




BMJ Open Reporting of and explanations for under-recruitment and over-recruitment in pragmatic trials: a secondary analysis of a database of primary trial reports published from 2014 to 2019

Pascale Nevins ^{1,2} Stuart G Nicholls,² Yongdong Ouyang,^{2,3} Kelly Carroll,² Karla Hemming,⁴ Charles Weijer ⁵, Monica Taljaard ³

To cite: Nevins P, Nicholls SG, Ouyang Y, *et al.* Reporting of and explanations for under-recruitment and over-recruitment in pragmatic trials: a secondary analysis of a database of primary trial reports published from 2014 to 2019. *BMJ Open* 2022;**12**:e067656. doi:10.1136/bmjopen-2022-067656

► 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-2022-067656>).

Received 23 August 2022
Accepted 16 November 2022



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to
Dr Monica Taljaard;
mtaljaard@ohri.ca

ABSTRACT

Objectives To describe the extent to which pragmatic trials underachieved or overachieved their target sample sizes, examine explanations and identify characteristics associated with under-recruitment and over-recruitment.

Study design and setting Secondary analysis of an existing database of primary trial reports published during 2014–2019, registered in ClinicalTrials.gov, self-labelled as pragmatic and with target and achieved sample sizes available.

Results Of 372 eligible trials, the prevalence of under-recruitment (achieving <90% of target sample size) was 71 (19.1%) and of over-recruitment (>110% of target) was 87 (23.4%). Under-recruiting trials commonly acknowledged that they did not achieve their targets (51, 71.8%), with the majority providing an explanation, but only 11 (12.6%) over-recruiting trials acknowledged recruitment excess. The prevalence of under-recruitment in individually randomised versus cluster randomised trials was 41 (17.0%) and 30 (22.9%), respectively; prevalence of over-recruitment was 39 (16.2%) vs 48 (36.7%), respectively. Overall, 101 025 participants were recruited to trials that did not achieve at least 90% of their target sample size. When considering trials with over-recruitment, the total number of participants recruited in excess of the target was a median (Q1–Q3) 319 (75–1478) per trial for an overall total of 555 309 more participants than targeted. In multinomial logistic regression, cluster randomisation and lower journal impact factor were significantly associated with both under-recruitment and over-recruitment, while using exclusively routinely collected data and educational/behavioural interventions were significantly associated with over-recruitment; we were unable to detect significant associations with obtaining consent, publication year, country of recruitment or public engagement.

Conclusions A clear explanation for under-recruitment or over-recruitment in pragmatic trials should be provided to encourage transparency in research, and to inform recruitment to future trials with comparable designs. The issues and ethical implications of over-recruitment should be more widely recognised by trialists, particularly when designing cluster randomised trials.

STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ This analysis included a broad range of randomised controlled trials with pragmatic orientation across diverse clinical areas.
- ⇒ Some trial characteristics used as explanatory factors in the analysis were poorly reported and may have been vulnerable to misclassification.
- ⇒ As verifying all sample size calculations was impossible, we had to assume that the target sample size in the report had been appropriately determined under valid assumptions for the primary objectives of the trial.

INTRODUCTION

An essential step in designing a randomised controlled trial (RCT) is calculating the required sample size. Reporting guidelines require authors to report their planned sample size and how it was determined, whether interim analyses were used to determine early stopping or continuation of the recruitment beyond the planned study end, and the explanation ‘if the actual sample size differed from the originally intended sample size for some other reason (eg, because of poor recruitment or revision of the target sample size)’.¹ Recruitment difficulties may lead to increased costs, delays in findings becoming available or even premature closure of the trial, which could render it unable to detect a potentially important treatment effect.^{2 3} Strategies that have been proposed to facilitate recruitment and retention include use of novel trial designs, open-label designs, novel approaches to informed consent, enhanced information provided to prospective participants and incentives for participation^{4 5}; reducing barriers such as stringent eligibility criteria and demands on

participants and staff^{6,7}; and the use of routinely collected data for outcome assessment.⁸ Many of these strategies are consistent with features of ‘pragmatic trials’, which are trials designed deliberately to promote applicability of results to patients, clinicians and decision makers in usual care conditions.^{9,10}

Although under-recruitment in RCTs is a recognised challenge,⁴ over-recruitment (ie, exceeding the planned number of participants) is rarely measured, although both under-recruitment and over-recruitment have ethical implications. If a study under-recruits, the contingent benefits of the research may not be realised and patients may have been exposed to research risks and burdens without the consequent benefits to society, undermining the social value of the research. Under-recruitment also represents an opportunity cost: resources might have been better directed towards other socially valuable research. An opportunity cost may also apply in the case of over-recruitment if the additional inclusion of participants is unjustified. Equally, if not adequately justified, over-recruitment raises the possibility that patients are exposed unnecessarily to research risks and burdens. Over-recruitment may occur inadvertently, especially in cluster randomised trials (CRTs)—a design often chosen to advance pragmatic aims.¹¹

Ethical implications of excessive cluster sizes in CRTs have been previously discussed.¹² Due to the presence of intracluster correlation, CRTs generally require larger sample sizes than comparably designed individually randomised trials, yet, once a certain level of saturation is reached, any further increases in the number of participants per cluster has minimal if any contribution to study power.¹³ However, over-recruitment may also occur more explicitly: with CRTs, power depends to a greater extent on the number of clusters than the number of participants. Thus, sample size calculation procedures may be focused on the required number of clusters given an anticipated number of eligible individuals per cluster over the planned duration of the trial. If more than the anticipated number of individuals are available, and especially when routinely collected data are used for outcome assessment, all available patients over the duration of the study may be included without re-estimation of the sample size.¹³ Furthermore, many CRTs do not have formal interim analyses and even when such interim analyses are conducted, investigators may be reluctant to reduce the target sample size partway through the trial.¹⁴

Within a large sample of self-labelled pragmatic trials, our objectives were to (A) describe ‘recruitment outcomes’, that is, the extent to which trials underachieved or overachieved their target sample sizes, (B) compare recruitment outcomes between cluster randomised and individually randomised trials, (C) examine any provided explanations for under-recruitment or over-recruitment and (D) identify characteristics associated with under-recruitment or over-recruitment in pragmatic RCTs.

