"To whom do the results of this trial apply?"

External validity of a randomized controlled trial involving 130 patients scheduled for primary total hip replacement

Mette K Petersen[1], Karen V Andersen[1], Niels T Andersen[2] and Kjeld Søballe[1]

Departments of [1]Orthopaedics, Aarhus Sygehus, Aarhus University Hospital, [2]Biostatistics, Aarhus University, Aarhus C, Denmark
Correspondence MKP: mekp@mail.dk
Submitted 06-06-30. Accepted 06-09-18

Background **Although the randomized controlled trial (RCT) is regarded as the gold standard for evaluation of the effect of an intervention, its external validity has been questioned. RCTs cannot be expected to produce results that are directly relevant to all patients and all settings, but they should at least allow patients and clinicians to judge to whom trial results can reasonably be applied.**

We assessed the external validity of an RCT investigating the efficacy of a fast-track program after total hip replacement.

Methods **130 patients were identified as potential participants.18 patients were excluded, 33 enrolled patients declined to participate, and 79 patients were enrolled and randomized. We studied the distribution of preoperative characteristics and postoperative clinical variables in these 3 groups.**

Results **A significant difference was found in both preoperative characteristics and clinical outcome variables. The non-consenters were older, less healthy, and needed more help from the home care system. Furthermore, they were hospitalized longer and were more often transferred to a rehabilitation ward.**

Interpretation **Our findings demonstrate the importance of patient inclusion criteria in RCTs. Moreover, they may account for the lack of reproducibility of RCT results in clinical practice dealing with fast-track programs.**

In an unblinded RCT, we assessed the efficacy of a fast-track program after elective primary unilateral total hip replacement (THR) and perioperative epidural analgesia (Petersen et al. 2006). Randomized controlled trials (RCTs) are frequently considered to be the gold standard of study designs for determining the efficacy of different interventions (Gross et al. 2002, Rothwell 2005). They must be internally valid (i.e. design and conduction) in order to minimize bias, but to be clinically useful the result must also be relevant to a definable group of patients in a particular clinical setting; this is generally termed external validity or generalizability (Rothwell 2005). Even if the randomized comparison in clinical trials is not biased by exclusion per se, external validity of trial results depends on the representativeness of the study sample (Swanson and Ward 1995, Britton et al.1999). The beneficial effects of some interventions can be very dependent on factors such as the characteristics of patients (Altman et al. 2001, Gross et al. 2002, Rothwell 2005). If only a proportion of potentially eligible patients is enrolled in a trial, it is important to evaluate how participants differ from non-participants as a result of eligibility criteria or other factors (Charlson and Horwitz 1984, Gross et al. 2002). Knowledge of dissimilarities in the prevalence of risk factors in the population of participants and non-participants may therefore help to evaluate possible sources of bias, and show whether the effects of treatment, as observed in the trial, may be applied to the general population (Smith and Arnesen 1988, Smith and Arnesen 1990, van Bergen et al. 1995). A critical question is whether non-participants have similar susceptibility compared to participants for the out-

DOI 10.1080/17453670610013367

Table 1. Trial recruitment terminology

Term	Definition [a]	The population under investigation
Target population	Location and characteristics of potentially eligible individuals; represents the individuals to whom the trial results are expected to apply.	Patients scheduled for elective primary unilateral THR and perioperative epidural analgesia (n = 130).
Eligibility fraction	Proportion of potential participants who undergo screening and are eligible to enroll.	Reason for exclusion of enrollment (n = 18): Rheumatoid arthritis (7) Contraindications for epidural analgesia (5) Daily use of opioids (1) Missed for enrollment (3) Unable to communicate in Danish (2)
Enrollment fraction	Proportion of patients who are eligible for participation and who actually enroll.	Patients asked for informed consent (n = 112)
Recruitment fraction	Proportion of potential participants who are actually enrolled and randomized.	Enrolled and randomized patients (n = 79)

[a] According to Gross et al. 2002

come events under study. If not, trial results may be difficult to extrapolate from the population outlined by the eligibility criteria (Charlson and Horwitz 1984). We assessed the external validity and generalizability of an RCT investigating the efficacy of a fast-track program after primary THR.

