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

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 published[16] 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 (multicentre, 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 ±30% and ±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 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 (±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±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

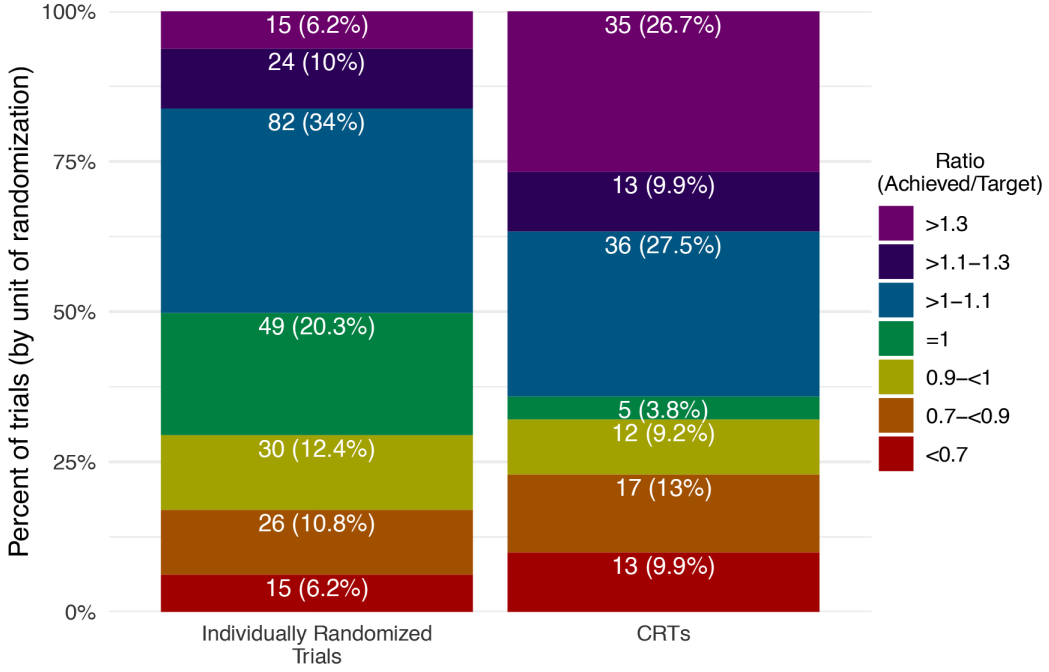


Figure 1 Distribution of ratio (achieved/target) sample size among the N=372 included trials. Percentage (of 241 individually randomised or 131 cluster randomised trial (CRT), respectively) of trials in each ratio category is indicated.

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 1994–2002[24] and 78% in 73 trials published during 2002–2008[2] (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 termination[6] (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 %)			
	Under<90%	Within±10%	Over>110%	
	N=71	N=214	N=87	P value
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 underrecognised 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 affectingv recruitment and may identify new approaches to recruitment and trial design developed in light of the global pandemic.

Author affiliations
[1]Department of Chemistry and Biomolecular Sciences, University of Ottawa Faculty of Science, Ottawa, Ontario, Canada
[2]Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Ontario, Canada
[3]School of Epidemiology and Public Health, University of Ottawa, Ottawa, Ontario, Canada
[4]Institute of Applied Health Research, University of Birmingham, Birmingham, UK
[5]Departments 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.

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

1 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.
2 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.
3 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.
4 Caldwell PHY, Hamilton S, Tan A, *et al*. Strategies for increasing recruitment to randomised controlled trials: systematic review. *PLoS Med* 2010;7:e1000368.
5 Treweek S, Pitkethly M, Cook J, *et al*. Strategies to improve recruitment to randomised trials. *Cochrane Database Syst Rev* 2018;2:MR000013.
6 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.
7 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.
8 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.
9 Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. *J Chronic Dis* 1967;20:637–48.
10 Patsopoulos NA. A pragmatic view on pragmatic trials. *Dialogues Clin Neurosci* 2011;13:217–24.
11 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.
12 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.
13 Hemming K, Eldridge S, Forbes G, *et al*. How to design efficient cluster randomised trials. *BMJ* 2017;358:j3064.
14 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.
15 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.
16 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.
17 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.
18 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.
19 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.
20 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.
21 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.
22 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.
23 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.

Appendix A: Supplementary Forms and Information

Appendix A1. Search filter, screening, and flow diagram of included trials

A1.1 Electronic search filter to identify pragmatic trials in Ovid MEDLINE[1]

#	Search Statement
	Trial design terms
1	(((pragmatic$ OR naturalistic OR real world OR real life OR unblinded OR unmasked OR cluster OR step$ wedge$ OR point of care OR factorial OR switchback OR switch back OR phase 4 OR phase IV) adj10 (study OR trial)) OR (practical trial OR effectiveness trial OR ((cluster$ or communit$) adj2 randomi$))).tw.
	Trial attribute terms
2	(general practice$ OR primary care OR registry based OR health record$ OR medical record$ OR EHR OR EMR OR administrative data OR routinely collected data OR (communit$ adj2 intervention$) OR quality improvement OR implementation OR decision support OR health service$ OR health system$ OR comparative effectiveness OR CER OR usual care OR evidence based OR practice guideline$ OR (guideline$ adj1 recommend$) OR knowledge translation OR health technology assessment OR HTA OR cost effectiveness OR process evaluation OR economic evaluation OR patient oriented).tw.
	Limit to records likely to be RCTs
3	randomized controlled trial.pt. OR ((comparative effectiveness OR randomi?ed) adj10 (trial OR study)).ti.
4	(comment on OR phase 1 OR phase I OR phase 2 OR phase II OR non-randomi?ed OR quasi-randomi?ed OR pseudo-randomi?ed).ti. OR (clinical trial, phase I OR clinical trial, phase II OR systematic review OR meta-analysis OR review OR editorial).pt.
	Include records tagged as pragmatic trials
5	pragmatic clinical trial.pt.
	Sensitivity-maximizing search (combines trial design terms or attribute terms with RCT terms)
6	((1 OR 2) AND (3 NOT 4)) OR 5
7	exp Animals/ NOT Humans/
8	6 NOT 7