METHODS

Identification of trials

This was a secondary review and analysis of an existing database of trials established as part of a broader study of the ethical and design considerations of pragmatic trials.¹⁵ Details concerning the search, eligibility and screening of trials have been published¹⁶ and are summarised in online supplemental appendix A1. In brief: an electronic search filter was developed to identify 4337 primary reports of trials more likely to be pragmatic in Ovid MEDLINE and published January 2014–April 2019.¹⁷ As in two previously published analyses of this database,^{18,19} we focused on the subset of 415 that were registered in ClinicalTrials.gov (CT.gov), a registry of clinical studies run by the US National Library of Medicine and that were clearly labelled by trial authors as ‘pragmatic’ anywhere in the title, abstract or main text. To be eligible for the present review, both a target and achieved sample size had to be available.

Data elements

Data elements had been downloaded from CT.gov and MEDLINE, or manually extracted from the trial reports as part of previously published reviews.^{17,20,21} Additional items were extracted as part of the present review. The data extraction form used to guide manual extractions is available in online supplemental appendix A2.

Previously downloaded data were the type of intervention (drug, device, biological/vaccine, procedure/surgery, educational/behavioural or other) from CT.gov and the trial registration number, journal name, title, author list and year of publication from MEDLINE.

From our previously published review of informed consent in pragmatic trials,²¹ we obtained the trial design (individually or cluster randomised, region of study recruitment (reclassified for this study as: USA and/or Canada only, Europe only, other high-income countries only, at least one low-income and middle-income country (LMIC) or multiple high-income regions) and journal impact factor. We also obtained information about individual informed consent, classified as obtained, not obtained (or a waiver of consent) or no information. From our review examining how claims of pragmatism were justified,¹⁸ we obtained the number of centres (multi-centre, single centre or unclear); type of setting (primary care, hospital/specialist care, nursing homes/long-term care, communities/residential areas, workplaces, schools or other); the use of patient or public engagement in the research; and exclusive use of routinely collected data for outcome assessment. Patient or public engagement was defined as ‘meaningful and active collaboration in governance, priority setting, conducting research and knowledge translation’ and was identified by searching the full text of the manuscript, author affiliations and the acknowledgements and funding sections for evidence of engagement. Exclusive use of routinely collected data was classified as outcome assessment solely from registries, electronic health records and administrative databases.

Target and achieved sample sizes were extracted by two reviewers per trial (PN and YO). Disagreements between reviewers were resolved through discussion with MT. The target sample size was extracted from the sample size section of the final trial report and included adjustment for attrition, if reported. When the target sample size was not clearly stated in the final report, it was extracted from the protocol, if available. Protocols were previously identified as part of a separate review of these trials.²⁰ We chose to extract target sample sizes from the final report (as opposed to from the protocol or CT.gov registration) because protocols were not available for all reports and because in a preliminary investigation, target sample sizes registered in CT.gov were found to be unreliable (eg, counting the number of clusters rather than participants).

At the request of a reviewer, we additionally extracted information on the statistical significance of the results for the primary outcome(s), classifying each trial as all primary outcomes significant, no primary outcomes significant, mixture of significant and non-significant primary outcomes, no primary outcomes identified and unclear.

Classification of recruitment outcomes

The ratio of achieved sample size over target sample size was calculated for each trial as a measure of the degree to which the trial achieved its target sample size. If less than 90% of the target sample size was achieved this was considered 'under-recruitment' and if more than 110% of the target sample size was achieved this was considered 'over-recruitment'. These boundaries were chosen prior to analysis to be comparable to those used in previous reviews,^{6 22 23} and allow room for trivial under-recruitment and over-recruitment. We also examined cut-points of $\pm 30\%$ and $\pm 50\%$ of the target sample size to provide a more granular perspective on extreme recruitment outcomes.

For trials that recruited less than 90% or more than 110% of their target sample size, we extracted whether the final report acknowledged the respective under-recruitment or over-recruitment and captured any provided explanations as text. A statement about the trial size being 'large' or 'small' without reference to the target sample size was not considered an acknowledgement. Statements about 'recruitment challenges' or about inclusion of 'all eligible participants' without clarification or elaboration were not considered explanations, as these were used by under-recruiting and over-recruiting trials alike. For over-recruiting CRTs, the source of the over-recruitment (cluster size, number of clusters or both) was also extracted.

Analysis

Categorical variables were described with frequencies and percentages. Continuous variables were described with median and IQR (Q1–Q3) and/or sum and SD. A

component bar chart was used to compare prevalence of under-recruitment or over-recruitment between cluster and individually randomised trials. Explanations for under-recruitment and over-recruitment were grouped into common themes.

To describe variation in recruitment outcomes across trial characteristics, χ^2 tests of association were conducted between the three-level categorical outcome (under 90%, 90%–110% and over 110%) and each of the eight trial characteristics of interest. These characteristics, predefined based on availability, were publication year, unit of randomisation (cluster vs individual), geographical region of recruitment, type of intervention, use of routinely collected data, whether individual informed consent had been obtained, use of patient or public engagement and journal impact factor. The rationale for considering each of these characteristics is described in online supplemental appendix A3. Continuous characteristics were dichotomised as below or above the median for all trials. To preserve degrees of freedom, categorical variables were recoded for analysis: the geographical categories 'other high-income countries only' and 'multiple high-income regions' were combined, and intervention type was dichotomised as educational/behavioural versus clinical, mixture or other. Where the exclusive use of routinely collected data or patient/public engagement in the research was unclear, these were classified as no use. To analyse associations with obtaining consent, we compared studies which indicated that consent had been obtained with studies that either explicitly reported no consent or did not state anything about consent. A similar approach was used in our previous review.²¹ This was thought to be appropriate as it is likely that if consent had been obtained, authors would have stated so. A sensitivity analysis was conducted excluding trials in which no information about consent was reported. Three trials with missing journal impact factors were categorised as below the median, as journals with missing impact factors likely have lower impact.

