University Library

MINERVA ACCESS

A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Bernhardt, J;Raffelt, A;Churilov, L;Lindley, RI;Speare, S;Ancliffe, J;Katijjahbe, MA;Hameed, S;Lennon, S;McRae, A;Tan, D;Quiney, J;Williamson, HC;Collier, J;Dewey, HM;Donnan, GA;Langhorne, P;Thrift, AG

Title:
Exploring threats to generalisability in a large international rehabilitation trial (AVERT)

Date:
2015-01-01

Citation:
Bernhardt, J., Raffelt, A., Churilov, L., Lindley, R. I., Speare, S., Ancliffe, J., Katijjahbe, M. A., Hameed, S., Lennon, S., McRae, A., Tan, D., Quiney, J., Williamson, H. C., Collier, J., Dewey, H. M., Donnan, G. A., Langhorne, P. & Thrift, A. G. (2015). Exploring threats to generalisability in a large international rehabilitation trial (AVERT). BMJ Open, 5 (8), https://doi.org/10.1136/bmjopen-2015-008378.

Persistent Link:
https://hdl.handle.net/11343/55463

License:
CC BY-NC

BMJ Open

Exploring threats to generalisability in a large international rehabilitation trial (AVERT)

Julie Bernhardt,[1] Audrey Raffelt,[1] Leonid Churilov,[1] Richard I Lindley,[2] Sally Speare,[1] Jacqueline Ancliffe,[3] Md Ali Katijjahbe,[4] Shahul Hameed,[5] Sheila Lennon,[6] Anna McRae,[7] Dawn Tan,[8] Jan Quiney,[9] Hannah C Williamson,[10] Janice Collier,[1] Helen M Dewey,[11] Geoffrey A Donnan,[12] Peter Langhorne,[13] Amanda G Thrift,[14] on behalf of the AVERT Trialists' Collaboration

To cite: Bernhardt J, Raffelt A, Churilov L, *et al*. Exploring threats to generalisability in a large international rehabilitation trial (AVERT). *BMJ Open* 2015;**5**:e008378. doi:10.1136/bmjopen-2015-008378

▸ Prepublication history and additional material is available. To view please visit the journal (http://dx.doi.org/10.1136/bmjopen-2015-008378).

Received 31 March 2015
Revised 3 July 2015
Accepted 19 July 2015

For numbered affiliations see end of article.

Correspondence to
Dr Julie Bernhardt;
julie.bernhardt@florey.edu.au

ABSTRACT

Objective: The purpose of this paper is to examine potential threats to generalisability of the results of a multicentre randomised controlled trial using data from A Very Early Rehabilitation Trial (AVERT).

Design: AVERT is a prospective, parallel group, assessor-blinded randomised clinical trial. This paper presents data assessing the generalisability of AVERT.

Setting: Acute stroke units at 44 hospitals in 8 countries.

Participants: The first 20 000 patients screened for AVERT, of whom 1158 were recruited and randomised.

Model: We use the Proximal Similarity Model, which considers the person, place, and setting and practice, as a framework for considering generalisability. As well as comparing the recruited patients with the target population, we also performed an exploratory analysis of the demographic, clinical, site and process factors associated with recruitment.

Results: The demographics and stroke characteristics of the included patients in the trial were broadly similar to population-based norms, with the exception that AVERT had a greater proportion of men. The most common reason for non-recruitment was late arrival to hospital (ie, >24 h). Overall, being older and female reduced the odds of recruitment to the trial. More women than men were excluded for most of the reasons, including refusal. The odds of exclusion due to early deterioration were particularly high for those with severe stroke (OR=10.4, p<0.001, 95% CI 9.27 to 11.65).

Conclusions: A model which explores person, place, and setting and practice factors can provide important information about the external validity of a trial, and could be applied to other clinical trials.

Trial registration number: Australian New Zealand Clinical Trials Registry (ACTRN12606000185561) and Clinicaltrials.gov (NCT01846247).

Strengths and limitations of this study

- Use of a screening log which captured a broad range of reasons for non-recruitment, not just demographic data.
- Use of a model to explore generalisability that goes beyond describing patient characteristics.
- A large, comprehensive data set relevant to broad, pragmatic trials.
- A limited number of demographic and clinical factors. Other factors may also have influenced recruitment.
- Use of a large data set may demonstrate statistical significance where it is of little clinical importance.

INTRODUCTION

Randomised controlled trials (RCTs) are the gold standard for determining the efficacy of an intervention. In the ideal world, positive trials lead directly to implementation of the intervention into practice. However, trial results are most meaningful (and useful) when the trial has both internal and external validity. Internal validity is well understood[1] and can be controlled in the design stage of most RCTs by using strategies such as masking of patients and assessors, randomisation, stratification and block randomisation (to ensure balance between allocated groups), and standardisation of treatment protocols. In contrast, external validity, the extent to which the results of a study can be generalised to other situations and to other people,[2] is often under-recognised, under-reported and undervalued.[2–4]

Rothwell[5] describes the concept of external validity as 'slippery' and complex, easy to define in broad terms but problematic to quantify. However, when the results of a promising clinical trial are not incorporated into practice, a commonly cited reason is a lack of generalisability of the results.[6–8] In reality, the degree to which trial results can

be generalised is often a matter of judgement, as well as pragmatics. When it comes time to implement an intervention, costs, site logistics and administrator goals play a critical role in uptake. Nevertheless, understanding how well the participants included in the trial are representative of the population of interest is important,[9 10] but so too is the treatment setting in which the trial took place and the expertise of the intervention staff. It is also important to understand whether the results can be generalised beyond often restrictive eligibility criteria such as age and comorbidities.[11]

Our aims were to examine potential threats to generalisability of the results from an ongoing multicentre randomised trial, A Very Early Rehabilitation Trial (AVERT). AVERT is a large, pragmatic clinical trial of very early out of bed training (mobilisation) after stroke. It takes place in real-world clinical settings, with existing clinical staff delivering the intervention. The inclusion criteria are kept broad, in an effort to test whether the intervention might be widely applicable to patients with stroke.[9] Unlike many acute stroke trials, there is no restriction on upper age limit or stroke subtype (infarct or haemorrhage), and apart from excluding patients with significant disability prior to the index stroke admission, there is no restriction on comorbidities or previous stroke. Although these design characteristics should enhance the external validity of this trial, the question of how broadly trial results can be applied more generally has implications for implementation into practice.[7 8]

The Proximal Similarity Model is a useful framework for considering different generalisability contexts.[12] The term proximal similarity was suggested by Donald T Campbell as an appropriate relabelling of the term external validity. Under this model, different generalisability contexts, and the settings and circumstances in which people in a study may be different from, or similar to the population of interest, are considered and a gradient of similarity determined. Use of this framework encourages us to explore more deeply the potential critical factors that threaten external validity. For example, in the design phase for any trial, one of the goals is to ensure that the sample of patients involved in the study is representative of the population of interest. This goal is achieved by systematically addressing and, when possible, minimising or eliminating identifiable threats to external validity. Such threats can be specific to either the patient or to the study. Patient-specific threats include the systematic differences in important demographic and clinical characteristics between the patients in the study and in the population of interest. Study-specific threats include processes and systems of care that may or may not be encapsulated in the exclusion criteria, such as when the care processes in the study are not easily generalisable to different care sites or care systems.[7] Using the proximal similarity framework, study-specific factors to consider would include both those related to 'place' (site and, if relevant, country) and to the 'settings and practices' (eg, Are patients with acute stroke managed in a stroke unit or intensive care setting?) where the trial is undertaken. A schematic of the proximal similarity framework applied to the AVERT context is shown in figure 1.