Patients and methods

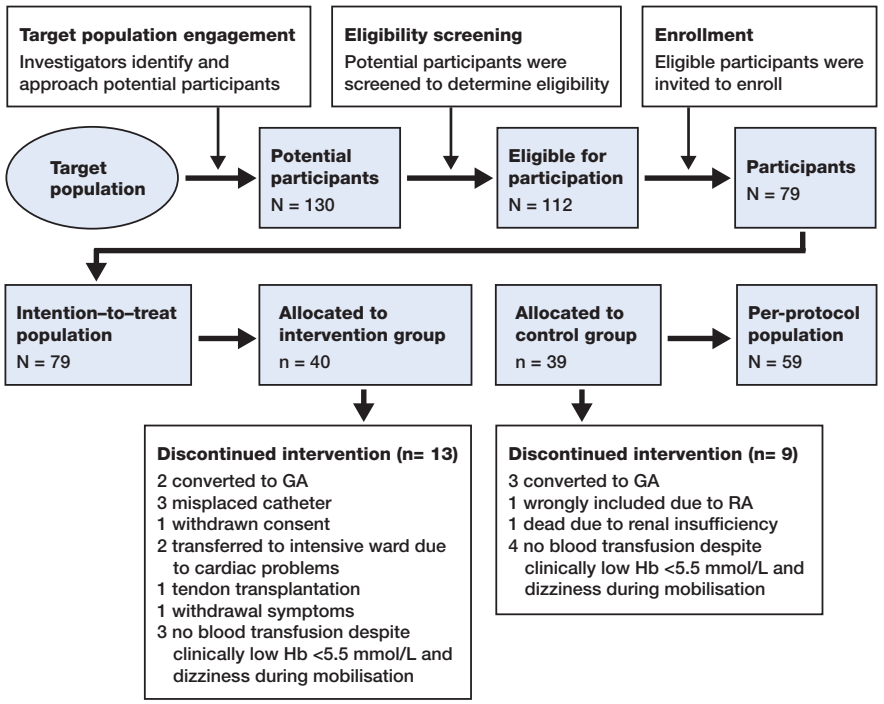
This study is a prospective cohort study with an embedded RCT. The study was approved by the local ethics committee and fulfilled the requirements of the Helsinki Declaration. Patients scheduled for elective primary unilateral THR and perioperative epidural analgesia were assessed for eligibility. Exclusion criteria were chronic opioid use, chronic pain syndrome, rheumatoid arthritis, and mental disorders. In order to estimate sample size for the RCT, we used data on length of hospitalization from the database register at Aarhus University Hospital. With a difference in hospitalization of 30%, expected standard deviation (SD) of 4.2 days, a power of 0.8 and significance level of 0.05, the minimum acceptable size of every group was calculated to be 25 patients.

130 patients were identified as potential participants.18 patients were excluded and 33 patients declined to participate (eligible non-consenters). 79 patients were enrolled and randomized (eligible consenters). The 3 groups (excluded, eligible consenters, and eligible non-consenters) represent the study population. Trial recruitment terminology and population data are described in Table 1.

In the recruitment period, all patients admitted for primary THR were identified as potential participants. Potential participants underwent eligibility screening to determine who was eligible for participation (eligibility fraction). Patients who were eligible for participation were asked to provide informed consent and to enroll in the study (enrollment fraction). The trial enrollment process and the progress through the phases of the RCT are illustrated in the Figure. In the event of ineligibility, the reason for exclusion was stated and described as in Table 1.

All patients in the eligible consenters group and eligible non-consenters group received standardized multimodal anesthesia and analgesia throughout the perioperative and postoperative periods. The epidural catheter was removed after 48 h. Disposal catheters were used when urine retention was > 350 mL, documented by a bladder scan. All patients received daily physiotherapy.

