

## Research Article

# Ensuring Homogeneous Study Groups for Randomized Trials in Spine

Kevin L Ju\*, Rachel M Deering, Dafang Zhang, Mitchel B Harris, Christopher M Bono

Department of Orthopaedic Surgery, Brigham and Women's Hospital, USA

## \*Corresponding author

Kevin L. Ju, Brigham and Women's Hospital, Department of Orthopaedic Surgery, 75 Francis St., Boston, MA 02115; Tel: 617-600-8498; Email: KJU@partners.org

Submitted: 03 November 2014

Accepted: 25 January 2015

Published: 27 January 2015

## Copyright

© 2015 Ju et al.

## OPEN ACCESS

## Keywords

- Power analysis
- Covariate balance
- Randomization

## Abstract

**Background:** Developing a randomized controlled trial requires a power analysis to calculate the number of patients needed to determine if a difference exists between two groups. While it is generally assumed that simple randomization will result in homogeneous groups, post hoc analysis is performed to compare demographical variables, comorbidities, and other covariables. In many cases, the experimental and control groups have significant differences in key covariables (despite adequate sample size) that can influence outcomes. The purpose of our study was to assess covariate frequency differences between mock randomized study groups comprised of patients seen in one spine clinic over a 12-month period.

**Methods:** A retrospective review was performed on all new patients seen in a spine clinic over the course of one calendar year. For each patient, demographical data and variables were recorded. Patients were categorized into 3 groups: 1) all new patients presenting to clinic, 2) new patients who underwent spinal surgery (a subgroup of Group 1), and 3) new patients who underwent lumbar surgery (a subgroup of Group 2). Each group was mock randomized into a control and experimental subgroup. Frequency differences between baseline variables in each subgroup were statistically compared.

**Results:** Group 1 showed an insignificant trend towards differences in the prevalence of diabetes ( $p=0.11$ ), osteoporosis ( $p=0.12$ ), and years smoked ( $p=0.09$ ); Group 2 had statistically significant differences in education level ( $p=0.026$ ) and marital status ( $p=0.022$ ); Group 3 showed an insignificant trend towards differences in age ( $p=0.12$ ) and prevalence of osteoarthritis ( $p=0.07$ ).

**Conclusion:** The risk of producing demographically inequitable groups via randomization is low. In the event that a particular covariable is considered critically influential (e.g. diabetes in a study of lumbar fusion), block randomization based on known confounders may be useful to minimize covariate imbalance in addition to enrolling enough patients based on the power analysis.

## INTRODUCTION

Randomized controlled trials (RCTs) are widely accepted as the most objective and unbiased method for evaluating the effects of two or more treatments on a particular disorder [1,2]. The key premise behind a well-designed RCT is that patients are assigned randomly and unpredictably to treatment and control groups, ideally minimizing selection bias and balancing known and unknown confounders [3]. When developing an RCT, an *a priori* power analysis is recommended to calculate the minimum sample size needed to detect an anticipated outcome difference between treatment and control groups.

Despite the fact that randomization assigns patients to control and experimental groups independent of their baseline characteristics, it does not guarantee that these groups will be balanced in terms of their baseline characteristics. Though more concerning with smaller studies, even large RCTs can have experimental and control groups that have significant differences in key covariables. Imbalance of these baseline covariables (i.e. covariate imbalance) and/or sample sizes between study groups decreases the power of the trial and can undermine the validity and credibility of the study's conclusions [4,5].

Based on these observations of previously published studies,

the authors hypothesized that simple randomization will not necessarily achieve covariate homogeneity between two study groups. We further hypothesized that a critical number of patients might exist beyond which randomization of key covariables is ensured. In following, the purpose of this study was to assess covariate balance of patients seen in one spine clinic over a 12-month period who were mock randomized.

