

ORIGINAL ARTICLE

Bias in animal studies of estrogen effects on cardiovascular disease: A systematic review and meta-analysis

Christopher Friis Berntsen MD^{1,2} | Pernille Rootwelt Medical Student³ |
Anders Erik Astrup Dahm MD, PhD^{4,5}  

¹Department of Medicine, Sykehuset Innlandet Hospital Trust, Gjøvik, Norway

²Institute for Health and Society, University of Oslo, Oslo, Norway

³Faculty of Medicine, University of Oslo, Oslo, Norway

⁴Department of Hematology, Akershus University Hospital, Lørenskog, Norway

⁵Institute of Clinical Medicine, University of Oslo, Oslo, Norway

Correspondence

Anders Erik Astrup Dahm, Department of Haematology, Akershus University Hospital, Postbox 1000, 1478 Lørenskog, Norway.
Email: aeadahm@gmail.com

Handling Editor: Suzanne Cannegieter

Abstract

Background: Randomized controlled trials on menopausal hormone therapy in humans have not confirmed the benefit of estrogens on cardiovascular disease found in animal studies. Flawed methodology or publication bias in animal studies may explain the discrepancy.

Objectives: The aim of this study was to investigate whether publication of the randomized controlled trials Heart and Estrogen/Progestin Replacement Study and Women's Health Initiative influenced study authors' assessment of research findings (confirmation bias) as well as to investigate publication bias and small-study effects in animal studies of estrogen effects on atherosclerosis.

Methods: The data source for this study was PubMed from inception to 2018. We selected animal studies with cardiovascular outcomes comparing 17- β -estradiol, its natural metabolites, or conjugated equine estrogen with control. Qualitative data were extracted on authors' conclusions about estrogen effects on cardiovascular disease, as well as quantitative data for atherosclerosis outcomes. Fixed- and random-effects meta-analyses were used. Publication bias/small-study effects were assessed using funnel plots and Egger's regression. Trim-and-fill plots and extrapolation from Egger's regression were used to adjust for publication bias. The main outcomes and measures were the primary study authors' interpretation of their own results, and estrogen effects on cardiovascular disease in general before and after publication of the Women's health Initiative study (2003). The effects of estrogens on atherosclerosis were measured as standardized mean difference between intervention and control.

Results: Of 1925 studies retrieved, 360 were eligible for analyses. Study-specific statements concluded that estrogens were protective against cardiovascular disease in 75% of studies before 2003 and 78% after, but the percentage of general statements about estrogens being cardioprotective changed from 70% to 40%. Meta-analyses showed less atherosclerosis in estrogen-treated animals. Extremely skewed funnel plots and $P < .01$ in Egger's regression suggested publication bias and/or exaggerated effects in small studies, which was more pronounced after 2002. There was substantial heterogeneity of effects ($I^2 = 68\%$ -86%) overall and in all subgroups

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. *Research and Practice in Thrombosis and Haemostasis* published by Wiley Periodicals LLC on behalf of International Society on Thrombosis and Haemostasis (ISTH)

except cynomolgus monkeys ($I^2 = 9\%$), the only animal subgroup without clear bias. Adjusting for publication bias, overall estimates of estrogen effects on atherosclerosis were close to null effect.

Conclusions: We found substantial evidence of publication bias but not of confirmation bias. Publication bias and flawed small studies may partly explain why findings differ between animal studies and clinical trials on the effect of estrogens on cardiovascular disease.

KEYWORDS

animals, cardiovascular disease, estrogens, publication bias, systematic review

Essentials

- Studies on animals show different effects of estrogens on cardiovascular disease than in humans.
- We reviewed studies of estrogen-treated animals with cardiovascular outcomes.
- We found strong indications of publication bias in these animal studies.
- This may partly explain the difference in estrogen effects between animal studies and clinical trials.

1 | INTRODUCTION

Before 2002, menopausal hormonal therapy (MHT) was recommended for prevention of coronary heart disease, based on the results of observational studies, in particular the Nurses' Health Study.¹⁻³ The observational studies showed that women using MHT had reduced risk of coronary heart disease compared with women not using MHT, and they were largely corroborated by animal studies that confirmed the positive effects of estrogen on the vascular bed.⁴⁻⁷

Controversy ensued in 1998 as the first randomized clinical trial (RCT) on MHT, the Heart and Estrogen/Progestin Replacement Study (HERS), failed to show that MHT protected against coronary heart disease.⁸ In 2002 and 2004, the Women's Health Initiative (WHI) studies^{9,10} confirmed the results from HERS. Moreover, the latter two were discontinued due to excess risk of cardiovascular disease in the MHT group. The WHI findings prompted a sharp decline in MHT use and left researchers struggling to explain why results from RCTs differed from observational studies and laboratory studies. Some researchers continued to argue that MHT protected against coronary heart disease, especially when it was started shortly after menopause, while others put more confidence in the RCTs and argued that MHT had no or very little effect on coronary heart disease.^{4,11,12} Additional publications from the WHI study and the Nurses' Health Study have shed further light on the difference between RCTs and observational studies, although a unifying conclusion has not been reached.^{11,13}

One of the main arguments against the HERS and WHI results was that most laboratory studies on atherosclerosis and vascular disease showed beneficial results of estrogens.^{14,15} Interventions showing promising results in animals have often proven ineffective or even harmful in subsequent clinical trials.¹⁶⁻¹⁸ Possible mechanisms include that inherent differences between species may impair

generalizability to humans^{16,17} and that many animal studies lack methodological rigor.^{16,17,19} Failure to publish negative results may further contribute to a flawed body of evidence.^{16-18,20}

In our study, we investigated two hypotheses regarding the diverging results in animal studies and RCTs on estrogens and cardiovascular disease. First, we hypothesized that confirmation bias may have led authors of animal studies to interpret findings on estrogen effects differently before and after the WHI study was published. We hence investigated if study conclusions were in alignment with the predominant scientific understanding at the time of publication. Second, we hypothesized the presence of publication bias, meaning that studies finding negative or neutral effects of estrogens remain unpublished.

2 | METHODS

2.1 | Literature search

We searched PubMed from inception through January 17, 2018, using relevant free-text and MeSH terms for applicable interventions combined with terms for cardiovascular outcomes. We applied a previously published filter for retrieving animal studies.²¹ The search strategy is detailed in Appendix A in the Supporting Information. Records were deduplicated using EndNote X8.2²² and the deduplication algorithm in the online literature screening tool Covidence.²³

2.2 | Study selection

We included primary reports of in vivo animal studies providing data on the effect of estrogens on any cardiovascular outcome, restricting

interventions to 17- β -estradiol and its natural metabolites or conjugated equine estrogen, administered by any route. Eligible studies must either have a time-series design or an intervention and a control group. The two groups should differ only by the presence or absence of estrogen administration. We applied no restrictions regarding animal species, sex, or age.