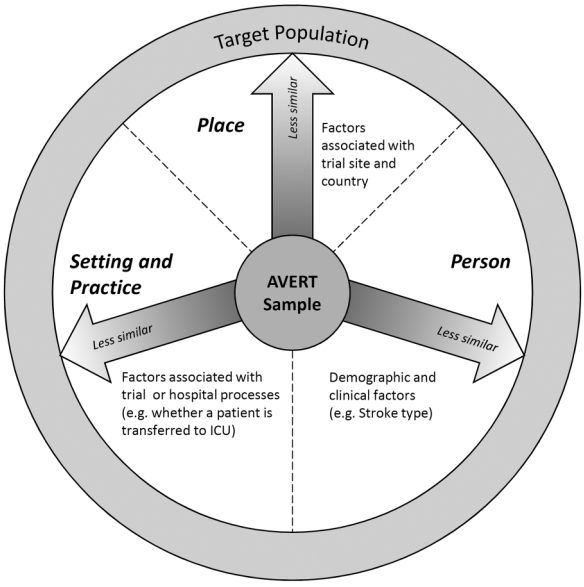


Figure 1 Proximal similarity framework applied to the AVERT trial: a model for conceptualising the dimensions along which the sample of patients may be similar to the target population. Each dimension (person, place and setting and practice) is affected by specific factors which may threaten external validity (AVERT, A Very Early Rehabilitation Trial; ICU, intensive care unit).

We aimed to (1) explore the potential threats to external validity using data from the first 20 000 patients screened in AVERT and (2) examine the person, place, and setting and practice related reasons for non-recruitment to the trial. Our four specific objectives were to:

1. Identify demographic and clinical differences between the patients randomised to AVERT and the general stroke population, using available community-based data;
2. Identify systematic differences between person, place, and setting and practice factors for those recruited and those screened but not recruited;
3. Examine the reasons for non-recruitment and explore the barriers to patient recruitment;
4. Explore whether time (both years of site involvement in the trial and years of study overall) is associated with differences in patient recruitment.

METHODS

Trial design in brief

AVERT (ACTRN12606000185561) is a prospective, parallel group, assessor-blind, randomised, multicentre, international clinical trial that has completed the primary outcome assessment for all randomised patients (a longer term follow-up continues). Patients admitted to a

stroke unit are randomised in a ratio of 1:1 to two groups: very early and frequent mobilisation out of bed (VEM) and usual care. The experimental intervention (VEM) is frequent, functional, out of bed sitting, standing and walking activity starting within 24 h of stroke onset and continued 6 days a week for 14 days or until discharge from acute stroke care (whichever is sooner). Patients are followed up at 3 and 12 months. The primary outcome is the modified Rankin Scale (mRS) score at 3 months poststroke, with secondary outcomes for safety, walking recovery and quality of life. This paper presents results from the first 20 000 patients who were screened, of whom 1158 patients were then randomised to the study.

Participating hospitals

At the time of analysis, investigators from sites in Australia, New Zealand, Malaysia, Singapore, England, Northern Ireland, Scotland and Wales had recruited patients to the trial. All participating sites had a dedicated stroke unit with a multidisciplinary stroke team. A detailed site questionnaire was collected from participating sites yearly. This provided a record of the number of admissions for stroke each year and the number of stroke beds available. Sites were classified, by investigators at each site, according to the type of stroke unit defined in the Australian National Stroke Audit, Acute Services Organisational Survey Report:[13] (1) intensive stroke unit care model involving short stay, high nurse patient ratio, life support facilities and no rehabilitation; (2) acute care model involving short stays, close physiological monitoring and limited rehabilitation or (3) comprehensive model involving both acute care and a strong rehabilitation focus, longer stays and broader staffing. Sites were also classified according to the geographic location: metropolitan (population >100 000), regional (25 000–100 000) or rural (<25 000).[13]

Patient eligibility

Inclusion and exclusion criteria (figure 2) were designed to optimise the diversity of the patients exposed to the intervention while considering patient safety and feasibility. There was no upper age limit for inclusion to the study and both ischaemic and haemorrhagic strokes were included. Patients treated with recombinant tissue plasmin activator (rt-PA) were eligible if the attending physician allowed. Patients were excluded if a comorbid condition (eg, lower limb fracture or amputation) would prevent the start of treatment within the first 24 h of stroke onset or if the outcome assessments were likely to be confounded by another serious comorbid medical illness. Patients admitted to a dedicated intensive care unit (ICU) are also excluded.

Screening and recruitment process

The intervention was designed to be delivered by trained nursing and/or physiotherapy staff 6 days/week, excluding Sunday. To meet the 24 h target for screening, recruitment and start of intervention, screening could take place Monday to Friday (8:00 to 17:00) and Saturday (8:00 to 12:00) to allow for the possibility of a recruited patient randomised to VEM on Saturday to start training on that day. No screening was conducted on Sundays. Screening and recruitment could be undertaken by stroke unit nurses, physiotherapists or research trials staff. A screening log was used to record all patients screened for AVERT. Age, sex, stroke severity and stroke type are important patient characteristics that can influence outcome after stroke and are therefore highlighted for attention on the screening logs.[14 15] The log also included reasons for exclusion as related to either trial inclusion and exclusion criteria, or to trial processes that might lead to exclusion (figure 2). Recruiting staff could also list the reason for exclusion as 'other'. This category was most commonly used for patients who were not admitted to a stroke unit. However, 'other' included patients who had a lower limb fracture (an exclusion criterion) and instances where a research therapist was unavailable to carry out the intervention, so recruitment was not possible. Multiple reasons could be listed for each patient. Recruiters were required to record all stroke admissions, even those that occurred on days or times outside of the set screening periods.

For recruited patients, stroke severity was measured using the National Institutes of Health Stroke Scale (NIHSS)[16] by accredited staff and then categorised into mild (NIHSS 1–7), moderate (NIHSS 8–16) and severe (NIHSS >16),[17] which were used to stratify patients in each group. To minimise the burden on recruiting staff, we did not require completion of a full NIHSS on patients who were excluded. Instead, the same trained staff were asked to estimate whether the patient would be likely to have an NIHSS score in the 'mild', 'moderate' or 'severe' range.

Data management and statistical analysis

We summarised the characteristics of the participating sites involved in the trial. Data presented include: number of stroke admissions, number of beds and whether the site was in a metropolitan or regional location.