## MATERIALS AND METHODS

Following institutional review board approval, a retrospective review of medical records of new patients seen in a single spine surgeon’s clinic over the course of one calendar year was performed. Demographical data was collected for each patient, including age, gender, race, education level, marital status, work status, and whether the patient was a manual laborer. In addition, other covariables that are known or have been suggested to influence the outcome of spinal procedures were also examined. This included BMI [6], smoking status and duration [7,8], previous spine surgery [9], drug use [10], and various other non-spine conditions [11] (e.g. depression, osteoarthritis, diabetes, psychiatric disorder). Finally, if the patient ultimately underwent surgery, the site and type of surgery was documented. Study data were collected and managed using the Research Electronic Data Capture (REDCap) electronic data capture tool.

Descriptive statistics were first performed on the whole cohort (Group 1). Patients who ultimately underwent spinal surgery constituted a subgroup of the whole cohort (Group 2). An additional subgroup (Group 3) was comprised of those who underwent lumbar spine surgery. All three groups were mock randomized into two subgroups (e.g. mock experimental and control groups) using Microsoft Excel 2007 (Microsoft, Redmond, WA), simulating three separate theoretical studies. Baseline characteristics for the groups in each of the three theoretical studies were compared using Spearman correlations, Chi-squared and Fisher’s exact tests, and Wilcoxon rank sums. All statistical analyses were performed using SAS version 9.2 (SAS Institute, Inc., Cary, NC). A p-value of less than 0.05 was considered to be significant. Institutional review board committee approval was obtained before initiating the study. There was no external funding source for this study, and the institutional funding did not influence the investigation.

## RESULTS AND DISCUSSION

In total, 589 new patients were seen in a single spine surgeon’s clinic over the course of the 2011 calendar year. For these 589 patients, summary demographic information is shown in **Table 1**, clinical data is shown in **Table 2**, and surgical data is shown in **Table 3**. Briefly, the mean age of all new patients was 55 years and the mean BMI was 28.86. There were roughly equal numbers of men and women, 50% of patients were employed at the time of initial evaluation, 39% were current or previous smokers, and 23% of patients had previously undergone spine surgery. Of these new patients, 28% went on to have spinal surgery.

These 589 patients (Group 1) were then mock randomized into two groups (Group 1A and Group 1B) to simulate our first randomized study (**Table 4**). When the two groups were compared with regards to baseline characteristics, substantial

**Table 1:** Demographical snapshot for *all* new patients presenting to clinic in 2011.

Variable	Mean	95% CI
Age	55.17	53.98-56.37
BMI	28.86	28.34-29.37
Years Smoked (if applicable)	19.64	17.48-21.80
<b>n (%)</b>		
Sex		
Male	274 (46.52)	
Female	315 (53.48)	
Race		
Caucasian	517 (87.78)	
African American	33 (5.60)	
Hispanic	20 (3.40)	
Asian	9 (1.53)	
Other	2 (0.34)	
Education		
Some High School	19 (3.23)	
High School Graduate/GED	129 (21.90)	
Some College/Vocational/Technical Program	111 (18.85)	
Graduate of College or Postgraduate School	279 (47.37)	
Marital Status		
Single	115 (19.52)	
Married	374 (63.50)	
Divorced	52 (8.83)	
Widowed	34 (5.77)	
Other	2 (0.34)	
Work Status		
Employed	296 (50.25)	
Unemployed	61 (10.36)	
Retired	111 (18.85)	
Disabled	28 (4.75)	
Worker’s Compensation	1 (0.17)	
Homemaker	20 (3.40)	
Manual Labor		
Yes	34 (5.77)	
No	456 (77.42)	

Some percentages do not add up to 100% as data was unavailable for some subjects.

(but not significant) differences were seen in the prevalence of diabetes (p = 0.11), osteoporosis (p = 0.12), and years smoked (p = 0.09). Of the Group 1 patients, 163 ultimately underwent spinal surgery. These 163 surgical patients (Group 2) were mock randomized into two groups (Group 2A and Group 2B) to simulate a second randomized study comprised of only surgical patients (**Table 5**). This yielded a statistically significant difference in education level (p = 0.026) and marital status (p = 0.022). Our third simulated study consisted of the 132 patients who underwent lumbar spine surgery (Group 3). When this subgroup was randomized into two groups (Group 3A and Group 3B), substantial (but not significant) differences were observed in age (p = 0.12) and the prevalence of osteoarthritis (p = 0.07) (**Table 6**).

