# Deficiencies of effectiveness of intervention studies in veterinary medicine: a cross-sectional survey of ten leading veterinary and medical journals

Nicola Di Girolamo, Reint Meursinge Reynders

The validity of studies that assess the effectiveness of an intervention (EoI) depends on variables such as the type of study design, the quality of their methodology, and the participants enrolled. Five leading veterinary journals and 5 leading human medical journals were hand-searched for EoI studies for the year 2013. We assessed (1) the prevalence of randomized controlled trials (RCTs) among EoI studies, (2) the type of participants enrolled, and (3) the methodological quality of the selected studies. Of 1707 eligible articles, 590 were EoI articles and 435 RCTs. Random allocation to the intervention was performed in 52% (114/219; 95%CI:45.2-58.8%) of veterinary EoI articles, against 87% (321/371; 82.5-89.7%) of human EoI articles (adjusted OR:9.2; 3.4-24.8). Veterinary RCTs were smaller (median: 26 animals versus 465 humans) and less likely to enroll real patients, compared with human RCTs (OR:331; 45-2441). Only 2% of the veterinary RCTs, versus 77% of the human RCTs, reported power calculations, primary outcomes, random sequence generation, allocation concealment and estimation methods. Currently, internal and external validity of veterinary EoI studies is limited compared to human medical ones. To address these issues, veterinary interventional research needs to improve its methodology, increase the number of published RCTs and enroll real clinical patients.

1  **Deficiencies of effectiveness of intervention studies in**

2  **veterinary medicine: a cross-sectional survey of ten**

3  **leading veterinary and medical journals**

4  **Nicola Di Girolamo, DVM, MSc(Evidence-Based Health Care);[1,2,3*] Reint Meursinge**

5  **Reynders, DDS, MS, MSc(Evidence-Based Health Care)[4,5]**

7  [1] Centro Veterinario Specialistico, Via Sandro Giovannini 53, 00137 Roma, Italy

8  [2] Department of Veterinary Sciences, University of Bologna, Via Tolara di Sopra 50,

9  40064, Ozzano Emilia (BO), Italy

10  [3] EBMVet, Cremona, Italy; nicoladiggi@gmail.com

11  [4] Private practice of orthodontics, Via Matteo Bandello 15, 20123 Milan, Italy

12  [5] Department of Oral and Maxillofacial Surgery, Academic Medical Center, University of

13  Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands;

14  reyndersmail@gmail.com

15  * Corresponding author; nicoladiggi@gmail.com; +393292003570    (ND)

## ABSTRACT

23 The validity of studies that assess the effectiveness of an intervention (EoI) depends on

24 variables such as the type of study design, the quality of their methodology, and the

25 participants enrolled. Five leading veterinary journals and 5 leading human medical

26 journals were hand-searched for EoI studies for the year 2013. We assessed (1) the

27 prevalence of randomized controlled trials (RCTs) among EoI studies, (2) the type of

28 participants enrolled, and (3) the methodological quality of the selected studies. Of 1707

29 eligible articles, 590 were EoI articles and 435 RCTs. Random allocation to the

30 intervention was performed in 52% (114/219; 95%CI:45.2-58.8%) of veterinary EoI

31 articles, against 87% (321/371; 82.5-89.7%) of human EoI articles (adjusted OR:9.2;

32 3.4-24.8). Veterinary RCTs were smaller (median: 26 animals versus 465 humans) and

33 less likely to enroll real patients, compared with human RCTs (OR:331; 45-2441). Only

34 2% of the veterinary RCTs, versus 77% of the human RCTs, reported power

35 calculations, primary outcomes, random sequence generation, allocation concealment

36 and estimation methods. Currently, internal and external validity of veterinary EoI

37 studies is limited compared to human medical ones. To address these issues,

38 veterinary interventional research needs to improve its methodology, increase the

39 number of published RCTs and enroll real clinical patients.

43

## Introduction

44

45    Evaluation of medical interventions may be biased by inferential reasoning. The

46    trustworthiness and applicability of outcomes of studies that assess the effectiveness of an

47    intervention (EoI) strongly depend on the (1) choice of the research design, (2) the

48    methodological quality, and (3) the characteristics of the included population (Schulz et al.,

49    1995; Moinpour et al., 2000; Moher et al., 2010). These issues apply to both veterinary and

50    human medicine. In this study we assessed how these three items differ between these fields.

51    A randomized controlled trial (RCT) is a study in which patients are randomly allocated

52    to either an intervention or a control group (Sackett et al., 1996; Sibbald & Roland, 1998). This

53    design avoids most of the biases that occur in observational studies and has the potential to

54    provide the highest quality of evidence (Moher et al., 2010). Ideally, the effectiveness of every

55    intervention should be tested in a RCT before its implementation in clinical practice (Moher et

56    al., 2010; Armitage, 1982). Although the importance of RCTs is universally acknowledged, the

57    prevalence of RCTs varies between different specialties. This prevalence could be used to assess

58    how robust is research on EoI in a specific field. For example, in periodontal research 10% of the

59    publications between 1980 and 2000 were RCTs (Sjögren & Halling, 2002). In nursing science,

60    3.8% of all articles published between 1986 and 2000 were RCTs (Lingna, 2002). In plastic

61    surgery journals, RCTs comprised only 1.8% of all published articles between 1990 and 2005

62    (Momeni et al., 2008). The prevalence of RCTs was not constant over the last 50 years (Sjögren

63    & Halling, 2002; Becker et al., 2008). In a survey including all specialties of internal medicine, a

64    common ascending trend in the publication of RCTs was found between 1966 and 2001

65    (Strippoli et al., 2004), while between 1998 and 2002 there was no increase. Similarly, in

66    periodontal research there was an increase in the annual number of RCTs published between

67   1980 and 1994, but the number remained approximately unchanged between 1994 and 2000

68   (Sjögren & Halling, 2002). Studies that assessed the prevalence of RCTs among EoI studies were

69   not identified in the published veterinary literature.

70       The label "randomized" is not sufficient to guarantee the methodological soundness of an

71   EoI study. In fact, RCTs should also adhere to a wide variety of other quality parameters (Schulz

72   et al., 1995; Moher et al., 2010). When the RCTs published in the veterinary literature were

73   evaluated, several methodological issues were identified (Elbers & Schukken, 1995; Lund et al.,

74   1998; Brown, 2006; Brown, 2007; Sargeant et al., 2009; Sargeant et al., 2010; Giuffrida et al.,

75   2012; Giuffrida, 2014). Lund et al. (1998) identified a lack of reporting of at least 2 of 6

76   evaluated domains (random sequence generation, informed consent, eligibility criteria, blinding,

77   power calculation and statistical analyses) in all RCTs on dogs and cats published between 1986

78   and 1990. Brown et al. (1998) found that only 11% of 97 RCTs published between 2000 and

79   2005 on dogs and cats reported both random sequence generation and allocation concealment. In

80   a subset of 63 RCTs on dogs and cats, Brown et al. (2007) reported that most RCTs with losses

81   to follow-up did not account for these losses in the data analysis and did not acknowledge the

82   potential impact on the outcomes of these RCTs. Another study found substantive deficiencies in

83   the reporting of key methodological domains in RCTs with dogs and cats published between

84   2006 and 2008 (Sargeant et al., 2010). RCTs performed in laboratory animal research showed

85   similar methodological problems: an overview of 31 systematic reviews found that only 29% of

86   studies reported randomization, 15% of studies reported allocation concealment, and 35% of

87   studies reported blinded outcome assessment (Hirst et al., 2014).

88       Even when RCTs are conducted according to high methodological standards (i.e., strong

89   internal validity), their applicability can be challenged by the type of patients enrolled. Outcomes

90  of RCTs have limited external validity when the participants of the study represent only a small

91  part of the population of interest (Pressler et al., 2013), e.g., if participants are healthier than

92  patients that are not recruited (Halbert et al., 1999; Moinpour et al., 2000). This issue can affect

93  the external validity of research outcomes and limits the translation of results from RCTs to

94  "true" clinical patients.

95      The veterinary literature shares several common factors with the medical literature (e.g.,

96  presence of generalist and specialist journals). Furthermore, opinion leaders have recently

97  underscored that animal trials should be more similar to human RCTs (Muhlhausler et al., 2013).

98  We therefore decided to exploit the similarities between these two disciplines, and to use human

99  medical literature as the comparator for the veterinary medical literature.

100     The purpose of this study is to compare the veterinary literature and human medical

101  literature regarding (1) the prevalence of RCTs among EoI studies; (2) the prevalence of RCTs

102  that did not enroll clinical patients; (3) the reporting of key methodological domains in RCTs.

103  Prior to undertaking this research study we conducted scoping searches of the literature to avoid

104  research replication (Chalmers et al., 2014). These searches showed that no previous cross-

105  sectional literature reviews compared the quality of EoI articles in veterinary medicine with

106  those in human medicine.

107

# Materials and methods

## Design and study outcomes

110     We performed a cross-sectional analysis of the literature comparing leading veterinary

111  medical and human medical journals in the year 2013.

112     Primary outcomes for this study were (1) the difference in the prevalence of RCTs

113     between veterinary and human medical literature; (2) the difference in the prevalence of

114     enrollment of participants that are not real clinical patients between veterinary and human

115     medical literature; (3) the difference in reporting of key methodological domains in RCTs

116     between veterinary and human medical literature. All other outcomes are considered secondary

117     outcomes.

