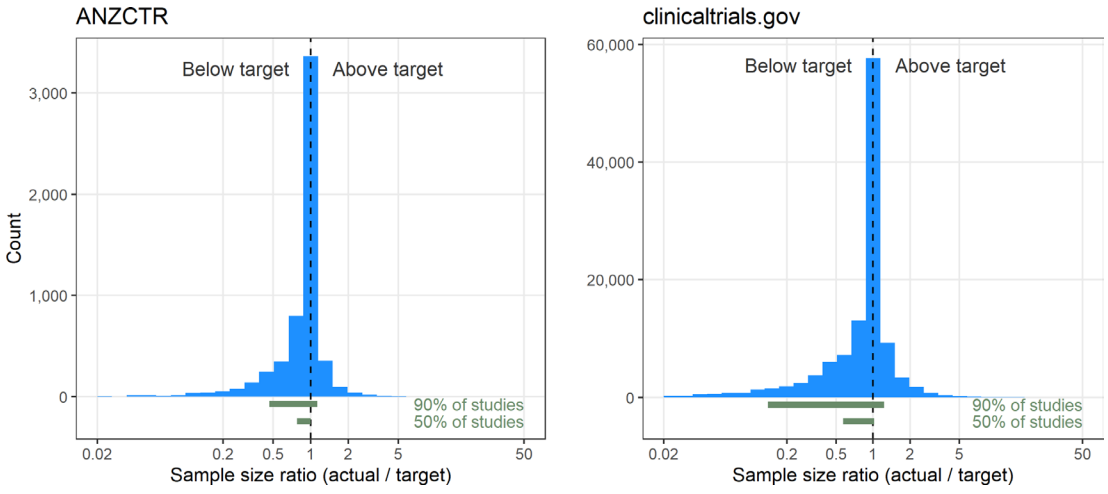


Figure 1 Histogram of the ratio of the actual-to-target sample size. The X-axes are on a log scale.

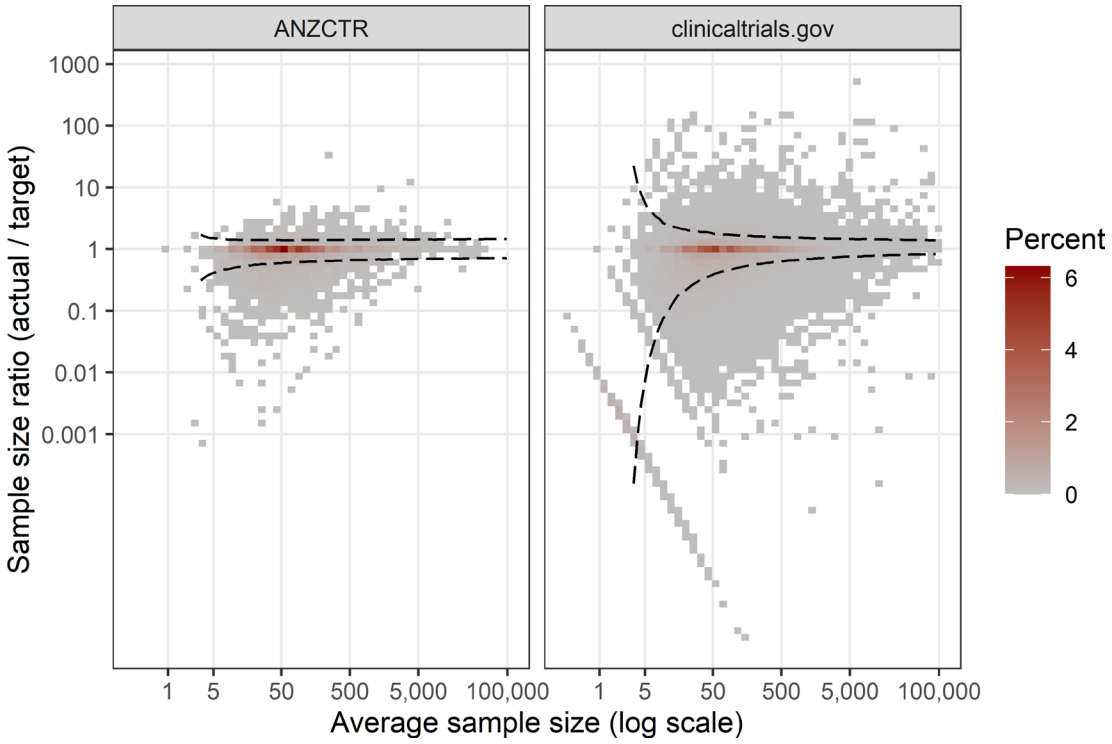


Figure 2 Bland–Altman tile plot of the sample size ratio against the average sample size. Both axes are on a log scale. Fifty-two studies (0.04%) with an average sample size over 100 000 were excluded from the plot as otherwise the plot area was compressed for most studies. The dotted lines are the estimated 95% limits of agreement using a t-distribution.

the range of this ratio using the central 50% and 90% of studies.

To estimate what study characteristics were associated with the sample size ratio, we used the same elastic net method as for the models of sample size. The ratio had a strong positive skew, so it was log-transformed (base e) for the modelling.

We used a Bland–Altman plot of the actual-to-target sample size against the average sample size ((actual +target)/ 2). The aim was to see whether the ratio narrowed for small and/or large sample sizes. We log-transformed (base e) the ratio because of the strong positive skew in sample sizes. Because of the very large sample size, a standard Bland–Altman scatter-plot using individual studies was too cluttered, hence we used a tile plot to summarise studies in bins.

We used the Bland–Altman limits of agreement to show the range in observed ratios that covers 95% of the data. However, the standard limits assume that the ratio is constant for all sample sizes which did not appear valid for these data. Hence, we used a Bayesian model and allowed the mean and variance of the limits of agreement to vary by the average sample size using a fractional polynomial approach with the eight powers: $\{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$[22] (see online supplemental appendix 2). We fitted 64 (8×8) separate models to cover all combinations for the mean and variance, and selected the best model using the deviance information criterion (DIC)[23] (see online supplemental figure 2). Because the ratio distribution had long tails, we used a t-distribution with 4 degrees of freedom instead of a Normal distribution, and this gave a far better fit to the data (DIC improvement of over 4000). For the clinicaltrials.gov data, we fitted these Bayesian models using a random sample of 10 000 studies (8% of the total) because of the time needed for the Markov chain Monte Carlo estimates.

The Bayesian models were fitted using the JAGS software (V.4.3.0).[24] We used vague Normal priors for all parameters. We used two chains thinned by three with a burn-in and sample of 2000. We visually checked the convergence and mixing of the chains (see online supplemental figure 3).

Patient and public involvement

No patients or members of the public were involved in the design, conduct or reporting of this study.

RESULTS

The number of included studies and reasons for exclusion are shown in online supplemental figure 1. The final analyses had 17 510 studies from ANZCTR and 272 160 from clinicaltrials.gov.

Some basic characteristics of the included studies are in table 1. The median target sample size was 66 for ANZCTR and 78 for clinicaltrials.gov. The median actual sample size was 60 for ANZCTR and 52 for clinicaltrials.gov for both databases. Additional summary statistics on the two databases are in online supplemental appendix 1.

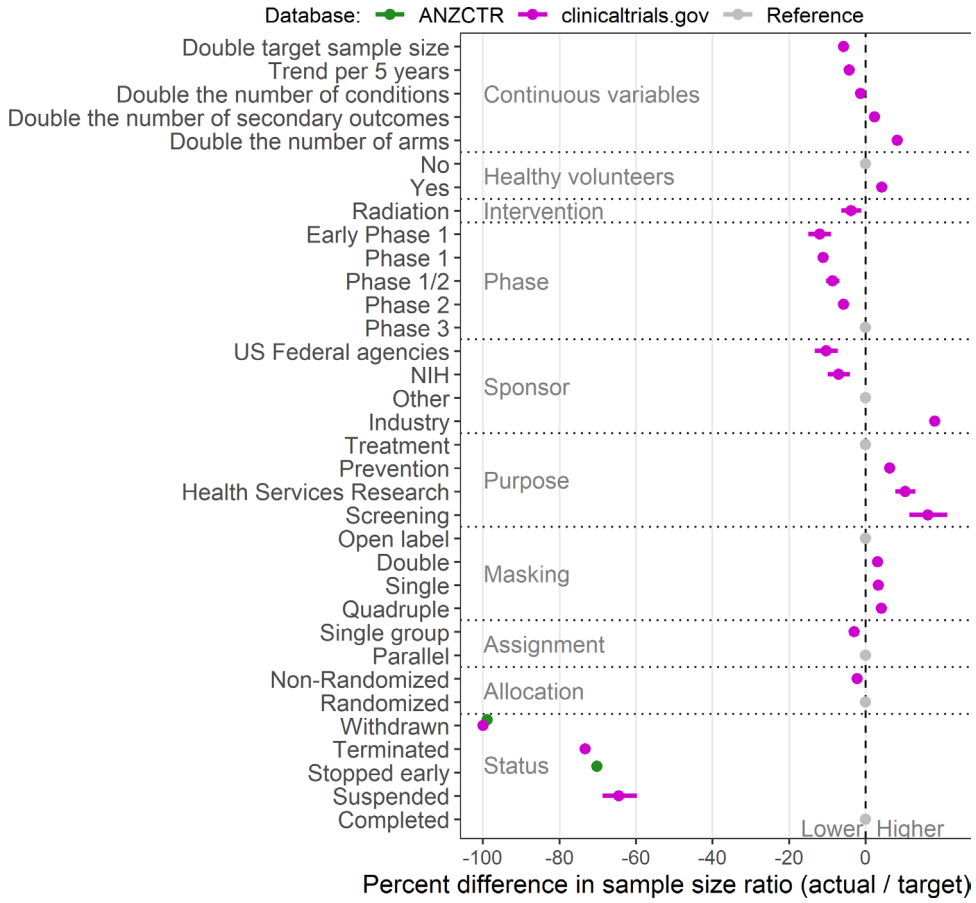


Figure 3 Percent changes in the actual-to-target sample size ratio. The dots are means and the solid horizontal lines are 95% CIs.

Target versus actual sample size

The number of studies with a target and actual sample size was 5712 in ANZCTR and 121 603 in clinicaltrials.gov.

The histograms of the ratios of the actual-to-target sample sizes are in figure 1. Many studies hit their target and a large proportion were also just below their target. The histograms are asymmetric around 1, with a larger 'shoulder' of studies missing their target compared with studies exceeding the target.

For ANZCTR, the central 50% of studies had a ratio of between 22% below their target to equalling their target. The central 90% of studies had a ratio of between 53% below their target to 13% above their target. For clinicaltrials.gov, the central 50% of studies had a ratio of 43% below their target to 2% over target. The central 90% of studies had a ratio of 86% below their target to 23% over target.

The Bland–Altman plot of the sample size ratio against the average sample size is in figure 2. Many studies with an average sample size between 10 and 200 hit their target sample size. The estimated limits of agreement narrowed for larger sample sizes in both databases.

For the ANZCTR data, the 95% limits of agreement for the sample size ratio were 0.58–1.38 for an average sample size of 50, narrowing slightly to 0.64–1.39 for an average sample size of 500. There are a small number of studies that are far above or below the limits of agreement, particularly studies in the 5–500 sample size range that were well below their target.

The 95% limits of agreement were generally wider for the clinicaltrials.gov data. The 95% limits of agreement were 0.37–1.84 for an average sample size of 50, narrowing to 0.63–1.54 for an average sample size of 500. The diagonal strip of studies in the bottom-left of the figure are studies with a small target sample size that recruited no participants.

Models of the actual-to-target sample size ratio

We used multiple variable regression to estimate what study characteristics were associated with the actual-to-target sample size ratio. The estimates are shown in figure 3, expressed as a percent change, and in online supplemental table 2.

Larger target sample sizes were associated with a lower actual-to-target ratio, meaning smaller actual sample sizes (5.7% lower per doubling of the target sample size). The actual-to-target ratio lowered over time (4.3% lower per 5 years).

Studies with more arms and more secondary outcomes were associated with a higher ratio, as were studies that included healthy volunteers.

Table 2 Percent changes in the target and actual sample size for the ANZCTR database

Group/variable*	Variable/category	Target PC (95% CI)	Actual PC (95% CI)
Continuous	Double the number of primary outcomes	–7.5 (–11.1 to –3.9)	–9.8 (–16.0 to –3.1)
	Double the number of secondary outcomes	14.2 (12.1 to 16.3)	10.3 (6.8 to 14.0)
	Number of funders	14.2 (11.5 to 16.9)	15.3 (10.4 to 20.4)
	Trend per 5 years	–9.6 (–11.7 to –7.6)	–21.1 (–25.0 to –17.1)
Gender	Females	4.3 (–1.9 to 10.8)	19.4 (7.0 to 33.2)
	Males	–27.6 (–32.7 to –22.2)	–15.9 (–26.0 to –4.4)
Maximum age limit	18+	–19.9 (–228 to –17.0)	–18.8 (–23.8 to –13.5)
	Not stated	17.5 (6.6 to 29.4)	30.3 (4.2 to 63.0)
	Under 18	–12.3 (–18.5 to –5.8)	
Minimum age limit	18+	–22.0 (–25.7 to –18.2)	–22.4 (–28.0 to –16.4)
	Missing		27.6 (–5.7 to 72.7)
	Not stated	–28.0 (–38.2 to –16.2)	
Healthy volunteers	Yes	12.3 (7.6 to 17.1)	26.2 (17.2 to 36.0)
Phase	Missing	–7.8 (–12.9 to –2.5)	–14.0 (–22.9 to –4.0)
	Phase 0	–55.1 (–65.9 to –41.0)	–73.5 (–85.3 to –52.2)
	Phase 1	–41.2 (–45.7 to –36.2)	–33.4 (–42.1 to –23.4)
	Phase 1/2	–40.6 (–47.2 to –33.1)	–40.0 (–53.4 to –22.9)
	Phase 2	–27.7 (–32.5 to –22.6)	–24.3 (–33.5 to –13.8)
	Phase 2/3	–20.5 (–30.2 to –9.5)	
	Phase 3/4	31.7 (15.1 to 50.7)	56.0 (20.5 to 102.1)
	Phase 4	4.9 (–1.6 to 11.9)	10.6 (-2.2 to 25.1)
Endpoint	Bioavailability	–48.5 (–59.0 to –35.3)	–32.2 (–53.5 to –1.2)
	Bioequivalence	32.6 (10.0 to 59.9)	60.7 (14.2 to 126.2)
	Missing	–7.9 (–13.3 to –2.1)	
	Pharmacodynamics	–27.3 (–41.1 to –10.2)	
	Pharmacokinetics	–33.9 (–43.1 to –23.2)	–16.6 (–34.9 to 6.7)
	Pharmacokinetics/pharmacodynamics	–33.4 (–44.3 to –20.5)	–23.9 (–44.6 to 4.6)
	Safety	–17.8 (–24.6 to –10.4)	
	Safety/efficacy	2.4 (–1.8 to 6.9)	9.3 (1.0 to 18.3)
Purpose	Diagnosis	14.0 (1.6 to 28.0)	12.9 (9.8 to 41.4)
	Educational/counselling/training	14.6 (7.6 to 22.1)	24.7 (11.7 to 39.2)
	Prevention	27.9 (20.9 to 35.2)	20.0 (8.6 to 32.7)
Masking	Missing	–2.2 (–11.1 to 7.5)	–12.7 (–26.5 to 3.6)
	Open	2.1 (-2.0 to 6.4)	
Assignment	Cross-over	–63.8 (–65.8 to –61.7)	–63.0 (–66.5 to –59.1)
	Factorial	24.6 (8.4 to 43.2)	39.5 (8.2 to 79.9)
	Missing	–11.1 (–19.0 to –2.3)	–7.2 (–21.6 to 9.8)
	Other	4.6 (–3.3 to 13.2)	
	Single group	–31.1 (–36.5 to –25.3)	–36.9 (–45.6 to –26.9)
Allocation	Non-randomised trial	–34.6 (–38.6 to –30.3)	–31.8 (–39.2 to –23.4)
Control	Dose comparison	–22.1 (–30.6 to –12.5)	–30.9 (–44.1 to –14.6)
	Historical	41.5 (23.9 to 61.6)	91.1 (50.3 to 142.9)
	Placebo	–9.8 (–14.0 to –5.4)	–8.7 (–15.9 to –0.8)
	Uncontrolled	–26.5 (–32.1 to –20.5)	–21.6 (–32.6 to –8.7)
Intervention	Behaviour		7.1 (–5.0 to 20.7)
	Diagnosis/prognosis	21.2 (5.8 to 38.8)	41.5 (7.8 to 85.7)
	Early detection/screening	134.0 (107.7 to 163.6)	193.5 (133.9 to 268.4)

