ORIGINAL ARTICLE

Management of Confounding in Controlled Orthopaedic Trials

A Cross-sectional Study

**Patrick Vavken MD, Georg Culen MD,
Ronald Dorotka MD**

Abstract Confounding occurs when the effect of an exposure on an outcome is distorted by a confounding factor and will lead to spurious effect estimates in clinical studies. Although confounding can be minimized at the design stage, residual confounding may remain. An argument therefore can be made for controlling for confounding during data analysis in all studies. We asked whether confounding is considered in controlled trials in orthopaedic research and hypothesized the likelihood of doing so is affected by participation of a scientifically trained individual and associated with the magnitude of the impact factor. We performed a cross-sectional study of all controlled trials published in 2006 in eight orthopaedic journals with a high impact factor. In 126 controlled studies, 20 (15.9%; 95% confidence interval, 9.5%–22.3%) studies discussed confounding without adjusting in the analysis. Thirty-eight (30.2%; 95% confidence interval, 22.2%–38.2%) controlled for confounding, although we suspect the true proportion might be somewhat higher. Participation of a methodologically trained researcher was associated with (odds ratio, 3.85) controlling for confounding, although there was no association between impact factor and controlling for confounding. The

question remains to what extent the validity of published findings is affected by failure to control for confounding.

Introduction

In recent years, orthopaedic surgery has moved toward evidence-based patient management [3, 12, 15]. Numerous articles and reviews explore the basics of research methodology to provide a sound understanding of proper design and conduct of studies [1, 2, 5, 7, 8, 11, 15]. Associations found in trials, however, may derive from four sources: chance, bias, confounding, or a true effect. Although we strive for detecting a true effect, we can only assume a true effect after the four sources are demonstrably unlikely enough to be neglected.

As noted, an association may be attributable to a confounder [14]. A confounder is a variable explaining, in part or in total, the association between an exposure and an outcome (Fig. 1). This poses a major threat to the validity of evidence. The risk of confounding can be minimized during the design stage of a trial, but the possibility of residual confounding remains, especially with flaws in study design such as insufficient sample size. Although sometimes regarded as a special form of bias, confounding and bias are distinct entities, and unlike bias, confounding can be evaluated quantitatively and controlled for during the analysis of the study. Even in carefully designed studies, this should be done to ensure the validity of findings. Although there is a wealth of literature on this issue, we do not know whether these suggestions are followed in practice.

We therefore asked what percentage of controlled trials published in eight highly ranked orthopaedic journals during 2006 considered confounding and its effects.

P. Vavken, R. Dorotka (✉)
Department of Orthopedic Surgery, Medical University
of Vienna, Währinger Gürtel 18-20, 1090 Vienna, Austria
e-mail: ronald.dorotka@meduniwien.ac.at

G. Culen
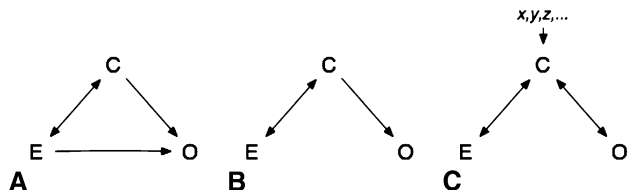UKH Meidling, Kundratstraße 37, Vienna, Austria

Fig. 1A–C Possible scenarios of confounding are shown. One-way arrows describe a causal relationship; two-way arrows describe an association. (**A**) The association between exposure (E) and outcome (O) is partially confounded by the confounder (C). (**B**) The confounder (C) causes the outcome (O). Because exposure (E) and C are associated, looking at E and O only, there is a spurious association. An example for this is the association of coffee drinking and pancreatic cancer, which actually is the result of the associations between coffee drinking and smoking and smoking and pancreatic cancer. (**C**) Again, there is a spurious association between exposure (E) and outcome (O) resulting from the confounder (C). In this case, however, there is no causal relationship between C and O; C is a surrogate for the causal variables x, y, z. An example would be area of residence as a confounder, which is a surrogate for lifestyle, environmental exposure, and access to medical treatment.

Furthermore, we asked whether participation of an individual specifically trained in research methods affected the likelihood of controlling for confounding and whether this likelihood is associated with the magnitude of the impact factor. Finally, we searched for the variables controlled for and methods used.