To meet the goals of objective 1, we summarised the demographic and clinical factors from all patients with stroke, including those who were recruited and those screened but not recruited. We intended to compare our data with both population-based data and data from other acute stroke trials in which demographic data are generally described in more detail. However, as many acute stroke trials have pharmacological interventions, the inclusion criteria were typically more narrow compared with AVERT.[18] We therefore compared demographic and clinical factors against world stroke data. We used epidemiological data from Feigin *et al*,[19 20] where available, with gender and stroke severity from the Virtual International Stroke Trials Archive (VISTA).[21] We

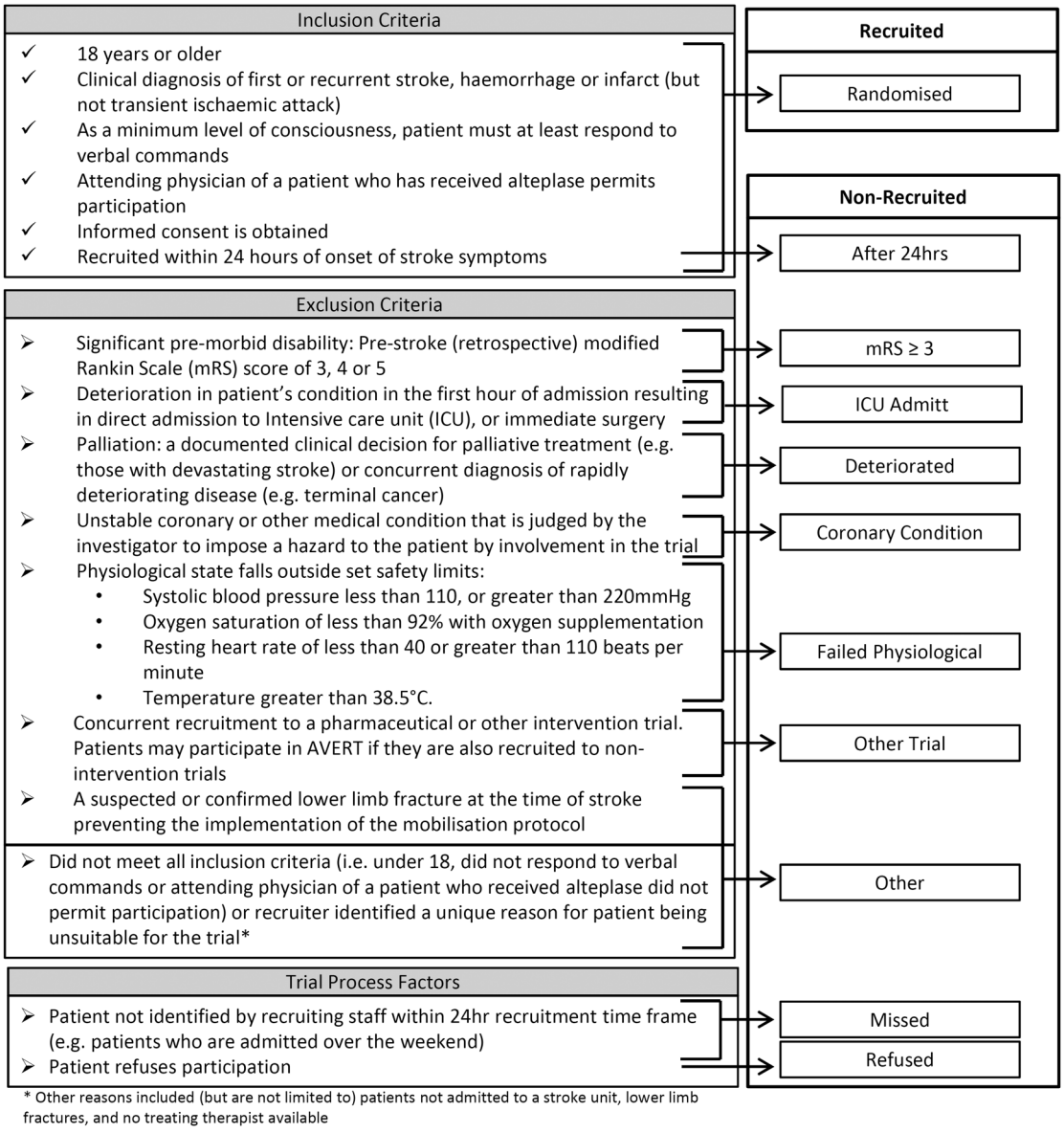


Figure 2 Relationship between trial inclusion/exclusion criteria and screening log categories (AVERT, A Very Early Rehabilitation Trial).

report 95% CIs, where available, in sample data to allow a broad comparison with population data. Differences in demographics between recruited versus non-recruited patients were examined using the Wilcoxon Mann-Whitney U test for continuous data (age) and Fisher's exact test for categorical data.

To explore differences between recruited and non-recruited patients (objective 2), we systematically examined the association between patient demographic and clinical factors (ie, age, gender, stroke severity, stroke type) and patient recruitment or non-recruitment. We used a random-effect multilevel logistic regression model with patient factors as independent variables, the recruitment status as the dependent variable, and treating site as a level variable. This enabled us to assess the association between demographic and stroke-related factors and the odds of recruitment in all patients. In a second set of analyses, using the same model, we explored each individual reason for non-recruitment in turn. We systematically compared the recruited patients to those not recruited due to a specific reason, and explored demographic and clinical factors associated with recruitment versus non-recruitment due to each given reason (figure 3, analysis 1). We report the estimated adjusted ORs of being recruited compared with the reference of non-recruitment for each specific reason (eg, the odds of being recruited vs non-recruitment due to late arrival). ORs>1 indicate the increased likelihood of recruitment and ORs<1 indicate the increased likelihood of non-recruitment due to that reason.

A similar random-effect multilevel logistic regression approach was also used to more closely examine the association between patient demographic factors and

specific reasons for non-recruitment (objective 3). In this analysis, for each reason in turn, we systematically compared the patients not recruited due to each reason to the patients not recruited for all other reasons combined (eg, non-recruitment due to late arrival, vs non-recruitment for all other reasons; see figure 3, analysis 2). ORs>1 indicate the increased likelihood of non-recruitment due to a specific reason and ORs<1 indicate the decreased likelihood of non-recruitment due to that reason.

All statistical analyses were performed using STATA-IC.