Continued

Table 2 Continued

Group/variable*	Variable/category	Target PC (95% CI)	Actual PC (95% CI)
	Lifestyle	–22.8 (–28.6 to –16.6)	
	Missing	–31.9 (–52.8 to –1.6)	
	None	–19.8 (–31.1 to –6.6)	
	Prevention	25.4 (17.0 to 34.3)	57.1 (39.1 to 77.5)
	Rehabilitation	–39.9 (–44.3 to –35.1)	–24.2 (–33.9 to –13.2)
	Treatment: devices	–25.9 (–30.2 to –21.3)	–14.7 (–23.8 to –4.6)
	Treatment: other	–20.2 (–23.9 to –16.3)	–9.8 (–17.2 to –1.7)
	Treatment: surgery	–13.9 (–22.2 to –4.7)	
Area	Alternative and complementary medicine	–27.9 (–37.1 to –17.3)	–26.4 (–41.7 to –7.0)
	Anaesthesiology	–19.9 (–26.2 to –13.0)	12.0 (–2.4 to 28.5)
	Cancer	2.7 (–4.3 to 10.3)	
	Diet and nutrition	–17.9 (–24.6 to –10.7)	
	Emergency medicine	80.9 (29.0 to 153.8)	134.0 (24.7 to 339.1)
	Eye	–28.2 (–36.3 to –19.1)	
	Human genetics and inherited disorders	–47.0 (–57.3 to –34.1)	–37.0 (–57.2 to –7.3)
	Infection	43.5 (30.7 to 57.6)	98.7 (69.7 to 132.5)
	Inflammatory and immune system	–17.5 (–26.8 to –6.9)	
	Injuries and accidents	–1.5 (–13.3 to 11.8)	
	Mental health	–3.6 (–9.7 to 3.0)	
	Metabolic and endocrine	–29.0 (–34.6 to –22.9)	–17.7 (–28.6 to –5.2)
	Musculoskeletal	–17.3 (–23.3 to –10.8)	–12.9 (–22.7 to –1.7)
	Neurological	–29.3 (–34.7 to –23.3)	–19.2 (–28.8 to –8.2)
	Oral and gastrointestinal	–21.0 (–28.1 to –13.1)	
	Physical medicine/rehabilitation	–29.7 (–36.2 to –22.5)	–23.9 (–34.8 to –11.2)
	Public health	61.2 (48.3 to 75.2)	72.1 (52.3 to 94.5)
	Reproductive health and childbirth	34.1 (20.7 to 48.9)	47.4 (19.8 to 81.3)
	Respiratory	–16.7 (–23.1 to –9.8)	
	Skin	–25.5 (–34.5 to –15.2)	–18.9 (–35.4 to 1.7)
	Surgery	–17.2 (–25.9 to –7.6)	
Status	Active, not recruiting	–	49.3 (35.3 to 64.7)
	Stopped early	–	–62.8 (–66.6 to –58.7)
	Withdrawn	–	–97.7 (–99.1 to –93.9)

The cells show the percent change (PC) and 95% CI.
*Reference groups are: gender=all; maximum/minimum age limit=no limit; healthy volunteers=no; phase=phase 3; endpoint=efficacy; purpose=treatment; masking=blinded; assignment=parallel; allocation=randomised controlled trial; control=active; intervention=treatment: drugs; area=cardiovascular; status=completed. All other variables are continuous or binary meaning the reference group is 'no'.

Studies sponsored by the National Institutes of Health (NIH) or US Federal agencies (including the Food and Drug Administration) had a lower average actual-to-target ratio of 7.0%–10.3%, whereas industry funded studies had a 18.1% higher average ratio.

In terms of study design, studies with some type of masking had a slightly higher actual-to-target ratio, whereas single group studies had a slightly lower ratio.

Compared with completed studies, studies that stopped early had a 73.2% smaller ratio (95% CI –73.5 to –72.9) and withdrawn studies had a 99.9% smaller ratio (95% CI –99.9 to –99.9).

One reason the actual sample size can be smaller than the target sample size is an adaptive trial that may require fewer patients than originally planned. However, there were only 168 (<0.1%) adaptive trials in the clinicaltrials.gov data and hence this variable is unlikely to impact the overall results and was not selected in the elastic net.

Models of sample size for ANZCTR

Here we examine the non-paired data on the target and actual sample size. The estimated percent differences in sample sizes for the ANZCTR database are shown in table 2 and plotted in figure 4.

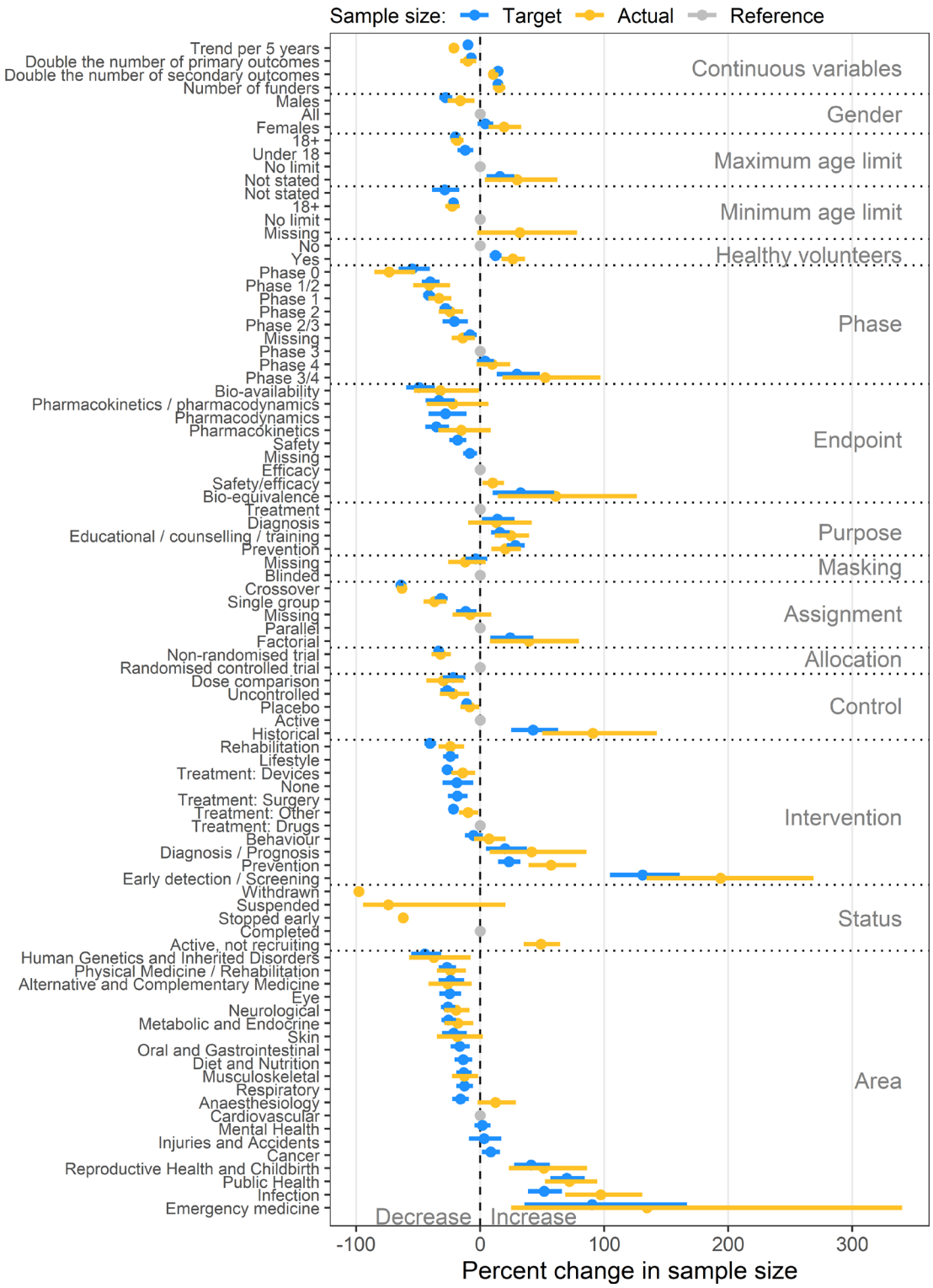


Figure 4 Percent changes in the target and actual sample size for the ANZCTR database. The dots are means and the solid horizontal lines are 95% CIs.

Some associations were as expected. More funders—and hence more resources—meant larger sample sizes. Studies with no age limits were larger than those with any limits. There was a generally increasing sample size for later phases. Bioequivalence studies had an over 30% larger sample size as demonstrating equivalence generally needs more participants than demonstrating efficacy. Studies that allowed healthy volunteers were larger, likely because it increases the available pool of participants. Factorial designs were over 20% larger than parallel studies, to account for the additional comparisons, while cross-over studies were over 60% smaller because of the key comparison is within-participants. Prevention studies were over 25% larger than treatment studies, and

Table 3 Percent changes in the target and actual sample size for the clinicaltrials.gov database

Group/variable*	Variable/category	Target PC (95% CI)	Actual PC (95% CI)
Continuous*	Double the number of arms	37.8 (36.2 to 39.4)	28.6 (27.0 to 30.1)
	Double the number of conditions	7.5 (6.6 to 8.5)	4.8 (3.7 to 5.9)
	Double the number of primary outcomes	–4.4 (–5.2 to –3.7)	–1.8 (–2.6 to –0.9)
	Double the number of secondary outcomes	7.0 (6.6 to 7.4)	7.9 (7.4 to 8.3)
	Trend per 5 years	–7.0 (–7.5 to –6.4)	–18.0 (–18.5 to –17.4)
Gender	Females	15.3 (13.6 to 17.1)	15.2 (13.2 to 17.2)
	Males	–21.4 (–23.1 to –19.6)	–20.6 (–22.3 to –18.8)
Maximum age limit	18+	–20.6 (–21.4 to –19.8)	–18.3 (–19.1 to –17.4)
	Under 18	–20.8 (–22.7 to –18.9)	–16.6 (–18.7 to –14.4)
Minimum age limit	18+	–14.9 (–16.2 to –13.6)	–15.4 (–16.8 to –14.0)
Healthy volunteers	Missing	41.7 (21.3 to 65.6)	
	Yes	5.0 (3.7 to 6.3)	8.3 (6.9 to 9.8)
Intervention	Behavioural	27.3 (25.2 to 29.4)	29.5 (27.2 to 31.9)
	Device	–15.0 (–16.2 to –13.7)	–13.7 (–15.2 to –12.2)
	Diagnostic	32.3 (26.6 to 38.3)	32.3 (23.3 to 42.0)
	Dietary	–23.4 (–25.3 to –21.4)	–23.4 (–25.4 to –21.3)
	Drug		–6.0 (–7.4 to –4.6)
	Other	4.7 (3.4 to 6.1)	
	Radiation	8.2 (5.0 to 11.4)	
Phase	Early phase 1	–72.9 (–74.0 to –71.8)	–70.5 (–72.0 to –68.9)
	Not applicable	–50.9 (–51.8 to –50.0)	–50.7 (–51.7 to –49.6)
	Phase 1	–77.3 (–77.8 to –76.9)	–75.3 (–75.9 to –74.8)
	Phase 1/2	–70.4 (–71.2 to –69.6)	–70.0 (–70.9 to –69.1)
	Phase 2	–61.1 (–61.9 to –60.4)	–60.6 (–61.4 to –59.8)
	Phase 2/3	–40.4 (–42.5 to –38.3)	–42.6 (–44.8 to –40.3)
	Phase 4	–42.2 (–43.4 to –41.0)	–42.3 (–43.5 to –41.0)
Sponsor	Industry	41.0 (39.3 to 42.7)	63.6 (61.5 to 65.7)
	NIH	38.1 (32.6 to 43.8)	17.0 (12.9 to 21.3)
	US federal agencies	11.3 (6.7 to 16.0)	
Purpose	Basic science	–17.2 (–19.0 to –15.4)	–14.6 (–16.7 to –12.5)
	Device feasibility	–55.5 (–59.0 to –51.7)	–44.3 (–49.7 to –38.4)
	Diagnostic	60.9 (57.3 to 64.6)	58.4 (54.1 to 62.8)
	Health services research	161.1 (153.4 to 169.1)	153.0 (144.5 to 161.7)
	Missing		7.1 (3.9 to 10.3)
	Other	13.5 (11.1 to 16.1)	18.7 (15.6 to 21.8)
	Prevention	81.0 (78.1 to 83.8)	75.6 (72.6 to 78.7)
	Screening	292.3 (273.4 to 312.1)	251.0 (230.8 to 272.5)
Masking	Double	–7.3 (–8.6 to –5.9)	
	Missing	32.3 (22.1 to 43.3)	
	Single	–7.5 (–8.8 to –6.2)	–6.0 (–7.5 to –4.5)
	Triple	–9.0 (–10.7 to –7.3)	–5.3 (–7.1 to –3.3)
Assignment	Cross-over	–56.2 (–57.0 to –55.4)	–53.7 (–54.6 to –52.9)
	Factorial	32.8 (27.7 to 38.1)	29.5 (24.2 to 35.0)
	Missing	33.4 (23.7 to 44.0)	
	Sequential	–13.7 (–16.6 to –10.7)	–18.4 (–22.3 to –14.2)
	Single group	–44.4 (–45.1 to –43.6)	–24.2 (–25.9 to –22.4)

Continued

Table 3 Continued

Group/variable*	Variable/category	Target PC (95% CI)	Actual PC (95% CI)
Allocation	Missing	–40.8 (–44.4 to –36.9)	–22.4 (–26.9 to –17.5)
	Non-randomised	–21.6 (–22.8 to –20.4)	–31.9 (–33.2 to –30.6)
	Not applicable		–33.9 (–35.6 to –32.2)
Design	Adaptive/platform trial	56.3 (34.4 to 81.6)	
	Longitudinal	24.6 (18.7 to 30.8)	28.7 (21.4 to 36.4)
Status	Active, not recruiting		43.5 (40.0 to 47.0)
	Suspended		–54.0 (–61.9 to –44.4)
	Terminated		–68.2 (–68.8 to –67.6)
	Unknown		18.6 (13.5 to 23.9)
	Withdrawn		–98.3 (–98.3 to –98.2)

The cells show the percent change (PC) and 95% CI.

*Reference groups for categorical variables are: gender=all; maximum/minimum age limit=no limit; healthy volunteers=no; phase=phase 3; sponsor=other; purpose=treatment; masking=open label; assignment=parallel; allocation=randomised; status=completed. All other variables are continuous or binary meaning the reference group is 'no'.
NIH, National Institutes of Health.

screening studies were over 130% larger. Public health studies were over 60% larger.

Surprisingly, more primary outcomes were associated with smaller sample sizes, although more secondary outcomes were associated with larger sample sizes.

Compared with studies in both genders, actual sample size for studies in men only were around 16% smaller, whereas studies in women only were 19% larger.

Many of the associations for the actual sample size mirrored those from the target sample size. A notable difference was that the decreasing trend in sample size was much larger for the actual sample size, at –21% per 5 years for the actual sample size compared with –10% for the target sample size.

The models of actual sample size included the study status which was a strong determinant of sample size when studies were stopped early or withdrawn.

Models of sample size for clinicaltrials.gov

The estimated percent differences in sample sizes for the clinicaltrials.gov database are shown in table 3 and plotted in figure 5.

As expected studies were larger if they had funding. Studies were also larger if they had more arms or more conditions. Surprisingly, studies with more primary outcomes were associated with a smaller sample size, although the reduction was small at under –4% per doubling in outcomes.

There was a decrease over time in the target sample size of –7% per 5 years, and this decrease was –18% for the actual sample size.

