Determining fragility and robustness to missing data in binary outcome meta-analyses , illustrated with conflicting associations between vitamin D and cancer mortality

David Robert Grimes[1,2]

[1]TCD Biostatistics Unit, Discipline of Public Health and Primary Care, School of Medicine, Trinity College Dublin, Dublin, Ireland.; [2]Retraction Watch, The Center For Scientific Integrity, New York, United States; [1]davidrobert.grimes@tcd.ie

---

Abstract Meta-analysis is a vital component in clinical decision making, but previous work found binary event meta-analytic results can be fragile, affected by only a small number of patients in specific trials. Meta-analyses can also miss literature, and a method for estimating how much additional unseen data would flip results would be a useful tool. This works establishes a complementary and generalisable definition of meta-analytic fragility, based on Ellipse of Insignificance (EOI) and Region of Attainable Redaction (ROAR) methods originally developed for dichotomous outcome trials. This method does not require trial-specific alterations to estimate fragility and yields a general method to estimate robustness of a meta-analysis to data redaction or addition of hypothetical trial outcomes. This method is applied to 3 meta-analyses with conflicting findings on the association of vitamin D supplementation and cancer mortality. A full meta-analysis of all trials cited in the 3 meta-analyses yielded no association between vitamin D supplementation and cancer mortality. Using the method outlined here, it was determined that meta-analytic fragility was high in all cases, with recoding of just 5 patients in the full cohort of 133,262 patients was enough to cross the significance threshold. Small amounts of redacted or non-included data also had substantial impact on each meta-analysis, with addition of just 3 hypothetical patients to an ostensibly significant meta-analyses (N = 38,538) enough to yield a null result. This method for analytical fragility is complementary to previous investigations that suggested meta-analyses are frequently fragile. It further shows that merely increasing the sample size is not an assurance against fragility. Caution should be advised when interpreting the results of meta-analyses and conflicting results may stem from inherent fragility and should be carefully employed.

Introduction

Meta-analyses are critical for synthesising medical evidence and providing robust estimates of treatment effects (***Murad et al., 2014***; ***Dechartres et al., 2013***; ***Nordmann et al., 2012***). However, there are concerns that meta-analyses are frequently improperly conducted, and in recent years there has been methodological concerns over the rigour of an increasing number of these publications (***Lawrence et al., 2021***; ***Siemens et al., 2021***; ***Hameed et al., 2020***), particularly when constituent Randomized Controlled Trials (RCTs) have data integrity issues which meta-analysis cannot solve (***Wilkinson et al., 2024***, ***2025***). There is an additional consideration that deserves mention in dichotomous outcome or binary outcome trials, which is the issue of fragility. In RCTs, many trial results are fragile, with recoding of a small number of events to non-events or vice versa in either the experimental or control arm often sufficient to convert a seemingly statistically significant result to a null, and vice versa (***Grimes, 2022***; ***Tignanelli and Napolitano, 2019***; ***Baer et al., 2021a***,b).

While this issue has generated much recent discussion in RCTs, this has not received much consideration for meta-analyses, which are often thought of as more robust to such manipulations. This however may not be the case. Previously, Atal et al (***Atal***

et al., 2019) suggested meta-analyses may be more fragile than might be presumed in binary outcome cases. In that work, they established an algorithm to find the minimum re-coding of patients between studies that would engineer a seemingly significant result from a null one or vice versa. Applying this to a large sample of meta-analyses, Atal et al found that many were inherently fragile. This is a powerful method, but one possible objection is that it has the downside of non-generalisability, finding very specific combinations of trials and patients that would have to be re-coded. For example, an Atal meta-analytic fragility of 4 pertains to a specific and often unique circumstance when 4 patients could be recoded from a specific study or combinations thereof to change outputs, but this does not generalise to any 4 patients in that meta-analysis. This makes this definition of meta-analytic fragility useful but not generalisable, and less intuitive to interpret than a typical RCT fragility metric.

In this work, we establish a generalizable meta-analytic fragility metric, based upon Ellipse of Insignificance (EOI) analysis for dichotomous outcome trials (***Grimes, 2022***). EOI is a refined and analytical fragility metric that simultaneously finds the degree of fragility in both control and experimental arms of dichotomous outcome trials, deriving its name from the geometrical reality relating to chi squared tests that for any seemingly significant result mapped on a plane where one axis corresponds to the experimental arm and the perpendicular the control, there is an associated ellipse inside which all results are null. EOI analysis analytically finds and resolves the minimum displacement (Fewest Experimental/Control Knowingly Uncoded Participants (FECKUP) vector) from this region, reporting fragility as proportion of group sizes in context. As it is analytical and based upon chi square testing, it handles samples of all size, reporting a readily interpreted contextualised measure.