118

## Sample size and included journals

119

120     We performed a pilot study to assess the prevalence of RCTs over the total number of

121     articles that described the effectiveness of interventions in veterinary medicine and human

122     medicine. All articles that were published in the first six months of 2006 in one veterinary

123     journal (JAVMA) and one human medical journal (JAMA) were assessed for the prevalence of

124     RCTs. A prevalence of 69,1% (29 RCTs/42 intervention articles) and 29,7% (8 RCTs/27

125     intervention articles) was identified for JAMA and JAVMA respectively. Using a formula for

126     two proportions and equal group size (Kirkewood & Sterne, 2003) we calculated that a minimal

127     sample of 45 articles per group was required to have 90% power to detect a difference at a level

128     of statistical significance of 1%. We estimated how many journals we had to search and for

129     which time span based on pertinent research data from a study by Giuffrida et al. (2012). Based

130     on the pilot study we expected to find roughly one RCT per 3 EoI articles and that each journal

131     should have published approximately 15 EoI articles per-year. To be conservative, we included 5

132     leading journals for each discipline. The "VETERINARY SCIENCES" and "MEDICINE,

133     GENERAL AND INTERNAL" categories of the 2013 ISI Journal Citation Report were sorted

134     by impact factor. All the journals that focused on sub-specialties or in non-English language

135   were excluded. All the journals that were not published before the year 2000 or had an impact

136   factor lower than 1.0 were excluded. Aims and scopes of the remaining journals were evaluated

137   on their websites until the first 5 journals presenting broad scope were identified. Veterinary

138   journals included were: '*Veterinary Journal*' (Vet J); '*Veterinary Record*' (Vet Rec); '*Journal of*

139   *Veterinary Internal Medicine*' (JVIM); '*Journal of the American Veterinary Medical*

140   *Association*' (JAVMA); '*American Journal of Veterinary Research*' (AJVR); Medical journals:

141   '*New England Journal of Medicine*' (NEJM); '*the Lancet*'; '*Journal of the American Medical*

142   *Association*' (JAMA); '*British Medical Journal*' (BMJ); '*Annals of Internal Medicine*'. Impact

143   factors ranged from 1.2 to 2.2 for veterinary medical journals and from 16.1 to 54.4 for human

144   medical journals. Details of the pilot study, sample size calculation and eligibility criteria for

145   journal inclusion are reported in Supplementary Note 1.

146

## Data extraction

148        We hand-searched all full-text articles of all the issues of the 10 selected journals

149   published between the 1st of January 2013 up to the 31st of December of 2013, including

150   supplements, through their online archives. The total number of full original articles, EoI articles,

151   and RCTs were recorded. EoI articles were subsequently classified based on their inclusion of

152   real patients and type of interventions (surgical/non-surgical). RCTs were further classified based

153   on their methodological characteristics.

154

### *Classification of the EoI articles*

156        We extracted the following data items:

157  *- Number of full original articles:* primary research, including subgroup analysis or follow-ups of

158  previous articles; case series, defined as original reports including more than one patient.

159

160  *- Number of articles evaluating effectiveness of interventions (EoI):* "*Effectiveness*" was defined

161  as "*evaluation of benefits*" of an intervention. "*Interventions*" were defined as "*acts used to*

162  *improve health, to treat a particular condition or disease in process or to prevent development of*

163  *a particular condition or disease*" (Farlex Medical Dictionary, 2014; Merriam-Webster

164  Encyclopaedia Britannica, 2014). Studies eligible as "EoI articles" included: Case series, case-

165  control studies, cohort studies, analytical cross-sectional studies, non-randomized controlled

166  trials, and RCTs.

167

168  *- Number of EoI articles that described surgical interventions*: Studies of surgical interventions

169  face different challenges regarding several aspects, including study design (McCulloch, 2009).

170  To account for this factor, the type of intervention was categorized as surgical/non-surgical. EoI

171  articles were considered "surgical" when (1) the intervention required cutting of the skin.

172  Needle-related procedures (e.g., amniocentesis, etc.) were not considered surgical procedures;

173  (2) The difference between the control and the experimental group was based on the type or

174  technique of the surgical procedure. A trial was not considered "surgical" when the difference

175  between the control and the experimental group was a medication given either before or after a

176  surgical procedure.

177

178  *- Number of RCTs:* Studies were defined RCTs based on the US National Library of Medicine

179  2008 definitions for the Publication Type terms 'Randomized Controlled Trial' and based on the

180    definition of the Cochrane glossary (Glossary of Cochrane terms, 2014). All the reports with

181    allocation to interventions described as "randomized" were included (Schulz et al., 1994). A

182    study was classified as "a RCT" when (1) at least two interventions were compared; (2) and

183    randomization was mentioned.

184

185    - *Number of RCTs that included real patients:* We evaluated if RCTs involved real clinical

186    patients or non-patients. Real clinical patients were defined as "*the population that presents the*

187    *condition that needs to be treated or prevented and that will benefit of the intervention once*

188    *established*". Non-patients refer to voluntary healthy participants or laboratory animals. Articles

189    were considered to include real clinical patients when these individuals or animals: (1) suffered

190    from a spontaneous disease; and (2) were exposed to real-life conditions.

191

192    The full definitions of each extracted data item are reported in Supplementary Note 2.

193    Categorization of the EoI articles in randomized and non-randomized trials was initially

194    performed through scrutinizing the title and the abstract of the article. If random allocation was

195    not mentioned in the title and the abstract, the full text was searched for the term "random". Full-

196    texts of all the articles classified as "RCTs" were retrieved for the second phase of the study.

197

198    ### *Evaluation of the RCTs*

199        We assessed all the RCTs retrieved from the 5 selected veterinary journals in 2013 for the

200    reporting of key methodological items (Chan & Altman, 2005). Random sampling stratified by

201    journal was performed on the medical RCTs in order to assess a number of medical RCTs in a

202    ratio of 1:2 compared with veterinary RCTs. Since 114 veterinary RCTs were retrieved, a total of

203   60 medical RCTs were sampled. Each medical RCT was sequentially numbered and 5 series of

204   12 random numbers (one for each journal) were generated with a random number generator.

205          Two operators independently assessed the electronic full-text of each RCT and eventual

206   supporting information. In case of disagreement, an arbiter was consulted. The following key

207   methodological domains were assessed: primary outcome, power calculation, random sequence

208   generation, allocation concealment, blinding of participants, blinding of personnel, blinding of

209   outcome assessors, intention-to-treat, effect size estimation methods (Table 1). Details regarding

210   research procedures and definitions are reported in the Supplementary Note 2.

211

## Statistical analysis

213   SPSS statistics (v22.0, IBM, Chicago, IL) was used for the statistical analysis. Statistical

214   significance was set at $P<0.05$, unless otherwise specified. Prevalence of RCTs in veterinary and

215   human medical journals was calculated as: number of RCTs/total number of EoI articles.

216   Confidence intervals for proportions were estimated with the Wilson procedure with continuity

217   correction (Newcombe, 1998). Articles were the unit of the primary analysis. In the unadjusted

218   analysis the strength of the association was described as odds ratios (ORs) with their 95%

219   confidence interval and was calculated in two-by-two tables. Chi-squared tests or Fisher tests

220   (depending on the number of expected values in each cell) were used to evaluate whether there

221   were significant differences in proportions. If one or more of the cells in the contingency table

222   were zero, a non-constant continuity correction was employed to account for the imbalance of

223   the group sizes. A factor of the reciprocal of the size of the opposite treatment group was added

224   to the cells (Sweeting et al., 2004).

225     Logistic regression and multilevel logistic regression models were developed to provide

226   odds ratio adjusted for confounders (Peng & So, 2002; Heck et al., 2013). In the multilevel

227   logistic regression with prevalence of RCT as the outcome, discipline (veterinary/general

228   medicine) and type of intervention (i.e., surgical/non-surgical) were included as fixed effects and

229   journal as a random effect (Heck et al., 2013). In the logistic regression models, confounders

230   included: type of intervention (i.e., surgical/non-surgical), type of trial (randomized/non-

231   randomized), type of patients enrolled (clinical patients/non-patients), and discipline

232   (veterinary/medicine). Variables were retained in the model based on statistical significance

233   (P<0.1) and based on the effect they had on the final model. To avoid overfitting of the model, a

234   minimum of 10 events per predictor variable were required (Peduzzi et al., 1996). Goodness of

235   fit was assessed with Hosmer-Lemshow test. Nagelkerke R squared was reported (Nagelkerke

236   1991). Multicollinearity was suspected with variance inflation factor > 3 and condition index

237   >30 (Midi et al., 2010).

238     To identify differences in the number of patients enrolled, the non-parametric Mann-

239   Whitney U test was used as indicated by the distribution (Lehmann, 1951).

240     Sensitivity and subgroup analyses planned a priori were conducted to support the

241   robustness and the generalizability of the association between a specific discipline and the

242   quality of reporting of key methodological domains. Sensitivity analyses were conducted by

243   excluding from the analysis certain types of randomized trials, in order to determine their effect

244   on the final results. In the first sensitivity analysis, we excluded all the RCTs that did not enroll

245   real clinical patients because an earlier study identified a lack of reporting of key methodological

246   items in RCTS on laboratory animals (Hirst et al., 2014). In an additional sensitivity analysis we

247   excluded all the surgical RCTs because previous research on such RCTs has found a lack of

248  adequate reporting of key methodological items (Sinha et al., 2009). In a further analysis, we

249  included only "explicit RCTs" (i.e., randomized trials that were explicitly defined in the text as

250  "randomized controlled trial" or trials registered in a repository) because we hypothesized that

251  such RCTs would have been conducted with a special focus on high methodological standards.