As per the results for the ANZCTR database, women only studies were larger, and men only studies were smaller than studies with both men and women.

Health services research studies were over 150% larger than treatment studies and screening studies were over 250% larger.

Studies using masking were smaller than studies using none, possibly because they are less prone to confounding. Somewhat surprisingly non-randomised studies were around 20% smaller than randomised studies, when these would be more prone to confounding and hence likely need a larger sample size. Adaptive or platform trials were over 50% larger. Longitudinal studies were over 24% larger.

Not surprisingly, studies that were suspended, terminated or withdrawn had a greatly reduced sample size. Those with an unknown study status had larger sample sizes compared with completed studies

Model checks

The cross-validations for the elastic net selections are plotted in online supplemental figure 4. Only one variable category (an allocation category of 'Missing') was removed due to colinearity, it was colinear with an assignment category of single group. The residuals for the final models are plotted in online supplemental figure 5 and are unimodal and approximately symmetric.

DISCUSSION

For the ratio of actual-to-target sample size, although the modal value was on target, the 90th percentiles were asymmetric with more studies below than above target (figure 1). This reflects the many challenges of achieving the target sample size, including difficulties with ethics and governance, difficulties finding and recruiting participants, and running out of time or funding. Larger studies

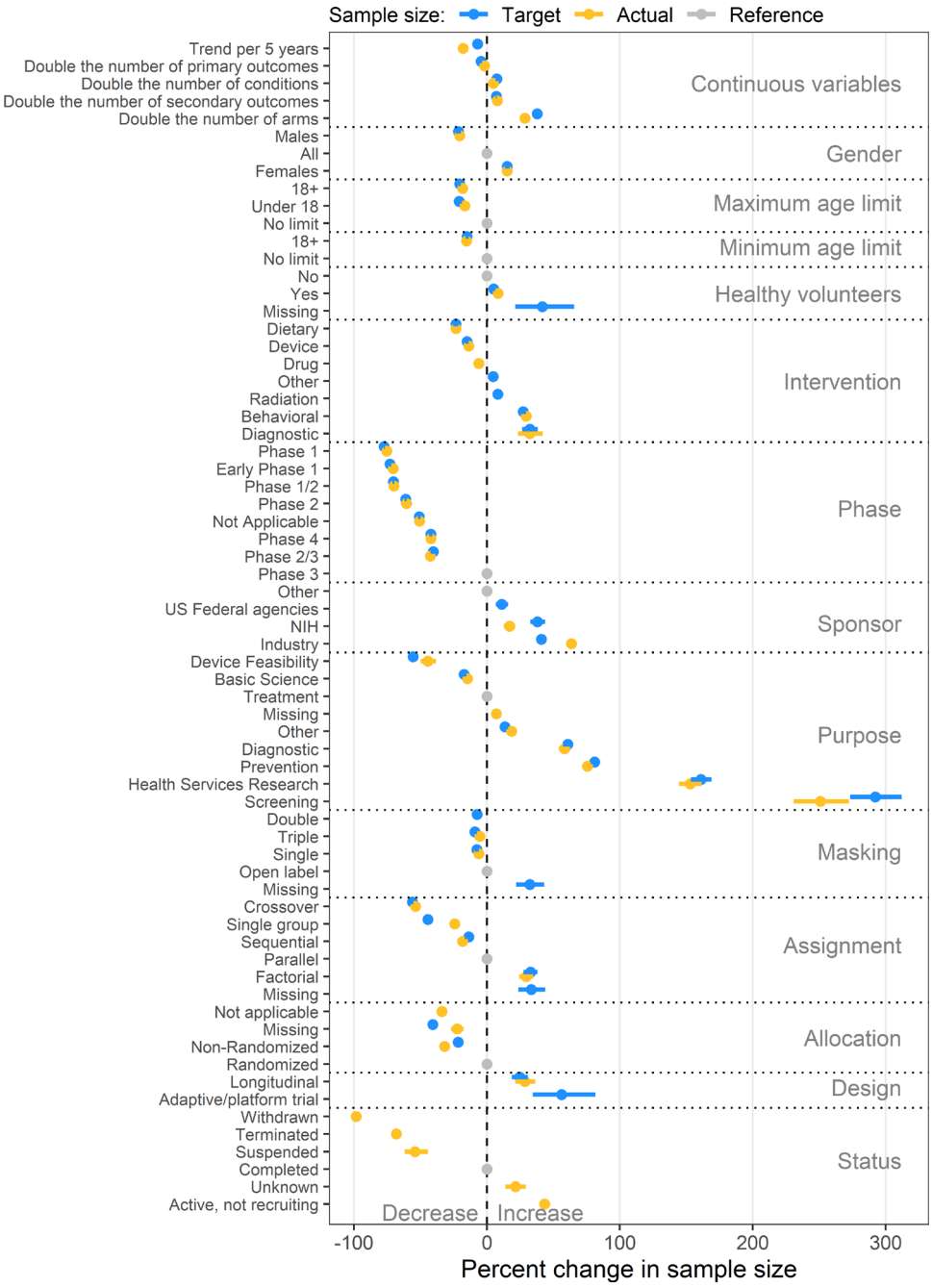


Figure 5 Percent changes in the target and actual sample size for the clinicaltrials.gov database. The dots are means and the solid horizontal lines are 95% CIs.

were generally closer to their target sample size (figure 2) but not by much.

Results from both databases showed a strong decrease in sample size over time. Interestingly, the target sample size decreased by 7%–10% per 5 years, whereas the actual decrease was 18%–21%, confirming the generally growing difficulty of recruiting research participants. The finding was confirmed in the lower actual-to-target sample size ratios over time. Smaller actual sample sizes mean studies may be underpowered with flow-on effects for the statistical power and uncertainty of meta-analyses.[25]

A recent observational analysis of the health literature shows a clear decrease in average effect sizes over time from 1990 to 2015.[26] We would expect larger average sample sizes over time to study these smaller effects with adequate statistical power. Our finding of smaller sample sizes (both actual and target) has implications for statistical power and strongly suggests that the problem of underpowered studies is ongoing. A study of the Cochrane database of systematic reviews from 1975 to 2014 estimated that the percentage of sufficiently powered studies increased from 5% in 1975–1979 to 9% in 2010–2014.[8] Another study of clinical trials from the Cochrane database estimated an increase in adequately powered studies over time with an OR of 1.02 per year.[27] Our results suggest this small previous increase in power

may now be at risk given the average decrease in sample sizes.

In both databases, there were more studies that were women only than men only (10% women only vs 5% men only), and in both databases, the women only studies were larger. This difference may partly be due to initiatives to fund women's health research to make up for the historical shortage of women in trials.[28] To examine other differences, we examined the top 10 words in the brief titles of the clinicaltrials.gov database in studies in women and men only (see online supplemental table 3). 'Cancer' and 'breast' were the two most common words in studies in women only, and 'study,' 'prostate' and 'cancer' were the top three words in studies in men only. Hence, the difference in sample sizes could be due to differences in the primary outcomes and effect sizes for these two cancers.

A previous study of the clinicaltrials.gov data examined clinical trials between 2007 and 2010 found 62% had 100 or fewer participants.[29] Another study of clinicaltrials.gov found that actual sample sizes for completed studies declined between 2000 and 2019.[30] A study of 114 trials found that only 31% achieved their target sample size.[31] A study of NIH funded clinical trials found that the proportion enrolling more than 500 or 1000 was relatively stable between 2005 and 2015.[32] Studies examining why trials are terminated early found that problems recruiting patients are the most common reason.[1–3 33–35] Trial characteristics that predicted smaller actual than target sample sizes were phase 2 studies compared with phase 3, more eligibility criteria, active control compared with placebo, fewer sites and public funding compared with industry funding.[33] These results match ours for study phase and industry funding, although we found active control did slightly better than placebo (figures 4 and 5).

Strengths and limitations

We analysed two databases and found generally consistent results in terms of what study characteristics were associated with sample size, which increases the robustness of our results.

A key strength is the large sample size available from the trial registry data. There are strong incentives for researchers to register trials before any participants are recruited, which means the registry data should be representative of the target population of all trials. However, there have been documented problems with trials not being updated to include the results and recruitment status.[15 36] The implication for our study is that actual sample sizes will be missing and there could well be an under-reporting bias for studies where the actual sample size was well below the target. Hence, our results may present a somewhat optimistic picture of the actual-to-target sample size ratio.

The databases record many trial features with little missing data. The completeness of studies on clinicaltrials.gov has increased over time, with over 90% completion since 2007 for key fields such as allocation, masking, gender, enrolment and study arms.[37] A study of the completeness of clinicaltrials.gov of phase 2–3 studies posted by pharmaceutical companies found incomplete data wwere generally below 3%.[38]

The registry data relies on researchers to correctly enter and update their study's details and there are likely to be data entry errors and poor reporting. For example, we found a study where an age limitation was mentioned in the descriptive text but not in the age limit field. We also found some cluster-randomised studies where the anticipated sample size was the number of clusters and the actual sample size was the number of participants (we excluded these six studies).

Data on the actual amount of funding for each study would have been useful so that an actual dollar value could have been modelled instead of the simpler variables of number of funders and funding class.

Conclusion

Registered studies are more often under-recruited than over-recruited and disappointingly both target and actual sample sizes appear to have decreased over time. If true, this is concerning and deserves attention by both researchers and funders, to examine causes and solutions of the problem. This could include understanding barriers to recruitment, the use of evidence-based recruitment processes[39] and incentives to increase the use of multicentre studies. We recommend ongoing implementation of evidence-based interventions to increase sample size and further monitoring of sample sizes.

Twitter Adrian Gerard Barnett @aidybarnett

Acknowledgements Thanks to the National Library of Medicine and ANZCTR for making the registry data available for research. Thanks to Nicholas De Vito for help with the clinicaltrials.gov data. Thanks to Andrew Althouse and Noah Haber for helpful comments on the first draft.

Contributors Study design and data interpretation: AGB and PG. Data analysis and initial manuscript drafting: AGB. Critical review of early and final versions of the manuscript: PG. AB is the guarantor.

Funding This work was supported by National Health and Medical Research Council (https://www.nhmrc.gov.au/) grant number APP1117784. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Competing interests Paul Glasziou is a member of the ANZCTR advisory committee.

Patient consent for publication Not required.

Ethics approval This study does not involve human participants.

Provenance and peer review Not commissioned; externally peer reviewed.

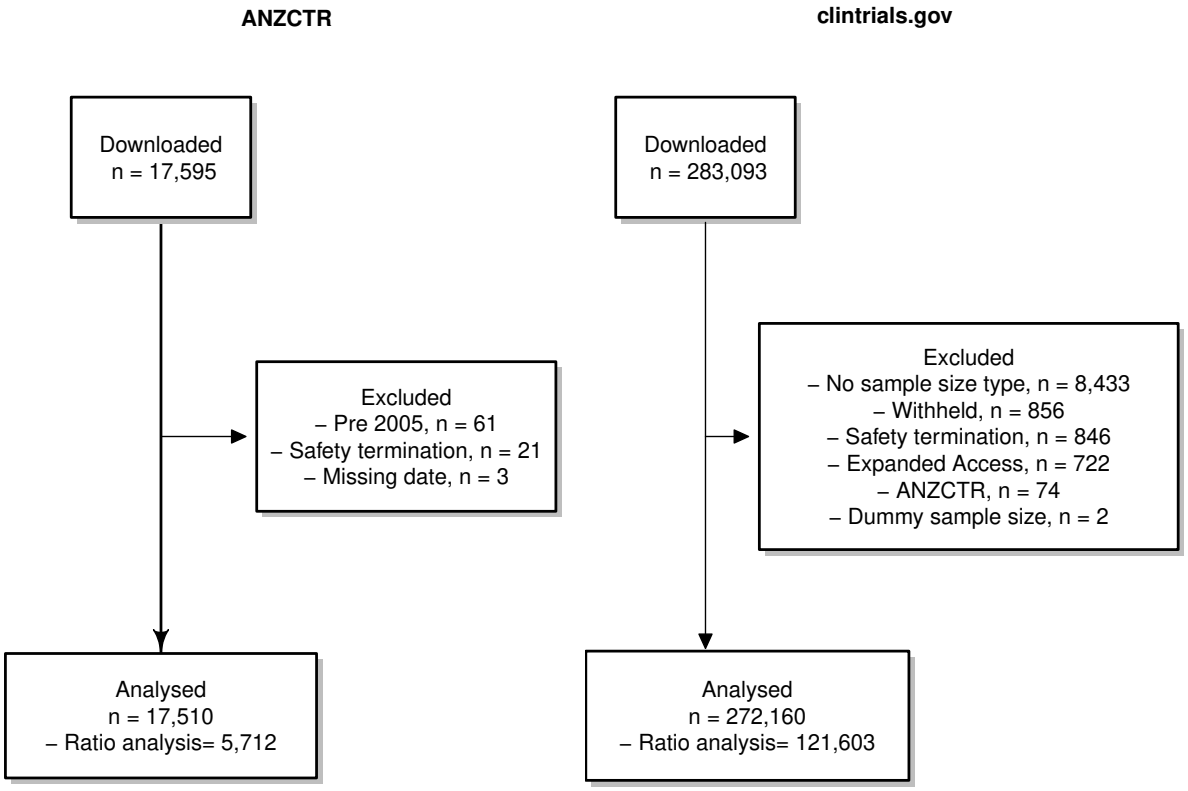
Data availability statement Data are available in a public, open access repository. All data and code are openly available from the github database: https://github.com/agbarnett/registries.

ORCID iDs
Adrian Gerard Barnett http://orcid.org/0000-0001-6339-0374
Paul Glasziou http://orcid.org/0000-0001-7564-073X

