Trustworthiness of randomized trials in endocrinology—A systematic survey

José Gerardo González-González1,2,3☯, Edgar Gerardo Dorsey-Treviño1,2,3☯, Neri Alvarez-Villalobos2,3,4, Francisco Jesús Barrera-Flores2, Alejandro Díaz González-Colmenero2, Carolina Quintanilla-Sánchez2, Victor M. Montori1,2,3*, Rene Rodríguez-Gutierrez1,2,3,4,5 *

1 Endocrinology Division, Department of Internal Medicine, University Hospital “Dr. José E. González”, Universidad Autónoma de Nuevo León, Monterrey, México, 2 Plataforma INVEST Medicina UANL—KER Unit Mayo Clinic (KER Unit México), Universidad Autónoma de Nuevo León, Monterrey, México, 3 Research Unit, University Hospital “Dr. José E. González”, Universidad Autónoma de Nuevo León, Monterrey, México, 4 Knowledge and Evaluation Research Unit in Endocrinology, Mayo Clinic, Rochester, MN, United States of America, 5 Division of Endocrinology, Diabetes, Metabolism, and Nutrition, Department of Medicine, Mayo Clinic, Rochester, MN, United States of America

☯ These authors contributed equally to this work.

* rodriguezgutierrez.rene@mayo.edu

Abstract

Background

Trustworthy (i.e. low risk of bias) randomized clinical trials (RCTs) play an important role in evidence-based decision making. We aimed to systematically assess the risk of bias of trials published in high-impact endocrinology journals.

Methods

We searched the MEDLINE/PubMed database between 2014 and 2016 for phase 2–4 RCTs evaluating endocrine-related therapies. Reviewers working independently and in duplicate used the Cochrane Risk of Bias Tool (CCRBT) to determine the extent to which the methods reported protected the results of each RCT from bias.

Results

We assessed 292 eligible RCTs, of which 40% (116) were judged to be at low risk, 43% (126) at moderate, and 17% (50) at high risk of bias. Blinding of outcome assessment was the least common domain reported 43% (125), while selective reporting of outcomes was the most common 97% (282). In multivariable analysis, RCTs with a parallel design (OR 2.4; 95% CI; 1.2–4.6) and funded by for-profit sources (OR 2.2; 95% CI; 1.3–3.6) were more likely to be at low risk of bias.

Conclusions

Trustworthy evidence should ultimately shape care to improve the likelihood of desirable patient outcomes. Six out-of 10 RCTs published in top endocrine journals are at moderate/high-risk of bias. Improving this should be a priority in endocrine research.
Introduction

Well conducted randomized clinical trials (RCTs), should help clinicians, patients, and policymakers make more confident decisions about care. Attention to this so-called core principle of evidence-based medicine (EBM),[1] has supported the critical appraisal of the methods used in RCTs, and has contributed to improve health care.[2–5] Such critical appraisal focuses on recognizing that RCTs often lack sufficient protection against bias that consequently reduces the confidence in their estimates.[6, 7] This confidence creates the trustworthiness for clinicians to apply evidence into patient care.

Trustworthiness in RCTs can be drawn by assuring the transparency of trial’s methods.[7] To do so, several strategies have been adopted to guide researchers in their reporting of methods.[8–11] However, despite these guidance, low quality methodological reporting seems to prevail among several fields of medicine.[12–15] These untrustworthy studies, which reliability is at most questionable, are frequently used by policymakers to develop clinical guidelines, promote an intervention, or generate recommendations often labeled as strong.[16, 17] If patient care should be stemmed from research evidence that mainly draws its recommendations from solid evidence aimed at discover, uncover, or invent treatments that improve their lives, relying in trials in which confidence appear to be obscure by untrustworthy methods opposes the true essence of EBM.[18] This incongruency of developing guidelines and recommendations based on quivery evidence, is prone to over- or underestimate the true effect of an intervention, and may ultimately cause harm to the patient or end up being research waste.[19–21] In this instance, conducting low quality clinical research translates into low quality of evidence that ultimately causes low quality of care for patients.

The extent to which the results of important RCTs of treatments for endocrine conditions are protected against bias and thus are trustworthy, however, remains uncertain. Consequently, we aimed to systematically evaluate the overall risk of bias of endocrine RCTs published in high-impact journals between 2014–2016.

Material and methods

This systematic review adhere to the Preferred Reporting Items for Systematic Review and Meta-analysis (PRISMA) (S1 Appendix).[22]

Study eligibility criteria

Eligible articles were phase 2 to 4 RCTs enrolling patients with an endocrinopathy (e.g., diabetes, thyroid, obesity, bone metabolism, cardiovascular (lipids)/metabolism, and pituitary-gonadal-adrenal axis) to estimate treatment efficacy, regardless of language of publication or number of participants included in the trial. As our intention was to evaluate potential bias of RCTs, we decided to only include the first report of the trial and exclude all follow-ups or any other observational designs (i.e. extensions from an RCTs) aimed at evaluating RCT population.

Search strategy

With the help of an experienced librarian (N.A-V.), we comprehensively searched MEDLINE/PubMed using the Cochrane Highly Sensitive Search Strategy[23] for RCTs published between January 2014 to September 2016—the time frame was chosen arbitrarily—in; a) top five impact-factor medical journals (New England Journal of Medicine, The Journal of the American Medical Association [JAMA], Lancet, Annals of Internal Medicine, and the British Medical Journal [BMJ]), b) two top diabetes journals (Diabetologia and Diabetes Care), c) two top general
endocrinology journals (Lancet Diabetes & Endocrinology and Journal of Clinical Endocrinology and Metabolism [JCEM]), d) two top cardiovascular journals (Journal of the American College of Cardiology and Circulation), and e) the five top impact-factor journals of thyroid, pituitary, bone, and obesity journals (Thyroid, Pituitary, Journal of Bone and Mineral Research, International Journal of Obesity). All journals were selected based on the 2015 Journal Citation Reports (JCR) [24]. The complete search strategy is provided in the S2 Appendix.

