Visualizing evidence-based disagreement over time: the landscape of a public health controversy 2002-2014

Tzu-Kun Hsiao*, Yuanxi Fu*, Jodi Schneider
School of Information Sciences, University of Illinois at Urbana Champaign, USA
{tkhsiao2, fu5, jodi}@illinois.edu
*Joint first authors

Abstract

Systematic reviews answer specific questions based on primary literature. However, systematic reviews on the same topic frequently disagree, yet there are no approaches for understanding why at a glance. Our goal is to provide a visual summary that could be useful to researchers, policy makers, and health care professionals in understanding why health controversies persist in the expert literature over time. We present a case study of a single controversy in public health, around the question: "Is reducing dietary salt beneficial at a population level?" We define and visualize three new constructs: the overall *evidence base*, which consists of the evidence summarized by systematic reviews (the *inclusion network*) and the unused evidence (*isolated nodes*). Our network visualization shows at a glance what evidence has been synthesized by each systematic review. Visualizing the temporal evolution of the network captures two key moments when new scientific opinions emerged, both associated with a turn to new sets of evidence that had little to no overlap with previously reviewed evidence. Limited overlap between the evidence reviewed was also found for systematic reviews published in the same year. Future work will focus on understanding the reasons for limited overlap and automating this methodology for medical literature databases.

Introduction

Systematic reviews are among the most influential study designs in the health literature, typically seen at the top of the evidence-based medicine pyramid. Commonly used to aid decision-making and to develop health recommendations, systematic reviews aim to provide a high-level synthesis of the evidence available on a given research question. Making sense of the primary literature is particularly challenging when systematic reviews of the evidence disagree—which can commonly happen, even for reviews published in the same year (Papatheodorou, 2019). In that case, we need a higher level of synthesis. Current approaches, such as umbrella reviews, are time-consuming, human-intensive, and relatively scarce: As of April 2020, we find fewer than 400 umbrella reviews in PubMed, compared to over 170,000 systematic reviews. We would like to compare systematic reviews in a way that helps us understand the evidence at a glance. For instance, systematic reviews are designed to avoid cherry picking the evidence, yet claims of citation bias have been leveled against them in some cases (e.g., Trinquart et al., 2016). We provide temporal visualizations of the change in the

evidence, and the change in scientific opinions synthesized from this evidence over a 13-year period, 2002-2014. We also demonstrate persistent differences in the types of evidence (study designs) taken into consideration.

Our goal is to provide a visual summary that could be useful in understanding why health controversies persist in the expert literature over time. We develop a case study of a single controversy in public health, on the question: "Is reducing dietary salt beneficial at a population level?" To analyze a body of evidence related to this question, we reuse and build on data from an epidemiological study's analysis of the controversy (Trinquart et al., 2016). Different from that study, we focus on a temporal network, and define new network constructs.

Our contributions are:
1. A visual summary that enables us to trace the dynamic evolution and division of the scientific opinion over time, for a 13-year period, 2002-2014.
2. We observe persistent differences in the set of evidence taken into consideration by different systematic reviews.

Background

Evidence synthesis is the "process of bringing together information and knowledge from many sources and disciplines to inform debates and decisions" (Donnelly et al., 2018). One approach, systematic reviewing, first operationalizes the problem it tries to answer by defining inclusion criteria. For instance, inclusion criteria may put restrictions on study population, intervention type, comparison, study time, study design, etc. Then, included articles are identified using a process of systematic search and screening papers for relevance. While these included articles are cited in the review, typically a review also cites articles that are not included and not synthesized as primary evidence—used for instance as background or discussion. After identifying included studies, additional analysis steps typically include extracting information, grouping related papers (such as those derived from the same research study), checking for risk of bias, and formally grading evidence quality. The final step is synthesizing the evidence; when evidence is sufficiently comparable, a meta-analysis may be used to provide a quantitative synthesis.

Evidence synthesis is especially important for public health controversies that influence policy, however, consensus formation is often challenging in these cases. On some topics, multiple systematic reviews have come to different conclusions such as whether e-cigarettes reduce the harms of tobacco use (Bareham et al., 2016). On other topics, conflicts of interest due to the predominance of industry funding impact what we know (Stanhope, 2016).