252  In the final analyses, we excluded all the cross-over RCTs and cluster RCTs because these trial

253  designs have also been associated with poor reporting (Walleser et al., 2011; Straube et al.,

254  2015). Results of sensitivity analyses were reported with forest plots generated with RevMan

255  (5.3; Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014).

256

## Results

## Effectiveness of intervention articles

259       A total of 1707 eligible articles were identified through hand-searching procedures in 10

260  selected journals for the year 2013 (Fig. 1). Of these articles, 990 were published in veterinary

261  medical journals, and 717 in human medical journals. The distribution of eligible articles among

262  the specific journals is depicted in Table 2. Overall, 35% of the eligible articles (590/1707) were

263  EoI studies. Effectiveness of an intervention was evaluated in 22% (219/990) of the veterinary

264  medical articles and in 52% (371/717) of the human medical articles.

265

## Randomized Controlled Trials

### *Prevalence of RCTs*

268    The prevalence of RCTs among the EoI articles was 52% (114/219; 95% CI: 45.2% to

269    58.8%) in veterinary medical journals, versus 87% (321/371; 95% CI: 82.5% to 89.7%) in

270    human medical journals. The prevalence of RCTs (number of RCTs/number of EoI articles) in

271    the various journals is reported in Table 2. EoI articles published in veterinary medical journals

272    had 5.9 times the odds of being non-randomized, compared with articles in human medical

273    journals (OR: 5.9; 95% CI: 4 to 8.8; P<0.001). In the adjusted analysis, including "type of

274    intervention" (surgical/non-surgical) as fixed effect and "journal" as random effect, the results

275    were substantially unchanged (OR: 5.6; 95% CI: 3.1 to 9.8). Human medical journals had 6 to 15

276    times the odds of publishing a RCT compared with one of the veterinary medical journals

277    (Supplementary Table S1). When only "explicit RCTs" were considered, the prevalence of RCTs

278    declined to 21% in veterinary medicine (47/219; 95% CI: 16.3% to 27.6%), while the prevalence

279    remained unchanged (87%) in human medicine.

280

281    **_Characteristics of RCTs_**

282    The number of subjects randomized in the articles ranged from 5 to 28244, with a median

283    of 59 subjects (IQR: 307 subjects), a mean of 857 subjects, and a SD of 3278 subjects. Sample

284    size was not normally distributed in veterinary and in human medical trials. Trials were

285    significantly smaller in veterinary medical journals having a median sample size of 26 subjects

286    (range ± IQR: 5-28244 ± 47) compared to 465.5 subjects in human medical journals (range ±

287    IQR: 32-27347 ± 1267)(Mann-Whitney U: 607.0; P<0.001). Furthermore, veterinary crossover

288    trials were significantly smaller, i.e., with a median sample size of 8 subjects (range ± IQR: 5-20

289    ± 5), compared with 36 subjects in other RCT designs (range ± IQR: 6-28244 ± 63)(Mann-

290    Whitney U: 131.5; P<0.001).

291        Only 1 out of 114 RCTs (0.9%) in the veterinary journals had a cluster design compared

292    with a 13.3% in the human medical journals (8/60). More than one fifth (21.9%; 25/114) of the

293    veterinary RCTs had a crossover design. Of the 60 randomly sampled human medical RCTs,

294    only one (1.7%; 1/60) had a crossover design, specifically a stepped-wedge design, i.e., cluster

295    and crossover design. Further characteristics of RCTs are reported in the Supplementary Data.

296

297    *Prevalence of surgical RCTs*

298        Surgical interventions accounted for 5% of the eligible articles (90/1707), for 15% of the

299    EoI articles (90/590) and for 9.6% of the RCTs (42/435). Surgical interventions had 4.2 times the

300    odds of being published as non-randomized articles than as randomized articles (OR: 4.2; 95%

301    CI: 2.6 to 6.7; P<0.001)(Fig. 2).

302        No statistically significant differences in the prevalence of surgical EoI articles were

303    found between veterinary and human medicine (18%; [40/219] vs 13%; [50/371]; OR: 0.69; 95%

304    CI: 0.44 to 1.09; P=0.11). However, veterinary RCTs had a notably lower prevalence of surgical

305    interventions compared with veterinary EoI studies (3.5%; 4/114), while the prevalence of

306    surgical interventions remained roughly constant among human medical EoI studies and RCTs

307    (11.8%; 38/321). The strength of the association between surgical intervention and lack of

308    randomization was stronger in veterinary medicine. Human medical journals had 3.7 times the

309    odds of publishing RCTs on surgical interventions than veterinary journals (OR: 3.7; 95% CI:

310    1.3 to 10.6; P=0.01).

311

312    *Enrollment of real clinical patients*

313     All the human medical RCTs focused on real patients, except one (99.7%; 320/321). Less

314     than a half of the veterinary RCTs (49.1%; 56/114) focused on real patients (OR: 331.4; 95% CI:

315     45 to 2441.9; P<0.001).

316     The vast majority of crossover trials enrolled non-clinical patients (84.0%; 21/25), while

317     41.6% of the remaining RCTs enrolled non-clinical patients (37/89). Crossover trials had 7.4

318     times the odds of enrolling non-clinical patients compared with other RCT designs (OR: 7.4;

319     95% CI: 2.3 to 23.3; P<0.001).

320

## Reporting of key methodological domains

322     The key methodological domains were consistently reported by most human medical

323     RCTs, but irregularly reported by veterinary medical RCTs (Table 3). Blinding procedures of

324     owners, personnel and outcome assessors were the most consistently reported domains in

325     veterinary RCTs. None (0/114) of the veterinary RCTs adequately reported all key

326     methodological domains, while 23% (14/60) of the human medical RCTs reported all those

327     methodological domains (OR: 70.5; 95% CI: 4.1 to 1208.4; P<0.001). Only 2% (2/114) of the

328     veterinary RCTs, versus 77% (46/60) of the human medical RCTs, adequately reported key

329     methodological domains that are always feasible (OR: 184; 95% CI: 40.2 to 842; P<0.001). The

330     number of key methodological domains adequately reported was positively correlated with the

331     number of patients enrolled (Spearman r: 0.73; P<0.001; Fig. 3).

332     A binary logistic regression model including "reporting of primary outcome", "reporting

333     of allocation concealment", "reporting of random sequence generation", "mentioning ITT" and

334     "use of estimation methods" as covariates, was well-fitted (Hosmer and Lemeshow test: P=0.87)

335     and useful to predict (Nagelkerke R-squared: 0.95) and discriminate (area under the curve: 0.99)

336   the source discipline (i.e., veterinary/medicine) of a RCT. "Reporting of power calculation" was

337   not included in the model because of collinearity between "reporting of the primary outcome"

338   and "reporting of the power calculation". Dependency between the variables after the removal of

339   "reporting of the power calculation" was acceptable for running the logistic regression model

340   (i.e., all condition index lower than 30.0).

341

342   ***Association between clinical patients and key methodological***

343   ***domains***

344         Veterinary RCTs enrolling non-clinical patients were more likely to lack adequate

345   reporting of each of the methodological domains evaluated in this study (Fig. 4) with the

346   exception of blinding of personnel and intention-to-treat. The lack of an association between

347   reporting and "intention-to-treat" should be interpreted with caution, because only 3 veterinary

348   RCTs reported on this domain.

349

350   ***Sensitivity and subgroup analyses***

351         Sensitivity analyses supported the robustness of the association between discipline and

352   lack of reporting of key methodological domains (Fig. 5). The amount of surgical RCTs and non-

353   real patient RCTs was not sufficient to perform sensitivity analyses that included these RCTs

354   only.

355

356   # Discussion

357    RCTs, introduced in the early 1940s (Stuart-Harris et al., 1943; Medical Research

358    Council, 1948), have been widely adopted in clinical research to reduce the risk of subjective

359    evaluation of effectiveness of interventions. Our study showed that the scientific rigor used to

360    evaluate effectiveness of interventions was suboptimal in leading veterinary journals compared

361    with general medical journals. Firstly, we observed a lower prevalence of RCTs in veterinary

362    journals compared with medical journals (research question 1). Secondly, more than half of the

363    RCTs in veterinary medicine did not enroll clinical patients (research question 2). Finally, key

364    methodological domains were underreported in veterinary RCTs compared with those in medical

365    RCTs (research question 3). This latter finding is critical as under-reporting of methodological

366    domains in veterinary RCTs has been associated with increased treatment effects (Sargeant et al.,

367    2010).

368    Observational non-randomized articles are fundamental in certain stages of the

369    development of interventional procedures (Schünemann et al., 2013). In veterinary medicine

370    these articles were overabundant and only about half of the articles that assessed the

371    effectiveness of interventions was randomized. A survey in 2006 by Kuroki et al. (2009)

372    calculated a prevalence of 34.7% (66/190) of RCTs on the total number of articles published in

373    JAMA, Lancet and NEJM. In the present study we observed a prevalence of 50.6% (260/514) of

374    RCTs on the total number of original research articles for the same 3 journals. These data should

375    be considered with caution, because they originated from 2 single articles that applied different

376    sampling strategies. However, there was a 16% increase in the prevalence of RCTs in these

377    journals between 2006 and 2013 (34.7% *vs* 50.6%; Risk difference: 0.16; 95% CI: 0.08 to 0.24).