To achieve objective 4, we examined the heterogeneity in the individual reasons for non-recruitment between recruiting sites and countries, as well as the effect of the time on recruitment patterns. The between-centre and between-country heterogeneity was estimated using intraclass correlation coefficients (ICCs) generated by the respective random-effect logistic regression models. In this analysis, England, Scotland, Northern Ireland and Wales were treated as individual countries, resulting

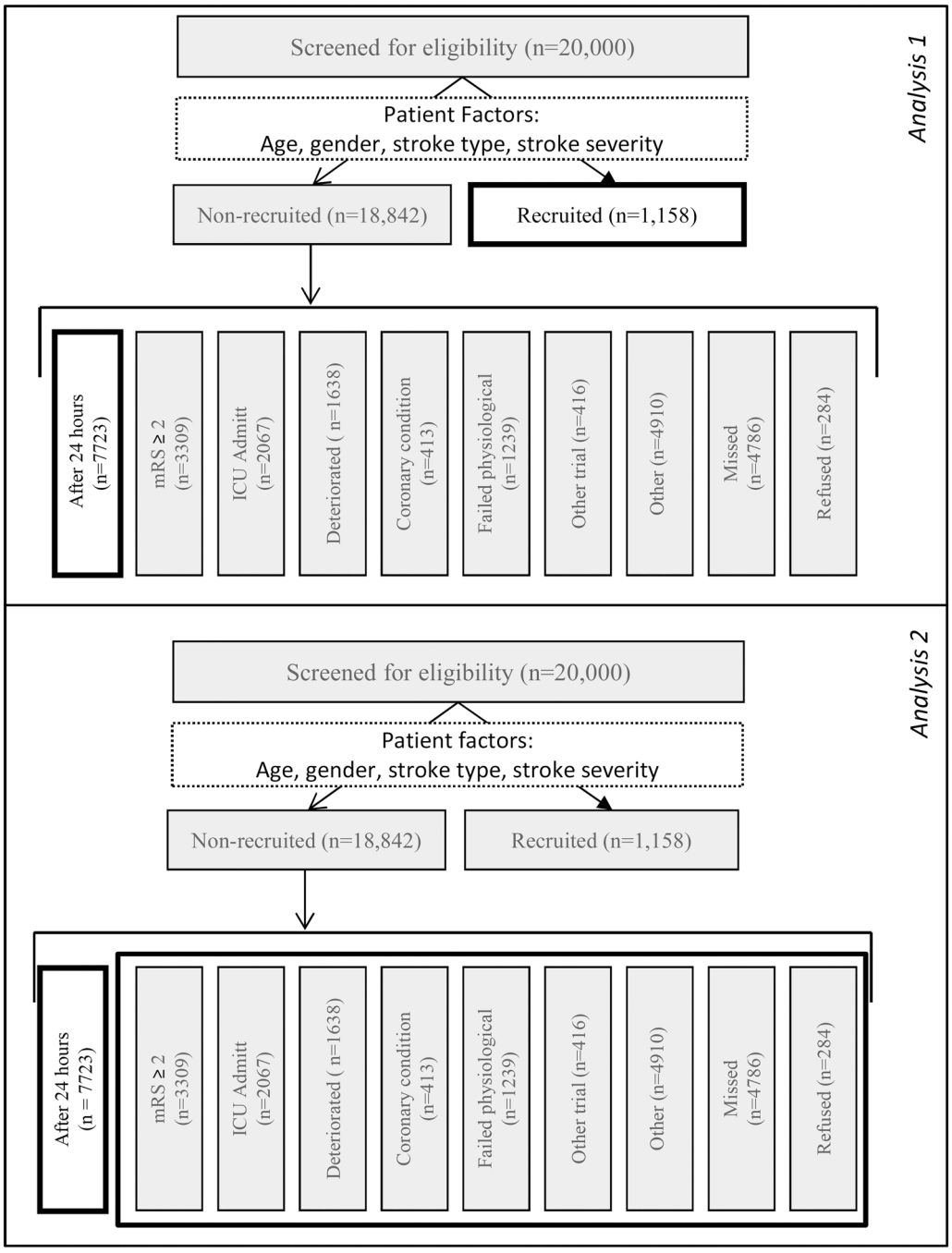


Figure 3 Methods of explorative analysis using the first reason for non-recruitment (arrived after 24 h) as an example. Bold boxes indicate data grouping. Analyses were repeated for all 10 reasons for non-recruitment, with four patient demographic and clinical factors. Trial site and month of trial were controlled for in each analysis (ICU, intensive care unit; mRS, modified Rankin Scale).

in eight countries included in the analysis. The possible values of ICC are between 0% and 100%. The ICC value indicates the proportion of the variance in the propensity of a specific outcome that can be attributed to sites or countries (eg, the proportion of the variance in the propensity of being non-recruited due to late arrival as opposed to other reasons that can be attributed to sites). In other words, higher values of the ICC signify larger between-site heterogeneity and lower values are indicative of the lower influence of site-specific or country-specific factors on recruitment patterns.

RESULTS

Site characteristics

Of the 44 sites recruiting, most sites included were metropolitan (n=37), with 6 regional and 1 rural. The number of admissions per year ranged from 33 to 793. Ward size (number of beds open at the time of review) ranged from 8 to 77, with between 1 and 54 dedicated stroke beds. All sites had a geographically defined stroke unit, a coordinated multidisciplinary stroke team and access to a CT scanner, with 20 sites having onsite access to neurosurgery. Eleven sites were classified as a stand-alone stroke unit, whereas the majority of sites (n=33) had dedicated stroke beds within a larger neurology or mixed medical ward. One site described itself as following an intensive care model, 22 as acute stroke unit models and 21 as comprehensive stroke unit models. Patients were screened between July 2006 and December 2011.

Sample characteristics of recruited and non-recruited patients

Table 1 presents demographic and stroke characteristics for non-recruited and recruited patients. There was a greater proportion of men in AVERT than in the non-recruited patients. When we compared recruited patients to non-recruited patients, we found that recruited patients were significantly younger than non-recruited patients (p<0.001), there was a greater proportion of men (p<0.001) and there were significantly fewer patients with severe stroke (p<0.001). The proportion of patients with haemorrhagic stroke was not different between the groups (p=0.504).

Baseline demographic data compared with world data

The characteristics of the recruited patients were broadly similar to world data.[19–21] In their most recent review, Feigin *et al*[19] identified 56 relevant studies across 28 countries (data from 1970 to 2008) totalling 37 016 strokes. The proportional frequency of stroke subtypes (ischaemic, intracerebral haemorrhage and subarachnoid haemorrhage) was reported for 12 242 strokes from 18 centres, ranging from 54% to 90% for ischaemic stroke and 6–27% for haemorrhagic stroke. The proportional frequency of ischaemic and haemorrhagic strokes in the recruited patients was within these ranges (87% and 13%, respectively). World median age was obtained from a previous Feigin *et al*[20] review, and was found to be older than for the recruited patients (world median is 75 years; recruited median is 73 years, with a 95% CI of 72 to 74). Data about the relative frequency of stroke in women were not available from either of the reviews by Feigin *et al*.[19 20] However, compared to the proportion of women in the VISTA database,[21] there were fewer women recruited to AVERT: world data[21] indicate 46% women, compared with the recruited sample which includes 37% women (95% CI of 34% to 40%).