Both our work and Trinquart's relate to controversies and challenges in consensus formation, and both use visualizations, but in different ways. Trinquart et al. (2016) primarily focused on statistical modeling of the structure of a claim-specific network; in passing, they provided a citation network, standard with two exceptions: the citing relationship has a polarity (agree/disagree on claim) and study design types are shown. In our work, the most important

difference from that of Trinquart is that we focus on the temporal aspect of the citation networks. Second, we define several new constructs, in order to provide a fine-grained, precise indication of the evidence base actually synthesized by each systematic review; the dynamic network of multiple systematic reviews over time; and the underlying evidence (used and not used in systematic reviews) over time.

Methods

We focus on the systematic reviews considered in an analysis of citation bias in a public health controversy about salt, produced by Trinquart et al. (2016). Different from Trinquart's method which considered all types of papers and all citations, we focus on systematic reviews and their citations to 'included studies', which are the studies used by the systematic reviews to form a scientific opinion on a certain topic. This subset of citations should be specified in a systematic review's included studies table or data supplement.

We define three new constructs: the **evidence base**—all studies included in one or more of a set of systematic reviews; the **inclusion network**—the network whose edges link systematic review nodes to the nodes of studies it includes; and the **isolated nodes**—studies not linked into the inclusion network at a given point in time. These constructs represent the evidence and its selection for, or omission from, a given set of systematic reviews.

We start with the 14 systematic reviews (SR1-SR14) (Trinquart et al., 2016), and the primary literature (68 articles grouped into 60 studies) included as evidence in those reviews.

To construct the evidence base, inclusion network, and isolated nodes for our case study, we drew on the supplemental materials of Trinquart et al. (2016). which list 14 systematic reviews published 2002-2014, shown in Table 1. For each systematic review, we retrieved available full-text and any supplemental materials. Then from each systematic review's included studies table and reference list, we manually identified the included studies. This resulted in an evidence base of 68 included studies, which we cross-checked with the Sankey diagram given as Web Figure 4 in Trinquart et al. (2016). During the process of identifying the included studies, we also collected the inclusion criteria of each systematic review. Four types of study design were mentioned in the inclusion criteria: randomized control trials (RCTs), prospective study, cohort study, and case-control study. Notably, SR10 is retracted, and although its included studies are listed in the retraction notice, information on its inclusion criteria is not available. Also, SR11 included articles with all four study designs. Finally, we collated Trinquart et al. (2016)'s categorization of each review, as supportive/contradictory/inconclusive of the hypothesis that reducing dietary sodium intake provides population-wide health benefits. Our data is summarized in our supplemental materials (Fu & Hsiao, 2020).

Table 1. Systematic reviews, included studies, and their scientific opinions

ID	SR	Scientific opinion about reducing dietary salt (Trinquart et al., 2016)	Included study design(s)	Incl$_{total}$	Incl$_{contradictory}$	Incl$_{supportive}$	Incl$_{inconclusive}$
SR1	Hooper et al. (2002)	Inconclusive	RCT	5	0	1	4
SR2	Hooper et al. (2003)	Inconclusive	RCT	6	0	2	4
SR3	Hooper et al. (2004)	Inconclusive	RCT	6	0	2	4
SR4	Strazzullo et al. (2009)	Supportive	Prospective Studies	13	7	6	0
SR5	Taylor et al. (2011a)	Inconclusive	RCT	8	1	3	4
SR6	Taylor et al (2011b)	Inconclusive	RCT	8	1	3	4
SR7	Li et al. (2012)	Supportive	Prospective Studies, Case-control Studies	12	4	8	0
SR8	Aburto & Ziolkovska (2012)	Supportive	Prospective	15	10	5	0
SR9	Aburto et al. (2013)	Supportive	RCT, Cohort Studies	18	10	6	2
SR10	DiNicolantonio et al. (2012)	Contradictory	NA	6	6	0	0
SR11	IOM (2013)	Contradictory	All Study Designs	25	14	11	0
SR12	Adler et al. (2013)	Inconclusive	RCT	7	0	1	6
SR13	Graudal et al. (2014)	Supportive	RCT, Cohort Studies	29	12	16	1
SR14	Poggio et al. (2014)	Contradictory	RCT, Cohort Studies	11	4	7	0

We visualized the evidence base, the inclusion network, and the isolated nodes over time to show how they evolved, over the 13-year period from 2002 to 2014.