378    In the present study all journals were hand-searched and all original research articles were

379    scrutinized. Instead, Kuroki et al. (2009) obtained a random sample of the original research

380  articles. Notwithstanding these differences in selection procedures, the definitions for "journal

381  article" and "RCT" were the same in both studies. If these research studies are considered

382  representative for their specific year of publication, the increase of the prevalence of RCTs in

383  these medical journals in 7 years was impressive. Veterinary journals should also aim to increase

384  the rate of published RCTs.

385      Experimental animal studies have been recently criticized for being inadequately

386  conducted and reported (Hirst et al., 2014; Sena et al., 2007; Landis et al., 2012). In 2010, the

387  REFLECT statement (Reporting guidelines for randomized controlled trials for livestock and for

388  food safety) and the ARRIVE guidelines (Animal research: reporting of in vivo experiments)

389  were developed for reporting RCTs of respectively livestock and laboratory animal research

390  (Kilkenny et al., 2010; O'Connor et al., 2010). However, these guidelines apply only partly to

391  veterinary medicine and often do not cover the needs of veterinary clinicians, i.e., they do not

392  provide guidelines for patient-specific interventions. The finding that just one quarter of the EoI

393  articles in the 5 veterinary journals included is a RCT on clinical patients should be a particular

394  warning sign for veterinary clinicians. Most of the veterinary RCTs are conducted in a pre-

395  clinical setting or in laboratory-controlled conditions and may not apply to the population of

396  interest.

397      In the present study we observed that crossover trials were rare in leading human medical

398  journals (1 out of 321 RCTs) but frequent in veterinary journals (more than one fifth of the

399  RCTs). Crossover trials in veterinary medicine had a smaller sample size (median sample size of

400  8 patients) and were more likely to enroll non-clinical patients. The association between non-

401  clinical patients, sample size and crossover design is straightforward in veterinary medicine: the

402  additional expenses related to the use of experimental subjects compared with the use of clinical

403    patients are mainly secondary to purchasing and maintaining them. Therefore, there is a tendency

404    to use the smallest "credible" sample sizes, and to apply crossover designs to provide additional

405    statistical power. The fact that only 1 out of 25 crossover trials presented a power calculation

406    supports this hypothesis. This information is not novel, as Macleod (2010) suggested *"In the face*

407    *of pressures to reduce the number of animals used, investigators often do studies that are too*

408    *small to detect a significant effect."*

409    In veterinary journals, parallel explicit RCTs were more likely to report power

410    calculation than crossover trials. However, only 25% of the parallel explicit RCTs reported a

411    power calculation. This is consistent with the results of a study by Giuffrida (2014), who

412    assessed trials of client-owned dogs and cats. The vast majority of medical RCTs reported power

413    calculations, which is consistent with the literature (Charles et al., 2009). Veterinary researchers

414    should apply stringent standards of statistical power when planning empirical research.

415    A statistically significant result does not provide information on the magnitude of the

416    effect and thus does not necessarily mean that the effect is robust (Landis et al., 2012). While

417    almost every RCT published in the selected human medical journals provided point estimates

418    with measures of uncertainty (i.e., 95% CI), most articles in veterinary medicine provided just

419    these measures of statistical significance, i.e., 'P values'. Over-reliance on 'P values' when

420    reporting and interpreting results of a RCT may be inappropriate and misleading (Rothman,

421    1978; Sterne & Davey Smith, 2001). Veterinary investigators should become more confident

422    with statistical procedures to generate point estimates and measures of uncertainty. Journal

423    editors should encourage authors to employ statistical techniques that maximize the clinical

424    interpretation of RCTs.

425    Reporting of allocation concealment in RCTs published in human medical journals

426    improved in the last 15 years from 34.4% to 64.7% (To et al., 2013). In our study, only one tenth

427    of the RCTs published in veterinary journals mentioned allocation concealment. This is worrying

428    as there is evidence that the lack of allocation concealment is empirically associated with bias

429    (Schulz et al., 1995). Researchers should be aware of these consequences and should understand

430    that concealing treatment assignments up to the point of allocation is always possible regardless

431    of the study topic (Schulz et al., 1994). Investigators should therefore implement this issue in

432    veterinary RCTs.

433        Previous studies evaluating veterinary RCTs focused only on one journal (Elbers &

434    Schukken, 1995), others only on some species (Lund et al., 1998; Sargeant et al., 2010) or only

435    on RCTs including client-owned animals (Brown, 2006; Brown, 2007; Giuffrida et al., 2012). To

436    maximize the external validity of this study, we included all the EoI studies. Nonetheless, the

437    present research study focused on a selected group of leading general journals. In 2006 the

438    proportion of RCTs published among JAMA, Lancet and NEJM was almost 3 times greater than

439    obstetric and gynecology journals (Kuroki et al., 2009). In addition, trials published in general

440    medical journals had higher quality scores than those in specialist journals (To et al., 2013).

441    Therefore, in this study biases are likely to overestimate the prevalence of RCTs and their

442    methodological quality compared with the remaining journals in human medicine, even if, it

443    cannot be excluded that some specialist journals have a higher prevalence of RCTs due to

444    editorial policies or other factors.

445        We found that surgical interventions were more likely investigated with non-randomized

446    trials than randomized trials in both disciplines. This is not unexpected, as surgical RCTs pose

447    more challenges than pharmaceutical RCTs (McCulloch 2009). However, the magnitude of this

448    effect may have been amplified by a selection bias. Journals included in this survey were

449    selected for their broad scope. However, it is possible that surgical teams prefer to submit the

450    results of surgical RCTs to specialist surgical journals instead of to broad scope journals.

451    Therefore, future research should include also specialist surgical journals to provide a fairer

452    estimate of the prevalence of these randomized trials.

453        A possible weakness of this study was the non-blinded assessment of studies, i.e.,

454    reviewers knew whether they assessed a veterinary or human medical article. Even though there

455    is some evidence that blind assessments of studies might result in more consistent ratings

456    compared with open assessments (Jadad et al., 1996), some other studies suggest that blind

457    assessments provide little benefit (West, 1997; Kjaergard et al., 2001). A further weakness is the

458    limited sample of included journals. Our selection criteria led to the inclusion of the American

459    Journal of Veterinary Research (AJVR), a journal that tends to focus more on experimental than

460    clinical research. This characteristic could have led to bias toward a lower number of "real

461    patient" RCTs. However, we have performed a sensitivity analysis removing AJVR to evaluate

462    the impact of this single journal on our primary outcomes (prevalence of RCTs, prevalence of

463    real patient RCTs). We observed only a modest change in real patient RCTs (i.e., from 26% to

464    28%), while there was a further drop in RCT prevalence in veterinary journals (i.e., from 52% to

465    45%). Therefore, we retain that the impact of this journal on the overall results is limited and it is

466    unlikely that it has significantly biased our conclusions. Another weakness of this study is that

467    we assessed how attrition was reported and dealt with only by means of a proxy (Chan &

468    Altman, 2005). This issue is difficult to address, because handling of attrition may be performed

469    in several different ways, and may be reported in different sections of the article (i.e., under

470    statistical analysis but also in the patient flow). As a proxy for this item, we searched for the

471    wording "intention-to-treat", based on the Cochrane glossary definition (Glossary of Cochrane

472    terms, 2014). Obviously mentioning "intention-to-treat analysis" does not ensure that attrition

473    was properly dealt with, because the term 'intention-to-treat' could have been misused.

474        A limitation in the generalizability of the study is that we had to obtain a sample of

475    journals due to the number of published journals in the two disciplines (n=133 in the 'veterinary

476    science' category; n=154 in the 'medicine, general & internal' category). This sample may not be

477    representative for the entire spectrum of journals in these fields, especially when considering

478    peculiar specialties (e.g., surgery), with exclusive issues in conduct and reporting of RCTs

479    (McCulloch, 2009). A larger sample of journals (e.g., inclusion of 10-20 journals per discipline)

480    would have been desirable to obtain results with greater generalizability.

481        The several limitations we have uncovered in the veterinary literature are probably

482    secondary to the intrinsic difficulties in this field. For example, RCTs are more expensive than

483    observational reports, and veterinary medicine has been historically allotted less funding

484    compared with the human medical field. Furthermore, in veterinary medicine the variety of

485    species makes performing RCTs on every species difficult. The goals in veterinary medicine are

486    also substantially different than in human medicine, e.g. in production medicine, productivity is

487    the primary metric of treatment success.

488    However, the aforementioned limitations of this study should be considered in the context of the

489    overwhelming magnitude of the identified differences between veterinary and human medical

490    EoI studies. Veterinary researchers should adopt medical research standards by introducing the

491    following pathways for improvement:

492    (1) *Education in basic evidence-based practices at all veterinary levels (i.e., undergraduates to*

493        *policy-makers).* Education on evidence-based practices effectively engages students

494        (Aronoff et al., 2013). Implementing tailored courses for veterinary professionals on

495     conducting, analyzing, and reporting of clinical research is an important starting point,.

496   (2) *Implementation of reporting guidelines at all research levels (i.e., researchers, peer-*

497       *reviewers, journal editors).* Although most veterinary journals adhere to reporting

498       guidelines, such as the CONSORT statement (Moher et al., 2010), journals do not require

499       submission of the pertinent checklist together with the manuscript. Implementing this

500       practice would be easy and costless.

501   (3) *Implementation of methodology reviewers in the peer-review process.* Methodology

502       reviewers can provide unique comments during peer-review, which are distinct from the

503       issues raised by regular reviewers (Day et al., 2002). As in several human medical

504       journals, veterinary journals should invite methodology reviewers to participate in the

505       peer-reviewing process.