Two authors (CFB and PR) screened all abstracts retrieved in the search independently and in duplicate using Covidence and scrutinized records deemed relevant in the primary screening for eligibility in full-text. Disagreements were solved by discussion. For full-text reports not available through our institution, we retrieved paper copies through the medical library of University of Oslo and its collaborating institutions. We used Google Translate²⁴ for assessing reports in languages that neither of the screeners knew (ie, Chinese and Japanese).

2.3 | Outcomes and data extraction

Using a standardized data extraction sheet, two authors (CFB and PR) independently extracted data from the included studies for the prespecified outcomes described below. Disagreements were resolved by discussion. The extraction sheet was piloted on a small subset of the data set and did not need any adjustments.

From each included article, we extracted qualitative information about whether authors made general statements about the effect of exogenous estrogens on cardiovascular risk, not limited to the scope of their study. General statements could be either extrapolations to other populations (including effect of MHT in humans), references to “well-known” or established effects on cardiovascular risk, explicit acknowledgment that evidence across studies is contradictory or inconclusive, or explicit assessments about the probable net direction of effect of estrogens across studies. If the estrogen-treated animals were included as a positive control in multiarm trials, we would consider this an implicit statement about beneficial cardiovascular effects. In addition, we also extracted qualitative information about the authors’ statements about the effect of estrogens within their own study (study specific statements), ie, the authors’ interpretation of their own results. All statements were categorized as follows: “protective” (exogenous estrogens reduce cardiovascular risk), “detrimental” (exogenous estrogens increase cardiovascular risk), “neutral” (exogenous estrogens do not affect cardiovascular risk), “ambivalent” (exogenous estrogens have both protective and harmful effects on cardiovascular disease) or “unclear” (we could not identify a clear conclusion about the effects of estrogen on cardiovascular risk).

For analysis of treatment effects in the studies and publication bias, we assessed whether each study contained information for a continuous outcome measuring the development of atherosclerosis. We chose this outcome because piloting indicated this to be the most commonly reported outcome, and one of the hypothesized effects of MHT before the HERS and WHI studies were published was that it reduced the risk of coronary heart disease in postmenopausal

women. Where data on atherosclerosis were reported for both intervention and control groups with appropriate measures of uncertainty, we extracted (i) the mean value for the outcome in each group; (ii) either the standard deviation, standard error of the mean, or confidence interval (CI) for the mean in each group (in order of preference); and (iii) the number of animals in each group. We extracted data for only one outcome per study.

As many studies report on multiple similar outcomes, we had prespecified a heuristic to choose between outcomes that were otherwise deemed equally relevant to the development of atherosclerosis:

1. We primarily extracted data for the outcome presented as the main or most important analysis by study authors.
2. If this was unclear, we extracted the first eligible outcome for which extractable data were presented in the text (either inline or in a table).
3. If numeric data were not presented in writing, we extracted data for approach 1 or 2 by reading values off graphs where available.
4. If data for several doses of estrogens were available for one study, we extracted data for the median dose, resorting to the lowest dose above the median if an even number of doses was tested.

If sufficient information was not identified in the report, the study was not included in the quantitative analysis of treatment effects. Eligibility for inclusion in this analysis was assessed independently by two authors (CFB and PR), and disagreements were resolved by discussion. Numerical accuracy of extracted data points was scrutinized by the reviewers together, and disagreements were resolved by joint reexamination of the study.

2.4 | Statistical analysis

For both the authors’ general statements about estrogen effects on cardiovascular risk and their statements about their own results, we computed the proportions of studies in each of the five categories. Studies were stratified by whether they were published in 2002 or earlier, or in 2003 or later. We chose this cutoff point because we believed studies published in 2003 or later could reasonably be expected to take the results of the WHI study into account. We also dichotomized the outcome by binning statements into positive and nonpositive, that is, whether estrogens were regarded as protective versus all other categories, and recomputed the proportions. For the two-by-two tables produced by dichotomizing both time of publication and assessments about outcome, we performed a χ^2 test to investigate whether author statements about estrogen effects changed after publication of the first WHI study.

As a supplementary post hoc analysis, we computed Cramer’s V for multilevel categorical data to assess how general statements correlated with study-specific conclusions, with bootstrap CIs.

For the quantitative analysis of the effect of exogenous estrogens on the development of atherosclerosis, we performed meta-analyses applying two alternative models: one fixed-effect

and one random-effects model (DerSimonian-Laird method). Studies were weighted with a generic inverse variance approach. The random-effects model is generally more conservative in terms of precision if there is large heterogeneity in treatment effects across studies. However, as it attributes comparatively more weight to smaller studies, it may be more sensitive to bias if small-study effects (ie, bias causing exaggerated effects due to lack of methodological rigor in small studies²⁵) are present. As we were investigating effects across species where one cannot expect one common fixed effect, we report the random-effects model as our primary model and give additional fixed-effects estimates in the appendix. As post hoc analyses, we additionally performed subgroup analyses per animal species and by general and study-specific statements about estrogen effects (dichotomized as protective vs all other groups).

Atherosclerosis outcomes are reported on a plethora of scales and using different approaches to quantify atherosclerosis. We therefore used the standardized mean difference (SMD) as the meta-analytic outcome measure, applying the Hedges' *g* estimator for the standardization.²⁶ When standard deviations of treatment group means were not found in study reports, we computed these values from the corresponding standard errors or confidence intervals and the group sizes. In a single study, a SMD of 1.0 means that the mean amount of atherosclerosis in the intervention group is shifted by one standard deviation compared to the control group, on whichever scale was used in the study.

We computed I^2 as a measure of statistical heterogeneity in all meta-analyses, and *P* values for the significance of heterogeneity. An I^2 of 0% indicates no heterogeneity. Higher numbers mean that variation in treatment effects across studies is less likely to be caused by sampling error alone. Values between 50% and 100% are commonly interpreted as indicative of important study-level differences, such as differing populations, impacting between-study variation of effect estimates.²⁷

To test for small-study effects and publication bias, we visually inspected funnel plots and applied Egger's regression test,^{28,29} considering a $P < 0.1$ as indicative of bias. If probable bias was detected, we performed trim-and-fill analyses, a nonparametric method imputing effects of presumed missing studies to reduce the skewness of the funnel plot and obtain adjusted treatment effect estimates.³⁰ We also obtained adjusted meta-analytic estimates from the Egger analysis, which corresponds to a weighted regression of study effects against their standard errors.³¹ The regression line intercept at zero standard variation may be construed as an estimate of the underlying mean effect if publication bias were not present.^{31,32} We performed separate analyses for publication bias in all subgroups with ≥ 10 studies; applying such tests with too few studies is not advisable both due to lack of power and the risk of spurious findings.³² Tests for small-study effects and publication bias may be at risk of type I errors when SMD is used as the outcome measure, particularly if groups are small.³³ We hence performed a sensitivity analysis using a normalized mean difference (NMD) as an alternative outcome measure in these tests. To obtain NMDs, individual study

outcomes were normalized by computing the raw mean difference between groups as a fraction of the control group mean,³⁴ that is, as percentage reduction or increase in atherosclerosis compared to control.