**Table 2:** Clinical snapshot for all new patients presenting to clinic in 2011.

Variable	n (%)
Previous Surgery	
No	443 (75.21)
Yes	137 (23.26)
Previous Surgery Location	
Cervical	32 (23.36)
Thoracic	5 (3.65)
Lumbar	97 (70.80)
Current or Previous Smoker	
Yes	229 (38.88)
No	360 (61.12)
Drug Use	
Yes	21 (3.57)
No	493 (83.70)
Comorbidities	
Osteoarthritis	100 (16.98)
Depression	65 (11.04)
Diabetes	61 (10.36)
Psychiatric Disorder	25 (4.25)
Inflammatory Arthritis	21 (3.57)
Migraines	17 (2.89)
Osteoporosis	3 (0.51)
Fibromyalgia	11 (1.87)
Non-Spinal Musculoskeletal Disorder	5 (0.85)
Systemic Neurological Disorder	10 (1.70)
Thoracic Outlet Syndrome	1 (0.17)
Ankylosing Spondylosis	1 (0.17)

Some percentages do not add up to 100% as data was unavailable for all subjects.

**Table 3:** Surgical snapshot for all new patients presenting to clinic in 2011.

Variable	n (%)
Surgery	
No	426 (72.33)
Yes	163 (27.67)
Surgery Location	
Cervical	29 (17.79)
Thoracic	2 (1.23)
Lumbar	132 (80.98)
Surgery Type	
ACDF	13 (7.98)
PCLF	12 (7.36)
Lumbar discectomy	30 (18.40)
Lumbar laminectomy and fusion	58 (35.58)
Other	50 (30.67)

Continuous data shown as means, and categorical data shown as n (%)

Though RCTs have long been seen as the gold standard for minimizing confounders [1,2], simple randomization does not guarantee covariate balance. However our study illustrates that the risk of this occurring in spinal surgery patients is generally low. We investigated the distribution of baseline characteristics

in three hypothetical RCTs in which new patients from a spine surgeon's practice were randomized into treatment and control groups. Mock randomization of the 132 patients who underwent lumbar spine surgery (Group 3) produced insignificant differences in age and osteoarthritis (Table 6), which are probably unlikely to influence the outcomes of a study. When all 589 new patients (Group 1) were assigned to two groups by simple randomization (Table 4), there was a slight trend, though statistically insignificant, towards a difference in the prevalence of diabetes and years smoked. Though insignificant, these differences might be problematic if the study was investigating surgical infection rates or fusion success, as diabetes and smoking are known risk factors [7,8,12,13].

The only statistically significant findings in the current study were found with mock randomization of the 163 patients who underwent spinal surgery (Group 2). This showed differences in the educational level and marital status between the two groups (Table 5). A patient's educational level has been shown to affect outcomes following spine surgery. Cobo Soriano *et al* demonstrated that individuals who were less educated had significantly less improvement in Oswestry disability index scores and less pain relief after lumbar decompression and fusion surgery [14]. Prior studies have found higher rates of depression in non-married individuals compared to their married