Intimately related to this is Region of Attainable Redaction (ROAR) analysis (***Grimes, 2024b***), a methodology for ascertaining potential influence of redacted data on reported results. There are many reasons why relevant data might be missing from an analysis, ranging from simple absent data, missed studies, inappropriate cut-offs and even deliberate cherry-picking (***Grimes and Heathers, 2021***). This is highly relevant to meta-analysis, and it would be useful to have a method to quantify the potential impact of redaction on reported results and gives an additional estimate of how robust reported results may be when new data is accrued. Accordingly in this work, we will extend these methodologies to handle meta-analysis, contrast them to Atal et al's metric, and apply them to conflicting Vitamin D trials to illustrate their epidemiological utility. This package of meta-analytic fragility tools we name *EOIMETA* for brevity.

Methods

Derivation of Meta-analytic fragility and redaction analysis (EOIMETA)

For both EOI and ROAR, full mathematical details have been previously published (***Grimes, 2022***, ***2024b***) and an online implementation is available at https://drg85.shinyapps.io/EOIROAR/ for any dichotomous outcome trial. Extending this to the meta-analytic situation, consider i studies, each with an experimental group with a_i events and b_i non-events, and control group with c_i events and d_i non-events per study. Consider for any given meta-analysis the crude unadjusted pooled relative ratio (RR_p) and a Cochran–Mantel–Haenszel risk ratio (RR_{CMH}), where confounding effects between studies are given by $|1 - \frac{RR_{CMH}}{RR_p}|$. When this is small ($< 10\%$), confounding is minimal and EOI analysis can be deployed on pooled meta-analysis results, by treating the entire meta-analysis as a single pooled sample. As meta-analysis weigh studies by variance, simple pooling does not inherently account for study heterogeneity. We can also adjust for this by using a generic inverse-variance weighed average model – in this case, for each included study, we compute relative risk and standard error as

$$\ln(RR_i) = \ln\left(\frac{a_i(c_i + d_i)}{c_i(a_i + b_i)}\right) \tag{1}$$

$$SE(\ln(RR_i)) = \sqrt{\frac{1}{a_i} + \frac{1}{a_i + b_i} + \frac{1}{c_i} + \frac{1}{c_i + d_i}} \tag{2}$$

where a continuity correction of 0.5 may be applied for any zero values. The weighted average of each trial is $w_i = SE(\ln(RR_i))^{-1}$ and total pooled risk ratio is given by

$$RR_{PG} = \exp\left(\frac{\Sigma w_i \ln(RR_i)}{\Sigma w_i}\right). \tag{3}$$

Once the pooled risk ratio is known, we can scale the event and non-event rate in the experimental group to match the adjusted relative risk while keeping the total number in the experimental arm constant. Letting $a = \Sigma a_i$,$b = \Sigma b_i$,$c = \Sigma c_i$ and $d = \Sigma d_i$, we allow $a^* = a + x$ and $b^* = b - x$ where x may be positive or negative to adjust the experimental group to the derived risk ratio. We can solve for this value to obtain

$$x = \frac{c(a+b)RR_{PG}}{c+d} - a. \tag{4}$$

For consistency, the values of a^* and b^* are rounded to integers. This modification accounts for study level heterogeneity, and a standard EOI analysis can then be applied to the vector (a^*, b^*, c, d). In addition, we can also employ ROAR analysis to the same vector, yielding the raw number of patients in either or both arm who could be added a given direction to change the result, and exact combination of control and experimental group redactions required to change the result from a significant finding to a null one. Caveats for implementation and interpretation are outlined in the discussion section.