We used R version 3.6.0³⁵ with the libraries meta,³⁶ metafor,³⁷ and ggplot2³⁸ for all analyses and plots.

Search strategies, inclusion and exclusion criteria, outcomes, methods for data extraction and analyses were pre-planned in a protocol available from the authors upon request. The analyses for adjusting meta-analytic estimates for publication bias and one subgroup analyses (by animal species) were conducted post-hoc in addition to the pre-specified analyses.

3 | RESULTS

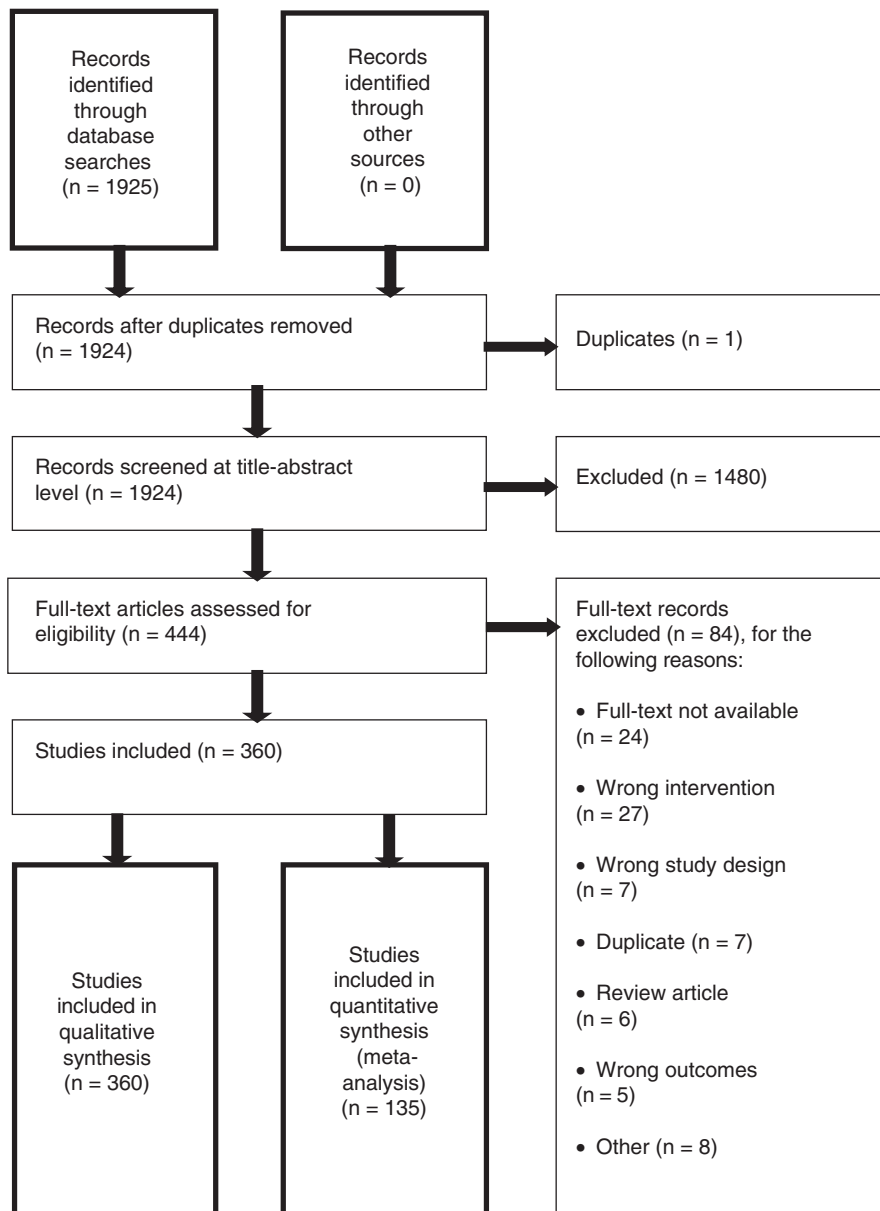
3.1 | Study conclusions on cardiovascular disease from animal studies before and after the WHI trial

A total of 1925 studies were retrieved in the initial search; 360 were included for the qualitative analysis of report conclusions, of which 135 also had data available for quantitative synthesis (included and excluded studies are listed in Appendices B and C in the Supporting Information). Study flow is illustrated in Figure 1.

Of the 360 studies included, 9 were published from 1950 to 1959, 9 from 1960 to 1969, 8 from 1970 to 1979, 18 from 1980 to 1989, 79 from 1990 to 1999, 152 from 2000 to 2009, and 85 from 2010 to 2018. The number of published studies stratified per decade is shown in Figure S1, where also the years of the publication of the Nurses' Health Study (1991) and the WHI study (2002) are marked.

The proportions of study-specific statements concluding that estrogen was effective, detrimental or had a neutral effect in preventing cardiovascular disease, as well as those being ambivalent or unclear about whether it was effective, are shown in Figure 2. Of the 360 studies included, 276 (77%; 95% CI, 72%-81%) concluded that estrogens had a protective effect against cardiovascular diseases in their experiments. We found no evidence that researchers concluded differently about their own findings before or after publication of the first WHI study (75%; 95% CI, 67%-81%), "protective" before WHI compared with 78% (95% CI, 71%-83%) after WHI, right panel Figure 2). The χ^2 test when dichotomizing into protective versus all other categories showed no indication of a difference ($P = .62$) in the proportion of study findings reported as positive. However, the authors' extrapolations to general statements about the properties of exogenous estrogens changed markedly: 70% (95% CI, 63%-76%) stated it to be cardioprotective before WHI, while 40% (95% CI, 34%-48%) considered it protective after. In Table 1, we show a cross-tabulation of general statements with study-specific statements for the entire period and for the periods before and after WHI. The correlation between study-specific statements and general statements was strong for the period before WHI (Cramer's *V*, 0.47; 95% CI,

FIGURE 1 Study flow for the inclusion and exclusion of studies



0.34-0.61), whereas it dropped in the period after WHI (Cramer's V , 0.25; 95% CI, 0.15-0.42). For the entire period, there was moderate to strong correlation between statements (Cramer's V , 0.37; 95% CI, 0.26-0.48).