Results

Figure 1 shows the evidence base, the inclusion network, and the isolated nodes as of the publication year of each systematic review (2002, 2003, 2004, 2009, 2011, 2012, 2013, 2014). In between these periods the evidence base grew, with new nodes (but no edges) added to the network.

Systematic reviews with 'inconclusive' scientific opinions emerged first (SR1, SR2 and SR3) between 2002 and 2004, as shown in Figures 1a, 1b, and 1c. Evidence to be synthesized was limited to RCTs, seen as the 'gold standard' for understanding the effects of medical treatments. It is also noteworthy that these three publications represent, in essence the same review, twice updated, with no change in the surrounding evidence base or in the included articles from 2003 to 2004; SR3 corrects statistical errors in SR2 but is otherwise identical.

It took five years for the next review to appear. SR4, the first systematic review with a 'supportive' scientific opinion, appeared in 2009, as shown in Figure 1d. Interestingly, its evidence is entirely disjoint from the evidence used by SR1, SR2 and SR3; this is because of a significant difference in their inclusion criteria, meaning that different study designs were considered. Rather than looking at RCTs, SR4 included only prospective population studies. This was an intentional choice, because the authors of SR4 thought it was "extremely unlikely" for an RCT to be undertaken to study long-term reduction in dietary salt "because of practical difficulties, the long duration required, and high costs" (Strazzullo et al., 2009). In their meta-analysis, they show that, while the studies they included provided indirect evidence, and "few had enough power to attain statistical significance" alone, combining them showed statistically significant evidence of increased risk of stroke and cardiovascular disease from high dietary salt intake (Strazzullo et al., 2009).

Two years later, in 2011, two new systematic reviews appeared: SR5 and SR6 both analyze the RCT-only literature. As shown in Figure 1e, they form a close cluster with SR1, SR2 and SR3, making an RCT-only network still disjoint from that of SR4. Synthesizing evidence from RCTs alone still results in an inconclusive scientific opinion.

In 2012, three new reviews were published as shown in Figure 1f. Two of them, SR7 and SR8, cluster with SR4 and its prospective evidence. Meanwhile, the first SR 'against' salt reduction appeared (SR10), which used a distinctive set of evidence, mostly RCTs. It is worthwhile to note the SR10 was later retracted because "two of the contributing studies likely contained duplicate data in tables reporting information on baseline characteristics and treatment effects" (Jun & Neal 2014). Despite this concern about study grouping, none of SR10's included studies have themselves been retracted.

In 2013, another systematic review with the 'against' scientific opinion emerged in 2013, SR11, as shown in Figure 1g. SR 11 has the most relaxed criteria for study designs. For the first time, the analysis of prospective population studies, SR4, became connected to the RCT studies.