REFERENCES

1 Schandelmaier S, von Elm E, You JJ, *et al*. Premature discontinuation of randomized trials in critical and emergency care: a retrospective cohort study. *Crit Care Med* 2016;44:130–7.
2 Baldi I, Lanera C, Berchialla P, *et al*. Early termination of cardiovascular trials as a consequence of poor accrual: analysis of ClinicalTrials.gov 2006-2015. *BMJ Open* 2017;7:e013482.
3 Janiaud P, Axfors C, Ioannidis JPA, *et al*. Recruitment and results reporting of COVID-19 randomized clinical trials registered in the first 100 days of the pandemic. *JAMA Netw Open* 2021;4:e210330.
4 Kearney A, Harman NL, Rosala-Hallas A, *et al*. Development of an online resource for recruitment research in clinical trials to organise and MAP current literature. *Clin Trials* 2018;15:533–42.
5 Healy P, Galvin S, Williamson PR, *et al*. Identifying trial recruitment uncertainties using a James Lind Alliance priority setting partnership – the PRioRiTy (prioritising recruitment in randomised trials) study. *Trials*2018;19.
6 Ioannidis JPA. Why most published research findings are false. *PLoS Med* 2005;2:e124.
7 Smaldino PE, McElreath R. The natural selection of bad science. *R Soc Open Sci* 2016;3:160384.
8 Lamberink HJ, Otte WM, Sinke MRT, *et al*. Statistical power of clinical trials increased while effect size remained stable: an empirical analysis of 136,212 clinical trials between 1975 and 2014. *J Clin Epidemiol* 2018;102:123–8.
9 Norman G, Monteiro S, Salama S. Sample size calculations: should the emperor's clothes be off the PEG or made to measure? *BMJ* 2012;345:e5278.
10 Bacchetti P. Current sample size conventions: flaws, harms, and alternatives. *BMC Med* 2010;8:17.
11 Ioannidis JPA, Fanelli D, Dunne DD, *et al*. Meta-research: evaluation and improvement of research methods and practices. *PLoS Biol* 2015;13:e1002264–7.
12 Lehman R, Loder E. Missing clinical trial data. *BMJ* 2012;344:d8158.
13 Goldacre B. How to get all trials reported: audit, better data, and individual accountability. *PLoS Med* 2015;12:e1001821–5.
14 Anderson ML, Chiswell K, Peterson ED, *et al*. Compliance with results reporting at ClinicalTrials.gov. *N Engl J Med Overseas Ed* 2015;372:1031–9.
15 Piller C. FDA and NIH let clinical trial sponsors keep results secret and break the law. *Science* 2020.
16 Grey A, Portch R, Gaby A, *et al*. Clinical trial registry documents and publication integrity. *Account Res* 2021;28:149–61.
17 R Core Team. R: A language and environment for statistical computing. Vienna, Austria: : R Foundation for Statistical Computing, 2021. Available: https://www.R-project.org/
18 Barnett A. agbarnett/registries: (v1.0.0). *Zenodo* 2021.
19 von Elm E, Altman DG, Egger M, *et al*. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ* 2007;335:806–8.
20 Pallmann P, Bedding AW, Choodari-Oskooei B, *et al*. Adaptive designs in clinical trials: why use them, and how to run and report them. *BMC Med* 2018;16.
21 Hastie T, Tibshirani R, Wainwright M. *Statistical learning with sparsity: the LASSO and generalizations*. CRC Press, 2015.
22 Royston P, Sauerbrei W. *Multivariable model - building: a pragmatic approach to regression anaylsis based on fractional polynomials for modelling continuous variables*. Chichester, England: John Wiley & Sons Ltd, 2008.
23 Spiegelhalter DJ, Best NG, Carlin BP, *et al*. The deviance information criterion: 12 years on. *J Royal Stat Soc* 2014;76:485–93.
24 Plummer M. JAGS version 4.3.0 user manual, 2014. Available: https://sourceforge.net/projects/mcmc-jags/files/Manuals/4.x/
25 Garcia-Alamino JM, Bankhead C, Heneghan C, *et al*. Impact of heterogeneity and effect size on the estimation of the optimal information size: analysis of recently published meta-analyses. *BMJ Open* 2017;7:e015888.
26 Monsarrat P, Vergnes J-N. The intriguing evolution of effect sizes in biomedical research over time: smaller but more often statistically significant. *GigaScience*2017;7.
27 Otte WM, Tijdink JK, Weerheim PL, *et al*. Adequate statistical power in clinical trials is associated with the combination of a male first author and a female last author. *eLife* 2018;7 doi:10.7554/eLife.34412
28 Liu KA, Mager NAD. Women's involvement in clinical trials: historical perspective and future implications. *Pharm Pract* 2016;14:708–8.
29 Califf RM, Zarin DA, Kramer JM, *et al*. Characteristics of clinical trials registered in ClinicalTrials.gov, 2007-2010. *JAMA* 2012;307:1838–47.
30 Gresham G, Meinert JL, Gresham AG, *et al*. Assessment of trends in the design, accrual, and completion of trials registered in ClinicalTrials.gov by sponsor type, 2000-2019. *JAMA Netw Open* 2020;3:e2014682.
31 McDonald AM, Knight RC, Campbell MK, *et al*. What influences recruitment to randomised controlled trials? A review of trials funded by two UK funding agencies. *Trials* 2006;7:9.
32 Gresham GK, Ehrhardt S, Meinert JL, *et al*. Characteristics and trends of clinical trials funded by the National Institutes of Health between 2005 and 2015. *Clin Trials* 2018;15:65–74.
33 Carlisle B, Kimmelman J, Ramsay T, *et al*. Unsuccessful trial accrual and human subjects protections: an empirical analysis of recently closed trials. *Clin Trials* 2015;12:77–83.
34 Pica N, Bourgeois F. Discontinuation and nonpublication of randomized clinical trials conducted in children. *Pediatrics* 2016;138:e20160223–3.
35 Rees CA, Pica N, Monuteaux MC, *et al*. Noncompletion and nonpublication of trials studying rare diseases: a cross-sectional analysis. *PLoS Med* 2019;16:e1002966.
36 Jones CW, Safferman MR, Adams AC, *et al*. Discrepancies between ClinicalTrials.gov recruitment status and actual trial status: a cross-sectional analysis. *BMJ Open* 2017;7:e017719.
37 Tasneem A, Aberle L, Ananth H, *et al*. The database for aggregate analysis of ClinicalTrials.gov (AACT) and subsequent regrouping by clinical specialty. *PLoS One* 2012;7:e33677.
38 Glass HE, Glass LM, DiFrancesco JJ. Clinicaltrials.Gov: an underutilized source of research data about the design and conduct of commercial clinical trials. *Ther Innov Regul Sci* 2015;49:218–24.
39 Trial Forge. Trial forge resources, 2021. Available: https://www.trialforge.org/trial-forge-resources/

Supplement Figure 1: Flow diagrams of the number of studies included and excluded



The numbers available for the actual-to-target ratio analysis are smaller because studies had to provide data for both the actual and sample size.

Barnett AG, Glasziou P. *BMJ Open* 2021; 11:e053377. doi: 10.1136/bmjopen-2021-053377

Supplement Table 1: Predictor variables in the two trial registries

The tables below show the predictor variables available for the two trial registries that were used in the multiple regression models.

The "categories" column gives the categories for factors.

The "mandatory" column indicates whether the variable has to be completed on the databases.

The "log" column indicates whether the variable was log-transformed (base 2).

Study status was only used for the models of actual sample size and not the target sample size, this is because the study status was likely to change after the target sample size was set.

clintrials.gov

Variable	**categories**	**mandatory**	**log**
Trend per 5 years		Yes	No
Gender	All, Male, Female	Yes	No
Maximum age limit		Yes	No
Minimum age limit		Yes	No
Healthy volunteers	No, Yes	Yes+	No
Behavioral	Yes, No	Yes	No
Biological	Yes, No	Yes	No
Combination	Yes, No	Yes	No
Device	Yes, No	Yes	No
Diagnostic	Yes, No	Yes	No
Dietary	Yes, No	Yes	No
Drug	Yes, No	Yes	No
Genetic	Yes, No	Yes	No
Procedure	Yes, No	Yes	No
Radiation	Yes, No	Yes	No
Other	Yes, No	Yes	No
Double the number of primary outcomes		Yes	Yes
Double the number of secondary outcomes		No	Yes
Double the number of conditions		Yes	Yes
Double the number of arms		Yes	Yes
Lead sponsor	Other, NIH, Industry, U.S. Fed	Yes	No
Longitudinal	Yes, No	No	No

Variable	categories	mandatory	logy
Phase	Phase 3, Phase 2/Phase 3, Phase 2, Phase 1, N/A, Phase 4, Phase 1/Phase 2, Early Phase 1	Yes	No
Purpose	Treatment, Prevention, Diagnostic, Screening, Device Feasibility, Basic Science, Other, Health Services Research, Supportive Care	Yes+	No
Masking	Triple, None, Double, Single, Quadruple, Missing	Yes+	No
Assignment	Parallel, Factorial, Single Group, Crossover, Sequential, Missing	Yes	No
Allocation	Randomized, N/A, Non-Randomized	Yes+	No
Adaptive_trial	Yes, No	No	No
Status	Completed, Withdrawn, Terminated, Recruiting, Active, not recruiting, Unknown status, Suspended, Enrolling by invitation, Not yet recruiting, Unknown status [Previously: Active, not recruiting], Unknown status [Previously: Enrolling by invitation]	Yes	No

+ = Mandatory after January 2017

ANZCTR

Variable	categories	mandatory	logy
Trend per 5 years		Yes	No
Gender	All, Females, Males	Yes	No
Maximum age limit		Yes	No
Minimum age limit		Yes	No
Control	Uncontrolled, Active, Placebo, Dose comparison, Historical	Yes	No
Double the number of primary outcomes		Yes	Yes
Double the number of secondary outcomes		Yes	Yes
Area	Cancer, Neurological, Cardiovascular, Stroke, Respiratory, Other, Human Genetics and Inherited Disorders, Reproductive Health and Childbirth, Blood, Infection, Inflammatory and Immune System, Skin, Musculoskeletal, Oral and Gastrointestinal, Alternative and Complementary Medicine, Public Health, Injuries and Accidents, Renal and Urogenital, Metabolic and Endocrine, Anaesthesiology, Diet and Nutrition, Mental Health, Surgery, Physical Medicine / Rehabilitation, Eye, Ear, Emergency medicine	Yes	No
Number of funders		Yes	No
Healthy volunteers	No, Yes	Yes	No
Allocation	Non-randomised trial, Randomised controlled trial	Yes	No
Phase	Phase 2, Phase 3, Phase 1 / Phase 2, Phase 3 / Phase 4, Phase 4, Phase 2 / Phase 3, Phase 1, Not Applicable, Phase 0	Yes	No
Endpoint	Safety/efficacy, Efficacy, Safety, Bio-equivalence, Pharmacokinetics, Pharmacodynamics, Bio-availability, Pharmacokinetics / pharmacodynamics	No	No
Purpose	Treatment, Prevention, Educational / counselling / training, Diagnosis	Yes	No
Masking	Open (masking not used), Blinded (masking used)	No	No
Assignment	Single group, Parallel, Factorial, Crossover, Other	No	No

Variable	**categories**	**mandatory**	**logy**
Intervention	Treatment: Drugs, Rehabilitation, Prevention, None, Treatment: Other, Treatment: Surgery, Diagnosis / Prognosis, Other interventions, Treatment: Devices, Behaviour, Lifestyle, Early detection / Screening, Not applicable	Yes	No
Study status	Completed, Active, not recruiting, Recruiting, Not yet recruiting, Stopped early, Withdrawn, Suspended	Yes	No

Supplement 1: Descriptive statistics and descriptions of missing data for the two databases

Table of Contents

Barnett AG, Glasziou P. *BMJ Open* 2021; 11:e053377. doi: 10.1136/bmjopen-2021-053377

clinicaltrials.gov

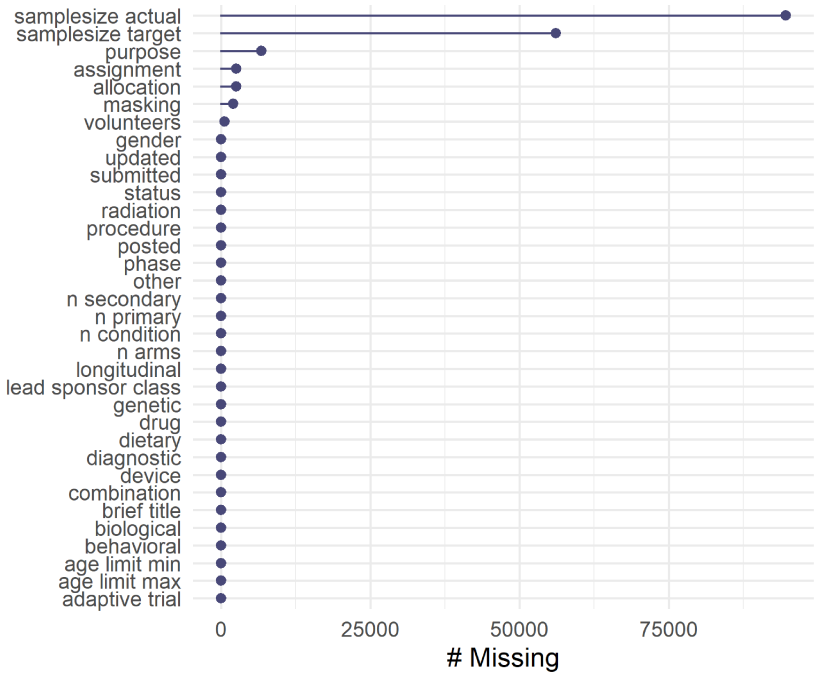
There were 10,100 studies excluded. The total number of remaining studies is 272,988. The data were extracted from clinicaltrials.gov on 01-Feb-2021.

Missing data (clinicaltrials.gov)

The plots below summarise the missing data. The columns are variables and the rows are observations. The grey areas show missing data. The coloured plots show a random sample of 10,000 studies as the data are too large to show every study.

Counts of missing data (clinicaltrials.gov)

The plots below shows the counts of missing data using all studies.



The total number of studies in 272,988.

Overall the amount of missing data was small. The most frequent missing variable was the study purpose which is the main objective of the intervention(s) being evaluated by the clinical trial. This was missing for 2% of studies.

The logical variables cannot be missing by design, for example 'longitudinal' was created by whether the word "longitudinal" was mentioned in the detailed study description or not.

Many variables were mandatory and so cannot be missing, e.g., study status.

Registration and dates (clinicaltrials.gov)

Registration dates (clinicaltrials.gov)

date	n	median	q1	q3
posted	272,988	Feb 2015	Nov 2010	May 2018
updated	272,988	Jun 2018	May 2015	Apr 2020

The summary statistics are the median and inter-quartile range.

Inclusions/Exclusions (clinicaltrials.gov)

Gender included (clinicaltrials.gov)

	Freq	% Valid	% Total
All	232734	85	85

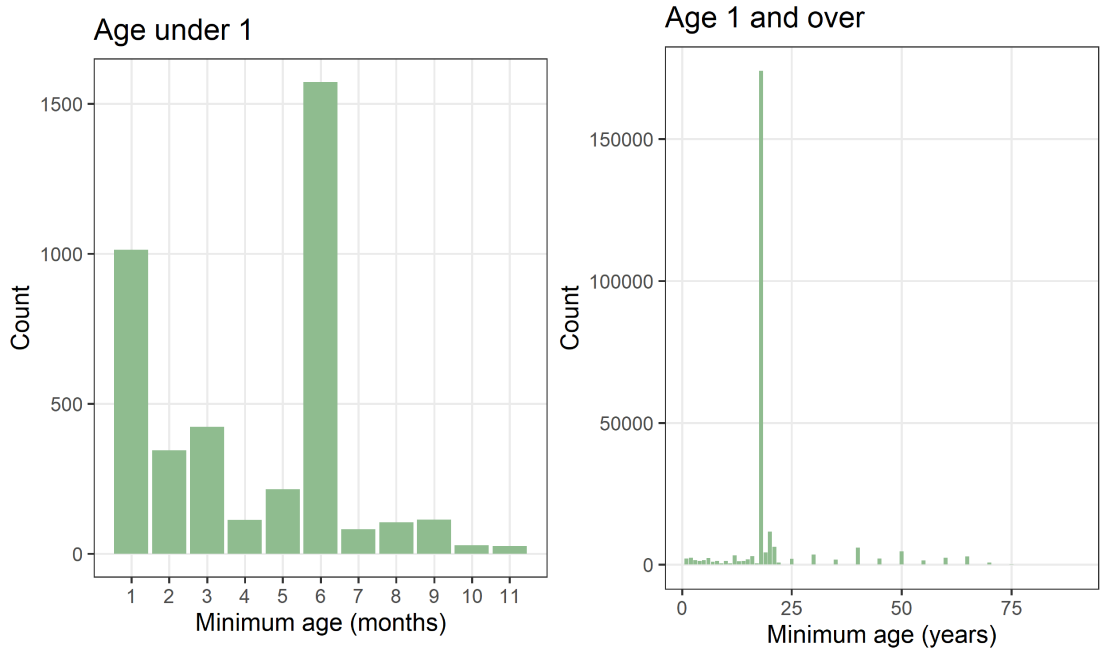
Female	26511	10	10
Male	13742	5	5
<NA>	1		0
Total	272988	100	100

Age included (clinicaltrials.gov)

Minimum age (clinicaltrials.gov)

	Freq	% Valid	% Total
Restricted	258396	95	95
No limit	14592	5	5
<NA>	0		0
Total	272988	100	100

Minimum age limit in years (clinicaltrials.gov)

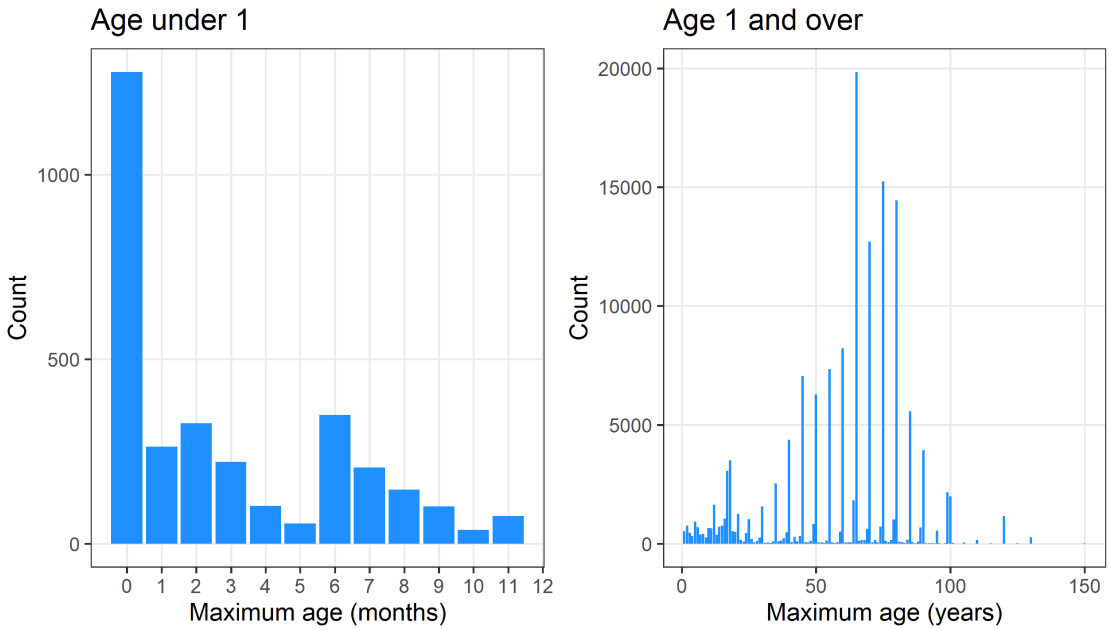


This plot is just for the 258,396 studies that had a minimum age limit. By far the most common minimum age is 18. The oldest exclusion for minimum age was 90 years.