3.2 | Effect of estrogens on atherosclerosis in animal studies

Results from the meta-analyses of the effect of exogenous estrogens on the development of atherosclerosis in animal studies are summarized in Figure 3 (full forest plots including individual study effects and fixed-effect model estimates are provided in Appendices D and G in the Supporting Information). Overall, a reduction in atherosclerosis was detected (SMD in the random-effects model -1.25 ; 95% CI -1.42 to -1.07). The SMD was -0.99 (95% CI, -1.21 to -0.78 in the

random-effects model) before 2003 versus -1.63 (95% CI, -1.91 to -1.36 in the random-effects model) after 2003. Thus, the effect of estrogens appeared to be weaker before 2003 compared with after 2003 ($P < .001$). There was substantial heterogeneity of treatment effects, both overall ($I^2 = 76\%$), before WHI ($I^2 = 77\%$) and after WHI ($I^2 = 68\%$) (all $P < .01$ for significant heterogeneity) (Figure 3).

3.3 | Treatment effects of estrogens for different species

Due to the large heterogeneity of treatment effects, we performed an additional subgroup analysis to investigate whether treatment effect differed by animal species (Figure 4, full forest plot and fixed-effects model estimates in Appendices D and G in the Supporting Information). Effects did vary across species ($P < .01$). Seven of the

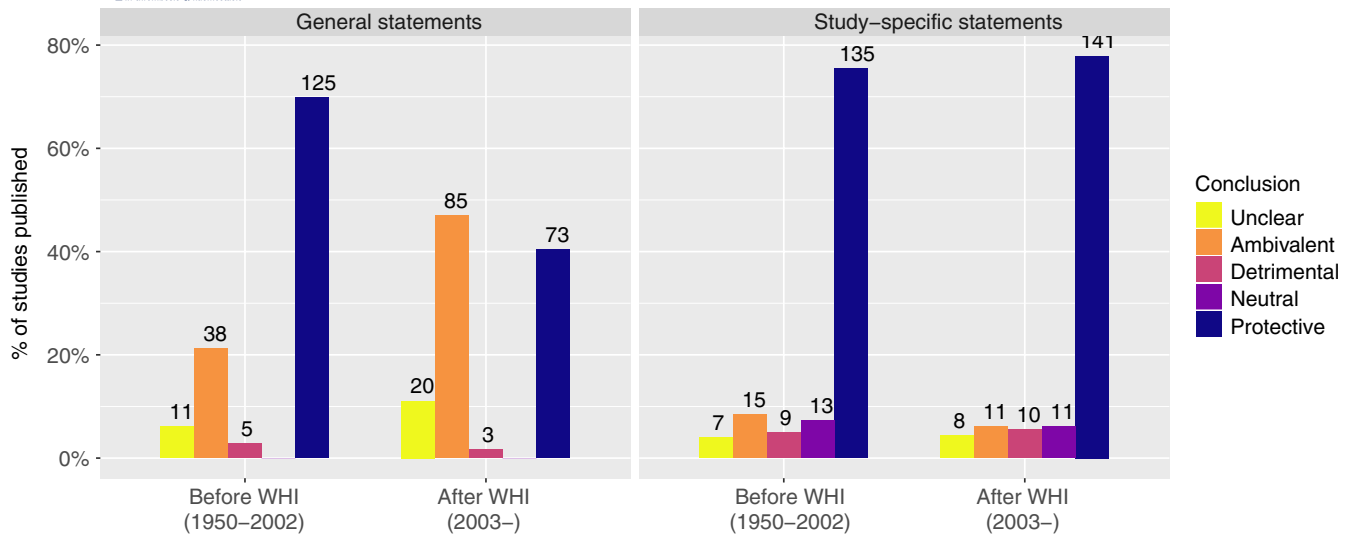


FIGURE 2 Study authors' statements about whether estrogen therapy protects against cardiovascular disease or not. The left panel shows general statements made by authors in their reports on the effect of estrogens on cardiovascular risk. The right panel shows how authors concluded about their own study findings in published study reports. The absolute numbers of reports in each category are given above the bars. WHI, Women's Health Initiative

TABLE 1 Authors' general statements on the effect of menopausal hormone therapy on cardiovascular risk compared with study-specific statements on the effect of estrogens in their own experiments

	Study-specific statements					Total
	Ambivalent	Detrimental	Neutral	Protective	Unclear	
General statements, n (%)						
All studies						
Ambivalent	15 (4.2)	12 (3.3)	9 (2.5)	83 (23.1)	4 (1.1)	123 (34.2%)
Detrimental	0 (0.0)	7 (1.9)	0 (0.0)	1 (0.3)	0 (0.0)	8 (2.2)
Protective	8 (2.2)	0 (0.0)	11 (3.1)	174 (48.3)	5 (1.4)	198 (55.0)
Unclear	3 (0.8)	0 (0.0)	4 (1.1)	18 (5.0)	6 (1.7)	31 (8.6)
Total	26 (7.2)	19 (5.3)	24 (6.7)	276 (76.7)	15 (4.2)	360 (100.0)
Before WHI (1950–2002)						
Ambivalent	6 (3.4)	4 (2.2)	3 (1.7)	24 (13.4)	1 (0.6)	38 (21.2)
Detrimental	0 (0.0)	5 (2.8)	0 (0.0)	0 (0.0)	0 (0.0)	5 (2.8)
Protective	7 (3.9)	0 (0.0)	9 (5.0)	106 (59.2)	3 (1.7)	125 (69.8)
Unclear	2 (1.1)	0 (0.0)	1 (0.6)	5 (2.8)	3 (1.7)	11 (6.1)
Total	15 (8.4)	9 (5.0)	13 (7.3)	135 (75.4)	7 (3.9)	179 (100.0)
After WHI (2003)						
Ambivalent	9 (5.0)	8 (4.4)	6 (3.3)	59 (32.6)	3 (1.7)	85 (47.0)
Detrimental	0 (0.0)	2 (1.1)	0 (0.0)	1 (0.6)	0 (0.0)	3 (1.7)
Protective	1 (0.6)	0 (0.0)	2 (1.1)	68 (37.6)	2 (1.1)	73 (40.3)
Unclear	1 (0.6)	0 (0.0)	3 (1.7)	13 (7.2)	3 (1.7)	20 (11.0)
Total	11 (6.1)	10 (5.5)	11 (6.1)	141 (77.9)	8 (4.4)	181 (100.0)

WHI, Women's Health Initiative.

subgroups were largely uninformative, containing only one to three studies. Moreover, six of these comprised <25 animals.