506   (4) *Creation of specific grants for pragmatic randomized controlled trials and predilection of*

507       *pragmatic randomized controlled trials in grants*. As pragmatic randomized controlled

508       trials provide higher quality evidence than other study designs (Sibbald & Roland, 1998),

509       it would be wise to establish purposely-designed grants for pragmatic RCTs, or to favor

510       publication of pragmatic RCTs.

511   (5) *Establishment of a register for veterinary randomized controlled trials.* Trial registration in

512       human medicine was promoted to enhance transparency and accountability in the

513       planning, execution, and reporting of RCTs (Dickersin & Rennie, 2003; DeAngelis,

514       2004). This example should also be adopted in veterinary medicine.

515   ### Conclusions

516   The objectives of this study were met, and most of the findings were alarming with regard to

517   both their number and magnitude. We observed that the veterinary literature was characterized

518   by a lower prevalence of randomized articles compared with the human medical literature. More

519   worrying, the vast majority of the RCTs in veterinary medicine lacked adequate reporting of key

520   methodological domains (i.e., primary outcome, power calculation, random sequence generation,

521   allocation concealment, intention-to-treat) and of currently recommended statistical reporting

522   (i.e., effect measures with confidence intervals). Furthermore, RCTs published in veterinary

523   journals were less likely to enroll clinical patients than RCTs published in human medical

524   journals. These outcomes are important, because they could slow down knowledge creation and

525   could result in inappropriate claims of effectiveness of interventions. These findings should be a

526   stimulus for future researchers to improve upon the soundness of their studies. Awareness of the

527   paucity of evidence behind interventions in veterinary medicine is the first step for improvement.

528

## 529   **References**

530   Armitage, P. 1982. The role of randomization in clinical trials. *Stat. Med.* **1**, 345-52.

531   Aronoff, S. C., Evans, B., Fleece, D., Lyons, P., Kaplan, L., & Rojas, R. 2010. Integrating

532   evidence based medicine into undergraduate medical education: combining online instruction

533   with clinical clerkships. *Teach Learn. Med.* **22,** 219-223.

534   Becker, A., Blümle, A., Antes, G., Bannasch, H., Torio-Padron, N., Stark, G. B., & Momeni, A.

535   2008. Controlled trials in aesthetic plastic surgery: a 16-year analysis. Aesthetic plastic surgery,

536   32(2), 359-362.

537   Brown, D. C. 2006. Control of selection bias in parallel-group controlled clinical trials in dogs

538   and cats: 97 trials (2000-2005). *J. Am. Vet. Med. Assoc.* **229**, 990-993.

539   Brown, D. C. 2007. Sources and handling of losses to follow-up in parallel-group randomized

540   clinical trials in dogs and cats: 63 trials (2000-2005). *Am. J. Vet. Res.* **68**, 694-698.

541   Chalmers, I., Bracken, M.B., Djulbegovic, B., Garattini, S., Grant, J., Gülmezoglu, A.M.,

542   Howells, D.W., Ioannidis, J.P. and Oliver, S., 2014. How to increase value and reduce waste

543   when research priorities are set. The Lancet, 383(9912), pp.156-165.

544   Chan, A. W., & Altman, D. G. 2005. Epidemiology and reporting of randomised trials published

545   in PubMed journals. *Lancet*. **365,** 1159-1162.

546   Charles, P., Giraudeau, B., Dechartres, A., Baron, G., & Ravaud, P. 2009. Reporting of sample

547   size calculation in randomised controlled trials: review. *BMJ*. **338,** b1732.

548   Day, F. C., Schriger, D. L., Todd, C., & Wears, R. L. 2002. The use of dedicated methodology

549   reviewers for peer review: a content analysis of comments to authors made by methodology and

550   regular reviewers. *Ann. Emerg. Med*. **40,** 329-333.

551   De Angelis, C., Drazen, J.M., Frizelle, F.A., Haug, C., Hoey, J., Horton, R., Kotzin, S., Laine,

552   C., Marusic, A., Overbeke, A.J.P. and Schroeder, T.V., 2004. Clinical trial registration: a

553   statement from the International Committee of Medical Journal Editors. New England Journal of

554   Medicine, 351(12), pp.1250-1251.

555   Dickersin, K., & Rennie, D. 2003. Registering clinical trials. *JAMA*. **290,** 516–523.

556   Elbers, A. R., & Schukken, Y. H. 1995. Critical features of veterinary field trials. *Vet. Rec*. **136**,

557   187-92.

558   Farlex Medical Dictionary. Definition: Intervention. Available: http://medical-

559   dictionary.thefreedictionary.com/intervention. (Accessed: June 5 2014)

560    Giuffrida, M. A. Type II error and statistical power in reports of small animal clinical trials. *J.*

561    *Am. Vet. Med. Assoc*. **244**, 1075-1080 (2014).

562    Giuffrida, M. A., Agnello, K. A., & Brown, D. C. Blinding terminology used in reports of

563    randomized controlled trials involving dogs and cats. *J. Am. Vet. Med. Assoc*. **241**, 1221-1226

564    (2012).

565    Glossary of Cochrane terms. Available: http://community.cochrane.org/glossary. (Accessed:

566    June 5 2014)

567    Halbert, J. A., Silagy, C. A., Finucane, P., Withers, R. T., & Hamdorf, P. A. Recruitment of older

568    adults for a randomized, controlled trial of exercise advice in a general practice setting. *J. Am.*

569    *Geriatr. Soc*. **47**, 477-81 (1999).

570    Heck RH, Thomas SL, Tabata LN. 2013. Multilevel Modeling of Categorical Outcomes Using

571    IBM SPSS. Routledge Academic.

572    Hirst, J.A., Howick, J., Aronson, J.K., Roberts, N., Perera, R., Koshiaris, C. and Heneghan, C.,

573    2014. The need for randomization in animal trials: an overview of systematic reviews.

574    Jadad, A. R., Moore, R. A., Carroll, D., Jenkinson, C., Reynolds, D. J. M., Gavaghan, D. J., &

575    McQuay, H. J. (1996). Assessing the quality of reports of randomized clinical trials: is blinding

576    necessary?. Controlled clinical trials, 17(1), 1-12.

577    Kilkenny, C., Browne, W. J., Cuthill, I. C., Emerson, M., & Altman, D. G. Improving bioscience

578    research reporting: the ARRIVE guidelines for reporting animal research. *PLoS Biol*. **8,**

579    e1000412 (2010).

580    Kirkewood, B. R., Sterne, J. A. *Essential Medical Statistics*. Oxford: Blackwell Science Ltd;

581    2003.

582   Kjaergard, L. L., Villumsen, J., & Gluud, C. Reported methodologic quality and discrepancies

583   between large and small randomized trials in meta-analyses. *Ann. Intern. Med*. **135,** 982-989

584   (2001).

585   Kuroki, L. M., Allsworth, J. E., & Peipert, J. F. Methodology and analytic techniques used in

586   clinical research: associations with journal impact factor. *Obstet. Gynecol*. **114,** 877-884 (2009).

587   Landis, S.C., Amara, S.G., Asadullah, K., Austin, C.P., Blumenstein, R., Bradley, E.W., Crystal,

588   R.G., Darnell, R.B., Ferrante, R.J., Fillit, H. and Finkelstein, R., 2012. A call for transparent

589   reporting to optimize the predictive value of preclinical research. Nature, 490(7419), pp.187-191.

590   Lehmann EL. Consistency and unbiasedness of certain nonparametric tests. The Annals of

591   Mathematical Statistics (1951): 165-179.).

592   Lingna, JXCYW. Literature evaluation on the randomised controlled trials and controlled

593   clinical trials published on the journal of nursing sciences. *J. Nurs. Sci*. **10**, 002 (2002).

594   Lund, E. M., James, K. M., & Neaton, J. D. Veterinary randomized clinical trial reporting: a

595   review of the small animal literature. *J. Vet. Intern. Med*. **12**, 57-60 (1998).

596   Macleod, M. Why animal research needs to improve. *Nature*. **477,** 511 (2011).

597   McCulloch, P. No surgical innovation without evaluation: the IDEAL recommendations. *Lancet*.

598   **374,** 1105-1112. (2009).

599   Medical Research Council. Streptomycin treatment of pulmonary tuberculosis: a Medical

600   Research Council investigation. *BMJ*. **2,** 769-782 (1948).

601   Merriam-Webster Encyclopaedia Britannica. Definition: Intervention. Available:

602   http://www.merriam-webster.com/medical/intervention. (Accessed: June 5 2014)

603   Midi, H., Sarkar, S. K., & Rana, S. 2010. Collinearity diagnostics of binary logistic regression

604   model. Journal of Interdisciplinary Mathematics, 13(3), 253-267.).

605   Moher, D., Hopewell, S., Schulz, K.F., Montori, V., Gøtzsche, P.C., Devereaux, P.J., Elbourne,

606   D., Egger, M. and Altman, D.G., 2010. CONSORT 2010 explanation and elaboration: updated

607   guidelines for reporting parallel group randomised trials. Journal of clinical epidemiology, 63(8),

608   pp.e1-e37.

609   Moinpour, C.M., Lovato, L.C., Thompson, I.M., Ware, J.E., Ganz, P.A., Patrick, D.L.,

610   Shumaker, S.A., Donaldson, G.W., Ryan, A. and Coltman, C.A., 2000. Profile of men

611   randomized to the prostate cancer prevention trial: baseline health-related quality of life, urinary

612   and sexual functioning, and health behaviors. Journal of clinical oncology, 18(9), pp.1942-1953.