Finally, in 2014, SR13 and SR14, the last two reviews, joined the network, as shown in Figure 1h. At this point, 52 of the 68 primary articles (48 of 60 studies) listed in the supplementary data section of Trinquart et al. (2016) had been included in at least one of the systematic reviews. However, it is not immediately clear why the remaining 18 isolated nodes were omitted from SR1-SR14.
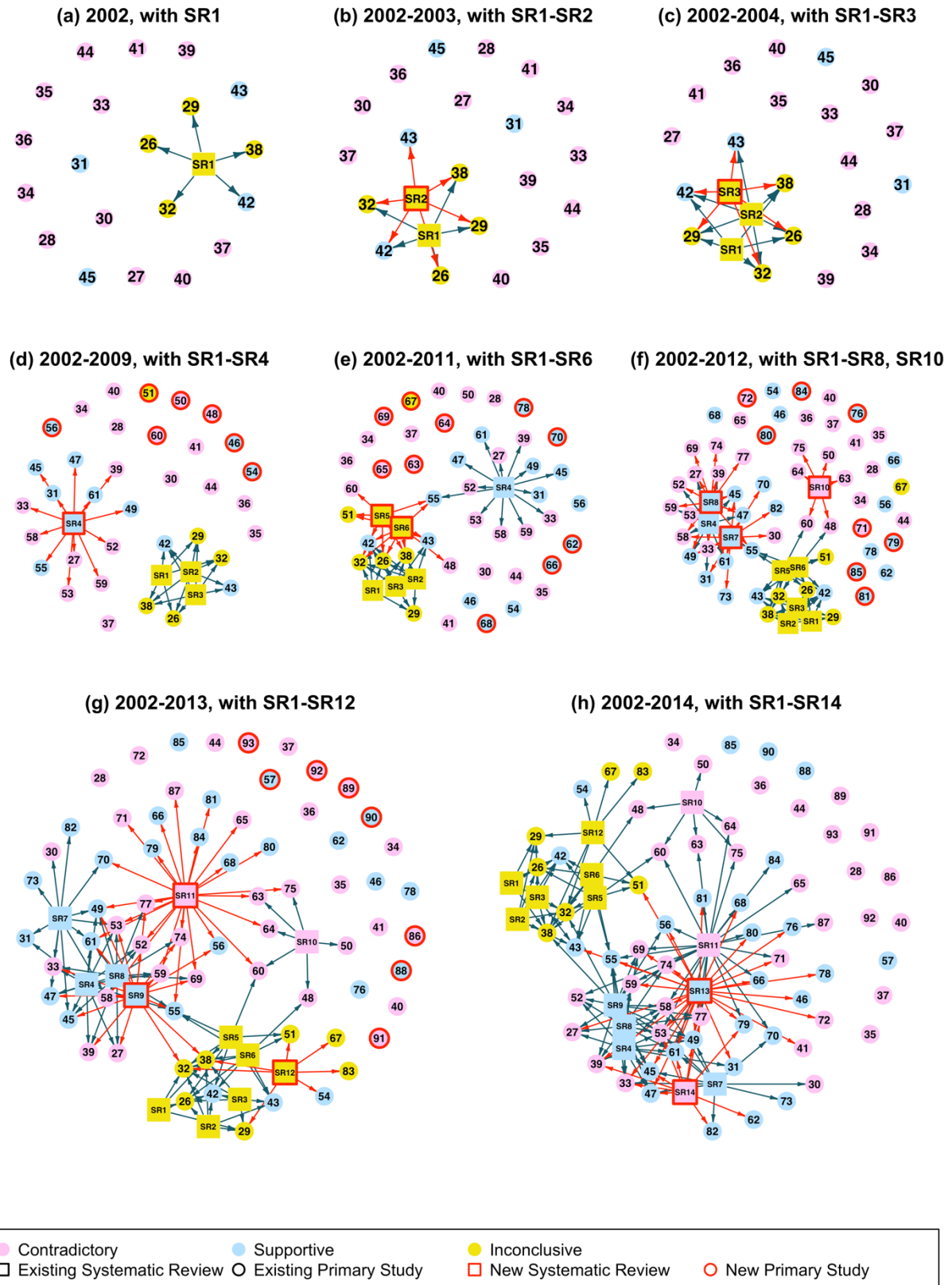
Figure 1: Visualization of the evidence base (included studies & isolated nodes) by year.