RCT=randomized controlled trial

[1]Search filter published as: Taljaard M, McDonald S, Nicholls SG, Carroll K, Hey SP, Grimshaw JM, Fergusson DA, Zwarenstein M, McKenzie JE. A search filter to identify pragmatic trials in MEDLINE was highly specific but lacked sensitivity. J Clin Epidemiol. 2020 Aug;124:75-84. doi: 10.1016/j.jclinepi.2020.05.003. Epub 2020 May 11. PMID: 32407765.

A1.2 Screening: Inclusion and exclusion criteria for creation of the large database of pragmatic trials[2]

INCLUSION CRITERIA		**SPECIFICATION**
Health RCT	A health RCT evaluates interventions aimed at changing subjective or objective measures of individual or group health status, or of processes which lead to changes in health status. Health status is defined as a state of human wellbeing, in individuals or groups, with physical or mental health correlates.	Trial must use randomization (as opposed to quasi-randomization) and must be comparative (i.e., at least one intervention and one control arm, or at least two intervention arms for "comparative effectiveness research").
Health care RCT	A health care RCT is one aimed at evaluating changes in the delivery of services to changing health status, or to change processes in the delivery of care that are known to lead to changes in health status. This includes RCTs of treatment, prevention, health promotion, health knowledge or behaviour as well as studies of the implementation, acceptability, efficiency, equity or cost of interventions for treatment, prevention, promotion, health knowledge or behaviour change. These studies may target individual patients, groups of patients, communities or populations, and/or carers from family, community and health care systems.	Trial must have a target enrolment of at least 100 individuals
EXCUSION CRITERIA		**SPECIFICATION**
1	Not a randomized controlled trial	Trial must use randomization (as opposed to quasi-randomization) and must be comparative (i.e., at least one intervention and one control arm, or at least two intervention arms for "comparative effectiveness research").
2	RCT but with <100 target enrolment	Trial must have a target enrolment of at least 100 individuals
3	Study protocol or design paper	Study does not report trial outcomes or reports only baseline data. Includes study protocols and published statistical analysis plans.
4	Methods paper – not a trial report	Exclude if the article doesn't present any results from a trial, e.g., it may only describe the design of a trial without presenting results or it may talk about recruitment difficulties or intervention development only or only a statistical analysis plan. Some studies may be a trial themselves e.g. the trial is about the best consent approach within a trial and they test two consent methods.

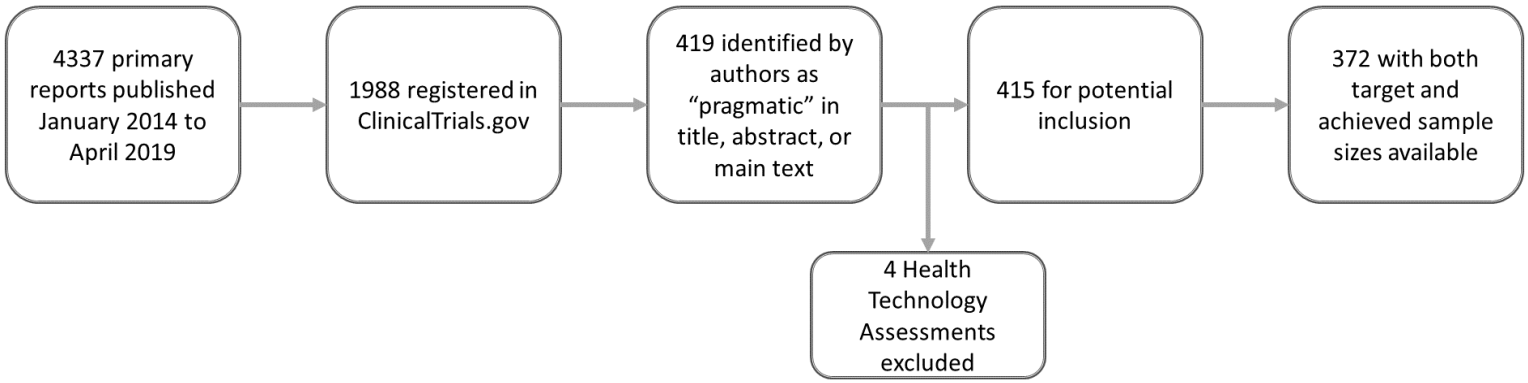
5	Not health research	Excluded trials include those of purely educational intervention in school-based settings that don't have a distinct health link. For example, a trial testing two different Mathematics curricula or trials assessing interventions to reduce bullying, but there is no direct health link (e.g., does not involve school nurses and does not assess health outcomes, e.g., anxiety.) There must be a clear link to health research. Other exclusion examples include trials testing new ways of sweeping the floor, or new administrative approaches of contacting patients, or recruiting patients into trials, or improving data quality, or new hiring practices. Unless there was a clear and direct route by which this would impact on patient health outcomes, the trial is not eligible (lack of direct relevance to patient health).
6	Pilot or feasibility study	If the trial is obviously labelled as a pilot or feasibility study, it is ineligible. If the conclusions simply refer to "feasibility", but also gives results for a primary health or health care outcome and it meets all the other criteria for inclusion the trial would be included.
7	Not a pragmatic intent	Trial that is obviously not of pragmatic orientation. For example, trials that focused on isolating a biological impact of an intervention without a clear clinical implication, or that did not assess clinical outcomes, were deemed more likely to not to have a pragmatic orientation.
8	Non-primary trial report	Any indication in the manuscript that it is not the primary trial report (meaning, the analysis of the primary trial outcome), indicates exclusion. Examples of non-primary reports include the analysis of an outcome clearly identified as a secondary trial outcome; subgroup analysis with primary trial results reported elsewhere; long-term follow-up (with primary endpoint reported previously), process analysis, mediation analysis, sensitivity analysis (if reported separately to primary outcome).**
9	Educational intervention of health professionals with no real patients or patient data	Studies randomizing clinicians to different vignettes and then surveying them to assess their response are excluded as they do not involve real patients. Other trials excluded here are: trials involving only manikins or simulations (e.g., for training providers to do CPR) or trials of purely educational (provider) interventions which do not measure patient data. For example, if a trial tests two different Medical curricula or outcomes are exclusively measures of clinical knowledge measured in clinicians or medical residents, the trial is not eligible.