**Table 4:** Demographical and clinical snapshot for all new patients, by mock randomization group.

Variable	Group 1A (mean)	Group 1B (mean)	p-value
Age	55.39	54.97	0.7325
Years Smoked (if applicable)	19.25	22.89	0.0909
	n (%)	n (%)	
Education			
Some High School	9 (3.32)	10 (3.75)	0.9567
High School Graduate/GED	68 (25.09)	61 (22.85)	
Some College/Vocational/ Technical Program	51 (18.82)	60 (22.47)	
Graduate of College or Postgraduate School	143 (52.77)	136 (50.94)	
Marital Status			
Single	61 (21.63)	54 (18.31)	0.4506
Married	181 (64.18)	193 (65.42)	
Divorced	23 (8.16)	29 (9.83)	
Widowed	15 (5.32)	19 (6.44)	
Other	2 (0.71)	--	
Comorbidities			
Osteoarthritis	50 (17.30)	50 (16.67)	0.8376
Depression	30 (10.38)	35 (11.67)	0.6185
Diabetes	24 (8.30)	37 (12.33)	0.1087
Osteoporosis	3 (1.04)	- (-)	0.1175 <sup>‡</sup>

Continuous data shown as means, and categorical data shown as n (%)

<sup>‡</sup>Fisher's exact test

**Table 5:** Demographical and clinical snapshot for all *surgical* patients, by mock randomization group.

Variable	Group 2A	Group 2B	p-value
	(mean)	(mean)	
Age	57.10	58.04	0.6787
Years Smoked (if applicable)	21.21	16.53	0.3253
	n (%)	n (%)	
Education			
Some High School	1 (1.32)	- (--)	0.0262 <sup>±*</sup>
High School Graduate/GED	11 (14.47)	21 (29.17)	
Some College/Vocational/Technical Program	23 (30.26)	11 (15.28)	
Graduate of College or Postgraduate School	41 (53.95)	40 (55.56)	
Marital Status			
Single	6 (7.32)	14 (18.18)	0.0217 <sup>±*</sup>
Married	68 (82.93)	48 (62.34)	
Divorced	3 (3.66)	8 (10.39)	
Widowed	4 (4.88)	7 (9.09)	
Other	1 (1.22)	- (--)	
Comorbidities			
Osteoarthritis	13 (15.66)	16 (20.00)	0.4692
Depression	8 (9.64)	7 (8.75)	0.8445
Diabetes	7 (8.43)	9 (11.25)	0.5458
Osteoporosis	- (--)	2 (2.50)	0.2393 <sup>±</sup>

Continuous data shown as means, and categorical data shown as n (%)

<sup>±</sup>Fishers exact test

<sup>\*</sup>Significant p-value

**Table 6:** Demographical and clinical snapshot for *lumbar surgical* patients, by mock randomization group.

Variable	Group 3A	Group 3B	p-value
	(mean)	(mean)	
Age	60.26	56.18	0.1190
Years Smoked (if applicable)	20.09	15.80	0.4355
	n (%)	n (%)	
Education			
Some High School	- (--)	- (--)	0.5551
High School Graduate/GED	11 (19.30)	11 (17.46)	
Some College/Vocational/Technical Program	15 (26.32)	12 (19.05)	
Graduate of College or Postgraduate School	31 (54.39)	40 (63.49)	
Marital Status			

Single	5 (8.06)	12 (18.18)	0.2580
Married	45 (72.58)	46 (69.70)	
Divorced	6 (9.68)	3 (4.55)	
Widowed	6 (9.68)	4 (6.06)	
Other	- (--)	1 (1.52)	
Comorbidities			
Osteoarthritis	16 (24.24)	8 (12.12)	0.0710
Depression	9 (13.64)	4 (6.06)	0.2420
Diabetes	9 (13.64)	5 (7.58)	0.2582
Osteoporosis	1 (1.52)	- (--)	1.0000

Continuous data shown as means, and categorical data shown as n (%)

counterparts [15-18], and patients with depression are known to have significantly poorer spinal surgery outcomes[11].

The authors' secondary hypothesis does not appear to be supported by this data. In other words, a critical range of the number of patients beyond which covariate imbalance is diminished (or eliminated) was not found. As indicated above, the data demonstrates that the only significant differences were found in group 2, which was comprised of 163 patients, while a smaller group of patients (group 3, who had undergone lumbar surgery) did not show similar differences. Thus, it would appear that covariate balance may be influenced by other factors in addition to patient numbers, such as underlying diagnosis or procedure performed.