Selection of studies
Two pairs of reviewers working independently and in duplicate reviewed all potentially eligible articles. In order to standardize the reviewers’ judgments based on the aforementioned inclusion and exclusion criteria, a pilot study reviewing 20 articles was performed with discussion until the pairs achieved optimal chance-adjusted inter-reviewer agreement (kappa ≥ 0.8). Disagreements between reviewers were initially resolved by consensus and, when needed, by adjudication by an endocrinologist and methodologist (R.R-G. or V.M.M.).

Data collection
Using a standardized web-based form (Online Microsoft Excel 2016, Microsoft, Redmont, WA, USA), reviewers working independently and in duplicate used the Cochrane Risk of Bias Tool (CCRBTO) to assess the protection against bias afforded by random sequence, allocation concealment, blinding of personnel and participants, blinding of outcome assessment, incomplete outcome data, selective reporting and use of the intention-to-treat analysis.[9] Additionally, we extracted data regarding year of publication, branch of endocrinology, funding, number of centers, type of outcomes (patient-important outcomes or surrogate or laboratory outcomes), analysis of data (intention-to-treat or per protocol), and type of journal, intervention, and design.

Risk of bias classification
Each of these seven domains was classified as indicative of high, moderate, or low-risk of bias based on specific criteria. For instance, we classified random sequence generation as placing a study at low risk of bias if the method of allocation was explicitly stated in the article (e.g., a computer-based program was used to randomly allocate patients); when the allocation was reported only as random, we classified the level of protection against bias as unclear. RCTs were also considered at low risk of bias when the investigator gathering the data or processing the data (e.g., trial statistician) were reportedly blind to trial allocation (blinded outcome assessor), when outcomes in trials showed no apparent sign of omission or reporting only positive outcomes (selective reporting), when loss to follow-up was <20% (incomplete outcome data) and when analyses adhered to the intention-to-treat principle. A full and detailed description of each domain is provided in the S1 Table.

When adequate protection was present across all seven domains or if only one domain was unprotected, we classified the study as at low risk of bias. When >3 domains were classified as having poor or unclear protections against bias, we classified the study as at high risk of bias. All other RCTs were classified as at being at moderate risk of bias.

Missing data
When data were missing, unclear or incomplete, we searched for this information in the registration record of the trial in clinicaltrials.gov, the Australian New Zealand Clinical Trial Registry (ANZCTR), or the University Hospital Medical Information Network Clinical Trial
Registry (UMIN-CTR). If the study was not registered or data was still unavailable, we contacted the corresponding author. After a lapse of 10 days, if no response was received, we excluded the article. Every contact was documented and reported. Additionally, we foresighted that some RCTs would fail to report the study phase. Whenever this happened, we evaluated each RCT and judged whether it had phase 1 properties, and if so, it was excluded, otherwise the study was included and labeled as “Not Reported”.

**Statistical analysis**

We used a descriptive analysis to report categorical variables with frequencies and percentages. We used multivariable analysis using a logistic regression model to assess the probability of a study being at low risk of bias (yes/no). Predictors were selected based on previous evidence and included type of intervention (pharmacological vs. nonpharmacological), trial design (parallel vs. other), type of outcome (patient-important outcomes vs. surrogate or laboratory outcomes), funding (non-profit sources vs. for-profit sources), and number of centers (single vs. multicenter).\[16, 25, 26\] Adjustments were made on these same variables and goodness-of-fit was determined by the c-statistic and Hosmer-Lemeshow test. Additionally, an univariate analysis was performed to assess the impact of adjusted variables. Trials in which data was missing, were excluded from the multivariable analysis. All variables were inputted in a stepwise backward manner and then excluded until a model that best fitted our data was identified. We took a \( p \) value < .05 as statistically significant; associations were described using odds ratios (ORs) and their associated 95% confidence interval (CI). We used SPSS version 22 (IBM Corp, Armonk, NY, USA) for all statistical analyses.

**Results**

We found 292 eligible RCTs (Fig 1). Most RCTs were industry funded (51%), single-center trials (44%) testing a drug vs. placebo (38%) in patients with diabetes (67%) and measuring its impact on surrogate or laboratory outcomes (70%) (Table 1). A complete list of included articles is provided as a supplementary file (S2 Table).

Chance-adjusted inter-observer agreement for risk of bias judgments ranged from moderate to acceptable (kappa = 0.5 to 0.8). Of the 292 trials, 116 (40%; 95% CI; 34–45) were judged to be at low risk, 126 (43%; 95% CI; 37–49) at moderate, and 50 (17%; 95% CI; 13–21%) at high risk of bias. Blinding of outcomes assessment was the least used protection (43%; 95% CI; 37–49%) and complete reporting the most commonly used protection against bias (97%; 95% CI; 94–99%) (Table 2). In the multivariable analysis, parallel design (OR 2.4; 95% CI; 1.2–4.6) and funded by profitable sources (OR 2.1; 95% CI; 1.2–4.6) were associated with better protection against bias. (Table 3). However, this association was found to be substantially small (\( R^2 = 0.1 \)) and the goodness-of-fit test was moderately inadequate (c-statistic = 0.6 and Hosmer and Lemeshow = 0.7).