10	Other (specify)	

** Trial registration information was used to facilitate identification of the primary trial report from among multiple publications from the same trial. Studies were also compared by first and senior authors and studies that reported the same trial were flagged. Articles flagged as potential non-primary reports were also scrutinized for any explicit statement referring to "primary results being published previously" or that the present article was presenting a secondary analysis. When multiple publications were associated with the same clinical trial registration, a decision had to be made about whether any could be considered the primary trial publication. In cases of uncertainty, the primary outcome in the registry was used to guide the decision. In the case of complex study designs, such as factorial designs with interventions reported separately, the article reporting on the first listed intervention or outcome reported in the registry was selected as the primary report. Finally, our search identified several Health Technology Assessment (HTA) reports, unique to studies funded by the UK National Institutes of Health Research. When the HTA report was the only report of the trial retrieved it was retained as the primary publication; otherwise, the associated journal publication of the trial was retained as the primary publication.

[2]Results of screening published in: Nicholls SG, Carroll K, Hey SP, Zwarenstein M, Zhang JZ, Nix HP, Brehaut JC, McKenzie JE, McDonald S, Weijer C, Fergusson DA, Taljaard M. 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 Sep;137:45-57. doi: 10.1016/j.jclinepi.2021.03.021. Epub 2021 Mar 28. PMID: 33789151.

A1.3 Flow diagram describing selection of studies for current secondary review from published work of 415 trial reports[3]



The search filter and screening (Appendix A1.1[1] and A1.2[2]) identified 4337 primary reports of trials more likely to be pragmatic published 2014-2019; of these, 1988 were registered in ClinicalTrials.gov. Trials that were explicitly labelled by authors as "pragmatic" accounted for 419 among this set. Excluding four trials published in Health Technology Assessment (and thus with larger word counts and scopes than other reports) provided a set of 415 trials for potential inclusion, as identified previously.[3]

[3]Vanderhout S, Fergusson DA, Cook JA, Taljaard M. Patient-reported outcomes and target effect sizes in pragmatic randomized trials in ClinicalTrials.gov: A cross-sectional analysis. PLoS Med. 2022;19(2):e1003896. Published 2022 Feb 8. doi:10.1371/journal.pmed.1003896

Appendix A2. Data extraction form

Extractor name

Trial characteristics

1. Unit of randomization

   1 Individual randomization
   2 Cluster randomization

2. Country of study recruitment (for identification of development status, use https://data.worldbank.org/country). *Select all that apply.*

   a. Canada
   b. USA
   c. UK
   d. European Union (EU) Country
   e. Australia or New Zealand
   f. Low- or Middle-Income Country (LMIC)
   g. Other High-Income Country

3. Multicenter or single center trial? Note: A "center" is a site (e.g., facility, hospital, clinic, or community) participating in the trial by recruiting participants into the trial; it may also be involved in delivering treatments and collecting data. Often, sites are stratification factors in the randomization, or may be the units of randomization. A multicenter trial can may have a smaller number of large centers, or a large number of small centers (e.g., if it is a rare disease).