(a) Systematic reviews with opposite conclusions in 2013

(b) Systematic reviews with opposite conclusions in 2014

● Contradictory   ● Supportive   ● Inconclusive   □ Systematic Review   ○ Primary Study
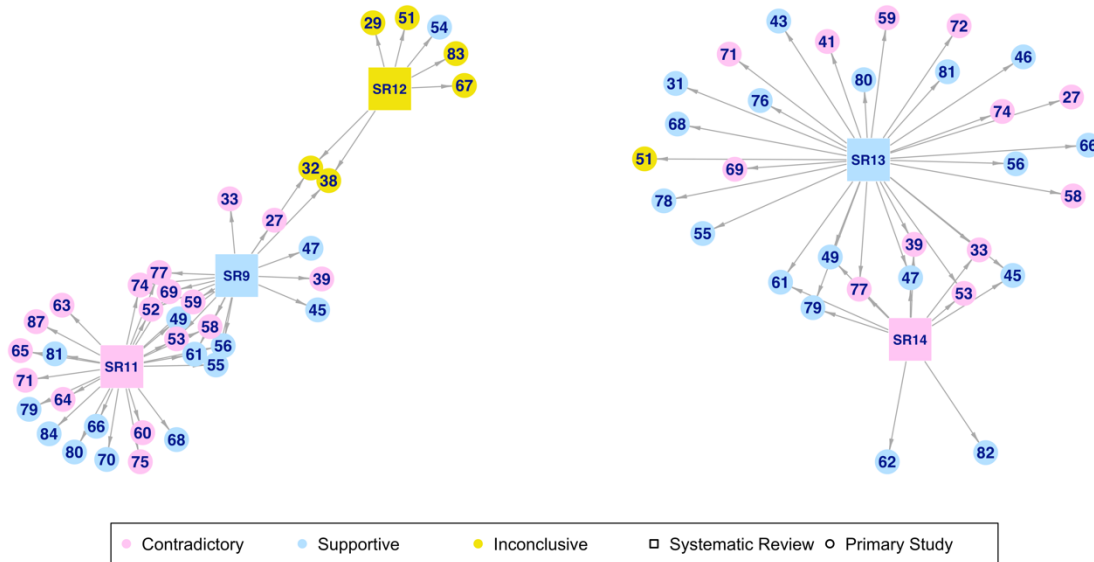
Figure 2: Some SRs published in the same year still have little overlap in their evidence, and come to different conclusions. Figure 2a: In 2013, three new reviews were published, with inconclusive (SR12), supportive (SR9), and contradictory (SR13) scientific opinions. Figure 2b: In 2014, two new reviews were published, with supportive (SR13) and contradictory (SR14) scientific opinions.

The evidence base grows by the year, so some difference in included evidence is natural. However, our visualization reveals the curious fact that some SRs published in the same year still have little overlap. Consider the two systematic reviews published in 2014, shown in Figure 2b, for instance: 31 articles are included in one or both reviews, with only 9 articles included in both. The remaining 22 articles (2/11 unique in SR14 and 20/29 unique in SR13) appear in only one review. Based on examining this different evidence, the reviews came to opposite scientific opinions.

In trying to explain this, we can look at the inclusion criteria, and, where available, the excluded articles list. The inclusion criteria differ in the main outcome measure required: both SRs considered mortality, but only SR13 accepted strokes and heart attacks as evidence. Otherwise, both SRs were broadly similar. While SR13 did not publish an excluded article list, SR14's excluded article list is informative, covering 11 of the 20 articles included in SR13 but not SR14[1]. Notably, while we can explain 11/22 differences in article choice, we cannot explain why SR14 did not include the remaining 9 articles unique to SR13, and why SR13 did not include the 2 articles unique to SR14.

---

[1] Reasons SR14 excluded 11 articles that SR13 included: Five articles (articles 27, 31, 55, 58, and 66) for "cardiovascular mortality not reported by levels of Na intake or only data of total mortality"; four articles (articles 33, 69, 74, and 78) for studying "high-risk populations"; two articles (articles 41 and 59) as "different studies analyzing the same cohort."

The SRs published in 2013 are even more perplexing. These three SRs took different scientific opinions towards the salt reduction and did not have a single article included in common, as shown in Figure 2a. The inclusion criteria are not enlightening. For example, SR12 and SR11 both include RCTs but share no included articles. And no excluded article list was available for SR9, SR11 or SR12. The strange situation with the reviews published in 2013 remains a question for future study.