**Discussion**

**Summary of findings**

About 4 in 10 endocrine RCTs in top medical journals are adequately protected against bias and thus warrant high confidence in their estimates. Industry-funded parallel design RCTs evaluating drugs exhibited the most methodological protections against bias. Fewer than a third of the reviewed RCTs assessed patient-important outcomes, and almost half of these produced results judged to be at low-risk of bias.
Strengths and limitations

To our knowledge, this is the most contemporary analysis evaluating the overall trustworthiness of RCTs in endocrinology. Although a newer version of the CCRBT has recently been published, perhaps our results may vary with these new rationales for assessing bias in RCTs. However, because we reviewed only top journals, these results may represent a best-case scenario, assuming that the peer-review process enriches the published record with more trustworthy trials. Conversely, our results may underestimate the protection against bias to the extent that RCT reports fail to report methods that the investigators did implement. Our protocol-driven methods and our reliance on multiple, independent, and reproducible judgments to select trials into the review and to analyze each RCT’s risk of bias using a standardized tool should warrant confidence in our findings.

Comparison with previous studies

Several fields of medicine have sought to evaluate the quality of reporting methods in a variety of fields of medicine. Regarding endocrinology, it appears that it has passed more than 20 years and the reporting of methods appears to have barely improve. For instance, in 1996,
Table 1. Risk of bias of the 292 included randomized trials.

| Trial Characteristics           | Total | High (%) | Moderate (%) | Low (%) |
|--------------------------------|-------|----------|--------------|---------|
| All Included Articles          | 292   | 50 (17)  | 126 (43)     | 116 (40) |
| Year                           |       |          |              |         |
| 2014                           | 119   | 23 (19)  | 45 (37)      | 51 (42) |
| 2015                           | 118   | 18 (15)  | 57 (48)      | 43 (36) |
| 2016                           | 55    | 9 (16)   | 24 (43)      | 22 (40) |
| Endocrinology Branch           |       |          |              |         |
| Diabetes                       | 195   | 33 (17)  | 81 (41)      | 81 (41) |
| Cardiovascular (Lipids) or Metabolism | 42 | 3 (7) | 22 (52) | 17 (40) |
| Bone                           | 37(13)| 8 (21)   | 16 (43)      | 13 (35) |
| Thyroid                        | 10(3) | 2 (20)   | 6 (60)       | 2 (20)  |
| PGA                            | 8(3)  | 4 (50)   | 1 (12)       | 3 (37)  |
| Type of Intervention           |       |          |              |         |
| Drug vs. Placebo               | 110   | 18 (16)  | 42 (38)      | 50 (46) |
| Active Drug                    | 76    | 14 (18)  | 36 (47)      | 26 (34) |
| Usual Care                     | 13    | 2 (15)   | 6 (46)       | 5 (39)  |
| Other                          | 47    | 1 (25)   | 2 (50)       | 1 (25)  |
| Education or Lifestyle Changes | 17    | 2 (11)   | 7 (41)       | 8 (47)  |
| Nutrition                      | 25    | 4 (16)   | 12 (48)      | 9 (36)  |
| Other Intervention             | 47    | 9 (19)   | 21 (44)      | 17 (36) |
| Type of Journal                |       |          |              |         |
| Specialty                      | 265   | 0 (0)    | 8 (29.6)     | 19 (70.4) |
| General                        | 27    | 50 (18.9)| 118 (44.5)  | 97 (36.6) |
| Phase                          |       |          |              |         |
| 2                              | 54    | 9 (16)   | 26 (48)      | 19 (35) |
| 3                              | 77    | 6 (7)    | 34 (44)      | 37 (48) |
| 4                              | 47    | 5 (10)   | 22 (46)      | 20 (42) |
| Not Reported                   | 114   | 30 (26)  | 44 (38)      | 40 (35) |
| Trial Design                   |       |          |              |         |
| Cluster                        | 6     | 1 (16)   | 2 (33)       | 3 (50)  |
| Crossover                      | 55    | 20 (16)  | 26 (47)      | 9 (16)  |
| Factorial                      | 3     | 0 (0)    | 1 (33)       | 2 (66)  |
| Parallel                       | 227   | 29 (12)  | 97 (42)      | 101 (44) |
| Other                          | 1     | 0 (0)    | 0 (0)        | 1 (100) |
| Number of Centers              |       |          |              |         |
| Single Center                  | 129   | 36 (28)  | 53 (41)      | 40 (31) |
| Few (<10 centers)              | 56    | 7 (13)   | 25 (45)      | 24 (43) |
| Many (>10 centers)             | 101   | 7 (7)    | 45 (45)      | 49 (49) |
| Not Reported                   | 6     | 0 (0)    | 3 (50)       | 3 (50)  |
| Type of Outcome                |       |          |              |         |
| Patient-Important Outcomes     | 89    | 7 (7)    | 40 (44)      | 42 (47) |
| Surrogate or Laboratory Outcomes | 203 | 43 (21) | 86 (42)     | 74 (36) |
| Type of Analysis               |       |          |              |         |
| Intention-to-Treat             | 139   | 8 (5)    | 43 (30)      | 88 (63) |
| Per Protocol                   | 153   | 42 (28)  | 83 (54)      | 28 (18) |
| Region of Work                 |       |          |              |         |
| Asia                           | 25    | 1 (4)    | 17 (68)      | 7 (28)  |