   1 Multicenter
   2 Single center
   3 Unclear (explain)

4. Type of setting in which intervention was delivered:

   1 Primary care (e.g., primary care clinics, primary care providers)
   2 Hospital or specialist care (e.g., hospitals, hospital wards, specialist providers)
   3 Nursing homes, long-term care
   4 Community or residential setting
   5 Schools
   6 Workplaces
   7 Other (doesn't fit any existing categories, multiple, or any additional details)

5. Does the trial exclusively use routinely collected data for outcome assessment? (Note: includes registries, electronic health records, and administrative databases)

   1 Yes
   2 No

3 Unclear or other (specify)

6. Statement of consent obtained?
   1 Yes, statement for at least some aspect of the trial
   2 Yes, statement indicated no consent (or a waiver of consent)
   3 No statement about individual consent provided

7. Does the trial explicitly report on patient or public engagement or elicitation of patient or public perspectives in the research? Note: Patient engagement is defined as "meaningful and active collaboration in governance, priority setting, conducting research and knowledge translation". Hint: search for "engagement", "consultation", "advisory", "perspective", "stakeholder", "committee" or "interview". Also please specifically search the acknowledgements and/or funding sections for evidence of engagement.
   1 Yes
   2 No
   3 Unclear or other (please explain)

Sample Size and Recruitment Outcomes

R1. Total number of eligible participants enrolled (trial size at the start of the trial) (NOTE: Look in abstract or flow diagram and indicate the number of eligible participants randomized. In most RCTs, eligibility status is determined and consent obtained before randomization in which case we simply extract the number randomized. For a select number of trials, it may not be possible to establish eligibility before randomization in which case we extract the number after eliminating those that were later determined to be ineligible or who did not consent.
For CRTs we want the total number of eligible individuals contributing to the trial at the lowest level of the hierarchy, i.e., patients or citizens. If the CRT only reports aggregate data, e.g., "event rates" we want the population size contributing to the denominator – not the numerator. This is often the "offset term" in a Poisson regression.
In CRTs with multiple phases (e.g., cross-over, stepped wedge, pre and post design), we want the total number of unique participants contributing to the trial in all phases.
Leave the item missing if trial size is unknown or not reported but add an explanation in the comment box).

R2. What was the target sample size (number of individuals)? Note: This should be after accounting for attrition and be after any updates to the sample size (most recent calculated sample size if an update was provided). If none is provided in the trial report, check if there is a protocol and whether the protocol has a target sample size.
Record the source (final report or protocol) of the target sample size.

R3. Record the source (final report or protocol) of the extracted target sample size.

R4. Any comments about R1 or R2 (e.g., explanation of why you cannot determine the trial size or trial target sample size)?

R5. Calculate recruitment success: Number of enrolled participants (R1) divided by target sample size (R2).

R6. If the trial under-recruited (R5 < 0.9), does the trial report acknowledge, mention, or refer to the fact that the target sample size was not achieved? Search final report for "power", "sample", "recruit", "enrol", "slow", "terminate", etc. Check the abstract, sample size calculation section, first paragraph of results, and discussion.

1 Yes
2 No
3 Unclear

R7. If R6 = 1, is an explanation for under-recruitment given? If yes, please quote.

1 Yes (Quote)
2 No
3 Unclear or other

R8. If the trial under-recruited, was it due to formal termination (e.g., by the DSMB)?

1 Yes, termination due to harm
2 Yes, termination to futility
3 No indication

R9. If the trial overrecruited (R5 > 1.1), does the trial report acknowledge, mention, or refer to the fact that the target sample size was achieved with excess? Search final report for "power", "sample", "recruit", "enrol", etc. Check the abstract, sample size calculation section, first paragraph of results, and discussion.

1 Yes
2 No
3 Unclear

R10. If R9 = 1, is an explanation for over-recruitment given? If yes, please quote.

1 Yes (Quote)
2 No
3 Unclear or other

R11. If the trial overrecruited (R5 > 1.1) and it is a cluster randomized trial (Q1 = 2), what was the source of the over-recruitment?

1 Number of clusters increased
2 Number of participants per cluster increased
3 Both number of clusters and of participants increased

☐ 4 Not enough information (e.g., no number of clusters listed only total recruitment target)

Appendix A3. Characteristics of interest for testing association with recruitment outcomes with supporting rationale for inclusion

Trial Characteristic	Data source	Rationale for inclusion	Categorization
Publication Year	Downloaded from PubMed	To determine whether there were changes in recruitment outcomes over time.	First three years (2014-2016) versus last three years (2017-2019) of studies.
Unit of randomization	Manual extraction from the final reports	Individually randomized and cluster randomized trials have different sample size requirements and may face different barriers in achieving their target sample sizes. Informed consent procedures and types of intervention also vary in individually versus cluster randomized designs.	Individually randomized versus cluster randomized trials.
Region of trial recruitment	Manual extraction from the final reports	Trials may encounter differing funding, regulations, and attitudes towards research dependent on the geographical region or income-level of the region, which may influence recruitment success.	Studies recruiting exclusively in a single region (USA/Canada or Europe/UK) were categorized as their region. Studies with recruitment in multiple regions (16) were classified as either recruiting in at least one LMIC country or exclusively recruiting in high income countries (including USA, Canada, UK, Europe). Studies conducted in a single high-income country not mentioned above were categorized as recruiting in high income countries.
Type of Experimental Intervention	Downloaded from ClinicalTrials.gov	Educational or behavioural interventions may be more likely to be conducted in a public health setting and less invasive than clinical	Specifically and exclusively educational/behavioural versus any other intervention type, including clinical interventions, interventions