Application to vitamin D studies

Results of vitamin D supplementation and cancer mortality have been markedly inconsistent. In recent years, several meta-analyses of published trials yielded markedly conflicting results. A review and meta-analysis by Zhang et al (***Zhang et al., 2019***) and correction (***Zhang et al., 2020***) of 5 RCTs involving 39,168 patients found a significant association between vitamin D supplementation and reduced cancer mortality, reporting a 15% decrease in the risk of cancer death with vitamin D supplementation. Similarly, a meta-analysis by Gou et al (***Guo et al., 2023***) of 11 RCTs involving 112,165 patients found a significant reduction in relative risk of cancer death with supplementation of 12%. By contrast, a subsequent study of 6 RCTS involving 61,853 patients (***Zhang et al., 2022***) found no significant reduction in cancer death risk with vitamin D supplementation. All 3 meta-analyses involved combinations drawn from just 12 trials.

To investigate this apparent case of meta-analytic fragility, we extracted data from all 12 studies included in the 3 meta-analyses. A small inconsistency between all 3 meta-analysis was found in relation to one RCT (***Scragg et al., 2018***); In Zhang et al, the rate of cancer deaths in the supplementation and control group were reported as 44/2558 and 45/2550 respectively, versus being reported as 28/2558 versus 30/2558 in Gou et al and 30/2550 versus 30/2550 in Zheng et al. Disagreement between these figures appears to stem from cancers detected in both groups prior to the randomisation process. Accordingly, the adjusted ITT figure used in this work is 30/2544 versus 30/2535 respectively for Scragg et al's work. A further nuance is that for another small RCT (***Ammann et al., 2017***), direct data was not given on the number of cases. This however can be estimated from minimal assumption through the hazard ratio. For all included studies, relative risk and 95% confidence intervals were calculated explicitly except for Lehouck et al (***Lehouck et al., 2012***) where zero recorded deaths in the supplementation sample required confidence intervals to be estimated by bootstrap methods.Initially, we ran fragility analysis on all of these meta-analysis, with the new metric outlined here as well as Atal et al's algorithm. After this, we combined data and performed a meta-analysis on all 12 studies to ascertain potential effects of vitamin D and cancer mortality. A full fragility and robustness analysis was then performed on this, and results discussed to elucidate findings.

Meta-analytic fragility package

We developed a custom R package (***Grimes, 2025***) (eoirroar, available at https://github.com/drg85/EOIMETA and on Zenodo) to perform the fragility and redaction analysis described in the methods, as well as Atal et al's study specific alogorithm method for comparison.