Main reasons for non-recruitment

Of the 20 000 people screened, 1158 (5.8%) were recruited. The most common reasons for non-recruitment were arrival 'after 24 h' (41.9%), 'missed' (25.2%) and premorbid 'mRS >2' (disability; 12.9%; table 2). Patients were marked as 'missed' if recruiting

Table 1 Baseline demographics for recruited versus non-recruited patients, including significance testing for difference between recruited and non-recruited patients

	AVERT		Difference (p value)
	Non-recruited	Recruited	Recruited: non recruited
N (%)	18 842 (94)	1158 (6)	
Age, median (IQR)	75 (64–82)	73 (63–80)	<0.001
Range	15–102	18–100	
Females age, median (IQR)	78 (61–80)	76 (66–82)	
Males age, median (IQR)	71 (68–85)	71 (61–79)	
Females % (95% CI)	47 (47 to 48)	37 (34 to 40)	<0.001
NIHSS, N (%)			<0.001
Mild (1–7)	10 012 (53)	619 (53)	
Moderate (8–16)	4934 (26)	358 (31)	
Severe (>16)	3896 (21)	181 (16)	
Stroke type, N (%)			0.504
Ischaemic	16 328 (87)	1012 (87)	
ICH	2514 (13)	146 (13)	

AVERT, A Very Early Rehabilitation Trial; NIHSS, National Institutes of Health Stroke Scale.

Table 2 Reasons for non-recruitment as a percentage of all non-recruited patients

Reason for non-recruitment	N	Median % (95% CI)
After 24 h	7723	41.9 (35.1 to 49.5)
mRS≥3	3309	12.9 (9.1 to 18.4)
ICU	909	1.5 (0.8 to 2.4)
Deteriorated	1638	8.6 (6.8 to 10.2)
Coronary condition	413	1.1 (0.5 to 1.9)
Failed physiological	1239	2.7 (1.9 to 4.4)
Other trial	416	1.3 (0.4 to 2.0)
Other	4910	16.1 (13.6 to 19.3)
Refused	284	0.9 (0.6 to 1.4)
Missed	4786	25.2 (19.0 to 33.6)

Reasons are not exclusive; multiple reasons may be listed for a single patient, and therefore sum to more than 100%. 'Other' was used when a patient was deemed not suitable for reasons other than those listed (eg, not admitted to a stroke unit, lower limb fractures) or when no treating therapist was available. 'Missed' was used when recruiters were on leave, unavailable or the patients arrived after hours or on weekends and by the time the recruiter had returned the patients were now outside of the 24 h recruitment window.
ICU, intensive care unit; mRS, modified Rankin Scale.

staff (predominantly ward therapists or nurses) were on leave, unavailable or the patients arrived after hours or on weekends, and by the time the recruiter had returned, the patients were now outside of the 24 h recruitment window. In total, 16.1% of reasons were reported as 'other'. Other was used when a patient was deemed not suitable for reasons other than those listed (eg, not admitted to a stroke unit, lower limb fractures) or when no treating therapists were available to deliver the intervention. Fewer than 1% of patients declined participation.

Association between patient factors and the likelihood of being recruited versus non-recruited due to a specific reason

When adjusting for the length of time the site participated in AVERT, we found that being older and female reduced the odds of recruitment to the trial (table 3). We then examined how these patient factors affected the odds of recruitment relative to each reason for non-recruitment in turn. Older patients were less likely to be excluded because of admission to an ICU than they were to be recruited. In contrast, they were more likely to be excluded (than recruited) because of prior disability, early deterioration or because they were missed. Having an intracerebral haemorrhage meant that the patients were less likely to be excluded (than recruited) because they were already involved in another trial, or had a coronary condition. In contrast, patients with intracerebral haemorrhage were more likely to be excluded than recruited because they deteriorated early, failed physiological criteria or were admitted directly to the ICU. Patients with severe strokes were more likely to be recruited than arrive late (and be excluded). Women were less likely to be recruited than men for all of the exclusion criteria. In other words, they were more likely to be excluded due to premorbid disability, a coronary condition, early deterioration, late arrival, refusing participation or being missed by recruiters.

Association between patient factors and the likelihood of non-recruitment due to a specific reason versus due to other reasons

For the third objective, we explored competing reasons for non-recruitment and how they related to the demographic characteristics of the excluded patients (table 4). Patients who were older, female and with severe stroke were more likely to be excluded because of premorbid disability (mRS≥3) and early deterioration. The odds of exclusion due to early deterioration were particularly high for those with severe stroke (OR=10.4, p<0.001, 95% CI 9.27 to 11.65). Patients with haemorrhagic stroke and severe stroke were more likely to be excluded because they failed physiological safety criteria or were admitted directly to the ICU. Patients with increasing age, haemorrhagic stroke and severe stroke were less likely to be excluded because of late arrival at hospital (although they could be excluded for other reasons), indicating that they were more likely to arrive early to hospital.

Site and country variability in recruitment

Differences between the sites (44 in total) accounted for approximately 8% of variability in the propensity to be recruited overall (ICC=8%, 95% CI 5% to 13%), while differences between countries (8 in total) accounted for only 3% (ICC=3%, 95% CI 1% to 10%). Exploring the between-site heterogeneity in recruitment patterns (see online supplementary table S1), the highest variability was found for non-recruitment due to the patient participating in another trial (ICC=43%, 95% CI 28% to 59%), followed by ICU admission (ICC=26%, 95% CI 17% to 38%), and refusal to participate (ICC=23%, 95% CI 13.7% to 36.1%). Heterogeneity by country for each exclusion reason varied between 5% (arrival >24 h) and 24% (ICU admission).

Effect of time

There was a small decline in participation rates as the trial progressed, with each extra month accounting for a 1% decline in the odds of recruitment (OR=0.99, 95% CI 0.99 to 1.00, p=0.01). However, overall we found that patient factors had a greater impact on the recruitment patterns rather than site, country or time.

DISCUSSION

Using the proximal similarity framework approach,[12] we have explored the reasons why patients did and did not participate in the trial, using person, place, and setting and practice factors that might explain participation. The most common reason for non-recruitment was late arrival to hospital (ie, >24 h), but being older and female also reduced the odds of recruitment to the trial.

Table 3 Odds of recruitment relative to exclusion overall (for all patients screened), and odds of recruitment relative to a specific reason for non-recruitment (subgroup analysis), according to age, gender, stroke type and severity*

		Recruited versus non-recruited due to									
	Recruited versus non-recruitment (n=20 000)	**After 24 h (n=8881)**	**mRS≥3 (n=4467)**	**ICU admit (n=2067)**	**Deteriorated (n=2796)**	**Coronary condition (n=1571)**	**Failed physiological (n=2397)**	**Other trial (n=1574)**	**Other (n=6068)**	**Refused (n=1442)**	**Missed (n=5944)**
Older age, years											
OR	0.99	1	0.94	**1.04**	0.96	0.99	1	1	1	1	0.99
95% CI	0.99 to 1.00	0.99 to 1.00	0.93 to 0.94	**1.03 to 1.05**	0.95 to 0.97	0.98 to 1.00	0.99 to 1.01	0.99 to 1.01	0.99 to 1.00	0.99 to 1.01	0.99 to 1.00
p Value	<0.01	0.37	<0.01	**<0.01**	<0.01	0.05	0.44	0.55	0.61	0.92	0.01
Female											
OR	0.67	0.67	0.47	0.89	0.57	0.76	0.64	0.85	0.68	0.67	0.73
95% CI	0.59 to 0.76	0.59 to 0.77	0.41 to 0.56	0.69 to 1.15	0.45 to 0.72	0.58 to 0.99	0.52 to 0.79	0.65 to 1.12	0.59 to 0.78	0.50 to 0.89	0.63 to 0.84
p Value	<0.01	<0.01	<0.01	0.370	<0.01	0.038	<0.01	0.251	<0.01	0.01	<0.01
Haemorrhagic stroke											
OR	0.93	1.06	1.15	0.28	0.30	**1.55**	0.64	**1.82**	**1.26**	1.10	1
95% CI	0.77 to 1.12	0.87 to 1.29	0.90 to 1.46	0.21 to 0.38	0.23 to 0.41	**1.01 to 2.37**	0.49 to 0.84	**1.19 to 2.79**	**1.02 to 1.56**	0.70 to 1.74	0.81 to 1.24
p Value	0.45	0.56	0.26	<0.01	<0.01	**0.04**	0.001	**0.01**	**0.03**	0.68	0.97
More severe stroke											
OR	0.94	**1.31**	0.61	0.23	0.11	0.54	0.47	0.91	1.07	1.10	1.08
95% CI	0.87 to 1.02	**1.20 to 1.43**	0.55 to 0.68	0.19 to 0.27	0.09 to 0.13	0.45 to 0.63	0.42 to 0.54	0.76 to 1.09	0.97 to 1.17	0.90 to 1.35	1.00 to 1.18
p Value	0.16	**<0.01**	<0.01	<0.01	<0.01	<0.01	<0.01	0.32	0.19	0.35	0.14