		interventions such as a drugs, surgeries and vaccines, and may be less burdensome on participants, thus facilitating recruitment. They may also be more likely to be CRTs.	that incorporated a mixture of educational/behavioural and a clinical intervention, and unspecified/other types (as chosen by trialists on CT.gov).
Use of routinely collected data	Manual extraction from the final reports	The use of routinely collected data is a less intrusive method of collecting data for research and may promote achievement of target sample sizes by reducing demands on participants and staff.	Exclusive use of routinely collected data versus clearly not, or unclear (e.g., not specified, impossible to determine whether it was routinely collected data in the setting of the trial or not)
Statement of consent obtained	Manual extraction from the final reports	Obtaining consent is often perceived as a barrier to successful recruitment.	Consent obtained for any part of the trial from participants versus a waiver of consent or no consent. When a mixture of consent approaches was used, categorization depended on whether consent was obtained for the intervention (= yes) or just for another part of the trial, e.g., analysis (= no).
Patient or public engagement in the research	Manual extraction from the final reports	Patient or public engagement in the research may promote recruitment approaches and interventions that are most relevant to patients or the public, and thus a more successful recruitment.	A clear indication of patient or public involvement in the research versus no indication, unclear, or specific acknowledgement of the non-use of patient/public engagement.
Journal Impact Factor	Downloaded from Journal Citation Reports (as of 2018). When unavailable, obtained from SCImago Journal and Country Rank (SJR) or Google searches.	Trials with perceived successful recruitment may be more likely to be published in high impact factor journals.	Above versus below the median journal impact factor in the sample of trials.

Appendix A4. Explanations for under-recruitment

Number	Explanation	Categorized Reason
1	Due to a slow recruitment rate, we have been unable to recruit the number of patients required by the power calculations	Infeasibility: slow recruitment rate
2	Data and Safety Monitoring Committee recommended the trial close due to low accrual. [...] Even with this pragmatic design, enrollment to the trial was terminated before reaching the planned sample size due to the slow rate of accrual. Two reasons may explain the lower than anticipated accrual. One, advances in the local and systemic management of primary breast cancer has lowered the incidence of ILRR in recent years, limiting the pool of eligible patients. Secondly, many oncologists were not genuinely uncertain and believed there was sufficient evidence from non-randomised series and randomised trials in other clinical settings to determine whether or not to administer chemotherapy for ILRR.	Termination: futility; Infeasibility: lack of eligible participants; other
3	The recruitment of patients was started on 5 April 2004 and terminated on 7 February 2012 despite insufficient statistical power at that point because the steering committee thought that further extension would not promote the enrolment of patients.	Termination: futility
4	An unplanned meeting of the Data Safety Monitoring Board (DSMB) was convened due to a lower than expected rate of the primary efficacy endpoint. The DSMB recommended terminating the trial because it was unlikely that the trial would show a difference between the access site strategies on the basis of the planned sample size.	Termination: futility
5	The study was underpowered as a result of the unexpected dropout in the F2C group. [...] As planned, we recruited 15 schools to permit randomization into three groups of five schools. Three schools in F2C withdrew after randomization, citing concerns about principal turnover and new mandatory academic programs. This occurred after the study had launched and trainings had begun. Attempts to recruit replacement schools failed primarily because the school year was underway.	Withdrawal of cluster(s)
6	We originally planned to include 1000 women, with an interim analysis after the first 500 women. Because recruitment was slower than expected and funding was ending, we revised this plan and decided to stop recruitment on a prespecified date, before we did any analysis; the trial results did not affect the decision to stop recruitment.	Infeasibility: Slow rate; resource constraints
7	Our projected sample size was not attained due to the loss of four health centers and decreased TB case notification rate during the trial period.	Sites withdrew after randomization; Infeasibility: lack of eligible participants

8	The first limitation of our study is that we did not achieve the target sample size. This was mainly because we ran out of funding since recruitment was slow as a result of clinicians' time limitations.	Slow rate; resource constraints (funding)
9	The desired sample size of 500 patients was not achieved. An investigation of the inclusion during a 2 month period showed that the majority of patients seeking treatment at the MHCs did not meet the trial inclusion criteria of our study because of the higher severity and burden of disease.	Infeasibility: Lack of eligible participants
10	We were not able to recruit a sufficient number of general practitioners to participate in the data collection within the time and resource limits that applied to our project. [...] The biggest challenge was to get in contact with the general practitioners due to them being guarded by the secretaries. When we had been able to talk to them, it was less difficult to get the general practitioner's consent to participate in the data collection. [...] Getting in touch with more general practitioners (and getting past their secretaries) would have required more resources, including incentives for the general practitioners, and time than we had available to us. [...] A further reduction [of number of patients practitioners were expected to recruit] to two, or even one, might possibly have increased the number of general practitioners who agreed to participate.	Resource constraints; Barriers related to obtaining consent
11	The number of individuals qualifying for and receiving FIT outreach alone was lower than anticipated during program planning and power computation.	Infeasibility: Lack of eligible participants
12	A weakness of the study is that 43 hospitals withdrew after randomization, which resulted in a smaller sample size than initially targeted: 26 hospitals decided to not participate because they found the workload associated with the study too high. Furthermore, all 14 Irish hospitals dropped out because of reorganization of the Irish health care system and 3 hospitals dropped out because of internal reorganization	External Events; Withdrawal of cluster(s)
13	the Steering Committee decided to stop recruitment after the target number of patients (n = 250) had been recruited in PRET-1, but before reaching the target number of patients for PRET-2, because the trial was already 2 years behind schedule, recruitment rates were decreasing, and provisions had to be made to cover compensations to participating sites for the 18-month follow-up data.	Infeasibility: Slow recruitment rate; Resource Constraints
14	The recruitment period of patients within each participating practice was originally planned to last a total of 6 months, but was extended by an additional 3 months, as the patient recruitment rate was lower than expected.	Infeasibility: Slow recruitment rate
15	Data and Safety Monitoring Board recommended to terminate the trial because of excess risk and the absence of any benefit of both experimental interventions	Termination: harm
16	Given a higher rate of discontinuation and lower rate of recruitment than anticipated, the board approved early termination of the study.	Infeasibility: Slow recruitment rate; Termination: futility