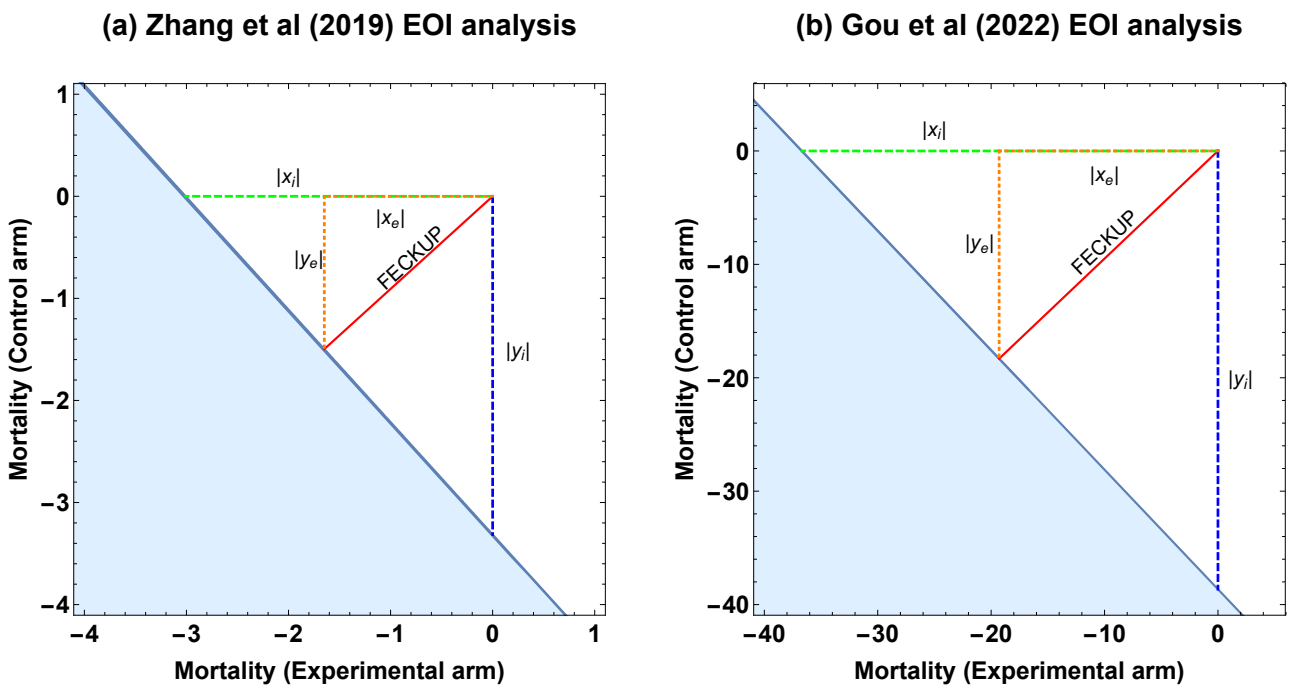

Figure 1. Meta-analytic EOI fragility analysis results for (a) Zhang et al and (b) Guo et al. See text for details

Results

Meta-analytic fragility estimation of Vitamin D Cancer Mortality studies

For all conflicting meta-analyses, the unadjusted crude pooled risk ratio and CMH risk ratio were calculated prior to the analysis being conducted. For the three meta-analyses, this ranged from $0.08\% - 1.1\%$ confounding, negligible differences which suggest the meta-analytic approach outlined in this work is applicable. Table 1 gives the results for EOI and ROAR meta-analysis as well as the Atal et al method. Figure 2 depicts the EOI analysis results for meta-analysis by Zhang et al (including 5 RCTs, 38,538 patients) and Gou et al (11 RCTs, 111,952 patients).

Table 1. Application of meta-analytic fragility / redaction sensitivity to conflicting meta-analyses

Meta-analysis	Total patients	Risk Ratio (95% CI)	I^2 (95% CI)	P-value	EOI fragility	Redaction fragility	Atal fragility
Zhang et al 2019	38,538	0.85* (0.74 - 0.97)	0% $(0\% - 79.2\%)$	0.0349*	4	3	5
Guo et al 2022	111,952	0.88* (0.80 - 0.96)	0% $(0\% - 60.2\%)$	0.004*	38	37	33
Zhang et al 2022	61,223	0.96 (0.80 - 1.16)	53.5% $(0\% - 81.4\%)$	0.6624	51	96	50

Figure 1 gives the EOI meta-analytic fragility vectors for both positive meta-analyses, showing the degree of recoding in both experimental and control groups required to flip results. The negative values on both axes arise because the relative risk of the experimental group (vitamin D supplementation) is ostensibly lower than the placebo group, so in the convention of EOI analysis, negative values imply additional of events to the experimental group and/or subtraction of events from the control. In the case of Zhang et al, it would take recoding of just 4 events / non-events in either experimental or control arms, or 4 recodings total in combination ($< 0.01\%$ of the total sample) to lose significance, even less than the fragility estimated by Atal et al's algorithm. In the case of Guo et al, recoding of 38 events / non-events ($< 0.038\%$ of sample) achieves the same result, whereas there is a even more extreme specific fragility detected by Atal et al's algorithm.

The redaction analysis is also informative, because even if there were no coding or data errors in any study or their synthesis, the addition of small numbers of patients can profoundly change the result. In the case of Zhang et al, adding just 3 patients to the cohort would be enough to lose significance, while in Guo et al it is 37. Despite Zheng et al having substantially fewer patients than Guo et al, it appears more robust to fragility by all metrics. This in turn leads to an important observation – the fact that a meta-analysis has more patients and studies does not inherently make it more robust or less fragile, as discussed in the next section.

Meta-analysis of all studies

Figure 2 depicts the meta-analysis forest plot of all 12 RCTs of 133,475 patients using a random effects model, while table 2 gives summary meta-analytic fragility statistics. Table 2 gives the statistics for when all 12 studies are combined, finding that both EOI fragility and Atal et al's fragility actually increase despite the increase in the number of subjects. Even with a large sample, the recoding of just 5 non-events to events, in this case in the placebo arm (additional deaths in the non-vitamin D group), would alter the results to cross the threshold of significance (p = 0.048) and likely change inferences.