613   Momeni, A., Becker, A., Antes, G., Diener, M. K., Blümle, A., & Stark, G. B. Evidence-based

614   plastic surgery: controlled trials in three plastic surgical journals (1990-2005). *Ann. Plast. Surg.*

615   **61**, 221-225 (2008).

616   Muhlhausler, B. S., Bloomfield, F. H., & Gillman, M. W. Whole animal experiments should be

617   more like human randomized controlled trials. *PLoS Biol.* **11**, e1001481 (2013).

618   Nagelkerke, N. J. (1991). A note on a general definition of the coefficient of determination.

619   Biometrika, 78(3), 691-692.

620   Newcombe, R. G. Two-sided confidence intervals for the single proportion: comparison of seven

621   methods. *Stat. Med.* **17,** 857-872 (1998).

622   O'Connor, A.M., Sargeant, J.M., Gardner, I.A., Dickson, J.S., Torrence, M.E., Dewey, C.E.,

623   Dohoo, I.R., Evans, R.B., Gray, J.T., Greiner, M. and Keefe, G., 2010. The REFLECT

624   Statement: Methods and Processes of Creating Reporting Guidelines for Randomized Controlled

625   Trials for Livestock and Food Safety by Modifying the CONSORT Statement†. Zoonoses and

626   public health, 57(2), pp.95-104.

627   Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., & Feinstein, A. R. A simulation study of the

628   number of events per variable in logistic regression analysis. *J. Clin. Epidemiol.* **49,** 1373-1379

629   (1996).

630   Peng, C. Y. J., & So, T. S. H. Logistic regression analysis and reporting: A primer. *Underst. Stat.*

631   **1,** 31-70 (2002).

632   Pressler, T. R., & Kaizar, E. E. The use of propensity scores and observational data to estimate

633   randomized controlled trial generalizability bias. *Stat. Med.* **32**, 3552-3568 (2013).

634   Rothman, K. J. A show of confidence. *N. Engl. J. Med.* **299**, 1362-1363. (1978).

635   Sackett, D. L., Rosenberg, W. M., Gray, J. A., Haynes, R. B., & Richardson, W. S. Evidence

636   based medicine: what it is and what it isn't. *BMJ.* **312**, 71-72 (1996).

637   Sargeant, J.M., Elgie, R., Valcour, J., Saint-Onge, J., Thompson, A., Marcynuk, P. and Snedeker,

638   K., 2009. Methodological quality and completeness of reporting in clinical trials conducted in

639   livestock species. Preventive veterinary medicine, 91(2), pp.107-115.

640   Sargeant, J.M., Thompson, A., Valcour, J., Elgie, R., Saint-Onge, J., Marcynuk, P. and Snedeker,

641   K., 2010. Quality of reporting of clinical trials of dogs and cats and associations with treatment

642   effects. Journal of veterinary internal medicine, 24(1), pp.44-50.

643   Schulz, K. F., Chalmers, I., Grimes, D. A., & Altman, D. G. Assessing the quality of

644   randomization from reports of controlled trials published in obstetrics and gynecology journals.

645   *JAMA.* **272**, 125-8 (1994).

646  Schulz, K. F., Chalmers, I., Hayes, R. J., & Altman, D. G. Empirical evidence of bias.

647  Dimensions of methodological quality associated with estimates of treatment effects in

648  controlled trials. *JAMA*. **273**, 408-412 (1995).

649  Schünemann, H.J., Tugwell, P., Reeves, B.C., Akl, E.A., Santesso, N., Spencer, F.A., Shea, B.,

650  Wells, G. and Helfand, M., 2013. Non-randomized studies as a source of complementary,

651  sequential or replacement evidence for randomized controlled trials in systematic reviews on the

652  effects of interventions. Research Synthesis Methods, 4(1), pp.49-62.

653  Sena, E., van der Worp, H. B., Howells, D., & Macleod, M. How can we improve the pre-

654  clinical development of drugs for stroke? *Trends Neurosci.* **30,** 433-439 (2007).

655  Sibbald, B., & Roland, M. Understanding controlled trials. Why are randomised controlled trials

656  important? *BMJ*. **316**, 201 (1998).

657  Sinha S, Sinha S, Ashby E, Jayaram R, Grocott MP. Quality of reporting in randomized trials

658  published in high-quality surgical journals. J Am Coll Surg. 2009 Nov;209(5):565-571.e1.

659  Sjögren, P., & Halling, A. Randomised controlled trials and publication trends in periodontal

660  research during 1980-2000. *J. Clin. Periodontol.* **29**, 1112-7 (2002).

661  Sterne, J. A., & Davey Smith, G. Sifting the evidence-what's wrong with significance tests?

662  *BMJ*. **322,** 226-231 (2001).

663  Straube S, Werny B, Friede T. A systematic review identifies shortcomings in the reporting of

664  crossover trials in chronic painful conditions. J Clin Epidemiol. 2015 Dec;68(12):1496-503.

665  Strippoli, G. F., Craig, J. C., & Schena, F. P. The number, quality, and coverage of randomized

666  controlled trials in nephrology. *J. Am. Soc. Nephrol.* **15**, 411-419 (2004).

667    Stuart-Harris, C. H., Francis, A. E., & Stansfeld, J. M. Patulin in the common cold. *Lancet.* **242,**

668    684 (1943).

669    Sweeting, M. J., Sutton, A. J., & Lambert, P. C. What to add to nothing? Use and avoidance of

670    continuity corrections in meta-analysis of sparse data. *Stat. Med*. **23,** 1351-1375 (2004).

671    To, M. J., Jones, J., Emara, M., & Jadad, A. R. Are reports of randomized controlled trials

672    improving over time? A systematic review of 284 articles published in high-impact general and

673    specialized medical journals. *PLoS One*. **8,** e84779 (2013).

674    Walleser S, Hill SR, Bero LA. Characteristics and quality of reporting of cluster randomized

675    trials in children: reporting needs improvement. J Clin Epidemiol. 2011 Dec;64(12):1331-40.

676    doi: 10.1016/j.jclinepi.2011.04.006. Epub 2011 Jul 20.

677    West, R. Does blinding of readers affect results of meta-analyses? *Lancet.* **350,** 892 (1997).

678

# Acknowledgments

680    The authors acknowledge Paolo Selleri for his critical review and comments and Lucy Nicole

681    Papa Caminiti for assistance with data extraction.

682

# Author contributions statement

684    N.D. conceived the study, N.D. and R.M.R. conducted the study, N.D. analyzed the results. All

685    authors reviewed the manuscript.

686

687 **Additional information**

688 This manuscript represents a portion of a thesis submitted by Dr. Di Girolamo to University of

689 Oxford in partial fulfillment of the requirement for the Master of Science in Evidence Based

690 Health Care degree. The authors disclose any financial interests with companies that

691 manufacture products that are the subject of the present research or with companies that

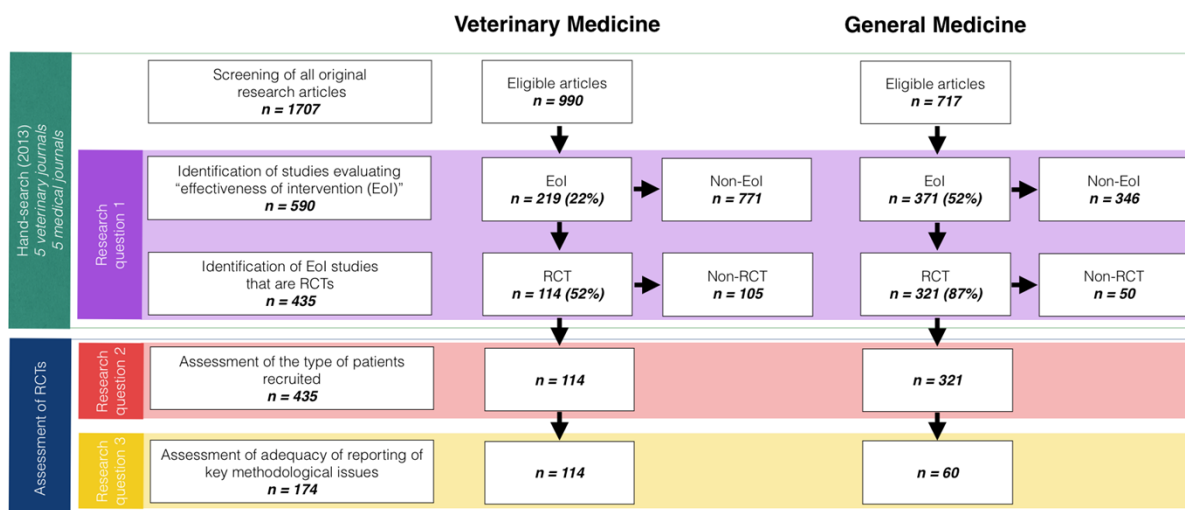
692 manufacture competing products.

