ELSEVIER

CrossMark

Journal of Clinical Epidemiology 67 (2014) 973–983

Journal of Clinical Epidemiology

Lack of blinding of outcome assessors in animal model experiments implies risk of observer bias

Segun Bello[a,*], Lasse T. Krogsbøll[a], Jan Gruber[b], Zhizhuang J. Zhao[c], Doris Fischer[d], Asbjørn Hróbjartsson[a]

[a]*The Nordic Cochrane Centre, Rigshospitalet, Department 7811, Blegdamsvej 9, 2100 Copenhagen Ø, Denmark*

[b]*Department of Biochemistry, Neurobiology and Ageing Programme, Centre for Life Sciences (CeLS), National University of Singapore, 28 Medical Drive, #04-21 Lab 2, 117456, Singapore*

[c]*Department of Pathology, University of Oklahoma Health Sciences Center 940, Stanton L. Young Blvd., BMSB 451, Oklahoma City, Oklahoma 73104, USA*

[d]*Department of Pediatrics, Division of Neonatology, J.W. Goethe - University Hospital Theodor Stern Kai 7, 60590, Frankfurt, Germany*

Accepted 4 April 2014; Published online 25 June 2014

Abstract

Objectives: To examine the impact of not blinding outcome assessors on estimates of intervention effects in animal experiments modeling human clinical conditions.

Study Design and Setting: We searched PubMed, Biosis, Google Scholar, and HighWire Press and included animal model experiments with both blinded and nonblinded outcome assessors. For each experiment, we calculated the ratio of odds ratios (ROR), that is, the odds ratio (OR) from nonblinded assessments relative to the corresponding OR from blinded assessments. We standardized the ORs according to the experimental hypothesis, such that an ROR <1 indicates that nonblinded assessor exaggerated intervention effect, that is, exaggerated benefit in experiments investigating possible benefit or exaggerated harm in experiments investigating possible harm. We pooled RORs with inverse variance random-effects meta-analysis.

Results: We included 10 (2,450 animals) experiments in the main meta-analysis. Outcomes were subjective in most experiments. The pooled ROR was 0.41 (95% confidence interval [CI], 0.20, 0.82; $I^2 = 75\%$; $P < 0.001$), indicating an average exaggeration of the nonblinded ORs by 59%. The heterogeneity was quantitative and caused by three pesticides experiments with very large observer bias, pooled ROR was 0.20 (95% CI, 0.07, 0.59) in contrast to the pooled ROR in the other seven experiments, 0.82 (95% CI, 0.57, 1.17).

Conclusion: Lack of blinding of outcome assessors in animal model experiments with subjective outcomes implies a considerable risk of observer bias.

Keywords: Blinding; Observer bias; Animal model; Translational medicine; Methods; Meta-analysis

1. Introduction

Translational medicine relies substantially on animal model experiments. Such experiments are used in preliminary investigations of potentially beneficial interventions, some of which are subsequently tested on humans in clinical trials. Other animal model experiments are used to investigate potentially harmful effects of interventions or chemical substances, sometimes subsequently evaluated in observational studies of human exposure, or directly impacting on environmental policy. Animal models are thus essential to translational medicine's ideal of bridging the ravine between the laboratory "bench and the bedside" of the patient [1].

It is not a standard practice in animal model experiments to blind outcome assessors [2–5], despite the suggestion of several guidelines [6–9]. Of 190 animal experiments published in *Journal of Cerebral Blood Flow and Metabolism* in 2008, only 15% reported using blinded outcome assessors [2]. Similarly, Kilkenny et al. reported that only 14% of animal experiments with qualitative scale outcomes reported blinded outcome assessors [4].

There is surprisingly little direct empirical evidence to support or refute the use of nonblinded outcome assessors in animal model experiments. Bebarta et al. found that animal experiments without blinding were three times more likely to report a positive finding as experiments with blinding [10]. A similar between-study analysis in stroke experiments [11] found no important difference between results

Conflict of interest: The authors declare that there is no conflicts of interest and financial disclosure to report.