Notably, we found consistent effect estimates without significant heterogeneity ($I^2 = 9\%$; $P = .34$) in studies on cynomolgus

monkeys, whereas within-group heterogeneity remained high for studies on all other species with more than one study ($I^2 = 68\%–86\%$; $P < .01$ for all). The estimate for cynomolgus monkeys suggests a small protective effect of estrogens on atherosclerosis is present

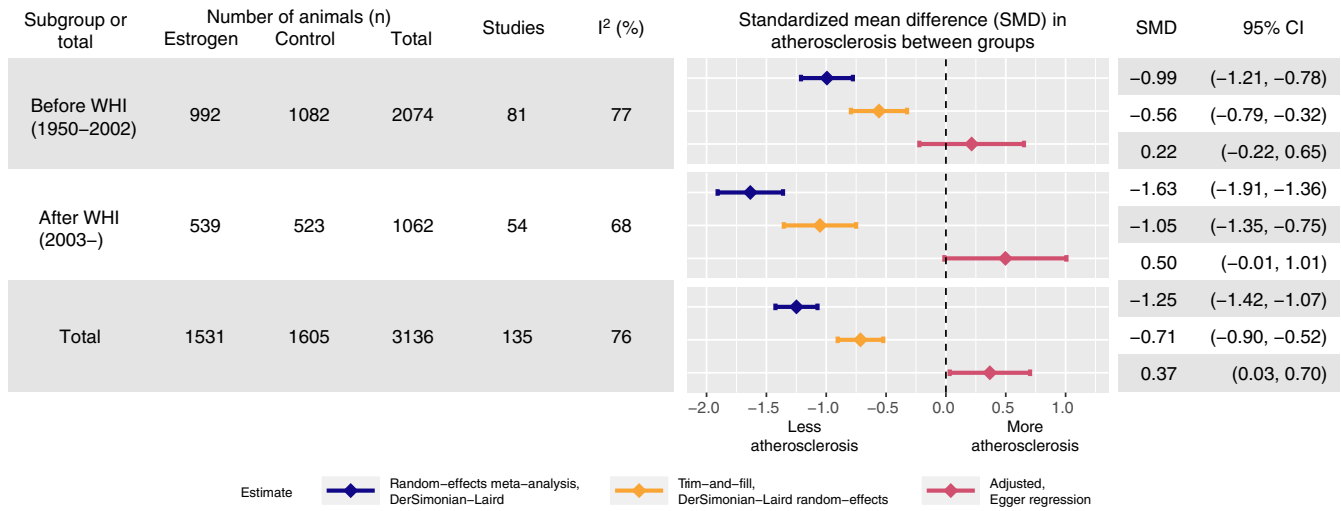


FIGURE 3 Pooled effect estimates of exogenous estrogen on the development of atherosclerosis in animal models. Horizontal lines represent 95% confidence intervals (CI). Negative standardized mean differences (SMD; left side of vertical dotted line) mean less atherosclerosis in the estrogen group. The trim-and-fill analyses and estimates adjusted using Egger regressions provide estimates corrected for the impact of publication bias and exaggerated effects in small studies. WHI, Women’s Health Initiative

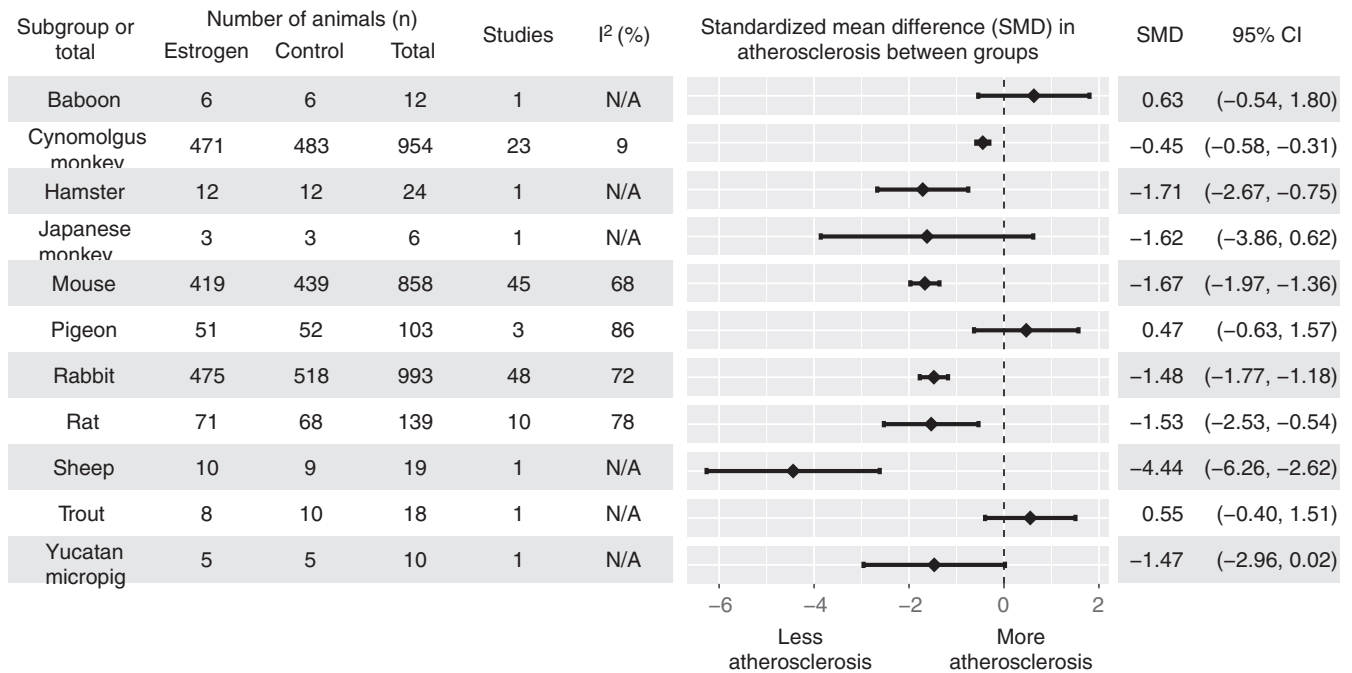


FIGURE 4 Subgroup analysis of treatment effect by species. Horizontal lines represent 95% confidence intervals (CI). Negative standardized mean differences (SMD; left side of vertical dotted line) mean less atherosclerosis in the estrogen group. I² is a measure of statistical heterogeneity; higher numbers mean more heterogeneity (not applicable for subgroups with only one study)

in this group (SMD, -0.45; 95% CI, -0.58 to -0.32 in random-effects model).

3.4 | Treatment effects according to general and study-specific statements about estrogen effects

There were no difference in standardized mean effects between studies where general statements indicated that estrogen had

protective effects and where they did not (SMD in random-effects model, -1.29; 95% CI, -1.50 to -1.09 vs -1.16; 95% CI, -1.48 to -0.85, Appendix D in the Supporting Information). A comparison between the study-specific conclusion of protective effects with all other effects found that studies concluding with protective effects from their own results indeed found less atherosclerosis than studies concluding with other effects (SMD in random-effects model, -1.45; 95% CI, -1.64 to -1.26 vs -0.22; 95% CI, -0.60 to 0.16; Appendix D in the Supporting Information).