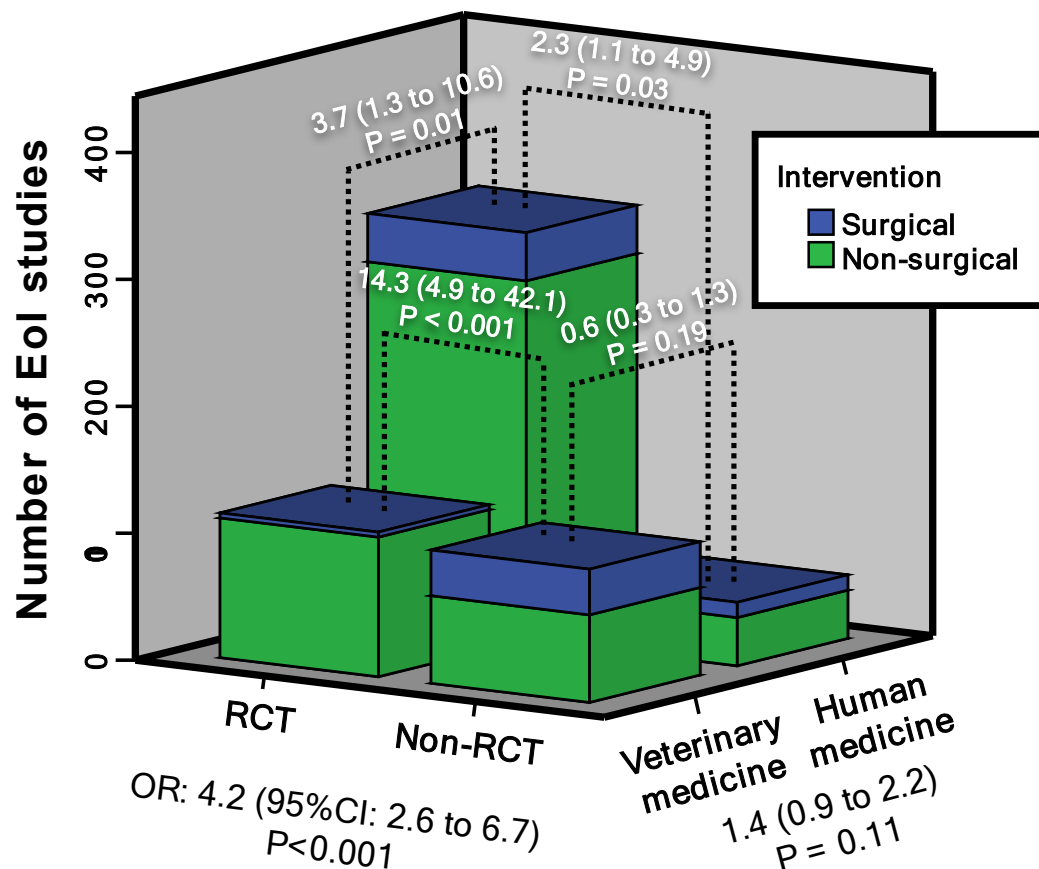
693 **Figures**

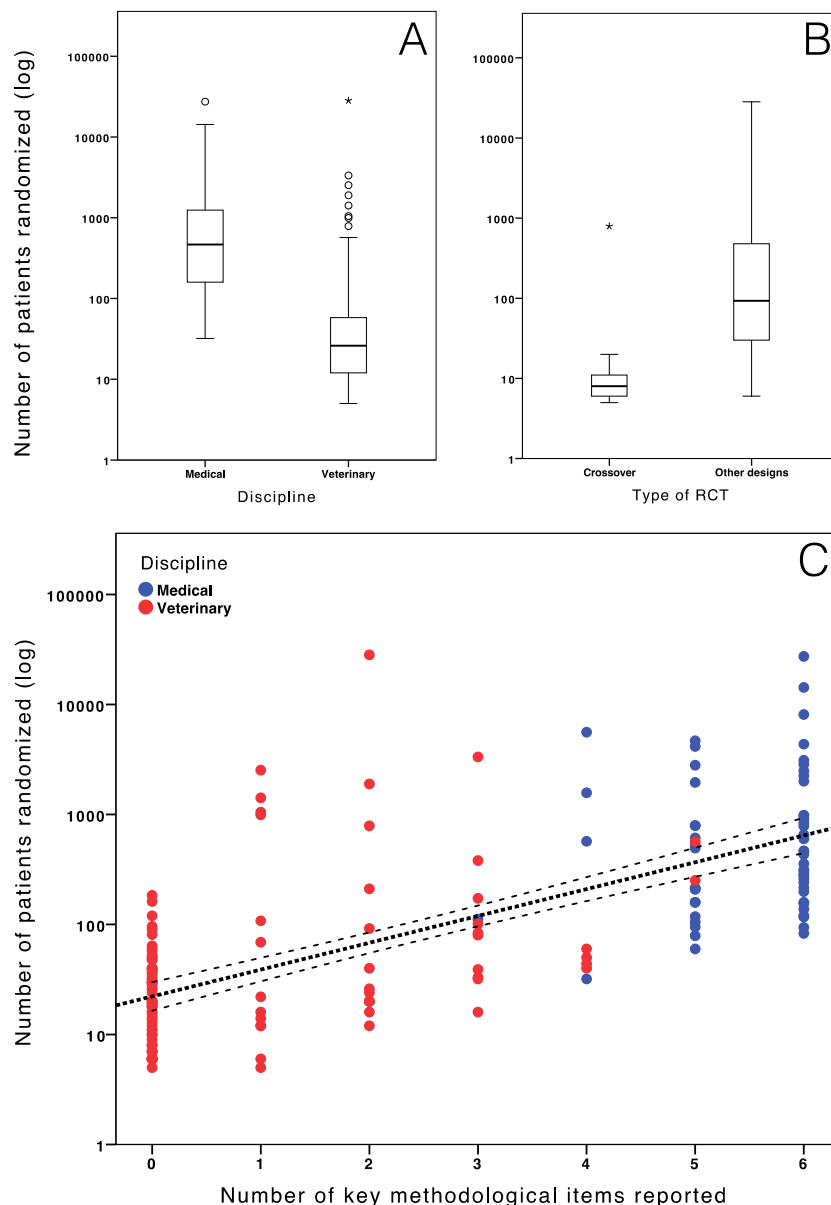
694

695 **Figure 1.** Study flow**.**

696

697
698
699

701 **Figure 2.** Association (odds ratio and 95% confidence intervals) between
702 randomization, discipline and type of intervention (surgical/non-surgical). Notice that the
703 overall prevalence of randomized controlled trials (RCTs) and prevalence of surgical
704 RCTs were lower in veterinary medicine (ORs: 4.2 and 3.7, respectively). However,
705 surgical interventions were more likely to be non-randomized in both disciplines (ORs:
706 14.3 and 2.3).

707
708
709
710

711    **Figure 3.** Number of patients randomized in RCTs and methodological quality. A-
712    Difference in number of patients randomized in veterinary and general medicine. B-
713    Difference in number of patients randomized in cross-over and other study designs. C-
714    Correlation between the number of methodological issues (primary outcome, power
715    calculation, random sequence generation, allocation concealment, estimation methods,
716    intention-to-treat) reported and number of patients enrolled in each RCT. Notice that the
717    y-axis were plotted on a logarithmic function.

718
719
720

721 **Figure 4.** Association between type of patients enrolled (real clinical patients *vs* non-
722 patients) and adequacy of reporting of key methodological issues in RCTs published in
723 leading veterinary journals (n = 114). M-H: Mantel-Haenszel; CI: Confidence interval.
724

| Study or Subgroup | Odds Ratio M-H, Fixed, 95% CI |
|---|---|
| Power calculations | 12.21 [2.67, 55.87] |
| Explicit primary outcome | 9.41 [2.60, 34.10] |
| Proper random sequence | 3.78 [1.36, 10.47] |
| Allocation concealed | 13.93 [1.73, 111.99] |
| Blinding owners/patients | 9.33 [2.01, 43.30] |
| Blinding personnel | 1.92 [0.84, 4.39] |
| Blinding assessors | 2.35 [1.11, 4.98] |
| Intention-to-treat | 2.11 [0.19, 23.96] |
| Estimation methods | 24.85 [3.17, 194.46] |