Notwithstanding the current findings, it is important to note the potential influence of demographical covariables on the outcomes of spinal surgery. In the aforementioned study, Katz et al. also found that patients who had musculoskeletal comorbidities such as osteoarthritis, lower subjective health ratings, or greater cardiovascular or overall comorbidities had significantly lower outcome scores after surgery [11]. Increasing age is not only associated with a higher prevalence of comorbidities, but it is also independently associated with lower patient-reported outcomes after lumbar spine surgery [19].

Covariate imbalance is not just a theoretical pitfall. Close inspection of the baseline characteristics between treatment groups of large randomized controlled trials in the spine literature reveals this phenomenon to varying degrees. The Spine Patient Outcomes Research Trial (SPORT) studies are a collection of well-known multicenter randomized controlled trials comparing nonoperative versus surgical treatments for lumbar spine conditions. Examination of the baseline characteristics for the 2008 SPORT paper on spinal stenosis reveals that the group undergoing surgery was younger (p = 0.004) and more likely to be employed (p = 0.05) and married (p = 0.06) compared to the non-operative group [20]. Additionally, the surgical group had more pain (p <0.001), a lower level of function (p <0.001), more psychological distress (p = 0.02), and more self-reported disability (p <0.001) than patients in the non-surgical group [20]. Among other possible factors, these differences were likely to the result of chance from randomization. The 2007 SPORT study on spondylolisthesis similarly demonstrated chance differences in age (p <0.001), prevalence of cardiovascular comorbidities (p =

0.055), and self-reported disability ( $p < 0.001$ ), pain ( $p < 0.001$ ), and level of function ( $p < 0.001$ ) [21]. Even though the authors recognized these differences and attempted to control for them in their multivariate statistical analysis, covariate imbalance nonetheless detracts from the study's power and increases the risk of confounding.

If deemed appropriate, one option for addressing covariate imbalance during univariate analyses is to conduct post-stratification tests, which involves classifying subjects into strata after enrollment and subsequently performing subgroup analyses. However smaller studies may not be amenable to this, as further dividing patients into subgroups will create smaller sample sizes, thus reducing statistical power. This method may also introduce bias into the study as the variables chosen for stratification can be done after one has already examined the actual trial results and data.

The more methodologically sound strategy is to employ a tactic that balances variables prior to assigning patients to specific groups. One such method is using block randomization based on covariables that may influence the study's primary outcome measure. This allocation technique segregates subjects into similar blocks based on influential covariables and then randomizes subjects within each block to treatment and control groups, thus ensuring that key outcome-related characteristics and group size will be more balanced between treatment and control groups. In fact, Xiao *et al.* showed that block randomization consistently yielded better balance and power when compared to other strategies such as simple randomization [22].

## CONCLUSION

The current study demonstrates that simple randomization carries a low, but present, risk for producing significant differences between groups of spine patients for most demographical covariables. In the end, it seems that the risk will vary with each randomization based on chance and does not have a critical threshold beyond which risk is substantially minimized. In the event that a certain variable is considered an important influence on the outcomes of a study, strategies such as block randomization may be considered.

## CONFLICT OF INTEREST

Each author certifies that he or she has no commercial associations (eg, consultancies, stock ownership, equity interest, patent/licensing arrangements, etc) that might pose a conflict of interest in connection with the submitted article.