Each cell indicates whether patients with a specific demographic factor are more/less likely to be recruited rather than fall into a particular exclusion category. Scores over 1 indicate that a patient is more likely to be recruited than be excluded for the reason identified (eg, patients with severe stroke are more likely to be recruited than arrive late), while scores under 1 indicate that a patient is less likely to be recruited than fall into that particular exclusion category (eg, female patients are less likely to be recruited than to arrive late).
Severity of stroke is based on an NIHSS estimate of mild, moderate or severe.
Grey cells indicate OR<1, p<0.05.
Bold cells indicate OR>1, p<0.05.
*Adjusted for time (month of trial).
ICU, intensive care unit; mRS, modified Rankin Scale; NIHSS, National Institutes of Health Stroke Scale.

Table 4 Odds of exclusion for a given reason versus non-recruitment for all other reasons, according to age, gender, stroke type and severity in the non-recruited group only (N=18 842 for all columns)*

	After 24 h	mRS≥3	ICU admit	Deteriorated	Coronary condition	Failed physiological	Other trial	Other	Refused	Missed
Increasing age, years										
OR	0.99	**1.08**	0.94	**1.04**	1	0.99	1	0.93	0.99	1
95% CI	0.99 to 0.99	**1.07 to 1.08**	0.94 to 0.95	**1.03 to 1.04**	0.99 to 1.01	0.99 to 1.00	0.99 to 1.00	0.86 to 0.99	0.98 to 1.00	1.00 to 1.00
p Value	<0.01	**<0.01**	<0.01	**<0.01**	0.87	0.01	0.28	0.03	0.07	0.24
Female										
OR	1	**1.46**	0.71	1.06	0.92	1.03	0.79	0.98	1.03	0.93
95% CI	0.94 to 1.07	**1.34 to 1.59**	0.61 to 0.83	0.94 to 1.21	0.75 to 1.13	0.91 to 1.17	0.65 to 0.97	0.92 to 1.06	0.81 to 1.31	0.86 to 0.99
p Value	0.907	**<0.01**	<0.01	0.33	0.44	0.62	0.026	0.62	0.81	0.03
Haemorrhagic stroke										
OR	0.87	0.69	**2.62**	**2.88**	0.53	**1.28**	0.64	0.77	0.82	0.87
95% CI	0.79 to 0.96	0.61 to 0.79	**2.23 to 3.08**	**2.50 to 3.32**	0.38 to 0.74	**1.10 to 1.49**	0.46 to 0.90	0.69 to 0.86	0.55 to 1.22	0.78 to 0.97
p Value	0.01	<0.01	**<0.01**	**<0.01**	<0.01	**<0.01**	0.009	<0.01	0.33	0.01
More severe stroke										
OR	0.59	**1.50**	**4.34**	**10.39**	**1.69**	**1.95**	1.01	0.86	0.76	0.80
95% CI	0.57 to 0.62	**1.42 to 1.58**	**3.89 to 4.85**	**9.27 to 11.65**	**1.49 to 1.92**	**1.81 to 2.11**	0.86 to 1.12	0.82 to 0.90	0.64 to 0.90	0.76 to 0.84
p Value	<0.01	**<0.01**	**<0.01**	**<0.01**	**<0.01**	**<0.01**	0.82	<0.001	<0.001	<0.001

Each cell indicates whether the demographic factor makes a patient more or less likely to be excluded for that particular reason. For example, for each year of age patients had an 8% increase in the odds of having a premorbid mRS≥3.
NIHSS is an estimate, grouped by mild, moderate or severe. Stroke type could be either ischaemic or haemorrhagic.
Grey cells indicate OR<1, p>0.05.
Bold cells indicate OR>1, p>0.05.
*Adjusted for time.
ICU, intensive care unit; mRS, modified Rankin Scale; NIHSS, National Institutes of Health Stroke Scale.

Our findings indicate that women had much greater odds of premorbid disability, were more likely to arrive late to hospital, were more likely to have early deterioration and fail the physiological safety criteria. The odds of exclusion due to early deterioration were particularly high for those with severe stroke. When looking at the between-site heterogeneity in the reasons for non-recruitment, the variability was highest for participation in another trial, admission to ICU and refusal to participate. These observations fit with expected variations in trial activity given that sites are likely to have different competing trials, different ICU admission protocols and different trials experience.

The lower proportion of women recruited to this trial is consistent with the acute stroke trials literature.[21–23] Sex disparities in stroke epidemiology, pathophysiology, treatment and outcomes are well documented in the stroke literature.[24 25] Stroke affects a greater number of women in old age,[25] largely because women have greater longevity and stroke which more commonly occurs in older people. Women are also more likely to be disabled at the time of their stroke.[26 27] Previous stroke disability is a major reason for exclusion from this study for the pragmatic need to optimise the number of patients able to contribute to the primary outcome, independence (score 0–2) on the mRS. While the intervention (rehabilitation) could potentially be suitable or even beneficial for patients with pre-existing disability, the chance of a participant with premorbid disability moving from being disabled (mRS>2) to having a good outcome is low. Our observation that women tended to have a delayed arrival to hospital has been reported in a number of studies examining prehospital delay, although this is not a uniform finding.[28] The most common explanation for this delayed arrival time is that more women than men live alone in older age and arrival times are faster when a stroke is witnessed by another person.[25] It is also possible that poorer recruitment of women reflects an unconscious bias on the part of the recruiter, in favour of recruiting men. Nevertheless, we have found significant relationships that help to explain the phenomenon.