Materials and Methods

To answer these questions, we conducted a cross-sectional study of a sample of controlled trials reported in the orthopaedic literature in 2006. According to the 2005 journal citation report, we included the top 20% of orthopaedic journals (eight journals) ranked by impact factor. Online and print issues of the respective journals were searched for articles on controlled trials on human subjects. We included the following journals: *Osteoarthritis and Cartilage, Journal of Orthopaedic Research, Clinical Journal of Sport Medicine, Journal of Bone and Joint Surgery–American Volume, Spine, Connective Tissue Research, European Spine Journal*, and *The Orthopedic Clinics of North America*. From January 1, 2006, to December 31, 2006, these journals published 2265 papers. Our review produced 126 (5.6% of all studies) controlled trials. Among those, 53 (42.1%) were randomized, 51 (40.5%) reported on losses to followup, but only 34 (27.0%) were at least single-blinded. The overall Jadad score [6] (see below) was 2 ± 0.97 points (95% confidence interval [CI], 1.8–2.2 points). Thirty (24.6%) studies report sample size or post hoc power calculations.

All data were gathered independently, digitally collected in spreadsheets, and crosschecked. Disagreement in classification between investigators was resolved by consensus or the senior author (RD) was consulted.

Two investigators (PV, GC) independently reviewed the Materials and Methods sections for the primary and secondary outcomes of interests, reporting of controlling for confounding, and discussion of confounding. Participation of an investigator with a degree in science (PhD, MSc) and the impact factor of the journal were recorded as risk factors of interest. For descriptive purposes, the investigators furthermore determined for which variables adjustments were done and which methods were used to do so and assessed the overall quality of the included trials using a modified Jadad score [6]. This score gathers information on randomization, blinding, and followup to calculate an estimate of the internal validity of the results of a trial. Giving 1 point for each variable, study quality may range from 0 (poor) to 3 (best).

We calculated the percentages of studies discussing confounding and reporting of controlling for confounding. The effect of participation of a scientifically trained individual on both outcomes was estimated by odds ratios (ORs). The statistical significance of these ORs was assessed using chi square tests. The influence of the magnitude of the impact factor on the outcome variables was estimated using logistic regression. Results are displayed as absolute numbers, percentages, and 95% binominal CIs; an α level of 5% was considered the threshold of significance. We used Stata® 9.2 (StataCorp LP, College Station, TX) for all analyses.

Results

Thirty-eight of the 126 (30.2%; 95% CI, 22.2%–38.2%) studies controlled for confounding. Eleven of these studies (8.7%) also discussed confounding, whereas 27 studies (21.4%) merely included suspected variables in their analyses. Twenty (15.9%; 95% CI, 9.5%–22.3%) studies explicitly discussed potential confounders (Table 1).

Confounding was controlled (OR, 3.9; 95% CI, 1.2–11.9; p = 0.0156) for more often if a scientifically trained individual participated in the study (Table 2). The

Table 1. Consideration of confounding in controlled trials

Study question	Number of studies	Percent	95% Confidence interval
All included trials	126	100	
Confounding discussed?	20/126	15.9	9.5–22.3
Confounding controlled?	38/126	30.2	22.2–38.2
Confounding controlled in literature [10]		32	

Table 2. Effect of participation of scientifically trained individual

Study question	Number of studies	Odds ratio	95% Confidence interval	p Value (Chi square)
Confounding discussed?	16/20	1.77	0.5–8.4	0.3706
Confounding controlled?	32/38	3.85	1.1–14.4	0.0156

magnitude of the impact factor of the journal was not related to the likelihood of controlling for confounding. Neither participation (OR, 1.8; 95% CI, 0.5–8.4; p = 0.3706) nor the magnitude of the impact factor (p = 0.951) was associated with discussion of confounding. There was no association between discussion of and controlling for confounding (OR, 0.996; 95% CI, 0.31–3.26; p = 0.9938).

Among the studies controlling for confounding, age was most commonly included as a potential confounder (22%) together with diagnosis-related variables (20%) and gender (17%). Twenty percent did not report which variables were used. Of all studies that did report on the variables used, only 40% explicitly did so in the Materials and Methods section.

The most frequently used techniques to control for confounding were stratification (40%) and multivariate logistic regression (40%). Seventeen percent used multiple regression, 3% used ordinal logistic regression, and 11% used nonspecified multivariate analyses. Numerous studies used more than one technique. Eleven percent did not report which technique was used but clearly stated the investigators controlled for confounding.