Data for the study were abstracted from evaluation charts that were completed for potential participants. No patients were lost to follow-up. The form included information on age, sex, type of



Flow diagram of the trial enrollment process and the phases of the study.

anesthesia, American Society of Anaesthesiology (ASA) classification, social and occupational factors, preoperative Harris hip score (HHS) (Soderman and Malchau 2001), pre- and postoperative need of home care service, length of hospitalization, transfer to rehabilitation ward, and prevalence of postoperative complications. Postoperative complications and readmission were registered within the first 30 days after surgery.

Length of stay was registered from the day of admission to the day of discharge. Discharge criteria were standardized according to departmental guidelines. In all cases, discharge was considered by departmental surgeons, who were blinded to randomization. Patients were considered for discharge when sufficient pain relief was obtained, estimated as VAS score of > 3 cm resting and > 5 cm moving, when patients were able to maintain personal hygiene, to walk with sticks, and to climb stairs.

The intervention in the RCT involved pre- and postoperative optimization strategies as described in Table 2; the control group received none of these optimized measures. Of 79 patients (the intention-to-treat population) 57 patients (per-protocol population) completed the study protocol (Figure).

Primary outcome of the RCT was length of stay. In the per-protocol analysis, more patients in the intervention group than in the control group were hospitalized for less than 8 days (16/27 (0.6) and 10/30 (0.3), respectively). The median length of stay in the intervention group was 7 (1–9) days, and it was 8 (1–10) days in the control group ($p = 0.02$). In the intention-to-treat analysis, the median length of hospitalization was 8 days in both groups ($p = 0.2$).

Secondary outcome of the RCT

The patients in the intervention group had an average energy intake of 103 kJ/kg (SD 26), compared to 76 kJ/kg (SD 24) in the control group ($p < 0.001$). The average protein intake was 1.25 g/kg (SD 0.35) in the intervention group and 0.74 g/kg (SD 0.25) in the control group ($p < 0.001$). The average total time out of bed was 37 h (SD 10) in the intervention group and 26 h (SD 14) in the control group ($p < 0.001$).

Safety endpoints

There was no difference in complications between groups. The relative risk of complications in the

Table 2. Pre- and postoperative optimization strategies

Preoperative optimization	
• Standard goals for mobilization and energy intake were described	Fixed standard goals for postoperative mobilization and nutrition were introduced and delivered to patients.
• Verbal and written supplementary information was standardized	Mutual expectations were discussed.
• Training in transfer and walking techniques required after surgery.	Transfer out of bed was trained. Walking aids were introduced and delivered. Training in walking with sticks.
Postoperative mobilization	
Aggressive and progressive structured mobilization plans.	Mobilization out of bed for 2 h on the day after surgery. Scheduled time out of bed increased by 2 h a day, from 2 h on the first to 12 h on the sixth postoperative day. Walking distance increased by 100 m a day from 100 m on the second postoperative day to 500 m on sixth postoperative day.
Postoperative nutrition	
Early and aggressive fluid and diet reintroduction. Eating and drinking despite lack of appetite was encouraged.	Registration and calculation of daily fluid and energy intake. Supplementary energy intake: 200 mL of a protein-rich drink (Fortimel; Nutricia, Zoetermeer, the Netherlands) 3 times a day between the main meals.
Postoperative rehabilitation	
Early aggressive rehabilitation program.	Sitting patients out of bed early on the first postoperative day Walking 100 meters on the second postoperative day. Early introduction to exercise program. Encouragement to follow fixed standard goals for mobilization and walking.

intention-to-treat analysis was 1.6 (0.6–4.0) ($p = 0.4$) and in the per-protocol analysis it was 1.7 (0.5–5.3) ($p = 0.5$). No patients were re-admitted.