Maximum age (clinicaltrials.gov)

	Freq	% Valid	% Total
Restricted	150717	55	55
No limit	122271	45	45
<NA>	0		0
Total	272988	100	100

Maximum age limit in years (clinicaltrials.gov)



This plot is just for the 150,717 studies that had a maximum age limit. The oldest exclusion for maximum age was 150 years.

Healthy volunteers included (clinicaltrials.gov)

	Freq	% Valid	% Total
No	202374	74	74
Yes	70079	26	26
<NA>	535		0
Total	272988	100	100

Study characteristics (clinicaltrials.gov)

Purpose (clinicaltrials.gov)

	Freq	% Valid	% Total
Treatment	177513	67	65
Prevention	28945	11	11
Basic Science	13568	5	5
Supportive Care	12463	5	5
Diagnostic	12317	5	5
Other	12028	5	4
Health Services Research	6412	2	2
Screening	2261	1	1
Device Feasibility	754	0	0

<NA>	6727		2
Total	272988	100	100

Masking (clinicaltrials.gov)

	Freq	% Valid	% Total
None	152877	56	56
Single	36915	14	14
Double	35433	13	13
Quadruple	26844	10	10
Triple	18873	7	7
Missing	5	0	0
<NA>	2041		1
Total	272988	100	100

Assignment (clinicaltrials.gov)

	Freq	% Valid	% Total
Parallel	159496	59	58
Single Group	78591	29	29
Crossover	23781	9	9
Sequential	4582	2	2
Factorial	3995	1	1
Missing	7	0	0
<NA>	2536		1
Total	272988	100	100

Allocation (clinicaltrials.gov)

	Freq	% Valid	% Total
Randomized	178772	66	65
N/A	61995	23	23
Non-Randomized	29732	11	11
<NA>	2489		1
Total	272988	100	100

Phase (clinicaltrials.gov)

	Freq	% Valid	% Total
Early Phase 1	3464	1	1
N/A	120353	44	44
Phase 1	32331	12	12
Phase 1/Phase 2	11054	4	4
Phase 2	44189	16	16
Phase 2/Phase 3	5307	2	2
Phase 3	30120	11	11
Phase 4	26170	10	10

<NA>	0		0
Total	272988	100	100

Adaptive/platform trial (clinicaltrials.gov)

	Freq	%
FALSE	272820	100
TRUE	168	0
Total	272988	100

Was an adaptive or platform trial mentioned in the study description.

Longitudinal (clinicaltrials.gov)

	Freq	%
FALSE	270702	99
TRUE	2286	1
Total	272988	100

Was longitudinal mentioned in the study description.

Number of outcomes (clinicaltrials.gov)

	n_primary	n_secondary
Min	0	0
Q1	1	1
Median	1	2
Q3	2	5
Max	214	459

Number of primary and secondary outcomes. A small number of studies had a huge number of outcomes.

Study status (clinicaltrials.gov)

	Freq	%
Completed	149721	55
Recruiting	39237	14
Unknown status	26671	10
Terminated	17641	6
Not yet recruiting	13659	5
Active, not recruiting	13370	5
Withdrawn	8019	3
Enrolling by invitation	2070	1
Suspended	1393	1
Unknown status [Previously: Active, not recruiting]	1195	0
Unknown status [Previously: Enrolling by invitation]	12	0
Total	272988	100

Funding (clinicaltrials.gov)

Lead sponsor type (clinicaltrials.gov)

	Freq	%
Other	190273	70
Industry	74612	27
NIH	4941	2
U.S. Fed	3162	1
Total	272988	100

Sample size (clinicaltrials.gov)

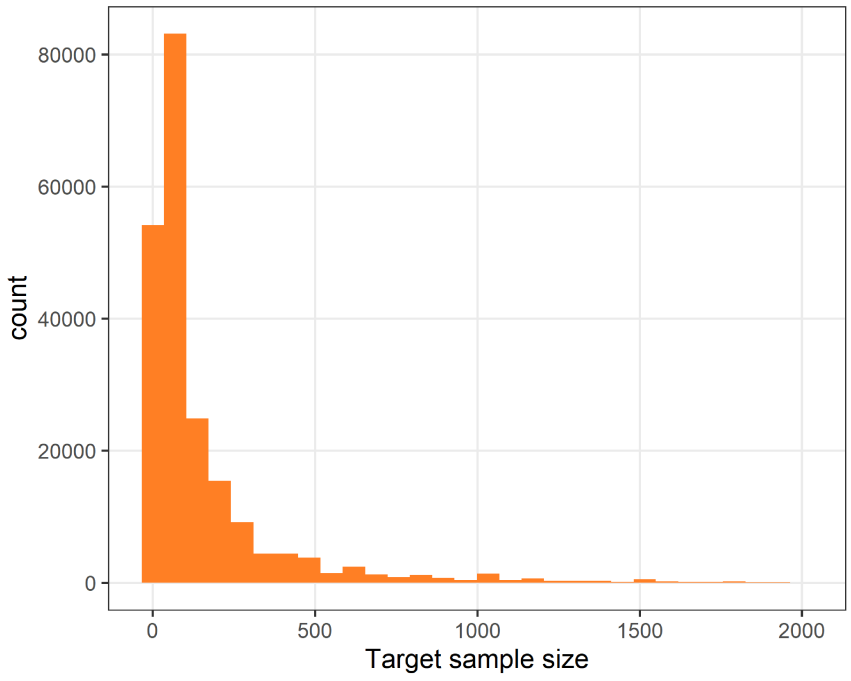
Target and achieved sample size (clinicaltrials.gov)

	Achieved	Target
Min	0	0
Q1	22	35
Median	52	70
Q3	139	178
Max	20121212	20000000

The table shows summary statistics.

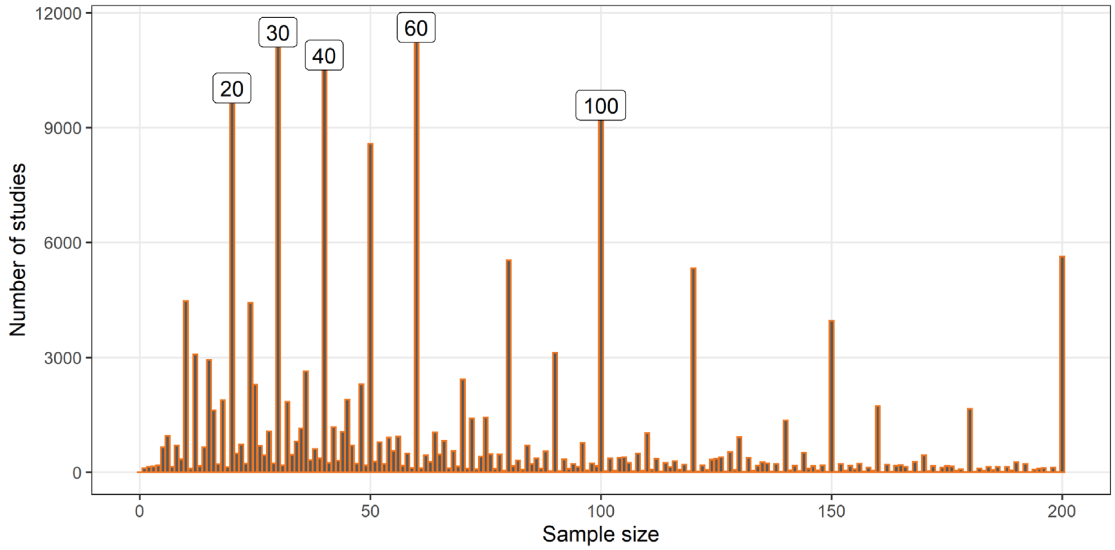
There were 5 studies with a zero target sample size.

Histogram of sample size (clinicaltrials.gov)



The plot above is for the 212,443 studies with a target sample size under 2,000.

Histogram looking for digit preference in target sample size (clinicaltrials.gov)



The plot above is for the 171,768 studies with a target sample size of 200 or fewer.

There is a strong digit preference at multiples of ten. The modal sample size is . The top five sample sizes are shown for each sample size type.

Barnett AG, Glasziou P. *BMJ Open* 2021; 11:e053377. doi: 10.1136/bmjopen-2021-053377

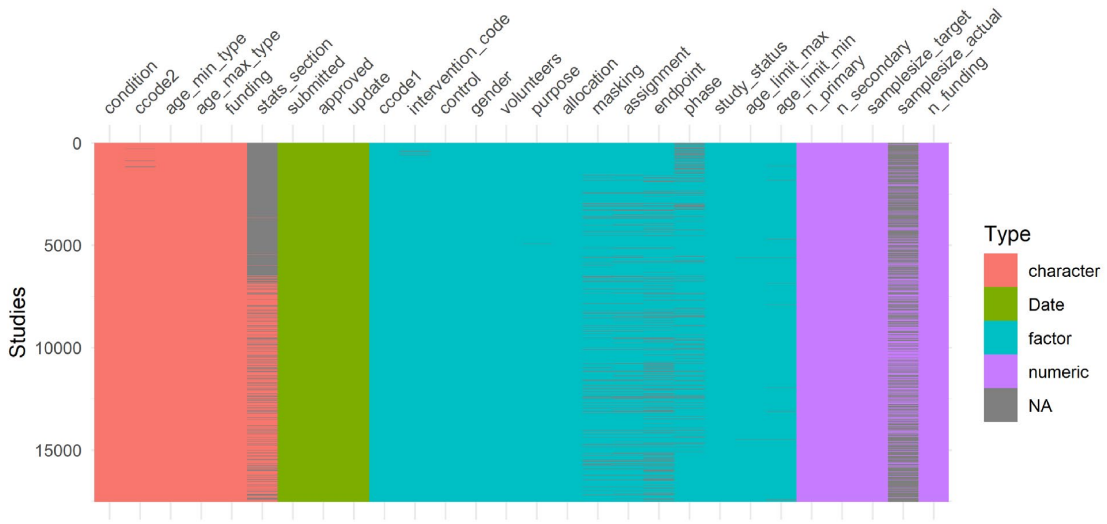
ANZCTR

These summary statistics are for trials registered in ANZCTR and not clinicaltrials.gov.

The total number of studies is 17,531. The data were extracted from ANZCTR on 01-Feb-2020.

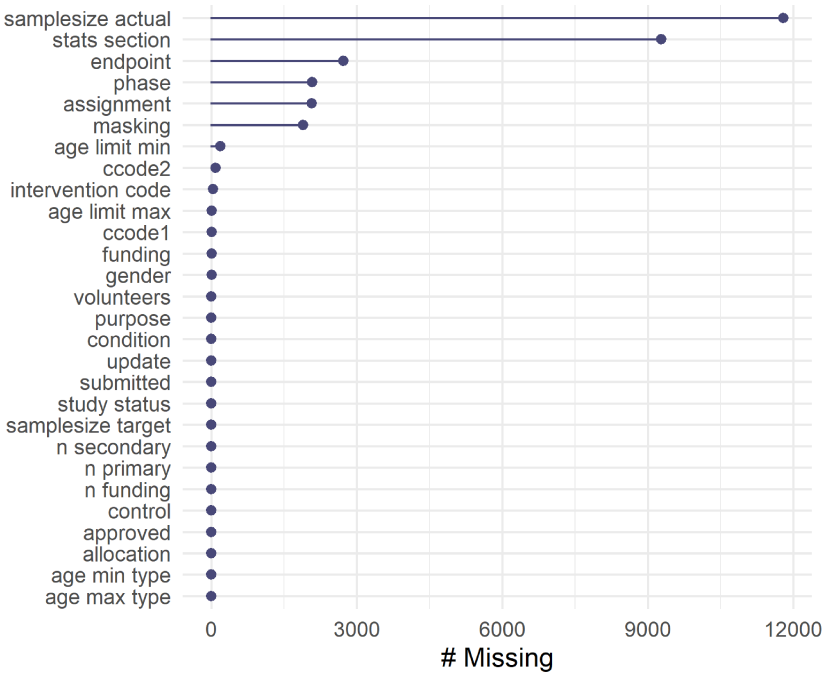
Missing data (ANZCTR)

The columns are variables and the rows are observations. The grey areas show missing data.



Counts of missing data (ANZCTR)

The plots below shows the counts of missing data.



Registration and dates (ANZCTR)

Registration dates (ANZCTR)

date	n	median	q1	q3
submitted	17,531	Oct 2014	Jan 2011	Dec 2017
approved	17,531	Nov 2014	Feb 2011	Jan 2018
update	17,531	Apr 2017	Apr 2013	Jun 2019

The summary statistics are the median and inter-quartile range.

Anticipated start date (ANZCTR)

min	q1	median	q3	max
Jun 1982	Jan 2010	Nov 2013	Oct 2017	Dec 2022

The summary statistics are the minimum, inter-quartile range, median and maximum.