Discussion

Confounding is a clear threat to the validity of research findings. We asked whether confounding is considered in orthopaedic trials and whether the participation of an investigator with formal scientific training increases the likelihood of doing so. Furthermore, we asked whether the magnitude of the impact factor would predict this likelihood, and how and for what variables controls were considered. We collected data on these issues from trials published in highly ranked orthopaedic journals in 2006.

We note two major limitations in our study. First, our data build on reported information, which is not necessarily complete or correct. This is a well-known problem when extracting data from publications [4, 13]. Not only errors but also considerations concerning publication length and print space may result in limited data, especially related to statistical detail. This means the assumption that not reported equals not performed might result in substantially biased estimates. Second, the decision to base our sample

on impact factors may limit the generalizability of our results. Nonetheless, we decided to use impact factors because we hypothesized trials from journals with high impact factors are more likely to control for confounding, which we found not to be true, and because the validity of the results of these frequently referred to articles is especially interesting because weaknesses might be perpetuated.

We found ⅓ of the studies in our sample controlled for confounding in the analysis. The odds of controlling for confounding are considerably higher if a scientifically trained individual participates in the study. We found no association between the magnitude of the impact factor and controlling for confounding. Only ½ of the studies controlling for confounding, however, discussed this explicitly. Although confounding frequently is discussed in the scientific literature, there are few studies scrutinizing its management. We believe there currently is no such study for any surgical specialty. However, a study presented in 2002 by Müllner et al. [10], reported on confounding in 537 articles from 34 highly ranked journals from mostly medical specialties and found 32% controlled for confounding compared with 30.2% in our sample. Our finding that participation of a methodologically trained researcher is associated with controlling for confounding is confirmed by this study. We stress this individual does not need to be a statistician but can be an orthopaedic surgeon as well. Given adequate training of the participating surgeon, this option may be preferable because such an investigator will be able to correctly judge the statistical meaning and clinical importance of a variable. We found no association between controlling for and reporting of confounding however. One possible explanation is that controlling was considered but not mentioned in the text, which would not affect the validity of the results, but seems worthwhile reporting nonetheless to allow the reader to appraise the validity of the findings or to publish newly identified confounders. The other explanation is that there was discussion in the text without control, which clearly jeopardizes the validity of results, and authors should comprehensively elaborate reasons for this.

A confounder is defined as a factor that partially or entirely explains an association between an exposure and an outcome, thus leading to a spurious effect estimate (Fig. 1). To do so, the confounder has to be associated with the exposure and outcome but must not be part of the causal pathway between exposure and outcome. The classic example is smoking confounding the association between drinking coffee and having cancer develop, which then develops in a patient as a result of the confounder smoking, not the exposure to coffee. Another example more relevant to orthopaedics is age as confounder of the association between physical activity and knee pain in osteoarthritis. People who exercise heavily tend to be

younger and in turn have a lower risk of knee pain; thus, any association between activity and knee pain would be exaggerated if the proportions of young individuals, who are more active with less pain, are unevenly distributed. Similar examples for confounding could be patient age in fracture treatment, weight/obesity in knee arthroplasty, or workers' compensation in revision shoulder arthroscopy. Unfortunately, there are no rules or test to identify confounders; they must be included based on prior knowledge on the subject or even best guesses. Subsequently, their meaning in the model can be validated by statistical tests comparing models with and without the confounder. Generally, there are two ways to deal with this confounding in a trial: either at the design stage or when analyzing the results. The former tries to neutralize confounding effects by evenly distributing risk factors among groups. This can be accomplished either by random patient allocation or by restricted or matched enrollment. Although we stress the importance of including the possibility of confounding at the design stage of a trial, we also point out these measures are subject to the play of chance and susceptible to inadequate sample size, a very common problem in surgical trials, which may lead to unevenly distributed confounders even in randomized samples. Thus, confounding may still pose a major threat to the validity of findings despite best efforts and should be included in the analysis of results. Given the problematic nature of sample sizes in surgical and orthopaedic research, this post hoc quality control becomes even more important [5, 8, 9]. This can be done either by stratification of the data by suspected variables or by multivariate analysis, usually using regression models. The latter seems favorable because it includes numerous variables in one model rather than establishing numerous submodels for each value of each confounder. Stratification, in turn, can be done only for categorical variables, thus not for weight or age, and will lead to a considerably high number of subgroups even with few confounders, thus posing a problem to adequate sample sizes. Confounding, however, is one problem of many in scientific methodology. Confounding frequently is thought of as a form of bias, which strictly speaking is not true because confounding can be controlled for and bias cannot be controlled for. Bias is a systematic deviation from the truth as a result of flaws in study design or data acquisition. Recruiting subjects for a trial of average survival of patients after THA from an intensive care unit would lead to selection bias resulting from inclusion of subjects that are nonrepresentative to the target population. Inquiry of subjects on events in the past can lead to recall bias, because most people forget, for example, minor trauma, but those with intractable pain will search their personal history relentlessly for reasons. Furthermore, confounding, which is a stable effect across groups, and interaction or effect modification, which is a varying effect across groups, should be separated to calculate valid effect estimates in trials. In light of these issues, we recommend participation of a clinician-scientist with expertise in the field studied and the methods used.