(Continued)
McIver et al. assessed the quality of methods in 79 trials enrolling patients with type 2 diabetes and found that the methods used for randomization and blinding were reported in only 15% and 46%, respectively.\[33\] Ten years later, Montori et al. evaluated 199 diabetes trials and judged 53% of these to be at high risk of bias.\[26\] Although they did not use the CCRBT, that study also found trials funded by industry to exhibit more protection against bias.\[26\] More recently, in 2008, Rios et al. broaden the scope and analyzed 89 trials published in three top endocrinology journals (e.g. Journal of Clinical Endocrinology and Metabolism [JCEM], European Journal of Endocrinology, and Clinical Endocrinology) and found that the reporting of allocation concealment (10%), blinding (20%), and intention-to-treat analysis (16%) was sub-optimal.\[34\] Although we found an improvement in the reporting of methods, when analyzing these domains as a whole to determine the overall trustworthiness rendered by trials, we found that endocrine trials warrant untrustworthy results due to insufficient protection against bias. This lack of confidence in endocrine trials resonates with what stated by Brito et al. in their systematic review, showing that most clinical guidelines recommendations in endocrinology were supported by evidence warranting low confidence.\[35\]

**Implications for clinical practice and research**

For healthcare, to assure that patients receive optimal care that ultimately seeks to improve their lives, that care must be supported by trustworthy evidence. Clinical recommendations are a way in which the healthcare enterprise assures that clinicians can make confident decisions regarding patients’ care.\[36, 37\] Thus, this so called evidence-based practice should be based on steadfast evidence that warrants confidence and aims at improving patients’ needs and preferences. Nonetheless, it appears, based on our results, that RCTs are predominantly being conducted without any proper protection against bias, which consequently causes clinical guidelines to draw recommendations based on estimates warranting low confidence and directed at improving intangible surrogate markers that render little or no benefit for patients.\[38, 39\] These incongruencies in clinical research and practice seems to obviate the main purpose of evidence-based medicine, assuring patient’s wellbeing. In light of this situation, clinicians should be more judicious in the confidence inputted in studies or recommendations they use to provide care.

---

**Table 1.** (Continued)

| Trial Characteristics | Risk of Bias n (%) |
|-----------------------|--------------------|
|                       | Total | High | Moderate | Low  |
| Australia             | 18    | 3 (16) | 5 (27) | 10 (55) |
| Europe                | 69    | 18 (26) | 30 (43) | 21 (30) |
| North America         | 78    | 20 (25) | 30 (38) | 28 (35) |
| South America         | 3     | 0     | 2 (66) | 1 (33)  |
| Multicenter           | 99    | 8 (8)  | 42 (42) | 49 (49) |

| Funding               | Risk of Bias n (%) |
|-----------------------|--------------------|
| Profit Sources        | 148 (51) | 17 (12) | 51 (34) | 80 (54) |
| Non-Profit Sources    | 109 (37) | 22 (20) | 57 (52) | 30 (28) |
| Mixed Sources         | 31 (11)  | 9 (29)  | 17 (55) | 5 (16)  |
| Not Reported          | 4 (1)    | 2 (50)  | 1 (25)  | 1 (25)  |

Data is presented as frequencies (percentages); PGA = pituitary-gonadal-adren al axis. Low Risk was labeled if seven or six domains were adequately protected. High risk when three or more domain were inadequately protected against bias. All other instances, were judged as Moderate Risk

https://doi.org/10.1371/journal.pone.0212360.t001
Table 2. Low risk assessment of the seven domains from the Cochrane Risk of Bias Tool.