To examine the independent contributions of these trial characteristics to recruitment outcomes, a multivariable exploratory multinomial logistic regression analysis was conducted. We included all eight variables of interest in the multivariable model regardless of statistical significance—no stepwise variable selection was used. A post hoc supplementary analysis, stratified by unit of randomisation, was conducted to examine whether these characteristics were differentially associated with recruitment outcomes in cluster versus individually randomised trials. This analysis was exploratory and did not adjust for multiplicity.

A level of significance of 5% was chosen a priori for all analyses. Analyses were performed using SAS Studio V.3.8 on SAS V.9.4 Software (SAS Institute).

Patient and public involvement

No patients or members of the public were involved in this review of published RCTs.

RESULTS

Identification of trials

Among the 415 previously analysed trials, 340 (82.1%) had target sample sizes available in the final reports and another 33 (8.0%) provided target sample sizes in an accessible protocol, however, one of these final reports only stated the target sample size and not the achieved sample size. Thus, in total, 372 trials (89.9%) had both a target and an achieved sample size available.

A flow diagram describing the identification of trials for the present review is presented in online supplemental appendix A1.

Characteristics of trials

Table 1 presents characteristics of the 372 included trials. More trials used individual randomisation (241, 64.8%) than cluster randomisation (131, 35.2%). Trials most often recruited in the USA and/or Canada (166, 44.6%) or in Europe (133, 35.8%); 56 (15.1%) took place in at least one LMIC. The most common settings were hospital or specialist care (174, 46.8%) with relatively fewer in public health settings such as communities or residential areas (37, 9.9%) and the majority (288, 77.4%) were multicentre trials. The most common type of intervention was educational or behavioural (144, 38.7%). Only 63 (16.9%) used exclusively routinely collected data. Individual informed consent was obtained in 289 (77.7%), and 35 (9.4%) reported patient or public engagement.

Sample size and recruitment

Sample size and recruitment outcome ratios are presented in table 2 and figure 1. Across all trials, the median (Q1–Q3) target sample size was 514 (250–1402) and the achieved sample size was 505 (250–1615). As expected, the median (Q1–Q3) target size was larger for CRTs than for individually randomised trials: 1200 (586–3960) vs 360 (220–800). The median ratio (achieved/target) was 1.00 (0.99–1.04) for individually randomised trials and 1.05 (0.94–1.33) for CRTs. Overall, 214/372 (57.5%) achieved their recruitment targets ($\pm 10\%$). The prevalence of under-recruitment was 71 (19.1%) overall: when comparing individually versus CRTs the prevalence was 41 (17.0%) vs 30 (22.9%). The prevalence of over-recruitment was 87 (23.4%): when comparing individually versus CRTs, the prevalence was 39 (16.2%) vs 48 (36.7%), respectively. Among the CRTs, 35 (26.7%) exceeded their recruitment target by more than 30%. Overall, 101 025 participants were recruited to trials that did not achieve at least 90% of their target sample size. When considering trials with over-recruitment, the total number of participants recruited in excess of the target was a median (Q1–Q3) 319 (75–1478) per trial for an overall total of 555 309 more participants than targeted.

Table 1 General trial characteristics of N=372 self-declared pragmatic trials included in this analysis

Characteristic	Frequency (%)
Publication year	
2014–2015	92 (24.7)
2016–2017	149 (40.1)
2018–2019	131 (35.2)
Unit of randomisation	
Individually randomised trial	241 (64.8)
Cluster randomised trial	131 (35.2)
Country or region of trial recruitment	
USA and/or Canada only	166 (44.6)
Europe only	133 (35.8)
At least one low-income or middle-income country	56 (15.1)
Other high-income countries only	9 (2.4)
Multiple high-income regions	8 (2.2)
Type of setting	
Hospital or specialist care	174 (46.8)
Primary care	113 (30.4)
Communities, residential areas	37 (9.9)
Nursing homes, long-term care	7 (1.9)
Workplaces, schools	5 (1.3)
Other (eg, online, mixture of the above)	36 (9.7)
Single centre or multicentre?	
Single centre	75 (20.2)
Multicentre	288 (77.4)
Unclear	9 (2.4)
Type of experimental intervention* (CT.gov)	
Educational or behavioural	144 (38.7)
Drug	50 (13.4)
Procedure/surgery	35 (9.4)
Device	31 (8.3)
Biological/vaccine	3 (0.8)
Other†	130 (34.9)
Exclusively routinely collected data collected?	
Yes	63 (16.9)
No	302 (81.2)
Unclear	7 (1.9)
Individual consent obtained?	
Yes, consent obtained	289 (77.7)
No, waiver of consent or consent not obtained	59 (15.9)
Not reported	24 (6.5)
Any patient or public engagement reported?	
Yes	35 (9.4)
No	332 (89.2)
Unclear	5 (1.3)

Continued

Table 1 Continued

Characteristic	Frequency (%)
Journal impact factor‡	
Median (Q1–Q3)	5.4 (3.5 to 19.0)
*Non-mutually exclusive categories.	
†Includes dietary supplement, radiation or genetic.	
‡Three journal impact factors were not available.	