Discussion

The visual summary captures moments when different scientific opinions towards the controversy emerged. The overall evidence base grew over time, and as time elapsed, reviewers had a larger body of evidence to draw upon. We attribute different scientific opinions in these SRs to differences in the sets of evidence they considered. This is shown most vividly in Figure 1d: SR4, the first SR to support salt reduction, considered none of the same evidence as the first three systematic reviews. Those three (SR1, SR2, and SR3) had only included RCTs, and had been unable to take an opinion either supporting or opposing salt reduction. This shows the importance of the inclusion criteria, and in particular the study designs selected, in influencing sensemaking of the evidence. Distinctive evidence was also used in SR10 (Figure 1f), the first SR to oppose salt reduction: 4 of the 6 included studies were used for the first time. Separation in the network—indicating lack of overlap in the evidence considered—reduced as the network grew and matured, but even as of 2014 (Figure 1h), clusters were still observable. Some of this difference is attributable to purposeful disagreements in what evidence should be taken into account (e.g., disease incidence, or only mortality for the new reviews published in 2013 and 2014, shown in Figure 2). However, some differences in the evidence used are still not explained.

Conclusions & Future Work

This study proposed a novel approach of using 'inclusion networks' as a way to study the persistence of controversies in the expert literature over time. Different from the traditional approach of using citation networks, the inclusion networks provide a more precise view of the relationship between the evidence base and the synthesized results. The visual summary has been surprisingly informative in terms of revealing some causes of division in scientific opinions towards a public health controversy. However, several questions remain to be answered. First of all, we must investigate the deeper cause for the limited overlap in which articles were included in different SRs. This investigation will need to take into consideration factors such as the growth of the evidence base over time, reviewers' goals, inclusion criteria, search strategies, quality appraisal, etc. Moreover, as the size of the network grows, interpreting the visual summary becomes more challenging, and may require filtering (i.e., narrowing to only new reviews as in Figure 2b, compared to showing all years in Figure 1h). In the future, we will use methods based on community detection and network statistics to help us understand large and complex networks. In the long run, we aim to build a tool to generate this type of network from medical literature databases. Researchers would then be able to identify the reasons for variations in scientific opinions and conceive the design of a study that might settle the debate.

Acknowledgements

Bibliography

Bareham, D., Ahmadi, K., Elie, M., & Jones, A. W. (2016). E-cigarettes: Controversies within the controversy. *The Lancet Respiratory Medicine*, *4*(11), 868–869. https://doi.org/10.1016/S2213-2600(16)30312-5

Donnelly, C. A., Boyd, I., Campbell, P., Craig, C., Vallance, P., Walport, M., Whitty, C. J. M., Woods, E., & Wormald, C. (2018). Four principles for synthesizing evidence. *Nature*, *558*(7710), 361–364. https://doi.org/10.1038/d41586-018-05414-4

Fu, Y., & Hsiao, T.-K. (2020). Dataset for "Visualizing evidence-based disagreement over time: The landscape of a public health controversy 2002-2014." https://doi.org/10.13012/B2IDB-9222782_V1

Jun, M., & Neal, B. (2014). Low dietary sodium in heart failure: A need for scientific rigour. *Heart*, 100(21), e2. https://doi.org/10.1136/heartjnl-2012-303266

Papatheodorou, S. (2019). Umbrella reviews: What they are and why we need them. *European Journal of Epidemiology*, *34*(6), 543–546. https://doi.org/10.1007/s10654-019-00505-6

Stanhope, K. L. (2016). Sugar consumption, metabolic disease and obesity: The state of the controversy. *Critical Reviews in Clinical Laboratory Sciences*, *53*(1), 52–67. https://doi.org/10.3109/10408363.2015.1084990

Strazzullo, P., D'Elia, L., Kandala, N.-B., & Cappuccio, F. P. (2009). Salt intake, stroke, and cardiovascular disease: Meta-analysis of prospective studies. *BMJ*, *339*, b4567. https://doi.org/10.1136/bmj.b4567

Trinquart, L., Johns, D. M., & Galea, S. (2016). Why do we think we know what we know? A metaknowledge analysis of the salt controversy. *International Journal of Epidemiology*, *45*(1), 251–260. https://doi.org/10.1093/ije/dyv184