| Trials Characteristics | Low Risk n (%) | \(n = 292\) | \(n = 214\) | \(n = 189\) | \(n = 248\) | \(n = 125\) | \(n = 225\) | \(n = 282\) | \(n = 139\) |
|------------------------|----------------|------------|------------|------------|------------|------------|------------|------------|------------|
| **All Included Articles** |                | 214 (73)   | 189 (65)   | 248 (85)   | 125 (43)   | 225 (77)   | 282 (97)   | 139 (48)   |
| **Endocrinology Branch** |                |            |            |            |            |            |            |            |            |
| Diabetes               |                | 142 (66)   | 124 (65)   | 168 (67)   | 70 (56)    | 159 (70)   | 189 (67)   | 108 (77)   |
| Cardiovascular (Lipids) or Metabolism | | 33 (15) | 32 (16) | 37 (14) | 24 (19) | 31 (13) | 41 (14) | 14 (10) |
| Bone                   |                | 26 (12)    | 25 (13)    | 31 (12)    | 25 (20)    | 24 (10)    | 36 (12)    | 9 (6)      |
| Thyroid                |                | 9 (4)      | 4 (2)      | 7 (2)      | 3 (2)      | 6 (2)      | 10 (3)     | 3 (2)      |
| PGA                    |                | 4 (1)      | 4 (2)      | 5 (2)      | 3 (2)      | 5 (2)      | 6 (2)      | 5 (3)      |
| **Type of Intervention** |                |            |            |            |            |            |            |            |            |
| Drug vs.               |                |            |            |            |            |            |            |            |            |
| Placebo                |                | 81 (37)    | 76 (40)    | 103 (41)   | 54 (43)    | 84 (37)    | 105 (37)   | 48 (34)    |
| Active Drug            |                | 53 (24)    | 45 (23)    | 59 (23)    | 24 (19)    | 59 (26)    | 74 (26)    | 42 (30)    |
| Usual Care             |                | 10 (4)     | 8 (4)      | 10 (4)     | 5 (4)      | 12 (5)     | 12 (4)     | 5 (3)      |
| Other                  |                | 3 (1)      | 3 (1)      | 3 (1)      | 2 (1)      | 3 (1)      | 4 (1)      | 0 (0)      |
| Education of Lifestyle Changes | | 13 (6) | 9 (4) | 15 (16) | 8 (6) | 14 (6) | 17 (6) | 9 (6) |
| Nutrition              |                | 22 (10)    | 14 (7)     | 22 (8)     | 12 (9)     | 21 (9)     | 25 (8)     | 9 (6)      |
| Other Intervention     |                | 32 (15)    | 34 (18)    | 36 (14)    | 20 (16)    | 32 (14)    | 45 (16)    | 26 (18)    |
| **Phase**              |                |            |            |            |            |            |            |            |            |
| 2                      |                | 54 (19)    | 35 (16)    | 33 (17)    | 46 (18)    | 21 (16)    | 45 (20)    | 54 (19)    | 30 (21)    |
| 3                      |                | 77 (26)    | 67 (31)    | 60 (31)    | 70 (28)    | 32 (25)    | 59 (26)    | 76 (27)    | 39 (28)    |
| 4                      |                | 47 (16)    | 37 (17)    | 36 (19)    | 39 (15)    | 22 (17)    | 35 (15)    | 46 (16)    | 22 (15)    |
| Not Reported           |                | 75 (35)    | 60 (31)    | 93 (37)    | 50 (40)    | 86 (38)    | 106 (37)   | 48 (34)    |
| **Trial Design**       |                |            |            |            |            |            |            |            |            |
| Parallel               |                | 227 (77)   | 175 (81)   | 161 (85)   | 193 (77)   | 108 (86)   | 173 (76)   | 221 (78)   | 115 (82)   |
| Cluster                |                | 6 (2)      | 4 (1)      | 2 (1)      | 5 (2)      | 4 (3)      | 5 (2)      | 6 (2)      | 5 (3)      |
| Crossover              |                | 55 (18)    | 31 (14)    | 22 (11)    | 46 (18)    | 10 (8)     | 43 (19)    | 51 (18)    | 17 (12)    |
| Factorial              |                | 3 (1)      | 3 (1)      | 3 (1)      | 2 (1)      | 3 (1)      | 3 (1)      | 2 (2)      |
| Other                  |                | 1 (0)      | 1 (0)      | 1 (0)      | 1 (0)      | 1 (0)      | 1 (0)      | 0 (0)      |
| **Type of Journal**    |                |            |            |            |            |            |            |            |            |
| General                |                | 26 (12)    | 22 (11)    | 26 (10)    | 15 (12)    | 21 (9)     | 27 (9)     | 20 (14)    |
| Specialty              |                | 265 (90)   | 188 (87)   | 167 (88)   | 222 (89)   | 110 (88)   | 204 (90)   | 255 (90)   | 119 (85)   |

(Continued)
Furthermore, our findings demonstrate that most investigator initiated RCTs, particularly those funded by federal agencies and foundations are associated with features that place their results at high risk of bias. There is evidence that industry funded trials, although more protected from bias, they are more likely to be affected by spin features—the distortion introduced by subtle features related to the trial question (e.g., selection of patients, interventions, outcomes, and methods) and to the presentation of its results—that mislead readers and are harder to ascertain.\[40\] Some of these concerns are being addressed by a series of efforts that include initiatives to promote better standards,\[41\] prospectively register trials,\[42, 43\] publish all the results from all trials,\[44\] and report trial results in adherence to the Consolidated Standards of Reporting Trials (CONSORT) reporting statement.\[45, 46\]. The extent to which these initiatives are improving the evidence base for endocrine practice appears limited at this point.

### Table 2. (Continued)

| Trials Characteristics | Random Sequence | Allocation Concealment | Blinding of Personnel and Participants | Blinding of Outcome Assessment | Incomplete Outcome Data | Selective Reporting | Intention-to-Treat |
|------------------------|----------------|------------------------|----------------------------------------|-------------------------------|------------------------|--------------------|------------------|
| **Number of Centers**  |                |                        |                                        |                               |                        |                    |                  |
| Single Center          | 129 (44)       | 84 (39)                | 70 (37)                                | 102 (41)                      | 55 (44)                | 100 (44)           | 122 (43)         |
| Few (<10 centers)      | 56 (19)        | 42 (19)                | 36 (19)                                | 51 (20)                       | 20 (16)                | 43 (19)            | 56 (19)          |
| Many (>10 centers)     | 101 (35)       | 82 (38)                | 78 (41)                                | 89 (35)                       | 47 (37)                | 76 (33)            | 99 (35)          |
| Not Reported           | 6 (2)          | 6 (2)                  | 5 (2)                                  | 6 (2)                         | 3 (2)                  | 6 (2)              | 5 (1)            |
| **Region of Work**     |                |                        |                                        |                               |                        |                    |                  |
| Asia                   | 25 (8)         | 23 (10)                | 19 (10)                                | 16 (6)                        | 13 (10)                | 21 (9)             | 24 (8)           |
| Australia              | 18 (6)         | 15 (7)                 | 14 (7)                                 | 16 (6)                        | 10 (8)                 | 12 (5)             | 17 (6)           |
| Europe                 | 69 (23)        | 46 (21)                | 36 (19)                                | 58 (23)                       | 20 (16)                | 50 (22)            | 65 (23)          |
| North America          | 78 (26)        | 47 (22)                | 42 (22)                                | 66 (26)                       | 35 (28)                | 62 (27)            | 76 (27)          |
| South America          | 3 (1)          | 3 (1)                  | 3 (1)                                  | 3 (1)                         | 1 (1)                  | 3 (1)              | 3 (1)            |
| Multicenter            | 99 (33)        | 80 (37)                | 75 (39)                                | 89 (35)                       | 46 (36)                | 77 (34)            | 97 (34)          |
| **Type of Outcome**    |                |                        |                                        |                               |                        |                    |                  |
| Patient-Important Outcomes | 89 (31)   | 74 (83)                | 64 (72)                                | 79 (89)                       | 39 (44)                | 65 (73)            | 87 (98)          |
| Surrogate or Laboratory Outcome | 203 (70) | 140 (69) | 125 (70) | 168 (83) | 86 (43) | 160 (79) | 194 (96) |
| Funding                |                |                        |                                        |                               |                        |                    |                  |
| Profit Sources         | 148 (51)       | 118 (55)               | 113 (59)                               | 136 (54)                      | 68 (54)                | 114 (50)           | 143 (50)         |
| Non-Profit Sources     | 109 (39)       | 76 (35)                | 62 (32)                                | 87 (35)                       | 48 (38)                | 85 (37)            | 106 (37) |
| Mixed Sources          | 31 (11)        | 17 (7)                 | 11 (5)                                 | 24 (9)                        | 7 (5)                  | 24 (10)            | 29 (10)          |
| Not reported           | 4 (1.4)        | 3 (1)                  | 3 (1)                                  | 1 (1)                         | 2 (1)                  | 2 (1)              | 4 (1)            |