17	The principal investigators, who obviously were still blinded with respect to treatment allocation, noted that the overall cure rate was 92%. This is comparable with the results of a recently published similar trial in women comparing 73 to 83 patients treated over 7 or 14 days, respectively [8]. As we had indeed included a larger sample size of 200 patients, we estimated that our study would likely have already met the criteria for non-inferiority while still having a power of 0.80 with a type 1 error of 0.05. As we were confronted with an almost empty budget and a dropping inclusion rate after almost 5 years of participation, we considered continuation of the trial was no longer realistic and thus we decided to stop the trial at this point.	Infeasibility: Slow recruitment rate; Resource constraints
18	This was not due to early termination or interim analyses but to 2 external causes: withdrawal of 1 large hospital 1 week before the start of the trial and the time constraints of obtaining informed consent. Also, the stepped-wedge design results in a fixed number and length of steps, reducing flexibility to add clusters or lengthen the inclusion period.	Barriers related to consent; Withdrawal of cluster(s)
19	Initially, the trial included six VA sites charged with recruiting 50 participants each, but early in the recruitment phase, one of the sites dropped out. [...] The loss of one study site reduced our statistical power by some 50 subjects.	Withdrawal of cluster(s)
20	The trial was stopped after recruiting 512 patients because of low recruitment rates. In an unplanned interim analysis, the estimation of the effect was robust enough to draw conclusions about the comparative effectiveness of the interventions. [...] We decided to stop the trial because of the slowing down of the recruitment rate.	Infeasibility: Slow recruitment rate
21	We did not achieve the planned sample size of 750 patients due to slower than expected recruitment.	Infeasibility: Slow recruitment rate
22	The total number of births across both arms (18,747) was slightly lower than the expected number used to calculate sample size.	Infeasibility: Lack of eligible participants
23	Because the final number of centers that recruited patients was lower than anticipated and because the mean number of patients with acute ischemic stroke was 13 fewer than estimated per cluster per period, the study power was reestimated with the observed degree of correlation inpatient characteristics between clusters and different periods.	Infeasibility: Lack of eligible participants
24	Recruitment was slower than anticipated, and the number of patients recruited was lower than the planned 240 despite a prolonged recruitment period. [...] Difficulty with recruiting patients to randomized trials on management of early pregnancy failure is not unique to our study. Our experience, in accordance with that of others, is that many patients have a strong preconceived treatment preference. Some might find the open-label design of our trial a limitation.	Infeasibility: Slow recruitment rate
25	However, the study was stopped after 66 patients had been randomized due to exhaustion of funds.	Resource constraints: funding
26	Originally, one interim sample-size calibration was planned to be conducted by the principle investigator; to verify whether the study remained sufficiently powered. In data provided by the clinical team, the observed outcomes were less	Other

	than expected and sample size recalculation demonstrated the need for the enrollment of many more thousands of patients per group. As a pragmatic clinical trial undertaken in two very active intensive care units, it was determined that the additional costs and resources now necessary to complete the trial was prohibitive. The study team deemed the study to no longer be viable. The study was abandoned and enrollment ceased. Soon thereafter it was brought to our attention by the clinical team that the clinical data provided and used in the interim sample calculation were not correct. An error had been made. Thus, the trial had incorrectly been abandoned after 73% of the target enrollment.	
27	The first limitation of this study is the dropout of hospitals (Appendix 4), leading to an underpowered study. There was a dropout of one country due to a health policy decision outside the project. Although at the beginning of the study a possible drop out was taken into account by randomizing 50% more hospitals than initially needed. we could not assess the total amount of patients as 7 hospitals (26.9%) submitted their exclusion logbook.	Withdrawal of cluster(s); External events
28	Recruitment was stopped after 2 years because of a limited time frame.	Resource constraints: time
29	The DSMB, the institutional review board, and FDA approved termination of the study for futility because outcomes had not differed in any of the interim analyses, indicating that no difference should be anticipated.	Termination: futility
30	Actual enrollment was about half of anticipated enrollment, due primarily to lower numbers of individuals being diagnosed with HIV during the study period than in the preceding years. This was a national phenomenon, reflecting the successful scale-up of ART and a corresponding reduction in HIV incidence.	Infeasibility: Lack of eligible participants
31	The trial was stopped because of a decreasing average per centre recruitment rate, despite considerable help and encouragement. [...] The severity of the acute conditions of our study population compared to previous studies suggests the inclusion of patients in real-life conditions. Demonstration of the positive impact of an intervention, such as a strategy to prevent oversedation, might be difficult in patients with particularly severe admission conditions and requires a larger sample of patients. In our study, less than half of the planned included patients were finally enrolled which undoubtedly makes the trial underpowered.	Infeasibility: Slow recruitment rate
32	Because of recruitment challenges such as lack of CR referral and the strike of cardiopulmonary technicians who were responsible for stress testing, the target sample size was not reached.	External events; Infeasibility: Lack of eligible participants
33	When using HRQoL measures as variables, as in this study, one can expect smaller changes over time and a larger variation between the persons, which means that more patients must be included. A larger sample is often desired, but it is often difficult to recruit a sufficient sample. In our study, the inclusion of participants was time-consuming.	Resource constraints: time
34	After the second interim analysis at 75% recruitment, the data and safety monitoring board recommended recruitment be stopped due to futility and concern for increased morbidity in the delayed pushing group. In addition, the data and	Termination: harm; Termination: futility