We found $\frac{1}{3}$ of the controlled human trials published in highly ranked orthopaedic journals in 2006 controlled for confounding, meaning the remaining $\frac{2}{3}$ might be distorted by untested variables. Participation of a scientifically trained individual substantially increased the odds of controlling for confounding. The magnitude of the journal's impact factor did not predict the likelihood of controlling for confounding. Although beyond the scope of our study, the question remains to what extent the validity of findings is affected. The effect of confounding, which might range from meaningless to crucial in the individual trials, cannot be estimated for the literature.

References

1. Bryant D, Havey TC, Roberts R, Guyatt G. How many patients? How many limbs? Analysis of patients or limbs in the orthopaedic literature: a systematic review. *J Bone Joint Surg Am.* 2006;88:41–45.
2. Bubbar VK, Kreder HJ. Topics in training: the intention-to-treat principle: a primer for the orthopaedic surgeon. *J Bone Joint Surg Am.* 2006;88:2097–2099.
3. Carr AJ. Evidence-based orthopaedic surgery: what type of research will best improve clinical practice? *J Bone Joint Surg Br.* 2005;87:1593–1594.
4. Emerson JD, Burdick E, Hoaglin DC, Mosteller F, Chalmers TC. An empirical study of the possible relation of treatment differences to quality scores in controlled randomized clinical trials. *Control Clin Trials.* 1990;11:339–352.
5. Freedman KB, Back S, Bernstein J. Sample size and statistical power of randomised, controlled trials in orthopaedics. *J Bone Joint Surg Br.* 2001;83:397–402.
6. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, McQuay HJ. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials.* 1996;17:1–12.
7. Kocher MS, Zurakowski D. Clinical epidemiology and biostatistics: a primer for orthopaedic surgeons. *J Bone Joint Surg Am.* 2004;86:607–620.
8. Lochner HV, Bhandari M, Tornetta P 3rd. Type-II error rates (beta errors) of randomized trials in orthopaedic trauma. *J Bone Joint Surg Am.* 2001;83:1650–1655.
9. McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D. Randomised trials in surgery: problems and possible solutions. *BMJ.* 2002;324:1448–1451.
10. Müllner M, Matthews H, Altman DG. Reporting on statistical methods to adjust for confounding: a cross-sectional survey. *Ann Intern Med.* 2002;136:122–126.
11. Petrie A. Statistics in orthopaedic papers. *J Bone Joint Surg Br.* 2006;88:1121–1136.
12. Poolman RW, Sierevelt IN, Farrokhyar F, Mazel JA, Blankevoort L, Bhandari M. Perceptions and competence in evidence-based medicine: are surgeons getting better? A questionnaire survey of members of the Dutch Orthopaedic Association. *J Bone Joint Surg Am.* 2007;89:206–215.

13. Poolman RW, Struijs PA, Krips R, Sierevelt IN, Lutz KH, Bhandari M. Does a "Level I Evidence" rating imply high quality of reporting in orthopaedic randomised controlled trials? *BMC Med Res Methodol*. 2006;6:44.

14. Weinberg CR. Toward a clearer definition of confounding. *Am J Epidemiol*. 1993;137:1–8.

15. Wright JG, Swiontkowski MF, Heckman JD. Introducing levels of evidence to the journal. *J Bone Joint Surg Am*. 2003;85:1–3.