Data is presented in frequencies (percentages); PGA = pituitary-gonadal-adrenal.

https://doi.org/10.1371/journal.pone.0212360.t002
and over the last decade. Even efforts to determine the questions that require better evidence derived from systematic guideline development appear futile: a systematic survey found that only 25% of the research gaps identified by The Endocrine Society guideline recommendations as based on evidence warranting low or very low confidence were being tested in ongoing RCTs.[47]

**Conclusion**

Most of the RCTs in endocrinology published in top medical journals seem insufficiently protected against bias. Improving the methodological quality of RCTs should be a top priority in endocrine research.

**Supporting information**

S1 Appendix. PRISMA checklist.
(DOC)

S2 Appendix. Full search strategy.
(DOCX)

S1 Table. Detailed description of the Cochrane Risk of Bias Tool domains.
(DOCX)

S2 Table. List of included articles.
(DOCX)

**Acknowledgments**

We thank Dr. Sergio Lozano for the revision of the manuscript’s language.

**Author Contributions**

**Conceptualization:** José Gerardo González-González, Edgar Gerardo Dorsey-Treviño, Neri Alvarez-Villalobos, Alejandro Díaz González-Colmenero, Victor M. Montori, Rene Rodríguez-Gutierrez.

**Data curation:** Edgar Gerardo Dorsey-Treviño, Neri Alvarez-Villalobos, Francisco Jesús Barrera-Flores, Alejandro Díaz González-Colmenero, Carolina Quintanilla-Sánchez.

**Formal analysis:** Edgar Gerardo Dorsey-Treviño.

| Predictor Variable | Odds Ratio (95% Confidence Interval) for Low Risk of Bias Trustworthiness |
|--------------------|-------------------------------------------------|
| Design (Other vs. Parallel) | 2.67 (1.4–5) | 2.40 (1.2–4.6) |
| Funding (Non-profit vs. Profit) | 2.39 (1.4–3.9) | 2.16 (1.2–3.6) |
| Center (Single vs. Multicenter) | 1.85 (1.1–3) | 1.05 (0.5–1.8) |
| Type of Intervention (No Drug vs Drug) | 1.09 (0.6–1.8) | 0.67 (0.3–1.2) |
| Type of Outcome (SorL vs PIO) | 1.55 (0.9–2.5) | 1.48 (0.8–2.5) |
| Phase (<4 vs 4) | 1.15 (0.6–2.1) | 1.2 (0.6–2.3) |
| Type of Journal (General vs. Specialty) | 0.24 (0.1–0.5) | 0.29 (0.1–1.7) |

SorL = Surrogate or Laboratory Outcomes; PIO = Patient-Important Outcomes

https://doi.org/10.1371/journal.pone.0212360.t003
Investigation: José Gerardo González-González, Edgar Gerardo Dorsey-Treviño, Neri Alvaréz-Villalobos.

Methodology: José Gerardo González-González, Edgar Gerardo Dorsey-Treviño, Neri Alvaréz-Villalobos, Alejandro Díaz González-Colmenero, Rene Rodriguez-Gutierrez.

Supervision: José Gerardo González-González, Rene Rodriguez-Gutierrez.

Writing – original draft: Edgar Gerardo Dorsey-Treviño, Rene Rodriguez-Gutierrez.

Writing – review & editing: José Gerardo González-González, Victor M. Montori, Rene Rodriguez-Gutierrez.