Table 2. Meta-analytic fragility of all 12 studies

Total patients	Risk Ratio (95% CI)	I^2 (95% CI)	P-value	EOI fragility	Redaction fragility	Atal fragility
133,262	0.92 (0.83 - 1.02)	23.2% (0% – 60.8%)	0.061	5	192	9

Data source	Population	Deaths (Vit-D)	Deaths (Placebo)	Relative Risk
Trivedi et al (2003)	People 65-85	63 / 1345	72 / 1341	0.87 (0.63 - 1.21)
WHI (2009)	Women 51-82	344 / 18176	382 / 18106	0.90 (0.78 - 1.04)
RECORD (2012)	People >70 years	151 / 2649	178 / 2643	0.85 (0.69 - 1.04)
Lehouck et al (2012)	COPD patients	0 / 91	1 / 91	0.33 (0.01 - 8.08)
Martineau et al (2015)	COPD patients	1 / 122	1 / 118	0.97 (0.06 - 15.29)
Ammann et al (2017)	Women 58-69	51 / 17411	66 / 17352	0.77 (0.53 - 1.11)
Akiba et al (2018)	NSCLC survivors	12 / 77	12 / 78	1.01 (0.49 - 2.11)
Scragg et al (2018)	People 50-84	30 / 2544	30 / 2535	1.00 (0.60-1.65)
Urashima (2019)	GI Cancer survivors	58 / 251	51 / 166	0.75 (0.55 - 1.04)
VITAL (2019)	Men >50 Women >55	154 / 12927	187 / 12944	0.82 (0.67 - 1.02)
Rake et al (2020)	People 65-84	14 / 802	5 / 813	2.83 (1.03 - 7.84)
Neale et al (2022)	People 65-84	242 / 10661	210 / 10649	1.15 (0.96 - 1.38)
Summary		**1113 / 66341**	**1202 / 66921**	**0.92 (0.82 - 1.02)**

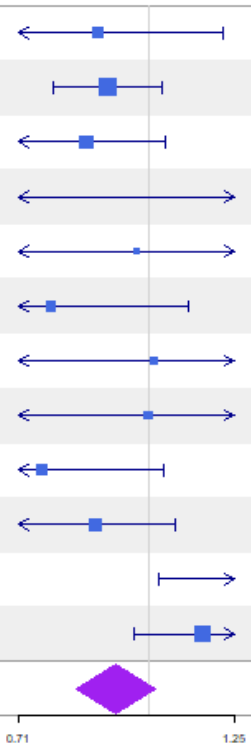

Figure 2. Meta-analysis of all 12 RCTS including study details and relative risk of vitamin D supplementation on cancer mortality

Discussion

The meta-analytic fragility tool presented here, EOIMETA, is generalisable and flexible, complementing Atal et al's study and patient specific method. The nuance between both approaches is that in Atal et al's algorithm identifies a specific combination of patients from specific studies that may be moved to alter results, whereas the current work establishes a method for pooling meta-analysis on ascertaining their group fragility and the potential impact of missing data. In this respect, they answer complementary and subtlety different questions specific to meta-analyses.

In RCT literature, a frequent objection to the use of fragility analysis methods is that the mere existence of a small FI might be an artefact of trial design. For clinical trials especially, a well-designed trial should be structured as to minimize patient exposure to unknown harms, striving to ensure that just enough participate to allow the detection of clinically relevant effects. He argument that RCTs at least might be fragile 'by design' is however convincingly countered by other authors, who find no evidence of p-value distributions clustering around significance thresholds after a sample size calculation and find additional that fragility in well-designed studies is not always low. More recent robust methods for fragility analysis like EOI and for redaction like ROAR also have application far beyond ostensibly fragile-by-design RCTs, but for cohort studies, preclinical work, and even ecological studies. The fragile-by-design argument is also a weaker on in the context of meta-analysis, where the marshalling of many studies to increase effect precision measurements should strengthen the evidence but results here and previously suggest that even meta-analysis of large groups of patients can be profoundly fragile.

There are a few subtle issues and limitations to consider additionally; Atal et al's meta-analytic fragility metric in principle identifies the precise combination of inter-study and patient recoding required to flip a result, a "worse-case" scenario, identifying the most fragile point in a meta-analyses. This usually results in the Atal et al fragility metric being typically smaller than the complement EOI fragility metric outlined here, but this is not always the case. Because Atal et al's method is a greedy algorithm, it is time-consuming to run on large collections of studies and can sometimes miss optimal and smaller solutions. Secondly, the adjustment of the events and non-events in the experimental group due to weighed inverse variance methods in the EOI fragility method subtly alters group composition, resulting in a smaller EOI fragility than a crude pooled estimate. What is evident from the examples shown in this work is that even in large cohorts with relatively low heterogeneity, the impact of recoding only miniscule numbers of patients can drastically alter interpretation; in the 12-study meta-analysis, recoding a mere 5 patients in 133,262 was sufficient to alter the result from a null to a positive finding.