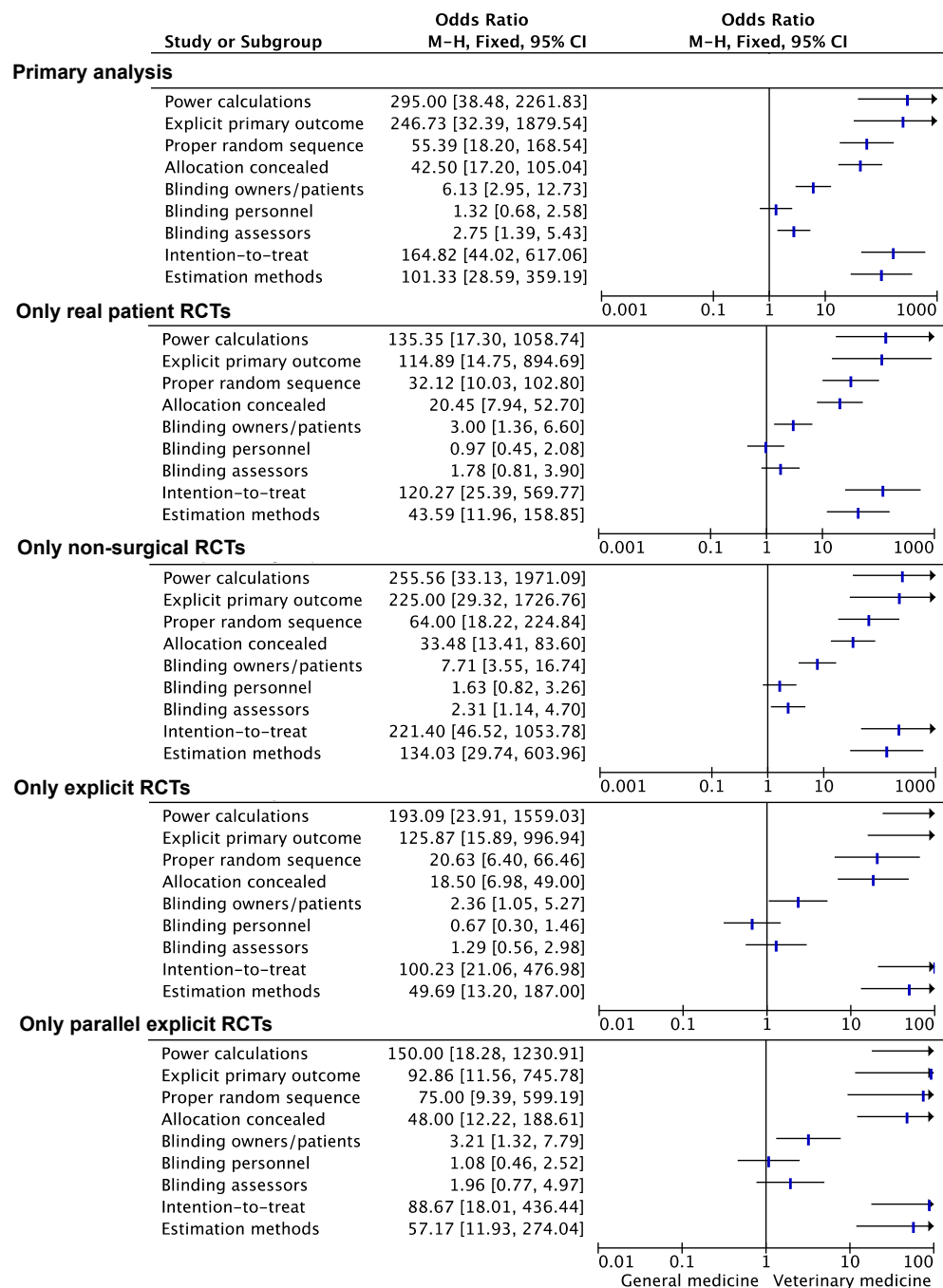
Odds Ratio
M-H, Fixed, 95% CI

0.02    0.1    1    10    50
Real patient RCTs worse   Non-patient RCTs worse

725
726
727
728
729

730 **Figure 5.** Sensitivity and subgroup analysis of the association between discipline and
731 reporting of key methodological items. Primary analysis. Inclusion of explicit RCTs.
732 Exclusion of surgical RCTs. Inclusion of parallel explicit RCTs. Exclusion of non-patient
733 RCTs. M-H: Mantel-Haenszel; CI: Confidence interval.
734

735
736
737

738

# Supplementary Information

740

741 **Supplementary Note 1.** Details on pilot study, sample size calculation and eligibility criteria for
742 journal.
743 **Supplementary Note 2.** Additional details on data extraction.
744 **Supplementary Data.** Other characteristics of RCTs evaluated.
745 **Supplementary Table S1.** Binary logistic regression outcome.

746
747
748
749
750
751

# Tables

753

754 **Table 1.** Definitions used to assess characteristics of publications from 5 leading veterinary and 5 leading
755 medical journals in 2013. Reporting of methodological domains was assessed in all the randomized
756 controlled trials (RCTs) extracted. Full definition of each item is given in the main text and
757 supplementary files. EoI: Effectiveness of intervention.

| | Terms | Descriptions |
|---|---|---|
| **Characteristics of articles** | *Original articles* | Primary research, including subgroup analyses, follow-ups of previous article and case series |
| | *Effectiveness of intervention (EoI) articles* | Primary research evaluating the benefits of an intervention |
| | *Randomized controlled trials (RCT)* | EoI studies with allocation to interventions reported as randomized |
| | *Real patients RCTs* | RCTs that included individuals or animals that suffered from a spontaneous disease and were exposed to real-life conditions |
| | *Surgical (RCT/EoI) articles* | Same as previous definitions, but evaluating the benefits of a surgical intervention |
| | *Explicit RCT* | Trials registered in a trial repository or self-defining "randomized controlled trial" |
| | *Explicit parallel RCT* | Same as pervious, but employing only two arms |
| | *Standalone RCT* | Lack of additional non-randomized work (i.e., in vitro or prospective data) reported in the same article of the RCT |
| | *Crossover RCT* | RCT in which participants receive a sequence of different |

| | | |
|---|---|---|
| | | treatments |
| | *Cluster RCT* | RCT in which groups of participants are randomized to different treatments |
| **Key methodological domains evaluated in RCTs** | *Primary outcome* | A primary outcome is explicitly reported in the published article |
| | *Power calculation* | A power calculation performed a priori to estimate the sample size is explicitly reported |
| | *Random sequence generation* | Methods employed to generate the random list and type of randomization are explicitly reported |
| | *Allocation concealment* | Methods used to prevent the individuals enrolling trial participants from knowing or predicting the allocation sequence in advance are explicitly described in the article |
| | *Blinding of participants* | Explicit description that participants/pet owners were unaware of participants' group allocation |
| | *Blinding of personnel* | Explicit description that operators involved in the care of participants were unaware of participants' group of allocation |
| | *Blinding of outcome assessors* | Explicit description that outcome assessors were unaware of participants' group of allocation |
| | *Intention-to-treat* | Explicit mention that the analysis was made on an "intention-to-treat" basis. |
| | *Effect size estimation methods* | Results are reported with methods that estimate the effect size with confidence interval. |

758

759 **Table 2.** Number of eligible articles, EoI articles and prevalence of RCTs in 10 leading veterinary and
760 medical journals in 2013. AJVR: American Journal of Veterinary Research; JAVMA; Journal of the
761 American Veterinary Medical Association; JVIM: Journal of Veterinary Internal Medicine; Vet J:
762 Veterinary Journal; Vet Rec: Veterinary Record; Annals: Annals of Internal Medicine; BMJ: British
763 Medical Journal; JAMA: Journal of the American Medical Association; NEJM: New England Journalof
764 Medicine.
765

| | **Veterinary Journals** | | | | | | **Medical Journals** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **AJVR** | **JAVMA** | **JVIM** | **Vet J** | **Vet Rec** | **Total** | **Annals** | **BMJ** | **JAMA** | **Lancet** | **NEJM** | **Total** |
| Eligible articles | 193 | 188 | 156 | 320 | 133 | 990 | 75 | 128 | 157 | 152 | 205 | 717 |
| EoI articles | 43 | 51 | 40 | 59 | 26 | 219 | 31 | 46 | 71 | 94 | 129 | 371 |
| | | | | | | | | | | | | |
| RCT (n) | 34 | 18 | 19 | 30 | 13 | 114 | 24 | 37 | 61 | 84 | 115 | 321 |
| (%) | 79% | 35% | 47% | 51% | 50% | 52% | 77% | 80% | 86% | 89% | 89% | 87% |
| Non-RCT (n) | 9 | 33 | 21 | 29 | 13 | 105 | 7 | 9 | 10 | 10 | 14 | 50 |
| (%) | 21% | 65% | 53% | 49% | 50% | 48% | 23% | 20% | 14% | 11% | 11% | 13% |

766
767
768
769

770    **Table 3.** Reporting of key methodological items in RCTs published in leading veterinary and human
771    medical journals. OR: Odds ratio; CI: Confidence interval.
772

| Methodological items | Reporting | Veterinary Count | % | Medical Count | % | OR | 95% CI | P value |
|---|---|---|---|---|---|---|---|---|
| **Power calculation** | Stated | 19 | 16.7 % | 59 | 98.3 % | | | |
| | Not stated | 95 | 83.3 % | 1 | 1.7% | 295 | 38.5 to 2261.8 | 0.001 |
| **Primary outcome** | Defined | 22 | 19.3 % | 59 | 98.3 % | | | |
| | Not defined | 92 | 80.7 % | 1 | 1.7% | 246.7 | 32.4 to 1879.5 | 0.001 |
| **Method of random sequence generation** | Reported | 23 | 20.2 % | 56 | 93.3 % | | | |
| | Not reported | 91 | 79.8 % | 4 | 6.7% | 55.4 | 18.2 to 168.5 | 0.001 |
| **Method of allocation concealment** | Reported | 12 | 10.5 % | 50 | 83.3 % | | | |
| | Not reported | 102 | 89.5 % | 10 | 16.7 % | 42.5 | 17.2 to 105.0 | 0.001 |
| **Detailed blinding of patients/pet owners** | Reported | 16 | 14.0 % | 30 | 50.0 % | | | |
| | Not reported | 98 | 86.0 % | 30 | 50.0 % | 6.1 | 2.9 to 12.7 | 0.001 |
| **Detailed blinding of personnel** | Reported | 33 | 28.9 % | 21 | 35.0 % | | | |
| | Not reported | 81 | 71.1 % | 39 | 65.0 % | 1.3 | 0.6 to 2.5 | 0.4 |
| **Detailed blinding of outcome assessors** | Reported | 57 | 50.0 % | 44 | 73.3 % | | | |
| | Not reported | 57 | 50.0 % | 16 | 26.7 % | 2.7 | 1.4 to 5.4 | 0.003 |
| **Intention-to-treat analysis** | Mentioned | 3 | 2.6% | 49 | 81.7 % | | | |
| | Not mentioned | 111 | 97.4 % | 11 | 18.3 % | 164.8 | 44.0 to 617.0 | 0.001 |
| **Effect size methods** | Used | 18 | 15.8 % | 57 | 95.0 % | | | |
| | Not used | 96 | 84.2 % | 3 | 5.0% | 101.3 | 28.6 to 359.1 | 0.001 |

773