* Corresponding author. Tel.: +45-3545-7110; fax: +45-3545-7007.

E-mail address: drsegunbello@yahoo.com or sb@cochrane.dk (S. Bello).

0895-4356/$ - see front matter
http://dx.doi.org/10.1016/j.jclinepi.2014.04.008

What is new?

Key findings

- Nonblinded assessors exaggerated odds ratios by approximately 59% in 10 animal model experiments including 2,450 animals.
- The degree of observer bias varied considerably and may be extreme in special circumstances involving strong predisposition. In three pesticide experiments, the ratio of odds ratios (ROR) was 0.20 (95% confidence interval [CI], 0.07, 0.59) in contrast to the pooled ROR in the other seven experiments, 0.82 (95% CI, 0.57, 1.17).

What this adds to what was known?

- Indirect comparisons between experiments with blinded observer and other experiments with nonblinded observers report conflicting results but may be confounded as the experiments could differ for many other reasons that blinding status.
- In this more reliable direct within-study comparison, we found a considerable degree of observer bias when subjective outcomes are assessed by nonblinded observers.

What is the implication and what should change?

- Use of nonblinded outcome assessors of subjective outcomes may cause considerable observer bias in animal model experiments.
- We suggest that outcome assessors in animal model experiments with subjective outcomes are routinely blinded to reduce the risk of observer bias.

in experiments with and without blinded assessors. For both studies, the compared blinded and nonblinded experiments differed in many other ways, implying a risk of confounding by, for example, disease model, animal type, and degree of outcome subjectivity. Other animal model meta-analyses within the fields of bone cancer pain [12], glioma [13], stroke, [14], and Parkinson disease [15] have reported inconsistent results on the impact of blinding outcome assessors on treatment estimates.

The most reliable design for evaluating the impact of not blinding outcome assessors involves a direct comparison of blinded and nonblinded assessments of the same outcome in the same experiment, as this design minimizes confounding considerably [16–18]. Randomized clinical trials with this within-study design have been identified and analyzed [16–18].

The common use of nonblinded outcome assessors in animal model experiments and the lack of empirical foundation for this practice are in sharp contrast to the widespread use of blinded outcome assessors in human experiments and the strong evidence of a considerable observer bias in randomized clinical trials [16–18].

We therefore conducted a systematic review of animal model experiments that used both blinded and nonblinded assessors. We wanted to determine the impact of not blinding outcome assessors on estimates of intervention effects in animal model experiments.

2. Materials and methods

2.1. Eligibility criteria

We included animal model experiments with the same outcome assessed independently by both blinded and nonblinded assessors. We defined an "animal model" as an experimental attempt to reproduce a human clinical condition in laboratory animals. We included experiments that allocated animals into experimental and control groups and with blinded and nonblinded assessments done (1) on the same animals (paired design) or (2) on different animals in two experimental runs (unpaired design). In the paired design, at least one blinded and one nonblinded assessors independently observed the same outcome on the same group of animals allocated to intervention and control groups in the same experiment. In the unpaired design, two or more parallel experimental runs, at least one assessed blinded and one assessed nonblinded, were conducted under the same experimental conditions.

We excluded animal model experiments that used both blinded and nonblinded outcome assessors if assessments were not done for the same outcomes or if either the blinded or nonblinded assessments were done on a subgroup of animals that was not randomly selected.

2.2. Search strategy

We searched PubMed, Biosis, Google Scholar, and HighWire Press from inception onward. Our core search string was "'blinded and unblinded' AND 'animal model'" (Supplementary Material 1/Appendix at www.jclinepi.com). Our last formal search was conducted on September 30, 2011, with supplementary searches in Google Scholar on May 10, 2012.

2.3. Data abstraction

One author (S.B. or A.H.) scanned titles and abstracts from standard databases, text fragments from full-text databases, and retrieved publications of potentially eligible studies. Two authors (S.B., A.H. or L.T.K.) read publications and decided on eligibility. Disagreements were resolved by discussion.

Two authors (S.B. and L.T.K.) extracted characteristics of included experiments and independently selected one outcome assessed blinded and nonblinded. If more than