While the method outlined here is analytic and rapid, it has some limitations that must be considered. Like Atal's method, it applies only to meta-analyses of trials with dichotomous outcomes and is not suitable for continuous outcomes. It is also appropriate only in contexts where internal confounding is low, and it may not be reliable in the presence of high between-study heterogeneity. Accordingly, careful application is required. At its most essential, this implementation reflects the meta-analytic relative risk by generating a synthetic pooled 2×2 table that preserves the total sample size in each arm. This can be interpreted as the best single-study approximation of all studies under fixed-effect assumptions. The resulting meta-analytic fragility metric is useful for modeling perturbations in results, such as the addition of patients (who were previously redacted) or the recoding of events and non-events. The advantage of this approach lies in its tractability, offering a simpler alternative to working with multiple studies individually and a deterministic and analytic resultant fragility metric. However, it is important to emphasize that it does not correspond to true patient-level pooling or standardisation.

The use of vitamin D meta-analyses in this work was chosen as illustrative rather than specific, but it is worth noting that there are methodological concerns with much Vitamin D research (***Grimes et al., 2024***). This goes beyond the scope of the current work, but serves as an example of the reality that meta-analysis is only as strong as its underlying data, and the conclusions drawn from them must always be seen in context. While meta-analysis is a powerful method for refining effect size estimates, it cannot overcome poorly conducted research or bad data (***Jané et al., 2025***), nor it is intrinsically robust. These limitations should be kept in mind when considering meta-analytic results, and scientists should consider them doubly when opting to perform a meta-analysis. Mass produced and unreliable meta-analyses are a recognised and growing problem (***Ioannidis, 2016***) , and we need be mindful not to add to increasingly issue of research waste and unreliable research (***Glasziou and Chalmers, 2018***; ***Grimes, 2024a***).

Data and code availability

All relevant code for this undertaking including the study data from the included trials is available at the linked Github repository and the Zenodo DOI.

Competing interests

The author declares no competing interests.

References

Ammann EM, Drake MT, Haraldsson B, Wallace RB, Johnson KC, Desai P, Lin EM, Link BK. Incidence of hematologic malignancy and cause-specific mortality in the Women's Health Initiative randomized controlled trial of calcium and vitamin D supplementation. Cancer. 2017; 123(21):4168–4177.

Atal I, Porcher R, Boutron I, Ravaud P. The statistical significance of meta-analyses is frequently fragile: definition of a fragility index for meta-analyses. Journal of Clinical Epidemiology. 2019; 111:32–40.

Baer BR, Gaudino M, Charlson M, Fremes SE, Wells MT. Fragility indices for only sufficiently likely modifications. Proceedings of the National Academy of Sciences. 2021; 118(49):e2105254118.

Baer BR, Gaudino M, Fremes SE, Charlson M, Wells MT. The fragility index can be used for sample size calculations in clinical trials. Journal of Clinical Epidemiology. 2021; 139:199–209.

Dechartres A, Trinquart L, Boutron I, Ravaud P. Influence of trial sample size on treatment effect estimates: meta-epidemiological study. Bmj. 2013; 346.

Glasziou P, Chalmers I. Research waste is still a scandal—an essay by Paul Glasziou and Iain Chalmers. Bmj. 2018; 363.

Grimes DR. The ellipse of insignificance, a refined fragility index for ascertaining robustness of results in dichotomous outcome trials. Elife. 2022; 11:e79573.

Grimes DR. Is biomedical research self-correcting? Modelling insights on the persistence of spurious science. Royal Society Open Science. 2024; 11(1):231056.

Grimes DR. Region of Attainable Redaction, an extension of Ellipse of Insignificance analysis for gauging impacts of data redaction in dichotomous outcome trials. Elife. 2024; 13:e93050.

Grimes DR. EOIMETA: EOIMETA V1 R code (eoimeta). Zenodo. 2025; doi: 10.5281/zenodo.15878923.

Grimes DR, Heathers J. The new normal? Redaction bias in biomedical science. Royal Society open science. 2021; 8(12):211308.