3.5 | Analysis of publication bias and small-study effects

The analyses for small-study effects strongly indicated publication bias in the overall body of animal studies ($P < 10^{-12}$ in Egger's test), as well as in other subgroups with 10 studies or more (before WHI: $P < 10^{-4}$, after WHI: $P < 10^{-8}$, mice: $P < 0.01$, rabbits: $P < 10^{-7}$) except cynomolgus monkeys ($P = .7$) and the small group of rat studies ($P = .97$). The tests may also indicate the presence of exaggerated effect estimates in small studies. The funnel plots in Figure 5 show gross pronounced skewness, with a clear sparsity of studies on the lower right side of the plots for the total cohort (upper left) and all subgroup analyses except for cynomolgus monkeys (lower left). Many studies fall outside the area of expected variability in effects, most of them showing large effects. For rabbits, however, skewness was not significant in the sensitivity analysis using NMD outcomes (see Figure S2).

An approximation of treatment effect estimates corrected for bias may be obtained by observing where the Egger regression lines intersect the x axis at the top of the plots in Figure 5.³¹ Notably, all intersect near null effect. Applying this method, the effects of estrogens shift from protective to slightly detrimental (estimates shown in Figure 4). The trim-and-fill analyses imputing the effect of unpublished studies also shifted the SMDs considerably toward null effect for the main analyses (Figure 4). Plots of the trim-and-fill analyses are given in Appendices D and E in the Supporting Information, and results for Egger regression sensitivity analyses using NMD outcomes are given in Appendix F.

4 | DISCUSSION

Our study confirmed that most authors of animal studies conclude that the results of their experiments show that estrogens are beneficial for cardiovascular disease. We investigated two explanations for the discrepancies between results from animal studies and RCTs, that is, confirmation bias and publication bias. The results showed that publication bias is a possible explanation, probably in combination with exaggerated effects in small studies.

To our knowledge, this is the first study to systematically appraise animal studies assessing the effects of estrogen on cardiovascular diseases. Small animal studies have shown to be prone to lack of rigor in other areas.^{16,17,19,39} As most identified studies were small, we believe that the heterogeneity observed is mainly caused by exaggerated effects in small studies.^{40,41} Laboratory studies have an inherent tendency for bias since experimental conditions can be adapted to what is assumed to be the "correct" result. Whereas the scientific community and regulatory authorities have invested large resources to improve the conduct and reporting of clinical studies, less attention has been devoted to the basic sciences. Preclinical researchers may consider that a negative result is a botched experiment, although it could in fact, if results are published, constitute a valuable scientific contribution.

We hypothesized that confirmation bias within the predominant scientific paradigm might have affected the study authors' conclusions about estrogenic effects on cardiovascular risk to become less positive after the publication of the first WHI study. Somewhat surprisingly, we found that the studies published after WHI showed stronger

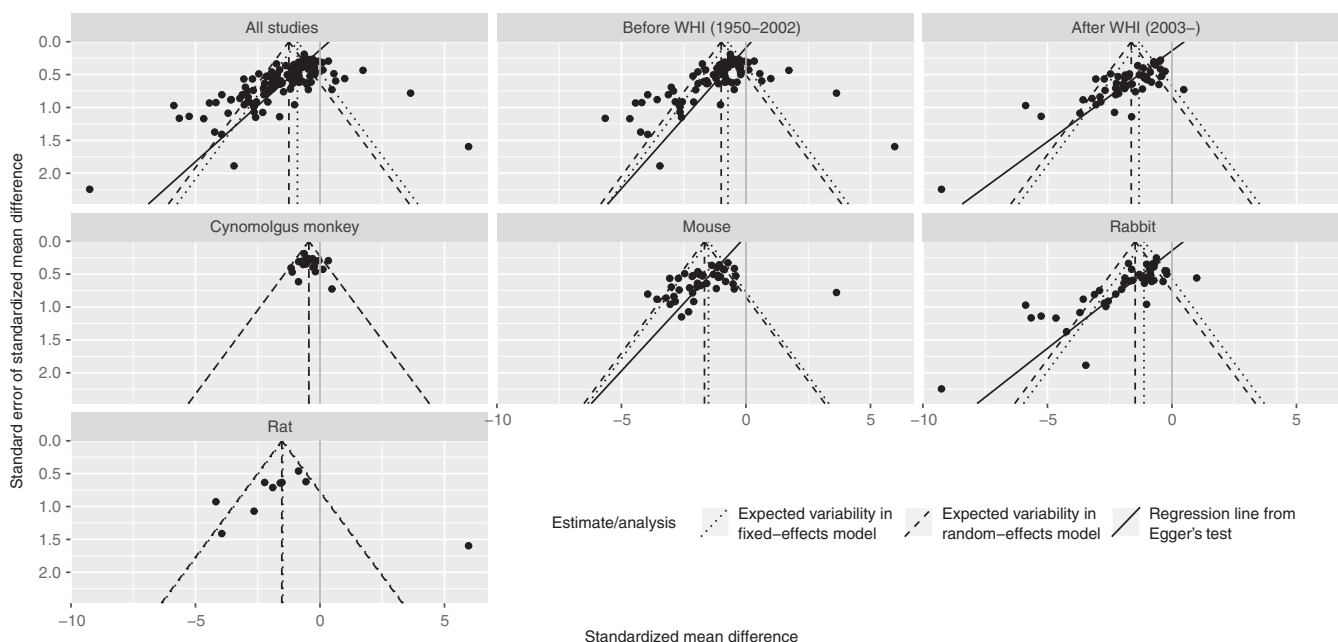


FIGURE 5 Funnel plots for the effect of exogenous estrogens on the development of atherosclerosis. If meta-analytic estimates (vertical dotted lines) are unbiased, an even spread of studies is expected within the triangular areas representing the treatment effect variability anticipated to arise by chance across studies. Larger studies with smaller standard errors appear toward the top. Regression lines from Egger's test are shown whenever significant skewness was detected. The regression line is not drawn for the cynomolgus monkey studies as no signs of publication bias were present in this group. WHI, Women's Health Initiative

cardioprotective effects of estrogens than before, with even more pronounced signs of publication bias. This was despite the fact that study authors seemed more cautious in their general statements regarding protective effects of estrogens on cardiovascular diseases after WHI. It is difficult to explain this observation. One possibility is that it was harder to publish ambivalent or neutral results compared with protective effects, even if such results were more in line with the WHI results.