Table 3 reports the prevalence of acknowledgements of under-recruitment and over-recruitment; quoted and classified explanations from both under-recruitment and over-recruiting trials are provided in online supplemental appendices A4 and A5. Under-recruiting trials commonly acknowledged that they did not achieve their planned targets (51, 71.8%), with the majority of these (38/51, 74.5%) providing an explanation. Common explanations were fewer eligible participants than anticipated (10, 26.3%) and resource constraints (9, 23.7%). On the other hand, over-recruiting trials did not commonly acknowledge exceeding their target sample size (only 11/87, 12.6%), with 10 providing an explanation. Power or sample size calculation details were not always complete enough to assess the source of over-recruitment, but where it was possible to determine (38/48; 79.2%), most CRTs with excessive sample sizes exclusively had a larger number of participants per cluster than targeted (26, 68.4%).

The post hoc analysis of statistical significance revealed that among the 71 under-recruiting trials, 31 (43.7%) obtained a statistically significant result on at least one of their primary outcomes, compared with 124 (57.9%) among 214 trials recruiting within 10% of their target sample size and 47 (54.0%) among the 87 over-recruiting trials.

Factors associated with recruitment outcomes

Table 4 presents the results of the χ^2 tests of association exploring factors associated with under-recruitment or over-recruitment. Unit of randomisation, type of intervention, use of routinely collected data, obtaining consent and journal impact factor were significantly associated with recruitment outcomes when considered on their own, but publication year, geographical region and use of patient or public engagement were not. The post hoc sensitivity analysis excluding studies not reporting consent did not result in any substantive changes to our results.

The results of the multivariable multinomial logistic regression analysis are presented in **table 5**. After accounting for all other characteristics, CRTs had significantly higher odds of under-recruitment (OR 2.68 (95% CI 1.39 to 5.15)), while trials published in higher impact factor journals had significantly lower odds of under-recruitment (OR 0.36 (95% CI 0.20 to 0.64)). When considering over-recruitment, CRTs (OR 2.8 (95% CI 1.51 to 5.17)), trials of educational/behavioural interventions

(OR 2.27 (95% CI 1.28 to 4.01)) and trials using routinely collected data (OR 2.74 (95% CI 1.36 to 5.54)) had significantly higher odds of over-recruitment. Publication year, region of trial recruitment, informed consent and use of patient or public engagement in the research had no significant association with recruitment outcomes in either direction. Trials published in higher impact factor journals had lower odds of over-recruitment, but the CI slightly overlapped with 1.

The results from the supplementary analyses stratified by unit of randomisation are presented in online supplemental appendix B. Due to quasi-complete separation of points (resulting from small frequencies in some cells), the analysis of individually randomised trials excluded patient or public engagement as a covariate, while the analysis of CRTs collapsed regions of trial recruitment into three categories: USA/Canada only, Europe only or other. Although CIs around the estimated ORs were wider, results were consistent with those obtained from the overall analyses and substantive conclusions did not change.

DISCUSSION

Statement of principal findings

Among 372 self-declared pragmatic trials with target and achieved sample sizes available, over half recruited to within $\pm 10\%$ of their target sample size, approximately one in five failed to achieve at least 90% of their target, while close to one in four recruited more than 110% of their target. While prevalence of under-recruitment was similar in cluster randomised and individually randomised trials, over-recruitment was substantially more prevalent in CRTs. Most under-recruiting trials provided an explanation, but few of the over-recruiting trials acknowledged or explained the excess. Most over-recruiting CRTs enrolled more than the planned number of patients per cluster (as opposed to more than the planned number of clusters). In multivariable analyses, cluster randomisation and lower journal impact factor were important characteristics associated with both under-recruitment and over-recruitment, while exclusive use of routinely collected data and educational/behavioural interventions were important characteristics associated with over-recruitment.

Strengths and weaknesses of the study

Important strengths of our study include the large sample size and wide range of pragmatic trials. By also including trials declared as ‘pragmatic’ only in the main text, we were able to access a greater and broader sample of pragmatic trials than if we had relied on the title and abstract alone. This range of trial designs and interventions allowed us to consider associations with several trial characteristics.

Our study had some limitations. The absolute number of trials with some characteristics was low and our results are therefore vulnerable to type II error. All statistical significance tests should be interpreted with caution: our

Table 2 Target and achieved sample sizes and prevalence of under-recruitment and over-recruitment among individually randomised and cluster randomised trials

Characteristic	Individually randomised trials (N=241)	CRTs (N=131)	All trials (N=372)
Sample sizes: median (Q1, Q3)			
Target sample size	360 (220, 800)	1200 (586, 3960)	514 (250, 1402)
Achieved sample size	353 (212, 806)	1463 (522, 4626)	505 (250, 1615)
Difference (achieved target)	1.0 (−7.0, 17.0)	37.0 (−87.0, 421.0)	2.5 (−14.5, 50.5)
Ratio (achieved/target)	1.00 (0.99, 1.04)	1.05 (0.94, 1.33)	1.01 (0.96, 1.09)
Recruitment: frequency (%)			
Recruited under target			
(Ratio <0.9)	41 (17.0)	30 (22.9)	71 (19.1)
Recruited to within 10% of target			
(0.9≤ratio≤1.1)	161 (66.8)	53 (40.5)	214 (57.5)
Recruited over target			
(Ratio>1.1)	39 (16.2)	48 (36.7)	87 (23.4)
No in under-recruiting trials			
	N=41	N=30	N=71
Median (Q1, Q3)	328 (182, 556)	636 (308, 1997)	396 (200, 1179)
Sum (SD)	31 989 (1521.5)	69 036 (3946.8)	101 025 (2889.4)
Absolute no over-recruited			
	N=39	N=48	N=87
Median (Q1, Q3)	98 (45, 334)	706.5 (225.5, 2883)	319 (79, 1478)
Sum (SD)	19 228 (1009.1)	536 081 (37 302.6)	555 309 (28 098.9)

CRTs, cluster randomised trials.

analyses were exploratory and we performed no adjustment for multiple comparisons. The characteristics of interest included in our analyses were limited by data availability, and some were vulnerable to misclassification due to poor reporting (eg, patient/public involvement,

routinely collected data, consent). Our ability to examine changes over time was limited to the small interval spanned by the available sample. As suggested by Schroen *et al*,²³ a trial's actual ability to address its primary endpoint (independent of percentage of target achieved) may be a