Grimes DR, Kimlin M, Glasziou P. Arbitrary vitamin D deficiency thresholds yield unreliable and potentially spurious results–a review and investigation. . 2024; doi: 10.31219/osf.io/su6cx.

Guo Z, Huang M, Fan D, Hong Y, Zhao M, Ding R, Cheng Y, Duan S. Association between vitamin D supplementation and cancer incidence and mortality: A trial sequential meta-analysis of randomized controlled trials. Critical Reviews in Food Science and Nutrition. 2023; 63(26):8428–8442.

Hameed I, Demetres M, Tam DY, Rahouma M, Khan FM, Wright DN, Mages K, DeRosa AP, Baltich Nelson B, Pain K, et al. An assessment of the quality of current clinical meta-analyses. BMC Medical Research Methodology. 2020; 20(1):105.

Ioannidis JP. The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. The Milbank Quarterly. 2016; 94(3):485–514.

Jané MB, Heathers J, Grimes DR. Major Flaws in Taylor et al.'s (2025) Meta-analysis on Fluoride Exposure and Children's IQ Scores. . 2025; doi: 10.31219/osf.io/zhm54_v3.

Lawrence JM, Meyerowitz-Katz G, Heathers JA, Brown NJ, Sheldrick KA. The lesson of ivermectin: meta-analyses based on summary data alone are inherently unreliable. Nature Medicine. 2021; 27(11):1853–1854.

Lehouck A, Mathieu C, Carremans C, Baeke F, Verhaegen J, Van Eldere J, Decallonne B, Bouillon R, Decramer M, Janssens W. High doses of vitamin D to reduce exacerbations in chronic obstructive pulmonary disease: a randomized trial. Annals of internal medicine. 2012; 156(2):105–114.

Murad MH, Montori VM, Ioannidis JP, Jaeschke R, Devereaux P, Prasad K, Neumann I, Carrasco-Labra A, Agoritsas T, Hatala R, et al. How to read a systematic review and meta-analysis and apply the results to patient care: users' guides to the medical literature. Jama. 2014; 312(2):171–179.

Nordmann AJ, Kasenda B, Briel M. Meta-analyses: what they can and cannot do. Swiss medical weekly. 2012; 142(0910):w13518–w13518.

Scragg R, Khaw KT, Toop L, Sluyter J, Lawes CM, Waayer D, Giovannucci E, Camargo CA. Monthly high-dose vitamin D supplementation and cancer risk: a post hoc analysis of the vitamin D assessment randomized clinical trial. JAMA oncology. 2018; 4(11):e182178–e182178.

Siemens W, Schwarzer G, Rohe MS, Buroh S, Meerpohl JJ, Becker G. Methodological quality was critically low in 9/10 systematic reviews in advanced cancer patients—A methodological study. Journal of Clinical Epidemiology. 2021; 136:84–95.

Tignanelli CJ, Napolitano LM. The fragility index in randomized clinical trials as a means of optimizing patient care. JAMA surgery. 2019; 154(1):74–79.

Wilkinson J, Heal C, Antoniou GA, Flemyng E, Alfirevic Z, Avenell A, Barbour G, Brown NJ, Carlisle J, Clarke M, et al. Protocol for the development of a tool (INSPECT-SR) to identify problematic randomised controlled trials in systematic reviews of health interventions. BMJ open. 2024; 14(3):e084164.

Wilkinson J, Heal C, Antoniou GA, Flemyng E, Ahnström L, Alteri A, Avenell A, Barker TH, Borg DN, Brown NJ, et al. Assessing the feasibility and impact of clinical trial trustworthiness checks via an application to Cochrane Reviews: Stage 2 of the INSPECT-SR project. Journal of Clinical Epidemiology. 2025; p. 111824.

Zhang R, Zhang Y, Liu Z, Pei Y, Xu P, Chong W, Hai Y, He L, He Y, Yu J, et al. Association between vitamin D supplementation and cancer mortality: a systematic review and meta-analysis. Cancers. 2022; 14(15):3717.

Zhang Y, Fang F, Tang J, Jia L, Feng Y, Xu P, Faramand A. Association between vitamin D supplementation and mortality: systematic review and meta-analysis. bmj. 2019; 366.

Zhang Y, Fang F, Tang J, Jia L, Feng Y, Xu P, Faramand A. Correction: Association between vitamin D supplementation and mortality: systematic review and meta-analysis. bmj. 2020; 370.