Statistics

Data with normal distribution were described by means, standard error, and 95% CI, and statistically tested using Student's t-test. If not normally distributed, data were described by medians and range, and the Mann-Whitney U-test was used. Frequency was compared using Fisher's exact test or Kruskal-Wallis test. The crude and adjusted odds ratio (OR; adjusted for age and sex) for consenting was estimated by logistic regression. The significance level was 0.05. We used SPSS version 11.0 for Windows.

Results

Eligible non-consenters were older than eligible consenters ($p = 0.01$), more often classified in ASA group 2 or 3 ($p = 0.01$), had a lower Harris hip score ($p = 0.05$), were more often on transfer income ($p < 0.001$), and received help from the home care service system more often preoperatively ($p = 0.001$) (Table 3). The crude and adjusted (sex and age) odds ratio for consenting is given in Table 4. The unadjusted OR was not significantly different from the adjusted OR. Although no significant difference was seen with regard to sex, the unadjusted OR showed that consenting was higher for men than for women: 1.98 (0.84–0.99). Social status did not appear to have any influence on whether or not a patient would consent.

As the intention-to-treat analysis in the RCT showed no differences between intervention and control patients on length of stay and postoperative complications, the two groups were analyzed together. The length of stay was significantly different in eligible consenters and eligible non-consenters ($p < 0.001$) and more patients in the eligible non-consenter group needed help from the home care service system after discharge from hospital ($p < 0.001$). A larger proportion of the eligible non-consenters were transferred to a rehabilitation ward ($p = 0.001$). More patients in the eligible non-consenter group had urinary tract infections ($p = 0.04$) (Table 5).

Table 3. Characteristics at baseline of eligible consenters, eligible non-consenters, and excluded individuals

Variables	Eligible consenters (n = 79)	Eligible non-consenters (n = 33)	Excluded (n = 18)	P-value [a]
Age	57 (26–84)	70 (27–90)	56 (23–80)	0.008 [b]
Sex				
Male	36/79 (0.5) [f]	10/33 (0.3)	3/18 (0.2)	0.08 [c]
Female	43/79 (0.5)	23/33 (0.7)	15/18 (0.8)	
ASA classification				
ASA class 1	40/79 (0.5)	9/33 (0.3)	3/18 (0.2)	0.01 [d]
ASA class 2	30/79 (0.4)	16/33 (0.5)	12/18 (0.7)	
ASA class 3	9/79 (0.1)	8/33 (0.2)	3/18 (0.2)	
Harris hip score	54.9 (SD 14.3)	44.9 (SD 18.6)	41.9 (SD 14.6)	0.05 [e]
Social factors				
Married	40/79 (0.5)	21/33 (0.6)	9/18 (0.5)	0.2 [c]
Single	39/79 (0.5)	12/33 (0.4)	9/18 (0.5)	
Occupational factors				
Employed	48/79 (0.6)	8/33 (0.2)	7/18 (0.4)	<0.001 [c]
Old-age pensioner	23/79 (0.3)	17/33 (0.5)	6/18 (0.3)	
Invalidity pensioner	8/79 (0.1)	8/33 (0.2)	5/18 (0.3)	
Preoperative home care service				
Yes	3/79 (0.03)	9/33 (0.3)	3/18 (0.2)	<0.001 [c]
No	76/79 (0.1)	24/33 (0.7)	15/18 (0.8)	

[a] Eligible consenters and eligible non-consenters were compared statistically. Excluded patients are described.
[b] Mann-Whitney test; [c] Fisher's exact test; [d] Kruskal-Wallis; [e] Student's test
[f] Proportions are given in parentheses.