Inclusions/Exclusions (ANZCTR)

Gender included (ANZCTR)

	Freq	% Valid	% Total
All	14900	85	85
Females	1701	10	10
Males	926	5	5

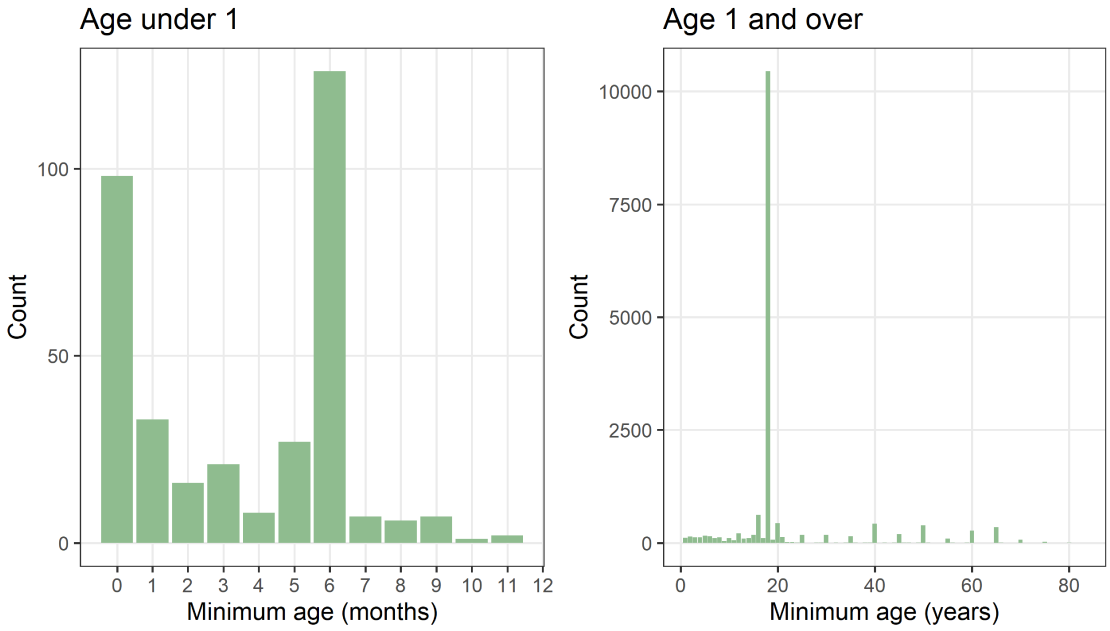
<NA>	4		0
Total	17531	100	100

Age included (ANZCTR)

Minimum age (ANZCTR)

	Freq	% Valid	% Total
Restricted	16623	95	95
No limit	632	4	4
Not stated	271	2	2
<NA>	5		0
Total	17531	100	100

Minimum age limit in years (ANZCTR)

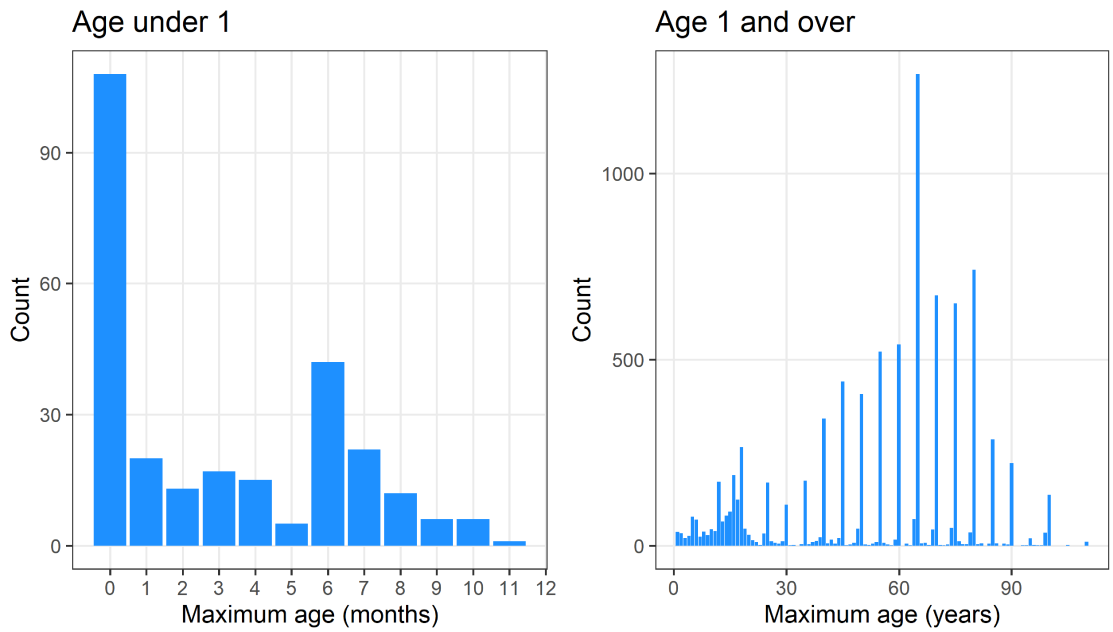


This plot is just for the 16,438 studies that had a minimum age limit. By far the most common minimum age is 18. The oldest exclusion for minimum age was 83 years.

Maximum age (ANZCTR)

	Freq	% Valid	% Total
Restricted	9092	52	52
No limit	7613	43	43
Not stated	821	5	5
<NA>	5		0
Total	17531	100	100

Maximum age limit in years (ANZCTR)



This plot is just for the 9,081 studies that had a maximum age limit. The oldest exclusion for maximum age was 110 years.

There is a strong digit preference in the age limit with spikes in multiples of five years.

Healthy volunteers included (ANZCTR)

	Freq	% Valid	% Total
No	12907	74	74
Yes	4621	26	26
<NA>	3		0
Total	17531	100	100

Study characteristics (ANZCTR)

Purpose (ANZCTR)

	Freq	% Valid	% Total
Treatment	12056	69	69
Prevention	3257	19	19
Educational / counselling / training	1432	8	8
Diagnosis	784	4	4
<NA>	2		0
Total	17531	100	100

Assignment (ANZCTR)

	Freq	% Valid	% Total
Parallel	10020	65	57
Single group	2427	16	14
Crossover	1925	12	11
Other	863	6	5
Factorial	226	1	1
<NA>	2070		12
Total	17531	100	100

Allocation (ANZCTR)

	Freq	% Valid	% Total
Randomised controlled trial	13155	75	75
Non-randomised trial	4376	25	25
<NA>	0		0
Total	17531	100	100

Masking (ANZCTR)

	Freq	% Valid	% Total
Blinded (masking used)	8274	53	47
Open (masking not used)	7366	47	42
<NA>	1891		11
Total	17531	100	100

Intervention code (ANZCTR)

	Freq	% Valid	% Total
Treatment: Drugs	4473	26	26
Treatment: Other	3502	20	20
Prevention	1868	11	11
Treatment: Devices	1802	10	10
Behaviour	1260	7	7
Rehabilitation	1237	7	7
Lifestyle	970	6	6
Other interventions	667	4	4
Treatment: Surgery	555	3	3
Diagnosis / Prognosis	514	3	3
Early detection / Screening	408	2	2
None	230	1	1
Not applicable	11	0	0
<NA>	34		0
Total	17531	100	100

Endpoint (ANZCTR)

	Freq	% Valid	% Total
Efficacy	9489	64	54
Safety/efficacy	3868	26	22
Safety	745	5	4
Pharmacokinetics	226	2	1
Bio-equivalence	152	1	1
Pharmacokinetics / pharmacodynamics	145	1	1
Pharmacodynamics	99	1	1
Bio-availability	86	1	0
<NA>	2721		16
Total	17531	100	100

Phase (ANZCTR)

	Freq	% Valid	% Total
Not Applicable	9830	64	56
Phase 0	58	0	0
Phase 1	1161	8	7
Phase 1 / Phase 2	331	2	2
Phase 2	1239	8	7
Phase 2 / Phase 3	269	2	2
Phase 3	907	6	5
Phase 3 / Phase 4	251	2	1
Phase 4	1401	9	8
<NA>	2084		12
Total	17531	100	100

Control (ANZCTR)

	Freq	% Valid	% Total
Active	10584	60	60
Placebo	3527	20	20
Uncontrolled	2780	16	16
Dose comparison	337	2	2
Historical	303	2	2
<NA>	0		0
Total	17531	100	100

Number of outcomes (ANZCTR)

	n_primary	n_secondary
Min	0	0
Q1	1	1
Median	1	3
Q3	2	6

Max	12	57

Number of primary and secondary outcomes.

Health conditions (ANZCTR)

Condition code #1 (ANZCTR)

	Freq	%
Mental Health	2029	12
Cancer	1592	9
Musculoskeletal	1157	7
Cardiovascular	1075	6
Respiratory	985	6
Public Health	978	6
Neurological	936	5
Anaesthesiology	927	5
Diet and Nutrition	917	5
Metabolic and Endocrine	906	5
Physical Medicine / Rehabilitation	671	4
Infection	664	4
Reproductive Health and Childbirth	611	3
Oral and Gastrointestinal	609	3
Surgery	478	3
Renal and Urogenital	380	2
Eye	352	2
Inflammatory and Immune System	345	2
Other	341	2
Injuries and Accidents	302	2
Skin	294	2
Stroke	283	2
Alternative and Complementary Medicine	262	1
Blood	211	1
Human Genetics and Inherited Disorders	95	1
Ear	85	0
Emergency medicine	38	0
Total	17523	100

Condition code #2, top 10 (ANZCTR)

	Freq	%
Other	12634	72
Other muscular and skeletal disorders	665	4
Diabetes	643	4
Depression	502	3
Health promotion/education	502	3
Obesity	475	3
Other diet and nutrition disorders	435	2

Other diseases of the mouth, teeth, oesophagus, digestive system including liver and colon	430	2
Anaesthetics	423	2
Other respiratory disorders / diseases	420	2
Physiotherapy	402	2
Total	17531	100

There are 155 condition codes, hence we just show the top ten. Conditions outside the top ten have been grouped together as 'other'.

Condition as free text, top 10 (ANZCTR)

	Freq	%
Other	15873	91
obesity	291	2
depression	263	2
stroke	204	1
type 2 diabetes	176	1
breast cancer	143	1
malaria	127	1
prostate cancer	125	1
asthma	114	1
anxiety	108	1
osteoarthritis	107	1
Total	17531	100

There are 8,395 conditions as free text, hence we just show the top ten.

This field is free text and so relies on what researchers write. Some categories could likely be combined, e.g., "prostate cancer" and "prostate cancer patients".

Study status (ANZCTR)

Study status (ANZCTR)

	Freq	%
Completed	7175	41
Not yet recruiting	4259	24
Recruiting	4072	23
Active, not recruiting	1042	6
Stopped early	602	3
Withdrawn	303	2
Suspended	78	0
Total	17531	100

Sponsors and funding (ANZCTR)

Top institutes (ANZCTR)

	Freq	%

Other	15163	86
University of Sydney	494	3
University of Queensland	305	2
Monash University	252	1
Royal Adelaide Hospital	244	1
Curtin University	231	1
University of Melbourne	204	1
University of South Australia	163	1
Deakin University	159	1
Griffith University	158	1
Royal Melbourne Hospital	158	1
Total	17531	100

There are 7,831 institutes, hence we just show the top ten.

Funding (ANZCTR)

Number of funders (ANZCTR)

	Freq	%
0	767	4
1	14046	80
2	2044	12
3	456	3
4	133	1
5	43	0
6	16	0
7	7	0
8	9	0
9	6	0
11	2	0
12	1	0
14	1	0
Total	17531	100

Funding type (ANZCTR)

	Freq	%
Government body	4570	22
University	3930	19
Commercial sector/Industry	3706	18
Charities/Societies/Foundations	3561	17
Hospital	2602	13
Self funded/Unfunded	922	4
Other Collaborative groups	646	3
Other	630	3
Total	20567	100

This includes multiple results per study.

Funders, top 10 (ANZCTR)

	Freq	%
Other	17879	87
NHMRC	1736	8
ARC	189	1
Health Research Council of New Zealand	170	1
Monash University	124	1
University of Otago	119	1
Curtin University	77	0
The University of Sydney	76	0
University of Sydney	70	0
Deakin University	65	0
Health Research Council	63	0
Total	20568	100

This includes multiple results per study.

There are 10,588 funders, hence we just show the top ten.

Sample size (ANZCTR)

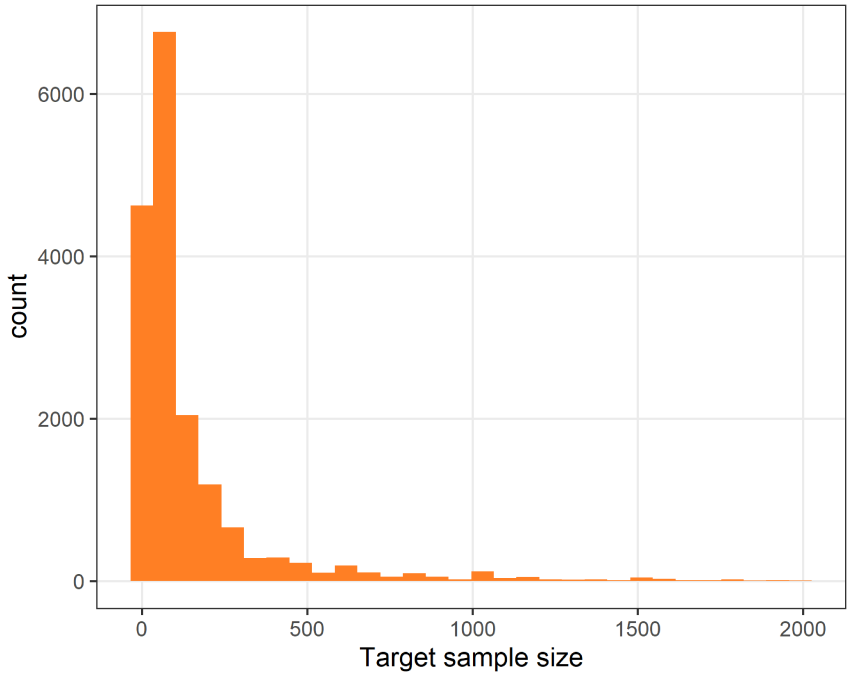
Target and achieved sample size (ANZCTR)

	Achieved	Target
Min	0	0
Q1	26	31
Median	60	66
Q3	140	158
Max	139977	714306

The table shows summary statistics.

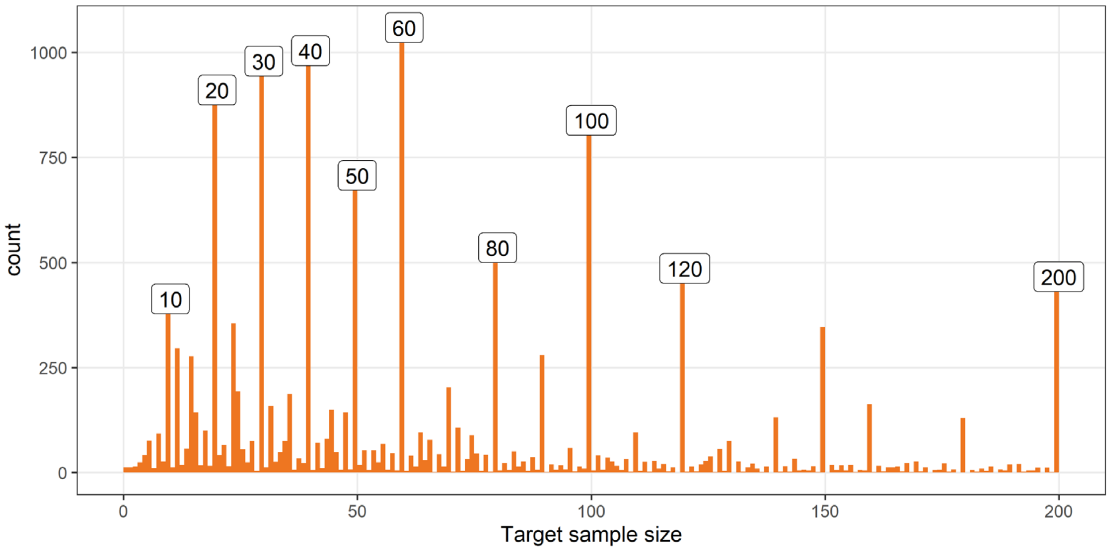
There were 2 studies with a zero target sample size.

Histogram of sample size (ANZCTR)



The plot above is for the 17,136 studies with a target sample size under 2,000.

Histogram looking for digit preference in target sample size (ANZCTR)



The plot above is for the 14,210 studies with a target sample size of 200 or fewer.

There is a strong digit preference at multiples of ten. The modal sample size is 60 and this is dominated by withdrawn studies.