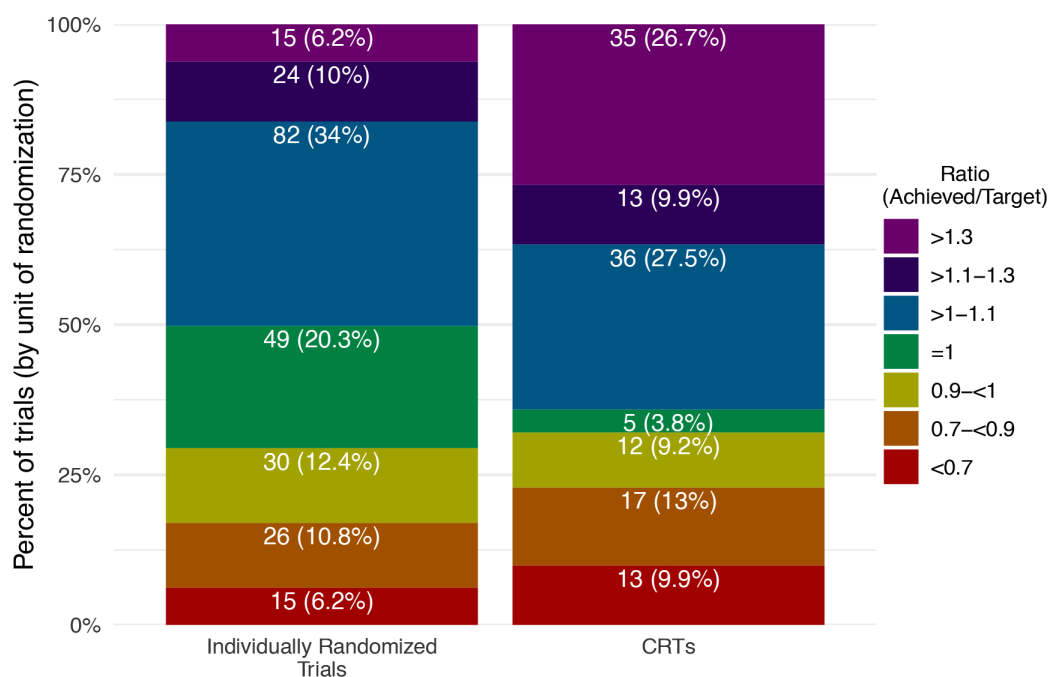


Figure 1 Distribution of ratio (achieved/target) sample size among the N=372 included trials. Percentage (of 241 individually randomised or 131 cluster randomised trial (CRT), respectively) of trials in each ratio category is indicated.

Table 3 Acknowledgement of under-recruitment and over-recruitment and explanations

Under-recruitment (N=71 trials recruiting <90% of target)	Frequency (%)
Does the final report acknowledge that the target sample size was not achieved?	
Yes	51 (71.8)
No	20 (28.2)
If yes, do they give an explanation? (N=51)	
Yes	38 (74.5)
No	13 (25.5)
Does the trial report formal termination as an explanation? (N=38)	
Yes, termination for futility	6 (15.8)
Yes, termination for harm	3 (7.9)
Over-recruitment (N=87 trials recruiting >110% of target)	Frequency (%)
Does the final report acknowledge that the target sample size was achieved with excess?	
Yes	11 (12.6)
No	76 (87.4)
If Yes, do they give an explanation? (N=11)	
Yes	10 (90.9)
No	1 (9.1)
If CRT, what was the source of over-recruitment? (N=48)	
No of clusters	3 (7.9)
No of participants per cluster	26 (68.4)
Both	9 (23.7)
Insufficient information	10 (26.3)

more valuable measure of ‘success’. As anticipated, the proportion of trials reaching statistical significance on the primary outcome(s) was lower among under-recruiting trials than among trials with recruitment to within 10% of their target sample size. However, it was not higher among over-recruiting trials relative to those recruiting within 10%. We did not determine the extent to which the planned sample size was appropriate for the primary objectives and thus, the extent to which over-recruitment or under-recruitment was harmful; our analysis assumes that the target sample size was determined appropriately. If target sample sizes were too small, for example, if they were determined based on an anticipated difference, as opposed to a true clinically important difference, over-recruitment may alleviate a concern of underpowered trials. Our analysis did not distinguish trials formally terminated after interim analysis for futility, safety or effectiveness. Finally, we focused on a set of trials in which authors explicitly used the label ‘pragmatic’ to describe their trial; however, ‘pragmatic’ is a dichotomous indicator of a concept that exists on a continuum, and as previously discussed with respect to our sample of trials, its use is frequently not explicitly justified.¹⁸

Comparison with other studies

Previous reviews have focused on under-recruitment although direct comparison with our results is challenging as many reviews considered RCTs in general, rather than specifically trials labelled as pragmatic. Definitions of recruitment outcomes have also varied. Reviews in the UK examining ‘recruitment success’ (defined as achieving >80% of recruitment target) found a prevalence of 55% in 114 multicentre trials published during 1994–2002²⁴ and 78% in 73 trials published during 2002–2008² (compared with 89% in our review at a corresponding cutpoint). Among 151 individually randomised Health Technology Assessments published during 2004–2016, 79% achieved at least 80% of their recruitment target.²⁵ Among phase 2 and 3 intervention trials closed in 2011, 80% (2051/2577) achieved 90% of their recruitment target before closing or termination⁶ (compared with 81% in our review at the corresponding cutpoint).