Table 4. Odds ratios (OR) for consenting

Variables	Unadjusted OR (95% CI)	Adjusted OR (95% CI) [a]
Age (year)	0.97 (0.94–0.99)	0.97 (0.94–0.99)
Gender		
Male	1.98 (0.84–4.7)	1.91 (0.79–4.60)
Female	1	1
ASA classification [a]		
ASA class 1	1	1
ASA class 2	0.42 (0.16–1.08)	0.61 (0.21–1.77)
ASA class 3	0.25 (0.08–0.84)	0.37 (0.10–1.43)
Harris hip score	1.04 (1.01–1.07)	1.03 (0.99–1.07)
Social factors [a]		
Married	0.59 (0.25–1.35)	0.49 (0.20–1.20)
Single	1	1
Occupational factors [a]		
Employed	1	1
Old-age pensioner	0.49 (0.16–1.47)	0.65 (0.13–3.30)
Invalidity pensioner	0.09 (0.03–0.26)	0.10 (0.02–0.45)
Preoperative home care service [a]		
Yes	0.10 (0.03–0.40)	0.17 (0.04–0.81)
No	1	1

[a] Adjusted for age and sex.

Discussion

We found that there was a significant difference between eligible consenters and eligible non-consenters with respect to important prognostic factors and subsequent clinical outcome variables.

Table 5. Postoperative clinical endpoint variables of eligible consenters, eligible non-consenters, and excluded individuals

Variables	Eligible consenters (n = 79)	Eligible non-consenters (n = 33)	Excluded (n = 18)	P-value [a]
Length of stay	8.00 (1–17)	11.00 (6–53)	9.50 (3–28)	<0.001 [b]
Transfer to rehabilitation ward				
Yes	3/79 (0.03) [d]	9/33 (0.3)	3/18 (0.2)	0.001 [c]
No	76/79 (0.1)	24/33 (0.7)	15/18 (0.8)	
Wound infections				
Yes	2/79 (0.02)	3/33 (0.09)	1/18 (0.05)	0.15 [c]
No	77/79 (0.1)	30/33 (0.9)	17/18 (0.9)	
Urinary tract infections				
Yes	12/79 (0.2)	11/33 (0.3)	6/18 (0.3)	0.04 [c]
No	67/79 (0.8)	22/33 (0.6)	12/18 (0.7)	
Luxation of the hip				
Yes	0/79 (0)	1/33 (0.03)	3/18 (0.2)	0.29 [c]
No	79/79 (1)	32/33 (0.)	15/18 (0.8)	
Postoperative home care service				
Yes	4/79 (0.05)	15/33 (0.4)	5/18 (0.3)	<0.001 [c]
No	75/79 (0.9)	18/33 (0.5)	13/18 (0.7)	

[a] Eligible consenters and eligible non-consenters were compared statistically. Excluded patients are described.
[b] Mann-Whitney test; [c] Fisher's exact test.
[d] Proportions are given in parentheses.

The non-consenters consisted of a subgroup of the screened population to whom the trial therapy could be applied, as this group did not include patients with contraindications to the trial therapy. Because of the low number of patients in our study population, possible bias as a result of lost information should be considered. The validity of length of stay as a clinical endpoint measure can be questioned. In order to minimize bias, we used standardized discharge criteria in all cases as previously recommended (Kehlet and Wilmore 2005).

Despite the potentially important implications of disparities between eligible consenters and eligible non-consenters, only a few studies dealing with fast-track programs have previously supplied postoperative clinical endpoint data in sufficient detail to allow a comparison. Our results are, however, in agreement with the findings of (Husted et al. 2004). Furthermore, it has been shown in population-based surveys (Bergstrand et al. 1983, Goldberg et al. 2001, Hasserius et al. 2002), in primary preventive trials (Wilhelmsen et al.1976), and in clinical trials (Smith and Arnesen 1988, 1990) that there is a higher frequency of non-consenters among subjects with increased risk of disease and mortality.

A basic prerequisite of clinical trials is that the study sample should be realistically representative of the target population for future treatment. Our data reinforce the need for those conducting clinical trials to provide additional information about the recruitment process, supplemented with readily available quantitative data as recommended in order to avoid misleading assessments regarding the degree to which the results may be generalized, and biased estimates of treatment effects (Altman et al. 2001). Our findings demonstrate the importance of patient inclusion criteria in RCTs. Moreover, they may account for the lack of reproducibility of results in clinical practice dealing with fast-track programs.