## REFERENCES

- D'Agostino RB, Kwan H. Measuring effectiveness. What to expect without a randomized control group. *Med Care*. 1995; 33: AS95-105.
- Ottenbacher K. Impact of random assignment on study outcome: an empirical examination. *Control Clin Trials*. 1992; 13: 50-61.
- Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ*. 1998; 317: 1185-1190.
- Atkinson AC1. The distribution of loss in two-treatment biased-coin designs. *Biostatistics*. 2003; 4: 179-193.
- Lachin JM. Properties of simple randomization in clinical trials. *Control Clin Trials*. 1988; 9: 312-326.
- Seicean A, Alan N, Seicean S, Worwag M, Neuhauser D, Benzel EC. Impact of increased body mass index on outcomes of elective spinal surgery. *Spine (Phila Pa 1976)*. 2014; 39: 1520-1530.
- Andersen T, Christensen FB, Laursen M, Høy K, Hansen ES, Bünger C. Smoking as a predictor of negative outcome in lumbar spinal fusion. *Spine (Phila Pa 1976)*. 2001; 26: 2623-2628.
- Glassman SD, Anagnost SC, Parker A, Burke D, Johnson JR, Dimar JR. The effect of cigarette smoking and smoking cessation on spinal fusion. *Spine (Phila Pa 1976)*. 2000; 25: 2608-2615.
- Frymoyer JW, Matteri RE, Hanley EN, Kuhlmann D, Howe J. Failed lumbar disc surgery requiring second operation. A long-term follow-up study. *Spine (Phila Pa 1976)*. 1978; 3: 7-11.
- Lawrence JT, London N, Bohlman HH, Chin KR. Preoperative narcotic use as a predictor of clinical outcome in degenerative anterior cervical arthrodesis. *Spine (Phila Pa 1976)*. 2008; 33: 2074-2078.
- Katz JN, Stucki G, Lipson SJ, Fossel AH, Grobler LJ, Weinstein JN. Predictors of surgical outcome in degenerative lumbar spinal stenosis. *Spine (Phila Pa 1976)*. 1999; 24: 2229-2233.
- Wimmer C, Gluch H, Franzreb M, Ogon M. Predisposing factors for infection in spine surgery: a survey of 850 spinal procedures. *J Spinal Disord*. 1998; 11: 124-128.
- Fang A, Hu SS, Endres N, Bradford DS. Risk factors for infection after spinal surgery. *Spine (Phila Pa 1976)*. 2005; 30: 1460-1465.
- Cobo Soriano J, Sendino Revuelta M, Fabregate Fuente M, Cimarra Diaz I, Martinez Urena P, Deglane Meneses R. Predictors of outcome after decompressive lumbar surgery and instrumented posterolateral fusion. *European spine journal: official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society*. 2010; 19: 1841-1848.
- Simon RW. Revisiting the relationships among gender, marital status, and mental health. *AJS*. 2002; 107: 1065-1096.
- Blumenthal MD. Mental health among the divorced. A field study of divorced and never divorced persons. *Arch Gen Psychiatry*. 1967; 16: 603-608.
- Inaba A, Thoits PA, Ueno K, Gove WR, Evenson RJ, Sloan M. Depression in the United States and Japan: gender, marital status, and SES patterns. *Soc Sci Med*. 2005; 61: 2280-2292.
- Pearlin L, Johnson JS. Marital status, life-strains and depression. *Am Sociol Rev*. 1977; 42: 704-715.
- Slover J, Abdu WA, Hanscom B, Weinstein JN. The impact of comorbidities on the change in short-form 36 and oswestry scores following lumbar spine surgery. *Spine (Phila Pa 1976)*. 2006; 31: 1974-1980.
- Weinstein JN, Tosteson TD, Lurie JD, Tosteson AN, Blood E, Hanscom B. Surgical versus nonsurgical therapy for lumbar spinal stenosis. *N Engl J Med*. 2008; 358: 794-810.
- Weinstein JN, Lurie JD, Tosteson TD, Hanscom B, Tosteson AN, Blood EA. Surgical versus nonsurgical treatment for lumbar degenerative spondylolisthesis. *N Engl J Med*. 2007; 356: 2257-2270.
- Xiao L, Lavori PW, Wilson SR, Ma J. Comparison of dynamic block randomization and minimization in randomized trials: a simulation study. *Clin Trials*. 2011; 8: 59-69.

### Cite this article

Ju KL, Deering RM, Zhang D, Harris MB, Bono CM (2015) Ensuring Homogeneous Study Groups for Randomized Trials in Spine. *Ann Orthop Rheumatol* 3(1): 1041.