A previous review examining under-recruitment in RCTs in general found under-recruitment to be associated with more eligibility criteria, using an active control and non-industry (public) funding, while multicentre trials had more recruitment meeting targets.⁶ We did not explore these characteristics in our review. Previous work has found that behavioural interventions are associated with recruitment to the target sample size^{26,27}; this is supported by our review, which found trials evaluating educational or behavioural interventions had lower odds of under-recruitment and higher odds of over-recruitment. Similarly, the use of exclusively routinely collected data in the research showed a strong association with exceeding the recruitment target in our review. There is some previous research showing patient engagement increases the odds of meeting enrollment goals, but we were unable to demonstrate this in our analysis.²⁸ A Cochrane systematic review examining methods to increase recruitment identified eight studies examining whether modified consent had an impact on recruitment rates: only one using an opt-out procedure showed increased recruitment.⁵ A study on obtaining consent in acute stroke trials found no significant improvement in recruitment yield using a waiver of consent.²⁹ Our bivariable tests of association show that trials that do not obtain consent are more likely to over-recruit, however, in an exploratory multivariable regression model, there was no significant association between consent and either under-recruitment or over-recruitment after accounting for other trial characteristics. Although CIs were wide, indicating considerable uncertainty, is it possible that the need to obtain individual informed consent is overstated as a barrier in pragmatic trials.

Recommendations and conclusions

Pragmatic trials aim to recruit diverse populations and yield results that are more applicable to the population who would receive the intervention outside the trial. Characteristics common to more pragmatic trials, such as using routinely collected data for outcome assessment,

Table 4 χ^2 tests of association with recruitment outcomes (N=372 trials)

	Recruitment (frequency, row %)			P value
	Under<90%	Within±10%	Over>110%	
	N=71	N=214	N=87	
Publication year				0.8438
2014–2016	31 (19.9)	87 (55.8)	38 (24.4)	
2017–2019	40 (18.5)	127 (58.8)	49 (22.7)	
Unit of randomisation				<0.0001
Individually randomised	41 (17.0)	161 (66.8)	39 (16.2)	
Cluster randomised	30 (22.9)	53 (40.5)	48 (36.6)	
Region of trial recruitment				0.3684
USA/Canada only	33 (19.9)	92 (55.4)	41 (24.7)	
Europe only	26 (19.6)	81 (60.9)	26 (19.6)	
At least one LMIC	7 (12.5)	31 (55.4)	18 (32.1)	
Other high-income countries	5 (29.4)	10 (58.8)	2 (11.8)	
Type of intervention				0.0001
Educational/behavioural	19 (14.0)	69 (50.7)	48 (35.3)	
Clinical, mixture or other	52 (22.0)	145 (61.4)	39 (16.5)	
Exclusively routinely collected data				<0.0001
Yes	9 (12.9)	29 (41.4)	32 (45.7)	
No or unclear	62 (20.5)	185 (61.3)	55 (18.2)	
Individual consent obtained				0.0002
Yes	56 (19.3)	179 (61.7)	54 (19.0)	
No or not reported	15 (18.3)	35 (42.7)	33 (39.0)	
Patient or public engagement				0.7829
Yes	6 (15.0)	24 (60.0)	10 (25.0)	
No or unclear	65 (19.6)	190 (57.2)	77 (23.2)	
Journal impact factor*				0.0011
Below median (<5.4)	46 (24.7)	90 (48.4)	50 (26.9)	
Above median (≥5.4)	25 (13.4)	124 (66.7)	37 (19.9)	

*Missing journal impact factors were classified as below the median. LMIC, low-income and middle-income country.

may facilitate achieving the target sample size (although the known limitations of routinely collected data should always be considered).^{30 31} Despite concerns about obtaining informed consent from participants being a barrier to achieving the target sample size, our analysis was unable to demonstrate such an association. We suggest that trialists and research ethics committees carefully weigh the benefits of foregoing informed consent in light of these results, as informed consent is central to respecting patient autonomy and upholding public trust in research.

We identified over-recruitment as a prevalent but under-recognised issue in pragmatic trials, especially in cluster randomised designs, trials using routinely collected data for outcome assessment and those evaluating educational/behavioural interventions. We recommend that trialists consult with an experienced statistician when

implementing a cluster randomised design, to ensure the trial is adequately powered without excessive inclusion of participants.¹³ Trialists should consider the nature of the intervention and likely recruitment rate when designing the trial and study duration. Availability of routinely collected data in advance of a trial also presents an opportunity to obtain more accurate estimates of the potential sample size available for the trial and should be considered in justifying the trial design and study duration. Clear reporting of recruitment and retention rates in trial publications is essential to inform the design and conduct of future RCTs. Data safety monitoring committees and trial steering committees monitoring the progress of a trial should not exclusively focus on whether the target sample size is achieved, but also consider the potential benefits and risks of overinclusion, with particular attention to trials with the above characteristics. If it seems

Table 5 Multivariable multinomial logistic regression analysis of recruitment outcomes (N=372)

Characteristic	<90% vs within target		>110% vs within target	
	OR (95% CI)	P value	OR (95% CI)	P value
Publication year				
2014–2016	1		1	
2017–2019	0.82 (0.46 to 1.44)	0.486	0.87 (0.50 to 1.50)	0.609
Unit of randomisation				
Individually randomised	1		1	
Cluster randomised	2.68 (1.39 to 5.15)	0.0032	2.79 (1.51 to 5.17)	0.0011
Region of trial recruitment				
USA/Canada only	1		1	
Europe only	0.79 (0.42 to 1.50)	0.4765	1.06 (0.55 to 2.02)	0.8689
At least one LMIC	0.53 (0.20 to 1.40)	0.1994	1.28 (0.58 to 2.80)	0.5453
Other high-income countries	1.72 (0.51 to 5.84)	0.3854	1.06 (0.20 to 5.51)	0.9498
Type of experimental intervention				
Clinical, mixture or other	1		1	
Educational/behavioural only	0.61 (0.32 to 1.17)	0.1374	2.27 (1.28 to 4.01)	0.0049
Exclusively routinely collected data				
No or unclear	1		1	
Yes	0.66 (0.27 to 1.59)	0.3536	2.74 (1.36 to 5.54)	0.0048
Individual consent obtained				
No or not reported	1		1	
Yes	0.90 (0.39 to 2.06)	0.8076	0.72 (0.34 to 1.52)	0.3874
Patient or public engagement				
No or unclear	1		1	
Yes	0.59 (0.22 to 1.56)	0.2876	0.99 (0.42 to 2.31)	0.9821
Journal impact factor				
Below median (<5.4)*	1		1	
Above median (≥5.4)	0.36 (0.20 to 0.64)	0.0005	0.58 (0.34 to 1.02)	0.0583