Publication bias tests must be interpreted with caution in the presence of large heterogeneity,^{40,41} and also when SMDs are used to pool studies.³³ The strength of the correlation between standard errors (a proxy for study size) and treatment effects in the overall Egger regression analysis in our study is, however, considerable. The correlation is also consistent across all subgroup analyses but one. We hence find it improbable that our main findings of publication bias and probable small study effects are spurious or significantly impacted by other sources of heterogeneity. It is, however, important to emphasize that our study does not rule out a true effect in animal models, and a plausible effect is present in the studies of cynomolgus monkeys, without signs of publication bias. A sensitivity analysis using NMD also failed to demonstrate significant skewness for rabbit studies, indicating that findings for this subgroup are less certain.

I^2 was high in all animal studies, except cynomolgus monkeys, suggesting large heterogeneity between study results. Possible explanations include variation in outcome measures or differences in design that is difficult to quantify, “nonobserved confounders.” Publication bias and small-study effects will also contribute to a high I^2 . It can sometimes be difficult to separate true heterogeneity from such effects. In the current article, however, with the skewed funnel-plots and the association between study size and effect suggested by Egger’s test, it is likely that publication bias causes a large part of the I^2 .

Many obstacles preclude generalization of findings from animals to humans. The models may inherently lack external validity, extrapolation to humans may be wrong due to differences between species, or the clinical studies may have shortcomings. However, a Cochrane review from 2015 found that most RCTs were of good methodological quality,⁴² and we believe major bias in the clinical study results to be less likely.

Whether the populations included in the early RCTs on MHT were the ideal target population has been debated.^{11,14,42–44} Indeed, subgroup analyses in the Cochrane review provide moderate-quality evidence that MHT started within 10 years of menopause may reduce the overall risk of death (relative risk [RR], 0.70; 95% CI, 0.52–0.95) and cardiovascular death (RR, 0.52; 95% CI, 0.29–0.96), at the expense of an increased stroke rate (RR, 1.37; 95% CI, 0.80–2.34).⁴² Considering the relatively low baseline risk of these events in the younger postmenopausal cohort, this still translates to small differences in absolute risk. Based on five studies^{9,10,45–47} with a median follow-up of 7.1 years (range, 3.4–10.1), the review authors suggest approximately six fewer deaths and four more strokes per 1000 women treated with estrogen for the duration of follow-up. Moreover, the estimates for effect on mortality were significantly

impacted by the inclusion of one study⁴⁵ deemed to be at high risk of bias.

Nevertheless, an effect in humans may be present. Considering our analyses indicating that treatment effects in published animal studies probably often provide biased and exaggerated estimates, the real discrepancy between findings in animals and humans does not seem inexplicably large. In fact, our analyses adjusted for possible publication bias suggest that the effect of estrogens on cardiovascular diseases may be small or close to null in animals as well.

5 | LIMITATIONS

The literature search was pragmatically limited to one database; hence, we may have missed some studies not indexed there. Given the large number of studies identified, we believe this would not impact significantly on our overall conclusions. Furthermore, our analysis of the study authors’ conclusions is inherently dependent on the reviewers’ discretion and interpretations, and hence at risk of misclassification bias. Although all assessments were performed in duplicate to mitigate this risk, we cannot be sure that we have not occasionally misinterpreted the study authors.

Our methods for assessing the impact of presumed unpublished studies on the meta-analytic estimates inherently involve extrapolating beyond the study data at hand. We do not know the true missing values, and the statistical corrections cannot be taken as providing certain adjusted estimates. However, the consistency of the findings indicates that the failure to publish negative or neutral studies decisively impacts the overall body of evidence from animal studies.

For Chinese and Japanese language studies, we relied on Google translate for abstracting data. Although a manual translation would arguably have been better, a recent study indicates this tool may now have reached sufficient accuracy for abstracting data for systematic reviews.⁴⁸

6 | CONCLUSIONS

Most publications on estrogen-treated animals interpret their own results as a beneficial effect of estrogens on cardiovascular outcomes. Most publications on estrogen-treated animals also contain general statements about the protective effects of estrogens on cardiovascular disease, although such statements became rarer after the publication of the WHI study at the same time as ambivalent general statements became more frequent. However, the change in general statements about estrogens and cardiovascular disease did not influence the authors’ interpretations of their results. The majority of studies of estrogen on atherosclerosis in animals showed indirect signs of large publication bias favoring results that showed less atherosclerosis in estrogen-treated animals. An important exception were the studies on cynomolgus monkeys. Our results indicate that the difference between animal studies and human studies on the

effect of estrogens on cardiovascular disease is perhaps less than previously assumed.

RELATIONSHIP DISCLOSURE

CFB and PR report no conflicts of interest. AEAD reports no conflicts of interests involving the work under consideration for publication but has the following conflicts of interests outside the submitted work: honoraria or payments from Pfizer, Bristol-Myers Squibb, Novartis, Bayer, and MSD, as well as research support from Pfizer Norway.

AUTHOR CONTRIBUTIONS

CFB performed the literature search, screened the abstracts, extracted the data, did the statistical analyses, made the first article draft, and approved the submitted manuscript. PR performed the literature search, screened the abstracts, extracted the data, wrote the article, and approved the submitted manuscript. AEAD had the idea for the study, did the literature search, wrote the article, and approved the submitted manuscript.