References
1. Montori VM. Evidence-based endocrinology. Treat Endocrinol. 2004; 3(1):1–10. PMID: 15743108
2. Cohen ND, Shaw JE. Diabetes: advances in treatment. Intern Med J. 2007; 37(6):383–8. https://doi.org/10.1111/j.1445-5994.2007.01369.x PMID: 17535382
3. Kelly MP, Capewell S. Relative contributions of changes in risk factors and treatment to the reduction in coronary heart disease mortality. Health Development Agency. 2004.
4. Chu EK, Drazen JM. Asthma: one hundred years of treatment and onward. Am J Respir Crit Care Med. 2005; 171(11):1202–8. Epub 2005/03/22. https://doi.org/10.1164/rccm.200502-2570OE PMID: 15778490.
5. Lau BD, Haut ER. Practices to prevent venous thromboembolism: a brief review. BMJ quality & safety. 2014; 23(3):187–95. Epub 2013/05/28. https://doi.org/10.1136/bmjqs-2012-001782 PMID: 23708438; PubMed Central PMCID: PMCPMC3932749.
6. Murad MH, Asi N, Alswas M, Alahdab F. New evidence pyramid. Evidence Based Medicine. 2016. https://doi.org/10.1136/ebmed-2016-110401 PMID: 27339128
7. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA. 1996; 276(8):637 –9. Epub 1996/08/28. PMID: 8773637.
8. Jackson JL, Srinivasan M, Rea J, Fletcher KE, Kravitz RL. The validity of peer review in a general medicine journal. PLoS One. 2011; 6(7):e22475. Epub 2011/07/30. https://doi.org/10.1371/journal.pone.0022475 PMID: 21798987; PubMed Central PMCID: PMCPMC3145147.
9. Higgins JPT, Altman DG, Getzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration’s tool for assessing risk of bias in randomised trials. BMJ; 2011; 343. https://doi.org/10.1136/bmj.d5926 PMID: 22008217
10. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized controlled trials: is blinding necessary? Control Clin Trials. 1996; 17(1):1–12. Epub 1996/02/01. PMID: 8721797.
11. Lim SM, Shin ES, Lee SH, Seo KH, Jung YM, Jang JE. Tools for assessing quality and risk of bias by levels of evidence. J Korean Med Assoc. 2011; 54(4):419–29.
12. Clark L, Schmidt U, Tharmananthan P, Adamson J, Hewitt C, Torgerson D. Poor reporting quality of key Randomization and Allocation Concealment details is still prevalent among published RCTs in 2011: a review. J Eval Clin Pract. 2013; 19(4):703–7. Epub 2013/05/08. https://doi.org/10.1111/jep.12031 PMID: 23648066.
13. Groves T. What makes a high quality clinical research paper? Oral Dis. 2010; 16(4):313–5. Epub 2010/03/18. https://doi.org/10.1111/j.1601-0825.2010.01665.x PMID: 20233312.
14. Mills E, Wu P, Gagnier J, Heels-Ansdell D, Montori VM. An analysis of general medical and specialist journals that endorse CONSORT found that reporting was not enforced consistently. J Clin Epidemiol. 2005; 58(7):662–7. Epub 2005/06/09. https://doi.org/10.1016/j.jclinepi.2005.01.004 PMID: 15939216.
15. Uetani K, Nakayama T, Ikai H, Yonemoto N, Moher D. Quality of reports on randomized controlled trials conducted in Japan: evaluation of adherence to the CONSORT statement. Intern Med. 2009; 48 (5):307–13. Epub 2009/03/03. PMID: 19252352.
16. Gandhi GY, Murad MH, Fujyoshi Å, Mullan RJ, Flynn DN, Elamin MB, et al. Patient-important outcomes in registered diabetes trials. JAMA. 2008; 299(21):2543–9. Epub 2008/06/05. https://doi.org/10.1001/jama.299.21.2543 PMID: 18523223.
17. Gartlehner G, Thieda P, Hansen RA, Morgan LC, Shumate JA, Nissman DB. Inadequate reporting of trials compromises the applicability of systematic reviews. Int J Technol Assess Health Care. 2009; 25(3):323–30. https://doi.org/10.1017/S0266462309990122 PMID: 19619351

18. Montori VM, Guyatt GH. Progress in evidence-based medicine. JAMA. 2008; 300(15):1814–6. https://doi.org/10.1001/jama.300.15.1814 PMID: 18854545

19. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. The Lancet. 374(9683):86–9. https://doi.org/10.1016/S0140-6736(09)60329-9

20. Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. Ann Intern Med. 2001; 135(11):982–9. Epub 2001/12/04. PMID: 11730399.

21. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, et al. Does quality of reports of randomized trials affect estimates of intervention efficacy reported in meta-analyses? Lancet. 1998; 352(9128):609–13. Epub 1998/09/24. https://doi.org/10.1016/S0140-6736(98)01085-X PMID: 9746022.

22. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ. 2009; 339:b2535. Epub 2009/07/23. https://doi.org/10.1136/bmj.b2535 PMID: 19622551; PubMed Central PMCID: PMCPMC2714657.

23. edited by Julian PTH, Sally G. Cochrane handbook for systematic reviews of interventions: Chichester, West Sussex; Hoboken NJ: John Wiley & Sons, [2008] 2008; 2008.

24. InCites Journal Citation Reports. Available from: http://jcr.incites.thomsonreuters.com.

25. Bhandari M, Busse JW, Jackowski D, Montori VM, Schunemann HJ, Sprague S, et al. Association between industry funding and statistically significant pro-industry findings in medical and surgical randomized trials. CMAJ. 2004; 170(4):477–80. Epub 2004/02/19. PMID: 14970094; PubMed Central PMCID: PMCPMC332713.

26. Montori VM, Wang YG, Alonso-Coello P, Bhagra S. Systematic Evaluation of the Quality of Randomized Controlled Trials in Diabetes. Diabetes Care. 2006; 29(8):1833–8. https://doi.org/10.2337/dc06-0077 PMID: 16873788

27. Higgins JPT SJ, Savović J, Page MJ, Hróbjartsson A, Boutron I, Reeves B, Eldridge S. A revised tool for assessing risk of bias in randomized trials. Cochrane Database Syst Rev. 2016;(10). https://doi.org/10.1002/14651858.CD016010

28. Devereaux PJ, Choi PT, El-Dika S, Bhandari M, Montori VM, Schunemann HJ, et al. An observational study found that authors of randomized controlled trials frequently use concealment of randomization and blinding, despite the failure to report these methods. J Clin Epidemiol. 2004; 57(12):1232–6. Epub 2004/12/25. https://doi.org/10.1016/j.jclinepi.2004.03.017 PMID: 15617948.