	safety monitoring board was concerned about the significantly higher rate of postpartum hemorrhage in the delayed pushing group.	
35	The use of DEPs [diabetes education programs] as the primary site of recruitment likely had a negative impact on enrollment, usage, and clinical impact.	Other
36	Enrollment was stopped in year five of the trial due to the end of funding. […] We were not able to enroll the planned sample size. From 2009–2015, the prevalence of delirium decreased at our institute for unclear reasons.	Resource constraints: funding; Infeasibility: Lack of eligible participants
37	Due to concurrent CRC screening initiatives at the practice, only 897 of the 1373 participants identified by the EHR pull were found to be eligible. [...] The study was initially powered to detect a difference of 11 percentage points, but due to concurrent clinic efforts to increase screening rates, fewer patients were included in the trial.	External events; Infeasibility: Lack of eligible participants
38	The trial was closed to recruitment due to harm as detected by a preplanned analysis of early deaths	Termination: harm

Appendix A5. Explanations for over-recruitment

Number	Explanation for over-recruitment	Categorized Reason
1	It should be noticed, that we enrolled more general practices and patients than suggested by our power calculation; there was a lot of interest among the practices for participating in the study and it turned out that there was much more patients with diabetes associated with the practices than expected.	Higher interest/eligibility than anticipated
2	the pragmatic emphasis in this trial resulted in sample sizes that were >6 times those samples, resulting in power estimates approaching 99%. [...] our sole criteria were purposefully minimized: age, colorectal cancer screening history, and bowel cancer diagnosis. Furthermore, the control participants and intervention participants who did not accept the navigation intervention were still assessed without direct consent.	Pragmatic emphasis
3	It emerged that complete healthcare personnel influenza immunization information became available for each organization, so the planned cluster analysis and sample size calculation was not needed or appropriate.	Use of routinely collected administrative data
4	A group size of at least 244 would provide 90% power to detect a 10% absolute difference in this outcome if the proportion in the control group were <7.5% and at least 80% power to detect a 10% difference in this outcome if it was <12%. Available resources permitted us to perform a study that included 646 participants.	Available resources
5	the original power calculation was for a standard randomized controlled trial, allowance for the number of steps and allowance for any repeated measures on individuals was not included in this sample size estimate.	Unadjusted sample size calculation
6	Many more subjects presenting with a stroke suspicion than expected were screened for eligibility during the study period.	Higher interest/eligibility than anticipated
7	To ensure adequate power, taking into consideration the potential for additional attrition and possible subgroup analyses based upon baseline characteristics, additional participants were recruited.	Recruitment for additional analyses

8	We did not stop enrolling patients once the minimum sample size was reached because an important aim of this study was to investigate to what degree the ARV community delivery model could decongest ART facilities. Our actual sample size was thus substantially larger.	Recruitment for additional analyses
9	This calculation was conservative in the sense that it concerned an analysis in which only outcomes from the first person with neuropathy for each randomized physician would be considered (cluster of size 1).	Unadjusted sample size calculation
10	We estimated 903 patients in each arm [...] We included all trial-eligible patients from our practice in the study and estimated this would provide a sufficient sample to stratify patients by sex.	Recruitment for additional analyses

Appendix B: Post-hoc Stratified Analyses by Unit of Randomization

Supplemental Table B1. Multivariable multinomial logistic regression analysis of recruitment outcomes (N = 241 individually randomized trials) odds ratios represent odds of either over or under-recruitment relative to recruitment within 10% of the target

	<90% versus within target		>110% versus within target	
Characteristic	**OR (95% CI)**	**p-value**	**OR (95% CI)**	**p-value**
Publication Year				
2014-2016	1		1	
2017-2019	0.90 (0.44, 1.87)	0.7867	0.51 (0.24, 1.09)	0.0810
Region of trial recruitment				
USA/Canada Only	1		1	
Europe Only	0.86 (0.38, 1.91)	0.7035	1.37 (0.58, 3.28)	0.4761
At least one LMIC	0.47 (0.10, 2.25)	0.3414	1.14 (0.29, 4.42)	0.8530
Other High-Income Countries	2.26 (0.61, 8.35)	0.2220	1.81 (0.30, 10.97)	0.5187
Type of Experimental Intervention				
Clinical, mixture, or other	1		1	
Educational/Behavioural only	0.73 (0.32, 1.68)	0.4537	2.15 (0.97, 4.76)	0.0587
Exclusively routinely collected data				
No or unclear	1		1	
Yes	1.19 (0.39, 3.64)	0.7551	3.43 (1.27, 9.26)	0.0149
Consent obtained				
No or not reported	1		ref	
Yes	0.85 (0.23, 3.20)	0.8128	0.56 (0.16, 1.97)	0.3666
Journal Impact Factor				
Below median (<5.4)[1]	1		ref	
Above median (≥5.4)	0.53 (0.25, 1.10)	0.0906	0.44 (0.20, 0.98)	0.0459

[1] Includes 3 studies with missing journal impact factors.