ORs represent odds of either over or under-recruitment relative to recruitment within 10% of the target .
 *Three missing journal impact factors were classified as below the median.
 LMIC, low-income and middle-income country .

that the target sample size is likely to be substantially exceeded, the benefits of stopping the trial or decreasing the trial duration should be considered. If stopping the trial is undesirable, continued recruitment should be adequately justified. For example, consideration can be paid to whether continued recruitment can contribute towards power for key secondary or safety outcomes or important prespecified subgroup analyses. It is also important to examine the original sample size calculation and target difference to ensure that it represents a ‘true’ minimum important or plausible difference, and that potential attrition and non-adherence are accounted for. These recommendations, which also apply to trials without prospective recruitment, can help improve the social value of the research.

Reviews across many disciplines have shown that sample sizes in RCTs are poorly justified, incompletely reported and often impossible to replicate.^{32–35} We recommend

that trialists report complete details of their sample size justification, referencing the original target sample size as well as any changes made during the conduct of the trial, with the goal of promoting transparency in research. For CRTs, not only the planned and achieved total sample size but also the number of clusters and size per cluster should be reported. Journal editors and peer-reviewers across all journals should insist that authors provide a clear explanation when the achieved sample size is either higher or lower than planned.

One key area for future methodological development is approaches for meaningful engagement with patients and members of the public in trial design and protocol development. While prioritisation exercises have identified recruitment and retention as important areas of focus,³⁶ there has been limited work on methods to involve patients in numerical aspects of clinical trials.^{37 38} Involving patients in discussions around primary outcome

selection and target differences can contribute to more appropriate sample size justification and help improve the social value of the research. Finally, we note that the time frame of the database excludes the COVID-19 pandemic which began in 2020. Our analysis, thus, reflects trials unaffected by the pandemic. Future reviews may see COVID-19 as a dominant issue affecting recruitment and may identify new approaches to recruitment and trial design developed in light of the global pandemic.

Author affiliations

¹Department of Chemistry and Biomolecular Sciences, University of Ottawa Faculty of Science, Ottawa, Ontario, Canada

²Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Ontario, Canada

³School of Epidemiology and Public Health, University of Ottawa, Ottawa, Ontario, Canada

⁴Institute of Applied Health Research, University of Birmingham, Birmingham, UK

⁵Departments of Medicine, Epidemiology & Biostatistics, and Philosophy, Western University, London, Ontario, Canada

Twitter Pascale Nevins @pascalenevins and Charles Weijer @charlesweijer

Acknowledgements We thank Shelley Vanderhout, Jennifer Zhe Zhang and colleagues for contributing previous data extraction to the present analysis. We also thank three peer reviewers whose comments have contributed to substantial improvements to our manuscript.

Contributors PN: writing—original draft, data curation, formal analysis, visualisation. MT: conceptualisation, funding acquisition, methodology, supervision, writing—original draft. YO: data curation, writing—review and editing. KC: project administration, data curation, writing—review and editing. KH: methodology, writing—review and editing. SN: data curation, writing—review and editing. CW: writing—review and editing. MT is the guarantor.

Funding This work was supported by the Canadian Institutes of Health Research through the Project Grant competition (competitive, peer-reviewed), award number PJT-153045. MT is supported by the National Institute of Aging (NIA) of the National Institutes of Health under Award Number U54AG063546, which funds NIA Imbedded Pragmatic Alzheimer's Disease and AD-Related Dementias Clinical Trials Collaboratory (NIA IMPACT Collaboratory).

Disclaimer The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The funders had no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

Competing interests CW receives consulting income from Cardialen, Eli Lilly & Company, and Research Triangle Institute International. The other authors declare that they have no competing interests.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. Data and statistical code available on reasonable request to the first author.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially,

and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Pascale Nevins <http://orcid.org/0000-0002-4488-7366>