29. Bridoux V, Moutel G, Roman H, Kianifard B, Michot F, Herve C, et al. Methodological and Ethical Quality of Randomized Controlled Clinical Trials in Gastrointestinal Surgery. J Gastrointest Surg. 2012; 16(9):1758–67. https://doi.org/10.1007/s11605-012-1952-0 PMID: 22777055

30. Chen X, Zhai X, Wang X, Su J, Li M. Methodological reporting quality of randomized controlled trials in three spine journals from 2010 to 2012. Eur Spine J. 2014; 23(8):1606–11. Epub 2014/04/22. https://doi.org/10.1007/s00586-014-3283-1 PMID: 24748442.

31. Lee JW, Chung JH, Jo JK, Lee SW. Assessing the quality of randomized controlled trials published in urology and urodynamics from 1993 to 2012. Neurol Urodyn. 2014; 33(5):472–4. Epub 2013/07/23. https://doi.org/10.1002/nau.22457 PMID: 23868818.

32. Jo JK, Chung JH, Kim KS, Song SH, Lee SW. Reporting of randomized controlled trials in andrology journals: a quality assessment. J Sex Med. 2015; 12(2):350–7. Epub 2014/12/30. https://doi.org/10.1111/jsm.12784 PMID: 25537865.

33. McIner B D S. An overview of randomized controlled trials in non-insulin dependent diabetes mellitus. Diabetologia. 1996; 39(A193).

34. Rios LP, Odueyungbo A, Moitri MO, Rahman MO, Thabane L. Quality of reporting of randomized controlled trials in general endocrinology literature. J Clin Endocrinol Metab. 2008; 93(10):3810–6. Epub 2008/06/28. https://doi.org/10.1210/jc.2008-0817 PMID: 18583463.

35. Brito JP, Domecq JP, Murad MH, Guyatt GH, Montori VM. The Endocrine Society guidelines: when the confidence cart goes before the evidence horse. J Clin Endocrinol Metab. 2013; 98(8):3246–52. Epub 2013/06/21. https://doi.org/10.1210/jc.2013-1814 PMID: 23783104.

36. Grimshaw JM, Russell IT. Achieving health gain through clinical guidelines II: Ensuring guidelines change medical practice. Qual Health Care. 1994; 3(1):45. PMID: 10136260

37. Montori VM, Brito J, Murad M. The optimal practice of evidence-based medicine: Incorporating patient preferences in practice guidelines. JAMA. 2013; 310(23):2503–4. https://doi.org/10.1001/jama.2013.281422 PMID: 24165826
38. Alexander PE, Brito JP, Neumann I, Gionfriddo MR, Bero L, Djulbegovic B, et al. World Health Organization strong recommendations based on low-quality evidence (study quality) are frequent and often inconsistent with GRADE guidance. J Clin Epidemiol. 2016; 72:98–106. Epub 2015/01/27. https://doi.org/10.1016/j.jclinepi.2014.10.011 PMID: 25618534.

39. Alexander PE, Bero L, Montori VM, Brito JP, Stoltzfus R, Djulbegovic B, et al. World Health Organization recommendations are often strong based on low confidence in effect estimates. J Clin Epidemiol. 2014; 67(6):629–34. Epub 2014/01/07. https://doi.org/10.1016/j.jclinepi.2013.09.020 PMID: 24388966.

40. Mahtani KR. 'Spin' in reports of clinical research. Evid Based Med. 2016; 21(6):201–2. Epub 2016/10/16. https://doi.org/10.1136/ebmed-2016-110570 PMID: 27737894.

41. Patient-Centered Outcomes Research Institute (PCORI) Methodology Standards [cited 2018 August, 13]. Available from: https://www.pcori.org/research-results/about-our-research/research-methodology/pcori-methodology-standards.

42. Scott A, Rucklidge JJ, Mulder RT. Is Mandatory Prospective Trial Registration Working to Prevent Publication of Unregistered Trials and Selective Outcome Reporting? An Observational Study of Five Psychiatry Journals That Mandate Prospective Clinical Trial Registration. PLoS One. 2015; 10(8). https://doi.org/10.1371/journal.pone.0133718 PMID: 26287998; PubMed Central PMCID: PMCPMC4546113.

43. Fredrickson MJ, Ilfeld BM. Prospective trial registration for clinical research: what is it, what is it good for, and why do I care? Reg Anesth Pain Med. 2011; 36(6):619–24. Epub 2011/10/13. https://doi.org/10.1097/AAP.0b013e318230fbc4 PMID: 21989152.

44. Zarin DA, Tse T, Williams RJ, Carr S. Trial Reporting in ClinicalTrials.gov—The Final Rule. N Engl J Med. 2016; 375(20):1998–2004. https://doi.org/10.1056/NEJMsr1611785 PMID: 27635471.

45. CONSORT: Transparent Reporting of Trials. Consort—Welcome to the CONSORT. Welcome to the CONSORT. [February 22, 2018]. Available from: http://www.consort-statement.org/.

46. Plint AC, Moher D, Morrison A, Schulz K, Altman DG, Hill C, et al. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. Med J Aust. 2006; 185(5):263–7. Epub 2006/09/05. PMID: 16948622.

47. Singh Ospina N, Rodriguez-Gutierrez R, Brito JP, Young WF Jr., Montori VM. Is the endocrine research pipeline broken? A systematic evaluation of the Endocrine Society clinical practice guidelines and trial registration. BMC Med. 2015; 13:187. Epub 2015/08/13. https://doi.org/10.1186/s12916-015-0435-z PMID: 26265226; PubMed Central PMCID: PMCPMC4533940.