Contributions of authors

MP: participated in the entire process—planning and execution of the study, data collection, statistical analysis and preparation of the manuscript. KA: assisted in execution of the study, data collection, and in preparation of the manuscript. NA: participated in the planning process, and in the statistical analysis and preparation of the manuscript. KS: performed many of the THRs and participated in the entire process, especially in preparation of the manuscript.

Altman D G, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, Gøtzsche P C, Lang T. The reviced consort statement for reporting randomized trials: Explanation and elaboration. Ann Intern Med 2001; 134: 663-94.

Bergstrand R, Vedin A, Wilhelmsson C, Wilhelmsen L. Bias due to non-participation and heterogenous sub-groups in population surveys. J Chronic Dis 1983; 36: 725-8.

Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C. Threats to applicability of randomised trials: exclusions and selective participation. J Health Serv Res Policy 1999; 4: 112-21.

Charlson M E, Horwitz R I. Applying results of randomised trials to clinical practice: impact of losses before randomisation. Br Med J 1984; 289: 1281-4.

Goldberg M, Chastang J F, Leclerc A, Zins M, Bonenfant S, Bugel I, Kaniewski N, Schmaus A, Niedhammer I, Piciotti M, Chevalier A, Godard C, Imbernon E. Socioeconomic, demographic, occupational, and health factors associated with participation in a long-term epidemiologic survey: a prospective study of the French GAZEL cohort and its target population. Am J Epidemiol 2001; 154: 373-84.

Gross C P, Mallory R, Heiat A, Krumholz H M. Reporting the recruitment process in clinical trials: Who are these patients and how did they get there ? Ann Intern Med 2002; 137: 10-6.

Hasserius R, Karlsson M K, Nilsson B E, Redlund-Johnell I, Johnell O. Non-participants differ from participants as regards risk factors for vertebral deformities; A source of misinterpretation in the European Vertebral Osteoporosis Study. Acta Orthop Scand 2002; 73: 451-4.

Husted H, Holm G, Sonne-Holm S. A small number of patient characteristics influence the length of hospital stay after total hip and knee arthroplasty. Ugeskr Laeger 2004; 166: 3197-200.

Kehlet H, Wilmore D W. Fast-track surgery. Br J Surg 2005; 92: 3-4.

Petersen M K, Madsen C, Andersen N T, Søballe K. Efficacy of multimodal optimization and nutrition in patients undergoing hip replacement: a randomized clinical trial. Acta Anaesthesiol Scand 2006; (in press).

Rothwell P M. External validity of randomised controlled trial: "To whom do the results of this trial apply?" Lancet 2005; 365: 82-93.

Smith P, Arnesen H. Non-respondents in a post-myocardial infarction trial: characteristics and reasons for refusal. Acta Med Scand 1988; 22: 537-42.

Smith P, Arnesen H. Mortality in non-consenters in a post-myocardial infarction. J Intern Med 1990; 228: 253-6.

Soderman P, Malchau H. Is the Harris hip score system useful to study the outcome of total hip replacement? Clin Orthop 2001; (384): 189-97.

Swanson G M, Ward A J. Recruiting minorities into clinical trials: toward a participant-friendly system. J Natl Cancer Inst 1995; 87: 1747-59.

van Bergen P F M M, Jonker J J C, Molhoek G P, van der Burgh [initials], van Domburg R T, Deckers J W, Hofman A. Characteristics and prognosis of non-participants of a multi-centre trial of long-term anticoagulant treatment after myocardial infarction. Int J Cardiol 1995; 49: 135-41.

Wilhelmsen L, Ljungberg S, Wedel H, Werko L. A comparison between participants and non-participants in a primary preventive trial. J Chronic Dis 1976; (29): 331-339.