Charles Weijer <http://orcid.org/0000-0002-5510-1074>

Monica Taljaard <http://orcid.org/0000-0002-3978-8961>

REFERENCES

- Moher D, Hopewell S, Schulz KF, *et al*. Consort 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c869.
- Sully BGO, Julious SA, Nicholl J. A reinvestigation of recruitment to randomised, controlled, multicenter trials: a review of trials funded by two UK funding agencies. *Trials* 2013;14:166.
- Al-Shahi Salman R, Beller E, Kagan J, *et al*. Increasing value and reducing waste in biomedical research regulation and management. *Lancet* 2014;383:176–85.
- Caldwell PHY, Hamilton S, Tan A, *et al*. Strategies for increasing recruitment to randomised controlled trials: systematic review. *PLoS Med* 2010;7:e1000368.
- Treweek S, Pitkethly M, Cook J, *et al*. Strategies to improve recruitment to randomised trials. *Cochrane Database Syst Rev* 2018;2:MR000013.
- 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.
- Prescott RJ, Counsell CE, Gillespie WJ, *et al*. Factors that limit the quality, number and progress of randomised controlled trials. *Health Technol Assess* 1999;3:1–143.
- Armstrong RA, Mouton R, Hinchliffe RJ. Routinely collected data and patient-centred research in anaesthesia and peri-operative care: a narrative review. *Anaesthesia* 2021;76:1122–8.
- Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. *J Chronic Dis* 1967;20:637–48.
- Patsopoulos NA. A pragmatic view on pragmatic trials. *Dialogues Clin Neurosci* 2011;13:217–24.
- Taljaard M, Goldstein CE, Giraudeau B, *et al*. Cluster over individual randomization: are study design choices appropriately justified? Review of a random sample of trials. *Clin Trials* 2020;17:253–63.
- Hemming K, Taljaard M, Forbes G, *et al*. Ethical implications of excessive cluster sizes in cluster randomised trials. *BMJ Qual Saf* 2018;27:664–70.
- Hemming K, Eldridge S, Forbes G, *et al*. How to design efficient cluster randomised trials. *BMJ* 2017;358:j3064.
- Hemming K, Martin J, Gallos I, *et al*. Interim data monitoring in cluster randomised trials: practical issues and a case study. *Clin Trials* 2021;18:552–61.
- Taljaard M, Weijer C, Grimshaw JM, *et al*. Developing a framework for the ethical design and conduct of pragmatic trials in healthcare: a mixed methods research protocol. *Trials* 2018;19:525.
- Nicholls SG, Carroll K, Hey SP, *et al*. A review of pragmatic trials found a high degree of diversity in design and scope, deficiencies in reporting and trial registry data, and poor indexing. *J Clin Epidemiol* 2021;137:45–57.
- Taljaard M, McDonald S, Nicholls SG, *et al*. A search filter to identify pragmatic trials in MEDLINE was highly specific but lacked sensitivity. *J Clin Epidemiol* 2020;124:75–84.
- Taljaard M, Nicholls SG, Howie AH, *et al*. An analysis of published trials found that current use of pragmatic trial labels is uninformative. *J Clin Epidemiol* 2022;151:113–21.
- Vanderhout S, Fergusson DA, Cook JA, *et al*. Patient-reported outcomes and target effect sizes in pragmatic randomized trials in ClinicalTrials.gov: a cross-sectional analysis. *PLoS Med* 2022;19:e1003896.
- Nevins P, Vanderhout S, Carroll K, *et al*. Review of pragmatic trials found that multiple primary outcomes are common but so too are discrepancies between protocols and final reports. *J Clin Epidemiol* 2022;143:149–58.
- Zhang JZ, Nicholls SG, Carroll K, *et al*. Informed consent in pragmatic trials: results from a survey of trials published 2014–2019. *J Med Ethics* 2021;medethics-2021-107765.
- Korn EL, Freidlin B, Mooney M, *et al*. Accrual experience of national cancer Institute cooperative group phase III trials activated from 2000 to 2007. *J Clin Oncol* 2010;28:5197–201.
- Schroen AT, Petroni GR, Wang H, *et al*. Achieving sufficient accrual to address the primary endpoint in phase III clinical trials from U.S. Cooperative oncology groups. *Clin Cancer Res* 2012;18:256–62.

- 24 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.
- 25 Walters SJ, Bonacho Dos Anjos Henriques-Cadby I, Bortolami O, *et al.* Recruitment and retention of participants in randomised controlled trials: a review of trials funded and published by the United Kingdom health technology assessment programme. *BMJ Open* 2017;7:e015276.
- 26 Pemberton VL, Evans F, Gulin J, *et al.* Performance and predictors of recruitment success in National heart, lung, and blood Institute's cardiovascular clinical trials. *Clin Trials* 2018;15:444–51.
- 27 Bernardez-Pereira S, Lopes RD, Carrion MJM, *et al.* Prevalence, characteristics, and predictors of early termination of cardiovascular clinical trials due to low recruitment: insights from the ClinicalTrials.gov registry. *Am Heart J* 2014;168:213–9.
- 28 Crocker JC, Ricci-Cabello I, Parker A, *et al.* Impact of patient and public involvement on enrolment and retention in clinical trials: systematic review and meta-analysis. *BMJ* 2018;363:k4738.
- 29 Feldman WB, Kim AS, Josephson SA, *et al.* Effect of waivers of consent on recruitment in acute stroke trials: a systematic review. *Neurology* 2016;86:1543–51.
- 30 Oortwijn W, Sampietro-Colom L, Trowman R. How to deal with the Inevitable: generating real-world data and using real-world evidence for HTA purposes - From theory to action. *Int J Technol Assess Health Care* 2019;35:346–50.
- 31 Beletsky A, Lu Y, Patel BH, *et al.* Should we question the external validity of database studies? A comparative analysis of demographics. *Arthroscopy* 2019;35:2686–94.
- 32 Rutterford C, Taljaard M, Dixon S, *et al.* Reporting and methodological quality of sample size calculations in cluster randomized trials could be improved: a review. *J Clin Epidemiol* 2015;68:716–23.
- 33 Copsey B, Thompson JY, Vadher K, *et al.* Sample size calculations are poorly conducted and reported in many randomized trials of hip and knee osteoarthritis: results of a systematic review. *J Clin Epidemiol* 2018;104:52–61.
- 34 Kaushal P, Adenwalla SF, Lightfoot CJ, *et al.* Evaluation of the design, conduct and reporting of randomised controlled trials in the haemodialysis population: a scoping review and interview study. *BMJ Open* 2022;12:e058368.
- 35 Lee PH, Tse ACY. The quality of the reported sample size calculations in randomized controlled trials indexed in PubMed. *Eur J Intern Med* 2017;40:16–21.
- 36 Kearney A, Williamson P, Young B, *et al.* Priorities for methodological research on patient and public involvement in clinical trials: a modified Delphi process. *Health Expect* 2017;20:1401–10.
- 37 Goulao B, Poisson C, Gillies K. Patient and public involvement in numerical aspects of trials: a mixed methods theory-informed survey of trialists' current practices, barriers and facilitators. *BMJ Open* 2021;11:e046977.
- 38 Goulao B, Bruhn H, Campbell M, *et al.* Patient and public involvement in numerical aspects of trials (point): exploring patient and public partners experiences and identifying stakeholder priorities. *Trials* 2021;22:499.