Barnett AG, Glasziou P. *BMJ Open* 2021; 11:e053377. doi: 10.1136/bmjopen-2021-053377

Bayesian fractional polynomial model used to model the Bland–Altman limits of agreement

The Bayesian model applied to the Bland–Altman limits of agreement was:

$$\begin{aligned}
\log_e(d_j) &\sim t(\mu_j, \sigma_j^2, 4), \qquad j = 1, \ldots, N, \\
\mu_j &= \alpha_0 + \alpha_1 f_1(a_j), \\
\sigma_j^{-2} &= \exp\left[\beta_0 + \beta_1 f_2(a_j)\right], \\
\alpha_k &\sim N(0, 1000), \qquad k = 0, 1, \\
\beta_k &\sim N(0, 1000), \qquad k = 0, 1.
\end{aligned}$$

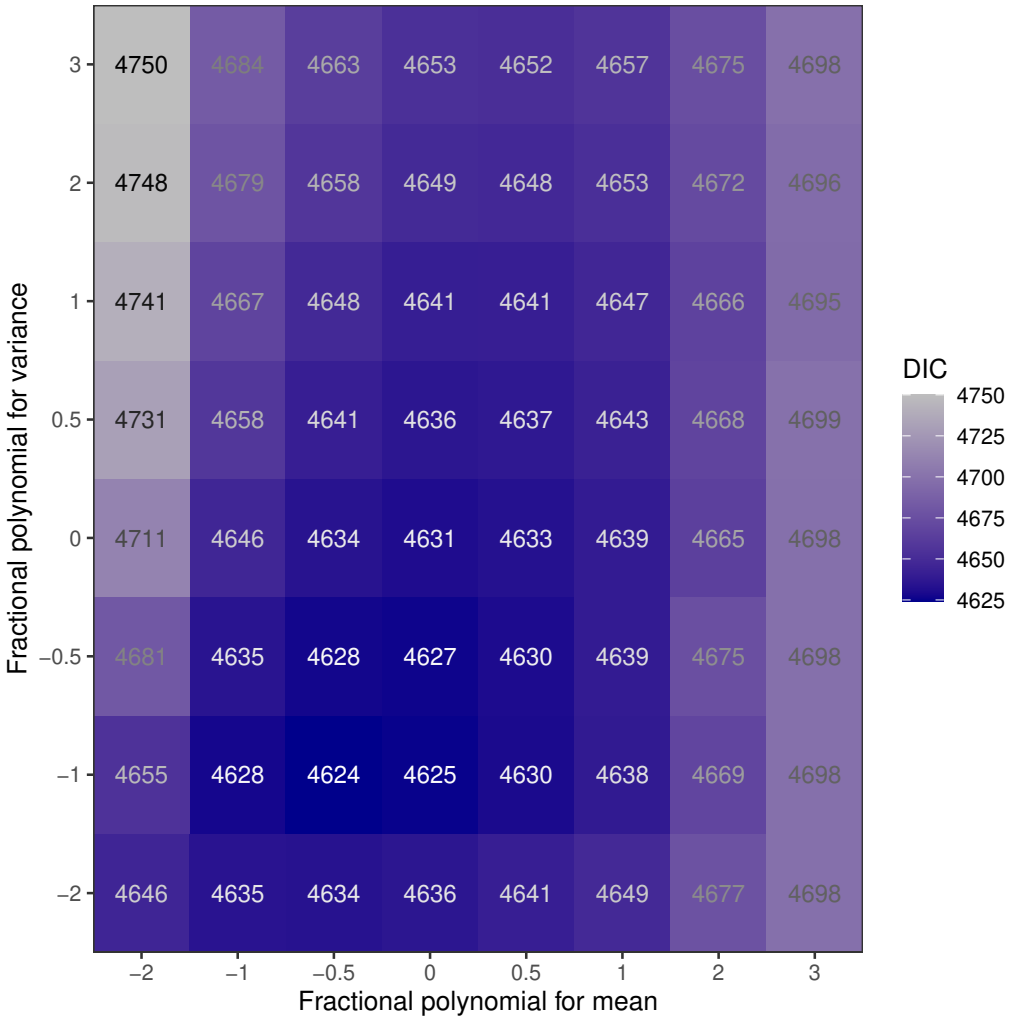
Where d_j is the difference in sample size for study j and there were N studies in total. This difference was modelled using a t-distribution with 4 degrees of freedom. The mean and variance were modelled using regression equations with an intercept and the study's average sample size a_j transformed using functions $f_1()$ and $f_2()$ which were fractional polynomials. We tested the same powers for the polynomials: $\{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$, but the mean and variance could have different best powers. The best pair of powers were chosen using the deviance information criterion.

The same model was used for the clinicaltrials.gov and ANZCTR databases.

Barnett AG, Glasziou P. *BMJ Open* 2021; 11:e053377. doi: 10.1136/bmjopen-2021-053377

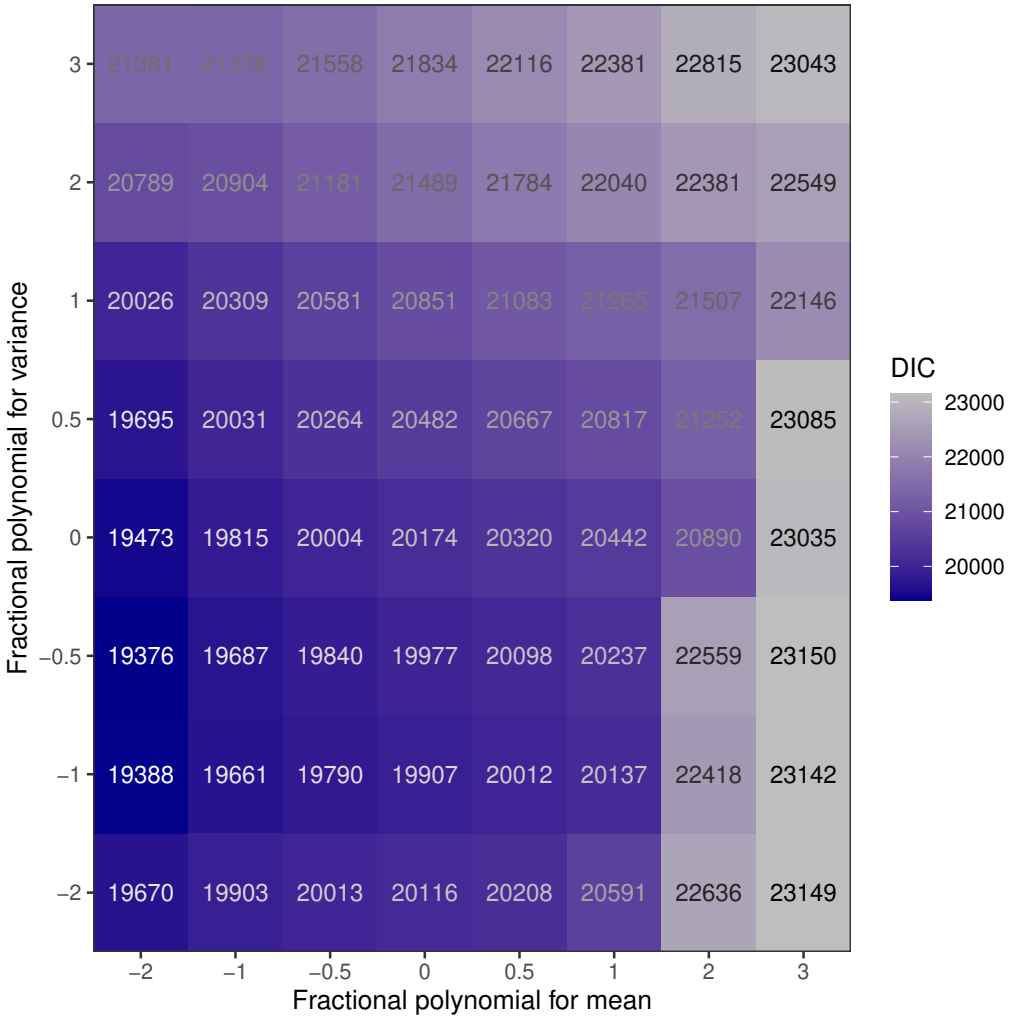
Supplement Figure 2: Deviance information criterion for the fractional polynomial models used to model the Bland–Altman limits of agreement

ANZCTR



The best model is for a fractional polynomial power –0.5 for the mean and –1 for the variance.

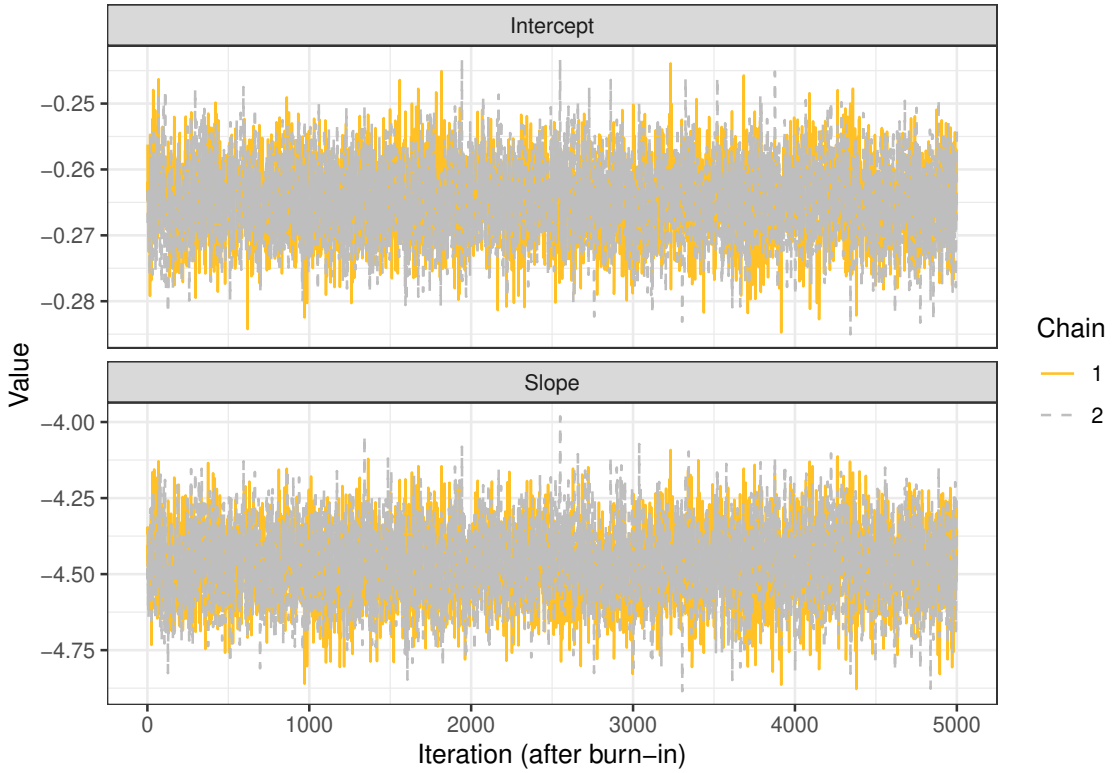
clinicaltrials.gov



The best model is for a fractional polynomial power –2 for the mean and –0.5 for the variance.

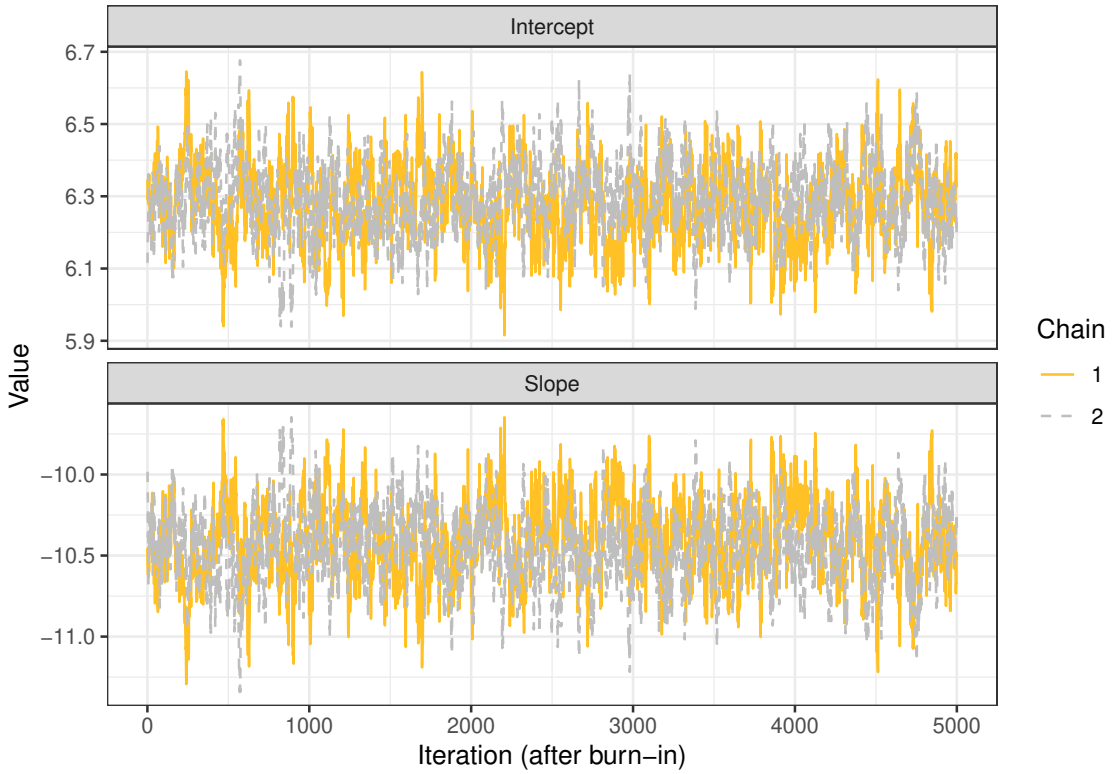
Supplement Figure 3: Plots of the Markov Chain Monte Carlo estimates for the Bland-Altman limits of agreement

Plot for the two parameters used to estimate the mean



Barnett AG, Glasziou P. *BMJ Open* 2021; 11:e053377. doi: 10.1136/bmjopen-2021-053377

Plot for the two parameters used to estimate the inverse-variance



All the plots show good convergence and mixing.

Barnett AG, Glasziou P. *BMJ Open* 2021; 11:e053377. doi: 10.1136/bmjopen-2021-053377

Supplement table 2: Percent changes (and 95% condidence intervals) in the ratio of the target to actual sample size.

Database	**group**	**label**	**PC**	**CI**
ANZCTR	Status	Stopped early	-70.2	-71.5 to -68.8
		Withdrawn	-98.8	-99.2 to -98.2
clinicaltrials.gov	continuous	Double target sample size	-5.7	-5.9 to -5.5
		Double the number of arms	8.4	7.3 to 9.4
		Double the number of conditions	-1.3	-2.1 to -0.6
		Double the number of secondary outcomes	2.4	2.1 to 2.7
		Trend per 5 years	-4.3	-4.8 to -3.8
	Healthy volunteers	Yes	4.2	3.3 to 5.2
	Intervention	Radiation	-3.7	-6.3 to -1.0
	Phase	Early Phase 1	-12.0	-14.9 to -9.0
		Phase 1	-11.0	-12.1 to -9.9
		Phase 1/2	-8.6	-10.3 to -6.8
		Phase 2	-5.7	-6.7 to -4.7
	Sponsor	Industry	18.1	17.1 to 19.1
		NIH	-7.0	-9.8 to -4.1
		US Federal agencies	-10.3	-13.3 to -7.3
	Purpose	Health Services Research	10.4	7.8 to 13.0
		Prevention	6.4	5.1 to 7.7
		Screening	16.3	11.5 to 21.4
	Masking	Double	3.2	2.0 to 4.4
		Quadruple	4.2	2.9 to 5.5
		Single	3.3	2.1 to 4.5
	Assignment	Single group	-2.9	-4.0 to -1.9
	Allocation	Non-Randomized	-2.2	-3.4 to -1.0
	Status	Suspended	-64.5	-68.7 to -59.7
		Terminated	-73.2	-73.5 to -72.9
		Withdrawn	-99.9	-99.9 to -99.9

PC = percent change. CI = confidence interval. Variables selected using elastic net.

Reference groups are: Healthy volunteers = No; Phase = Phase 3; Sponsor = Other; Purpose = Treatment; Masking = Blinded; Assignment = Parallel; Status = Completed. All other variables are continuous or binary meaning the reference group is 'No'.