ORCID

Anders Erik Astrup Dahm  <https://orcid.org/0000-0003-4477-3526>

TWITTER

Anders Erik Astrup Dahm  @coagulationAD

REFERENCES

- Stampfer MJ, Colditz GA, Willett WC, et al. Postmenopausal estrogen therapy and cardiovascular disease: ten-year follow-up from the Nurses' Health Study. *N Engl J Med*. 1991;325(11):756–762.
- Grady D, Rubin SM, Petitti DB, et al. Hormone therapy to prevent disease and prolong life in postmenopausal women. *Ann Intern Med*. 1992;117(12):1016–1037.
- Barrett-Connor E, Grady D. Hormone replacement therapy, heart disease, and other considerations. *Annu Rev Public Health*. 1998;19:55–72.
- Kim KH, Rapid BJR. Estrogen receptor-mediated signaling: why is the endothelium so special? *Sci STKE*. 2005;2005(288):pe28.
- Mendelsohn ME, Karas RH. The protective effects of estrogen on the cardiovascular system. *N Engl J Med*. 1999;340(23):1801–1811.
- Mendelsohn ME, Karas RH. Estrogen and the blood vessel wall. *Curr Opin Cardiol*. 1994;9(5):619–626.
- Farhat M, Lavigne M, Ramwell P. The vascular protective effects of estrogen. *FASEB J*. 1996;10(5):615–624.
- Hulley S, Grady D, Bush T, et al. Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. *JAMA*. 1998;280(7):605–613.
- Writing Group for the Women's Health Initiative I. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA*. 2002;288(3):321–333.
- Anderson GL, Limacher M, Assaf AR, et al. Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: the Women's Health Initiative randomized controlled trial. *JAMA*. 2004;291(14):1701–1712.
- Miller VM, Harman SM. An update on hormone therapy in postmenopausal women: mini-review for the basic scientist. *Am J Physiol Heart Circ Physiol*. 2017;313(5):H1013–H1021.
- Usselman CW, Stachenfeld NS, Bender JR. The molecular actions of oestrogen in the regulation of vascular health. *Exp Physiol*. 2016;101(3):356–361.
- Rossouw JE. Reconciling the divergent findings from clinical trials and observational studies of menopausal hormone therapy for prevention of coronary heart disease. *Semin Reprod Med*. 2014;32(6):426–432.
- Mendelsohn ME, Karas RH. HRT and the young at heart. *N Engl J Med*. 2007;356(25):2639–2641.
- Hulley S, Grady D. Postmenopausal hormone treatment. *JAMA*. 2009;301(23):2493–2495.
- Perel P, Roberts I, Sena E, et al. Comparison of treatment effects between animal experiments and clinical trials: systematic review. *BMJ*. 2007;334(7586):197.
- van der Worp HB, Howells DW, Sena ES, et al. Can animal models of disease reliably inform human studies? *PLoS Med*. 2010;7(3):e1000245.
- Sena ES, van der Worp HB, Bath PM, Howells DW, Macleod MR. Publication bias in reports of animal stroke studies leads to major overstatement of efficacy. *PLoS Biol*. 2010;8(3):e1000344.
- Sena E, van der Worp HB, Howells D, Macleod M. How can we improve the pre-clinical development of drugs for stroke? *Trends Neurosci*. 2007;30(9):433–439.
- Anglemyer AT, Krauth D, Bero L. Industry sponsorship and publication bias among animal studies evaluating the effects of statins on atherosclerosis and bone outcomes: a meta-analysis. *BMC Med Res Methodol*. 2015;15:12.
- Hooijmans CR, Tillema A, Leenaars M, Ritskes-Hoitinga M. Enhancing search efficiency by means of a search filter for finding all studies on animal experimentation in PubMed. *Lab Anim*. 2010;44(3):170–175.
- EndNote. *EndNote* (X8.2 ed). Philadelphia, PA: Clarivate Analytics; 2018.
- Veritas Health Innovation. Covidence Systematic Review Software. Melbourne, Australia: Veritas Health Innovation. www.covidence.org. Accessed January 17, 2018.
- Google LLC. *Google Translate for iOS*. 6.12.59202 ed. Mountain View, CA: Google LLC; 2020.
- Sterne JA, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *J Clin Epidemiol*. 2000;53(11):1119–1129.
- Hedges LV. Distribution theory for Glass's estimator of effect size and related estimators. *J Educ Stat*. 1981;6(2):107–128.
- Deeks JJ, Higgins JP, Altman DG. Analysing data and undertaking meta-analyses. In: *Cochrane Handbook for Systematic Reviews of Interventions* [Internet]. The Cochrane Collaboration. 5.1.0 (updated March 2011); 2011. www.handbook.cochrane.org
- Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. 1997;315(7109):629–634.
- Sterne JA, Egger M, Smith GD. Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis. *BMJ*. 2001;323(7304):101–105.
- Duval S, Tweedie R. A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *J Am Stat Assoc*. 2000;95(449):89–98.
- Moreno SG, Sutton AJ, Ades A, et al. Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study. *BMC Med Res Methodol*. 2009;9(1):2.
- Sterne JA, Sutton AJ, Ioannidis JP, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*. 2011;343:d4002.
- Zwetsloot P-P, Van Der Naald M, Sena ES, et al. Standardized mean differences cause funnel plot distortion in publication bias assessments. *Elife*. 2017;6:e24260.
- Vesterinen H, Sena E, Egan K, et al. Meta-analysis of data from animal studies: a practical guide. *J Neurosci Methods*. 2014;221:92–102.

35. R Core Team. *R: A Language and Environment for Statistical Computing*. 3.5.1 ed. Vienna, Austria: R Foundation for Statistical Computing; 2018.
36. Schwarzer G. Meta: an R package for meta-analysis. *R News*. 2007;7(3):40–45.
37. Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Softw*. 2010;36(3): 1–48
38. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. Switzerland: Springer Nature; 2016.
39. van der Worp HB, de Haan P, Morrema E, Kalkman CJ. Methodological quality of animal studies on neuroprotection in focal cerebral ischaemia. *J Neurol*. 2005;252(9):1108–1114.
40. Ioannidis JPA. Differentiating biases from genuine heterogeneity: distinguishing artifactual from substantive effects. In: Rothstein HR, Sutton AJ, Borenstein M, eds. *Publication Bias in Meta-analysis: Prevention, Assessment and Adjustments*. Wiley; 2005;287–302. <https://onlinelibrary.wiley.com/doi/abs/10.1002/0470870168.ch15>
41. Ioannidis JPA, Trikalinos TA. The appropriateness of asymmetry tests for publication bias in meta-analyses: a large survey. *CMAJ*. 2007;176(8):1091–1096.
42. Boardman H, Hartley L, Eisinga A, et al. Hormone therapy for preventing cardiovascular disease in post-menopausal women. *Cochrane Database Syst Rev*. 2015;3(3):1–164.
43. Grodstein F, Manson JE, Stampfer MJ. Hormone therapy and coronary heart disease: the role of time since menopause and age at hormone initiation. *J Womens Health*. 2006;15(1):35–44.
44. Clarkson TB, Meléndez GC, Appt SE. Timing hypothesis for postmenopausal hormone therapy: its origin, current status, and future. *Menopause*. 2013;20(3):342–353.
45. Schierbeck LL, Rejnmark L, Tofteng CL, et al. Effect of hormone replacement therapy on cardiovascular events in recently postmenopausal women: randomised trial. *BMJ*. 2012;345:e6409.
46. Veerus P, Hovi SL, Fischer K, Rahu M, Hakama M, Hemminki E. Results from the Estonian postmenopausal hormone therapy trial. *Maturitas*. 2006;55(2):162–173.
47. Nachtigall LE, Nachtigall RH, Nachtigall RD, Beckman EM. Estrogen replacement therapy II: a prospective study in the relationship to carcinoma and cardiovascular and metabolic problems. *Obstet Gynecol*. 1979;54(1):74–79.
48. Jackson JL, Kuriyama A, Anton A, et al. The accuracy of Google Translate for abstracting data from non-English-language trials for systematic reviews. *Ann Intern Med*. 2019;171(9):677–679.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Friis Berntsen C, Rootwelt P, Dahm AE. Bias in animal studies of estrogen effects on cardiovascular disease: A systematic review and meta-analysis. *Res Pract Thromb Haemost*. 2021;5:e12507. <https://doi.org/10.1002/rth2.12507>