In many acute stroke trials, intracerebral haemorrhage is an exclusion criterion.[21] In the VISTA trials database,[29] 6% of patients have haemorrhagic stroke, which is at the lower limit of estimates for the proportion of stroke due to haemorrhage from international epidemiological data.[19] We were pleased to note that this group was well represented in our trial. Once again, where these patients were excluded, the reasons for their exclusion were strongly aligned with what we already know about this stroke subgroup: early deterioration, failed physiological criteria and ICU admission. The sudden onset of the haemorrhage is often dramatic, resulting in rapid transport to hospital (ie, they are less likely to arrive within 24 h poststroke). Patients with intracerebral haemorrhage often experience early deterioration within hours of stroke and have greater mortality. Therefore, it is not uncommon for these patients to be managed in the ICU rather than on the general stroke unit.[30] Patients with intracerebral haemorrhage were also less likely to be excluded because of a coronary condition and less likely to be recruited to another trial, again consistent with current evidence.[23 31]

One of the unexpected barriers to recruitment to this trial has been the significant proportion of patients who were ineligible because they arrived late to hospital. This, rather than clinical characteristics of potential patients, was the major barrier. This should be considered a modifiable exclusion variable. Delay in hospital admission has implications for delivery of proven stroke therapies such as thrombolysis. There were surprisingly few data from sources, such as local stroke registries or large clinical trials, with a 24 h window, to inform our estimates. Studies of thrombolysis, often from single centres, indicated that between 25% and 59% of patients arrive within 3 h of stroke, with older age, ethnicity and gender (females) influencing arrival time.[32] The large proportion of patients delaying arrival to hospital resulted in the need to enrol a larger number of trial sites than originally planned, and this, in turn, extended the duration of the trial. Fortunately, the duration of a site's participation in the trial appeared to have little impact on the reasons for exclusion and variability between sites was generally low. When variation was observed by site (eg, ICU admission greater in some centres than others), it was easily explained by different care models and processes across the participating hospitals. For example, in some hospitals, it is routine for patients experiencing severe stroke to be admitted to the ICU, while in most hospitals these patients will be managed in the stroke unit. Variability in recruitment to other trials is also readily explained by the fact that some teaching hospitals conduct a large number of trials, while in smaller sites, AVERT was the only active trial in the stroke unit.

The strengths of our study include a large, international data set, collected over several years, that is relevant to the conduct of broad, pragmatic trials. Second, we have used a conceptual framework[12] to explore a broader range of factors than those commonly examined in reports of external validity.[4] We believe that this analytic approach shows promise and encourages a more in-depth exploration of the reasons why different patients are excluded. In the case of AVERT, with the exception of delayed arrival to hospital, most exclusions appeared to be a direct consequence of the trial question or related safety concerns.

The main weakness of our study is that, as a consequence of this being a large pragmatic trial, we have relatively limited information on each patient screened for eligibility for inclusion in the trial. Further, screening log data were only checked against source data at site visits, which occurred on average once per year. The data manager monitored variations in screening

log data on a more regular basis directly with site investigators, and all noted variations were followed up. Data were collected by different staff across eight countries, although all used the same standardised data collection form for screening. It is also important to remember that a large data set may demonstrate statistical significance where it is of little practical importance.[33]

A common complaint of clinicians is uncertainty about the external validity of trial results.[2 7 8] Our understanding of the external validity of RCTs could be improved by routine and standardised exploration of the critical drivers of exclusion from trials in different settings and conditions. In the current CONSORT statement, external validity is discussed, but there are no clear recommendations for reporting.[3] If an agreed framework for examining external validity could be found, standard recommendations could form part of future CONSORT statements, thereby increasing the practical value of clinical trial data reported.

CONCLUSION

In this study using data obtained in AVERT, we have demonstrated that a model of exploring generalisability that considers people, places and processes of care can provide important information about the external validity of a trial. This approach could be applied to other trial populations, but trialists would need to identify and plan for the collection of appropriate variables.

Author affiliations
[1]Florey Institute of Neuroscience and Mental Health, Heidelberg, Victoria, Australia
[2]Westmead Clinical School and The George Institute for Global Health, Westmead Hospital C24, Sydney, New South Wales, Australia
[3]Royal Perth Hospital, Perth, Western Australia, Australia
[4]Physiotherapy Unit, Medical Rehabilitation Services Department, UKM Medical Centre, Kuala Lumpur, Malaysia
[5]Singapore General Hospital, Singapore, Singapore
[6]School of Health Sciences, Flinders University, Repatriation General Hospital, Daw Park, South Australia, Australia
[7]Community and Long Term Conditions Directorate, Auckland District Health Board, Auckland City Hospital, Auckland, New Zealand
[8]Department of Physiotherapy, Singapore General Hospital, Singapore, Singapore
[9]Royal Melbourne Hospital, Parkville, Victoria, Australia
[10]Department of Physiotherapy, Austin Health, Austin Hospital, Heidelberg, Victoria, Australia
[11]Florey Institute of Neuroscience and Mental Health, and Faculty of Medicine, Nursing and Health Sciences, Monash University, Box Hill Hospital, Box Hill, Victoria, Australia
[12]Florey Institute of Neuroscience and Mental Health, University of Melbourne, Parkville, Victoria, Australia
[13]Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK
[14]School of Clinical Sciences at Monash Health, Monash University, Clayton, Victoria, Australia

Twitter Follow Julie Bernhardt at @AVERTtrial

Collaborators The AVERT Trialists' Collaboration: Available at the Lancet http://dx.doi.org/10.1016/S0140-6736(15)60690-0.

Contributors JB designed the trial, interpreted the data, and drafted and revised the paper. She is the guarantor. AR contributed to data interpretation, and drafted and revised the paper. LC analysed and interpreted the data, and revised the draft. RIL contributed to the conception of the trial and revised the draft. SS contributed to data interpretation and revised the draft. JA, MAK, SH, SL, AM, DT, JQ and HCW acquired the data and revised the draft. JC contributed to the conception of the study, monitored the data collection and revised the draft. HMD, GAD and PL contributed to the conception of the study and revised the draft. AGT assisted in the design of the screening tool, contributed to the conception of the study and revised the draft.

Funding This trial was funded through grants from: the National Health and Medical Research Council (project grant numbers: 386201, 1041401), the Stroke Association Australia, Chest Heart and Stroke Scotland (Res08/A114), and Sing Health (SHF/FG401P/2008). All involved researchers are independent of these funding bodies.

Competing interests JB reports grants from the National Health & Medical Research Council, during the conduct of the study; DT reports grants from the National Stroke Research Institute Australia, and grants from the Singhealth Foundation Research Grant, during the conduct of the study; HMD reports grants from the National Health & Medical Research foundation, Australia, during the conduct of the study; AGT reports grants from the National Health & Medical Research Council, during the conduct of the study.