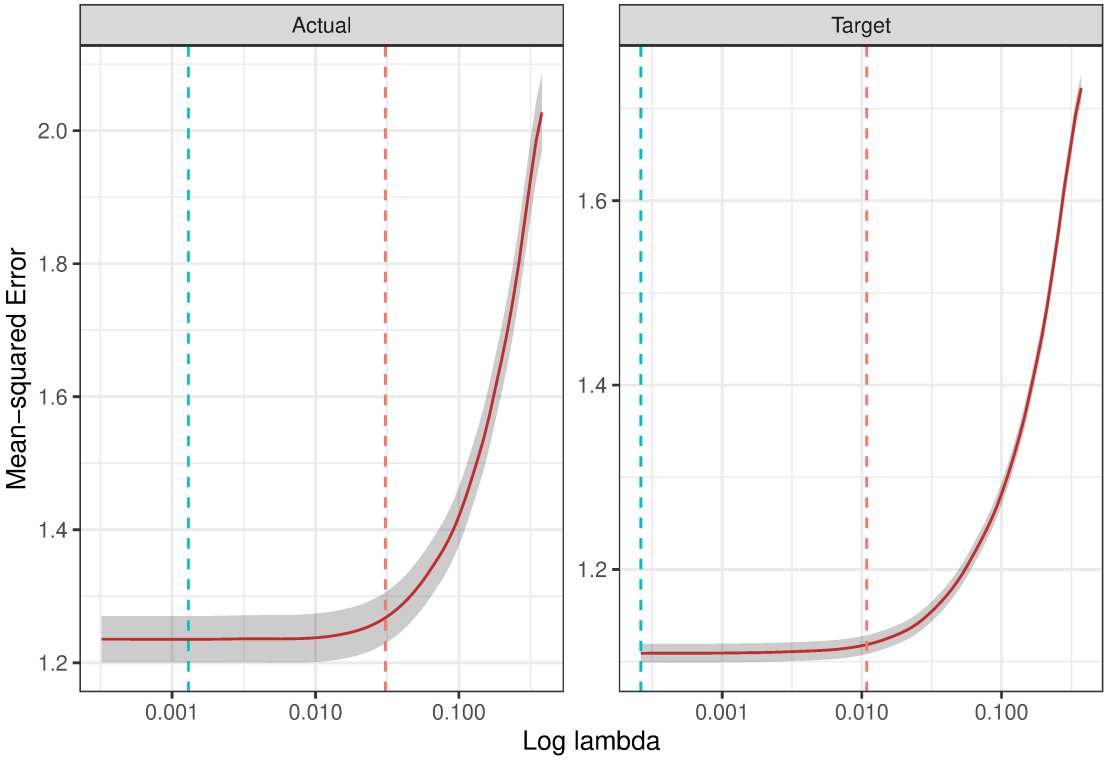
Barnett AG, Glasziou P. *BMJ Open* 2021; 11:e053377. doi: 10.1136/bmjopen-2021-053377

Supplement Figure 4: Cross-validated errors for the elastic net models

The models were fitted using the "glmnet" package in R. We used 10-fold cross-validation.

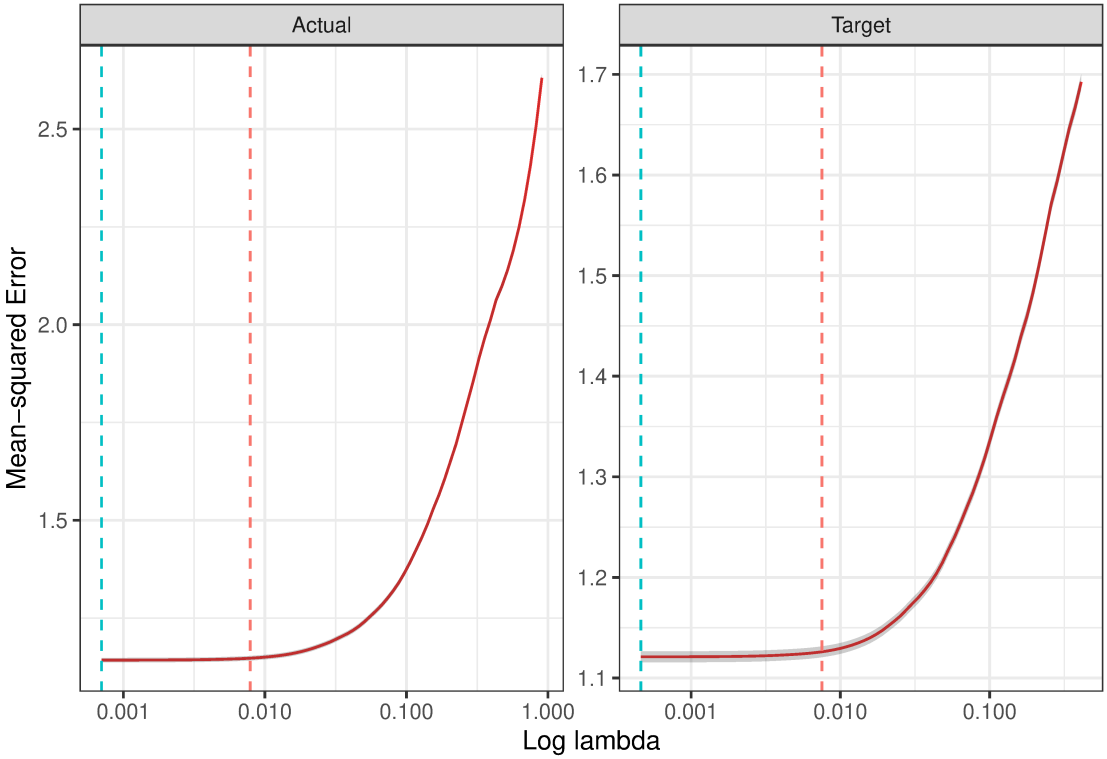
Models of the target and actual sample size

a) ANZCTR



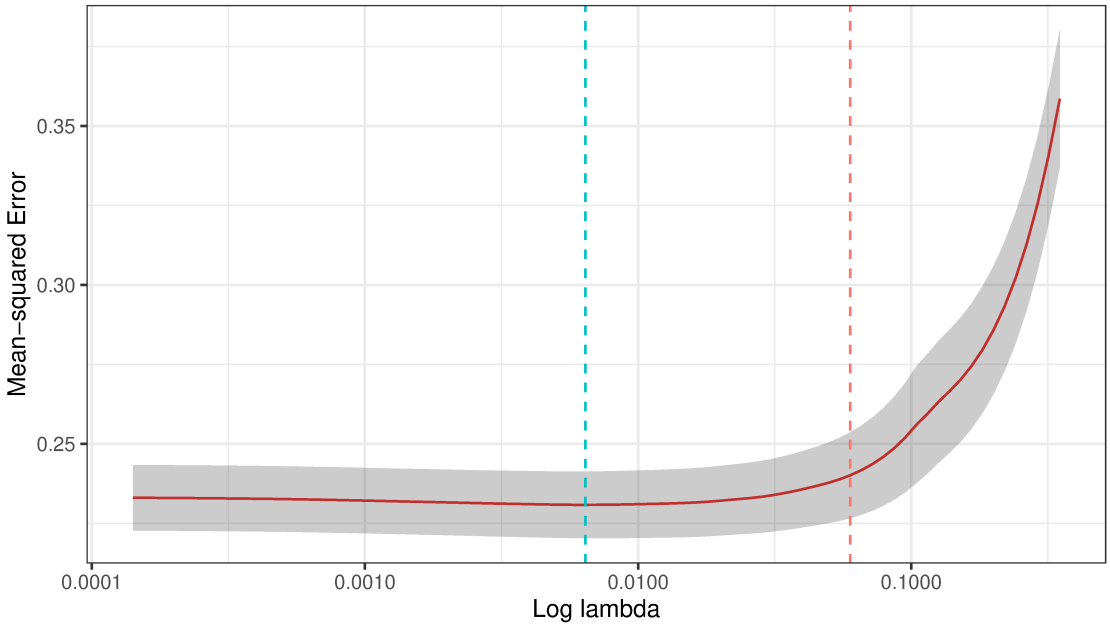
The shaded area is the 95% confidence interval for the error. The blue vertical dotted line is the minimum mean-squared error. The red vertical dotted line is the minimum mean-squared error plus 1 standard error.

b) clintrials.gov



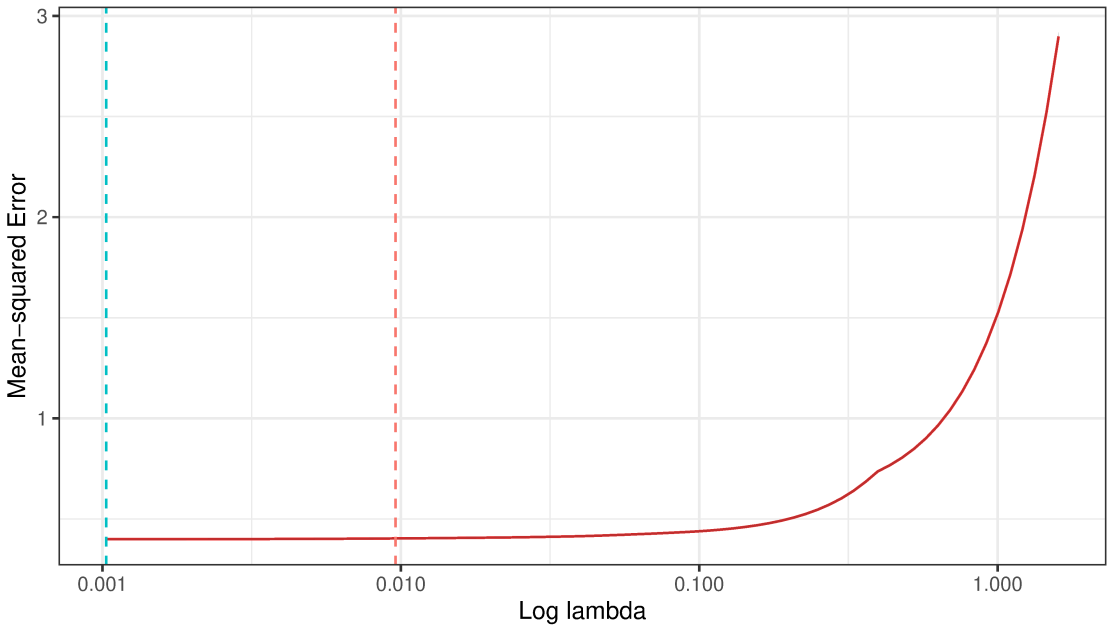
Models of the target to actual sample size ratio

a) ANZCTR



The confidence intervals are much wider here because the sample size for the ratio models is smaller.

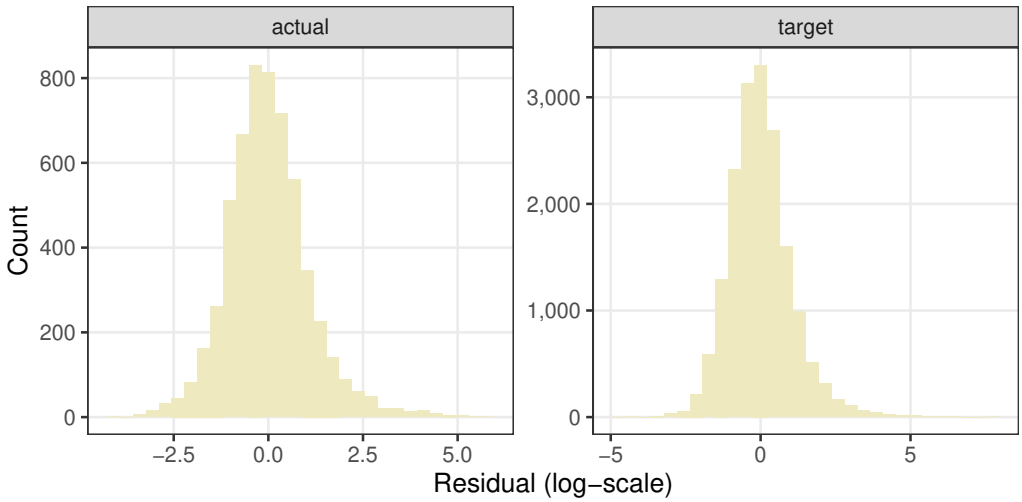
b) clinicaltrials.gov



Supplement Figure 5: Histograms of residuals

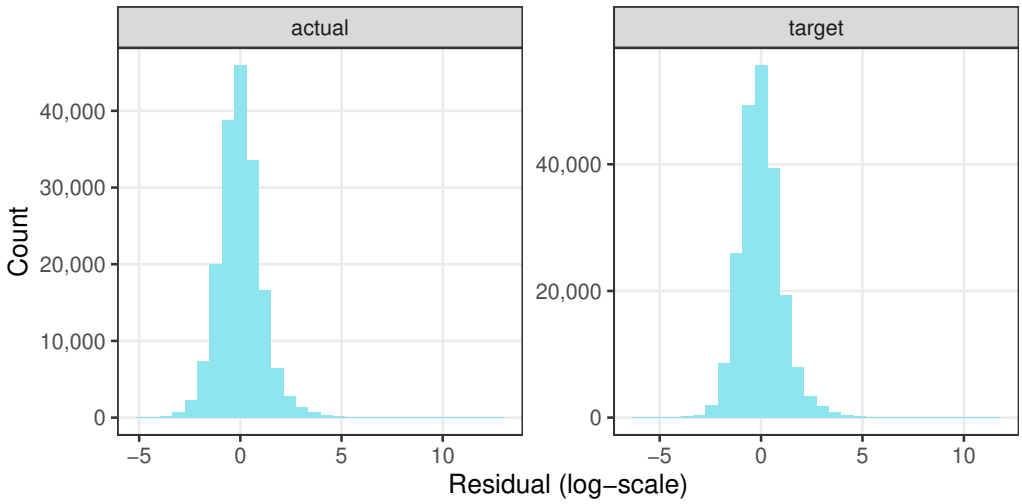
Models of target and actual sample size.

ANZCTR



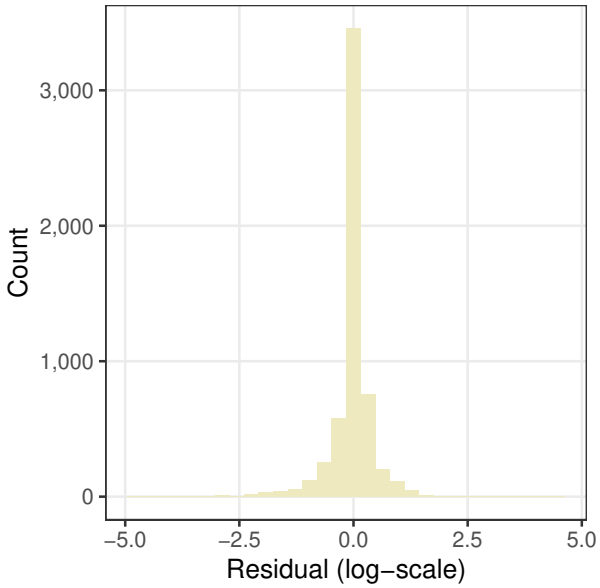
The plot shows the residuals for the actual and target sample size, and for the interventional and observational studies. The x- and y-axes vary by panel.

clintrials.gov

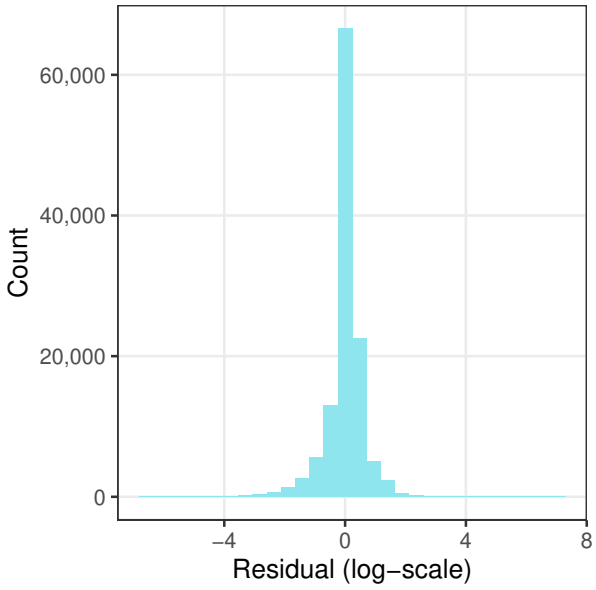


Models of the target to actual sample size ratio

ANZCTR



clinicaltrials.gov



Supplement table 3: Top ten words in the brief titles of clinicaltrials.gov for studies in women only and men only

gender	word	Count
Female	cancer	6,081
	breast	4,719
	women	4,591
	study	4,292
	patients	3,253
	treatment	2,420
	effect	1,748
	ovarian	1,674
	versus	1,626
	trial	1,596
Male	study	4,175
	prostate	3,022
	cancer	2,693
	healthy	2,311
	patients	1,920
	safety	1,604
	male	1,467
	men	1,192
	subjects	1,129
	effect	1,091

Stop words were removed and all text was converted to lower case.

Barnett AG, Glasziou P. *BMJ Open* 2021; 11:e053377. doi: 10.1136/bmjopen-2021-053377