Supplemental Table B2. Multivariable multinomial logistic regression analysis of recruitment outcomes (N = 131 cluster randomized trials): odds ratios represent odds of either over or under-recruitment relative to recruitment within 10% of the target

	<90% versus within target		>110% versus within target	
Characteristic	**OR (95% CI)**	**p-value**	**OR (95% CI)**	**p-value**
Publication Year				
2014-2016	1		1	
2017-2019	0.78 (0.29, 2.06)	0.6120	1.49 (0.63, 3.51)	0.3608
Region of trial recruitment				
USA/Canada Only	1		1	
Europe Only	0.63 (0.20, 2.02)	0.4410	0.82 (0.28, 2.40)	0.7187
Other	0.42 (0.11, 1.65)	0.2162	1.01 (0.36, 2.83)	0.9887
Type of Experimental Intervention				
Clinical, mixture, or other	1		1	
Educational/Behavioural only	0.51 (0.17, 1.50)	0.2205	2.48 (1.03, 5.97)	0.0422
Exclusively routinely collected data				
No or unclear	1		1	
Yes	0.34 (0.08, 1.47)	0.1474	1.95 (0.70, 5.37)	0.1986
Consent obtained				
No or not reported	1		1	
Yes	0.83 (0.28, 2.45)	0.7373	0.71 (0.27, 1.84)	0.4789
Patient or public engagement				
No or unclear	1		1	
Yes	2.18 (0.54, 8.72)	0.2719	1.84 (0.50, 6.75)	0.3571
Journal Impact Factor				
Below median (<5.4)	1		1	
Above median (≥5.4)	0.22 (0.08, 0.64)	0.0055	0.68 (0.28, 1.63)	0.3875

Supplemental Table B3. Frequencies of covariates by recruitment outcome (N = 241 individually randomized trials)

	Recruitment (Frequency, Row %)		
	Under <90% N = 41	**Within ±10% N = 161**	**Over >110% N = 39**
Publication Year			
2014-2016	16 (15.7)	65 (63.7)	21 (20.6)
2017-2019	25 (18.0)	96 (69.1)	18 (13.0)
Region of trial recruitment			
USA/Canada Only	20 (18.0)	72 (64.9)	19 (17.1)
Europe/UK Only	14 (15.4)	63 (69.2)	14 (15.4)
At least one LMIC	2 (8.3)	18 (75.0)	4 (16.7)
Other High-Income Countries	5 (33.3)	8 (53.3)	2 (13.3)
Type of Intervention			
Educational/Behavioural	31 (19.0)	112 (68.7)	20 (12.3)
Clinical, mixture, or other	10 (12.8)	49 (62.8)	19 (24.4)
Exclusively routinely collected data			
Yes	35 (17.0)	145 (70.4)	26 (12.6)
No or unclear	6 (17.1)	16 (45.7)	13 (37.1)
Consent obtained			
Yes	4 (16.7)	13 (54.2)	7 (29.2)
No or not reported	37 (17.1)	148 (68.2)	32 (14.8)
Patient or public engagement			
Yes	41 (18.7)	142 (64.8)	36 (16.4)
No or unclear	0 (0)	19 (86.4)	3 (13.6)
Journal Impact Factor			
Below Median (<5.4)[1]	24 (20.2)	69 (58.0)	26 (21.9)
Above Median (≥5.4)	17 (13.9)	92 (75.4)	13 (10.7)

[1] Includes 3 studies with missing journal impact factors.

Supplemental Table B4. Frequencies of covariates by recruitment outcome (N = 131 cluster randomized trials)

	Recruitment (Frequency, Row %)		
	Under <90% N = 30	**Within ±10% N = 53**	**Over >110% N = 48**
Publication Year			
2014-2016	15 (27.8)	22 (40.7)	17 (31.5)
2017-2019	15 (19.5)	31 (40.3)	31 (40.3)
Region of trial recruitment			
USA/Canada Only	13 (23.6)	20 (36.4)	22 (40.0)
Europe/UK Only	12 (28.6)	18 (42.9)	12 (28.6)
At least one LMIC	5 (15.6)	13 (40.6)	14 (43.8)
Other High-Income Countries	0 (0)	2 (100.0)	0 (0)
Type of Intervention			
Educational/Behavioural	21 (28.8)	33 (45.2)	19 (26.0)
Clinical, mixture, or other	9 (15.5)	20 (34.5)	29 (50.0)
Exclusively routinely collected data			
Yes	27 (28.1)	40 (41.7)	29 (30.2)
No or unclear	3 (8.6)	13 (37.1)	19 (54.3)
Consent obtained			
Yes	11 (18.7)	22 (37.3)	26 (44.1)
No or not reported	19 (26.4)	31 (43.1)	22 (30.6)
Patient or public engagement			
Yes	24 (21.2)	48 (42.5)	41 (36.3)
No or unclear	6 (33.3)	5 (27.8)	7 (38.9)
Journal Impact Factor			
Below Median (<5.4)	22 (32.8)	21 (31.3)	24 (35.8)
Above Median (≥5.4)	8 (12.5)	32 (50.0)	24 (37.5)