Ethics approval This trial was approved by the appropriate research ethics committee at all participating sites. Committee names (with reference codes) are as follows: the Austin Health Human Research Ethics Committee (HREC) (2006/0215); Royal Perth Hospital HREC (EC2006/123); Melbourne Health HREC (2006.136); Southern Adelaide Clinical HREC (25/067); Western Sydney Local Health District HREC (SAC2006/9/4.6) Northern Sydney Central Coast Health HREC (03/2007 (06–52)); Peninsula Health HREC (2007–08); West Gippsland Healthcare Group HREC (no project number); Hunter New England HREC (06/08/23/5.02); UnitingCare HREC (2007–23); Barwon Health Human Research Ethics Committee (07/40); Auckland District Healthboard Research Review Committee (NTY/07/08/094); Epworth HREC (50010); Royal Brisbane and Women's Hospital HREC (2007/120); Sir Charles Gairdner Hospital HREC (2007–177); St John of God Health Care HREC (02/2008); St Vincent's Hospital Sydney HREC (08/SVH/53); South Eastern Sydney Local Health District HREC (08/039); Coast HREC (06/52); Illawarra Shoalhaven Local Health District HREC (08/SVH/53); New South Wales HREC (08/SVH/53); Human Research Ethics Committee of Western Health (2008.144); Western Hospital HREC (2009.086); North West Wales Research Ethics Committee (09/WNo01/1); Scotland Multi-centre Research Ethics Committee (08/MRE00/38); Singapore General Hospital Institutional Review Board (185/2008); National University of Malaysia Research Ethics Committee (FF-227-2009).

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The authors are willing to share data, and would consult the appropriate ethical approval board depending on the nature of the request. Requests for data sharing or additional data can be emailed to julie.bernhardt@florey.edu.au.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

REFERENCES

1. Higgins JP, Green S. *Cochrane handbook for systematic reviews of interventions*. West Sussex, England: The Cochrane Collaboration, 2008.
2. Rothwell PM. External validity of randomised controlled trials: "To whom do the results of this trial apply?" *Lancet* 2005;365:82–93.
3. Moher D, Hopewell S, Schulz KF, *et al*. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *J Clin Epidemiol* 2010;63:e1–e37.
4. Glasgow RE, Green LW, Klesges LM, *et al*. External validity: we need to do more. *Ann Behav Med* 2006;31:105–8.

5. Rothwell PM. Commentary: external validity of results of randomized trials: disentangling a complex concept. *Int J Epidemiol* 2010;39:94–6.
6. Rothwell PM. Factors that can affect the external validity of randomised controlled trials. *PLoS Clin Trials* 2006;1:e9.
7. Glasgow R, Klesges L, Dzewaltowski D, *et al*. The future of health behavior change research: what is needed to improve translation of research into health promotion practice? *Ann Behav Med* 2004;27:3–12.
8. Glasgow RE, Emmons KM. How can we increase translation of research into practice? Types of evidence needed. *Annu Rev Public Health* 2007;28:413–33.
9. Dekkers OM, von Elm E, Algra A, *et al*. How to assess the external validity of therapeutic trials: a conceptual approach. *Int J Epidemiol* 2010;39:89–94.
10. Jüni P, Altman DG, Egger M. Assessing the quality of controlled clinical trials. *BMJ* 2001;323:42–6.
11. Bernhardt J, Churilov L, Dewey H, *et al*. Statistical analysis plan (SAP) for A Very Early Rehabilitation Trial (AVERT): an international trial to determine the efficacy and safety of commencing out of bed standing and walking training (very early mobilisation) within 24 h of stroke onset vs usual stroke unit care. *Int J Stroke* 2015;10:23–4.
12. Trochim WMK. *The research methods knowledge base*. 2nd edn. William M.K. Trochim, 2006 (cited 18 February 2015). http://www.socialresearchmethods.net/kb/
13. National Stroke Foundation. *National Stroke Audit—Acute Services Organisational Survey report 2013*. Melbourne, Australia: National Stroke Foundation.
14. Veerbeek JM, Kwakkel G, van Wegen EEH, *et al*. Early prediction of outcome of activities of daily living after stroke: a systematic review. *Stroke* 2011;42:1482–8.
15. van Almenkerk S, Smalbrugge M, Depla MFIA, *et al*. What predicts a poor outcome in older stroke survivors? A systematic review of the literature. *Dis Rehabil* 2013;35:1774–82.
16. Briggs D, Felberg R, Malkoff M, *et al*. Should mild or moderate strokes be admitted to an intensive care unit? *Stroke* 2001;32:871–6.
17. Brott T, Adams H, Olinger C, *et al*. Measurements of acute cerebral infarction: a clinical examination scale. *Stroke* 1989;20:864–70.
18. Ali M, Bath PM, Curram J, *et al*. The virtual international stroke trials archive. *Stroke* 2007;38:1905–10.
19. Feigin VL, Lawes CMM, Bennett DA, *et al*. Worldwide stroke incidence and early case fatality reported in 56 population-based studies: a systematic review. *Lancet Neurol* 2009;8:355–69.
20. Feigin VL, Lawes CMM, Bennettt DA, *et al*. Stroke epidemiology: a review of population-based studies of incidence, prevalence, and case-fatality in the late 20th century. *Lancet Neurol* 2003;2:43–53.
21. The VISTA Steering Committees. Development, expansion, and use of a stroke clinical trials resource for novel exploratory analyses. *Int J Stroke* 2012;7:133–8.
22. Schulz KF, Altman DG, Moher D; for the CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *Ann Intern Med* 2010;152:1–7.
23. Busija L, Tao LW, Liew D, *et al*. Do patients who take part in stroke research differ from non-participants? Implications for generalisability of results. *Cerebrovasc Dis* 2013;35:483–91.
24. Haast RA, Gustafson DR, Kiliaan AJ. Sex differences in stroke. *J Cereb Blood Flow Metab* 2012;32:2100–7.
25. Reeves MJ, Bushnell CD, Howard G, *et al*. Sex differences in stroke: epidemiology, clinical presentation, medical care, and outcomes. *Lancet Neurol* 2008;7:915–26.
26. Di Carlo A, Lamassa M, Baldereschi M, *et al*. Sex differences in the clinical presentation, resource use, and 3-month outcome of acute stroke in Europe: data from a multicenter multinational hospital-based registry. *Stroke* 2003;34:1114–19.
27. Gall S, Donnan G, Dewey H, *et al*. Sex differences in presentation, severity, and management of stroke in a population-based study. *Neurology* 2010;74:975–81.
28. Barr J, McKinley S, O'Brien E, *et al*. Patient recognition of and response to symptoms of TIA or stroke. *Neuroepidemiology* 2006;26:168–75.
29. Mishra NK, Diener HC, Lyden PD, *et al*. the VISTA Collaborators. Influence of age on outcome from thrombolysis in acute stroke: a controlled comparison in patients from the Virtual International Stroke Trials Archive (VISTA). *Stroke* 2010;41:2840–8.
30. Morgenstern LB, Hemphill JC, Anderson C, *et al*. Guidelines for the management of spontaneous intracerebral hemorrhage: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke* 2010;41:2108–29.
31. O'Donnell MJ, Xavier D, Liu L, *et al*. Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study. *Lancet* 2010;376:112–23.
32. Lacy CR, Suh DC, Bueno M, *et al*. Delay in presentation and evaluation for acute stroke: Stroke Time Registry for Outcomes Knowledge and Epidemiology (STROKE). *Stroke* 2001;32:63–9.
33. Sankoh AJ, Huque MF, Dubey SD. Some comments on frequently used multiple endpoint adjustment methods in clinical trials. *Stat Med* 1997;16:2529–42.