Data: Cochrane Database of Systematic Review

The Cochrane Database of Systematic Review (CDSR), which is now published by Cochrane Review Group, is a one of the leading journals publishing systematic review articles about studies on human health intervention and diagnosis to inform diverse sets of stakeholders such as policy makers, patients, health professionals, starting from April 1995[1]. The publisher, Cochrane Review Group which was originally found in the U.K. as a non-profit organization, now has 53 different groups[2] with over 30,000 expert volunteers in health science across the world (Cochrane, 2019). Even though the quality was questioned (e.g., Olsen et al. 2001, Büchter et al. 2016), the CDSR has been regarded as one of the most reliable medical resources that guide the evidence-based medical practice (Handoll et al. 2008) since its foundation.

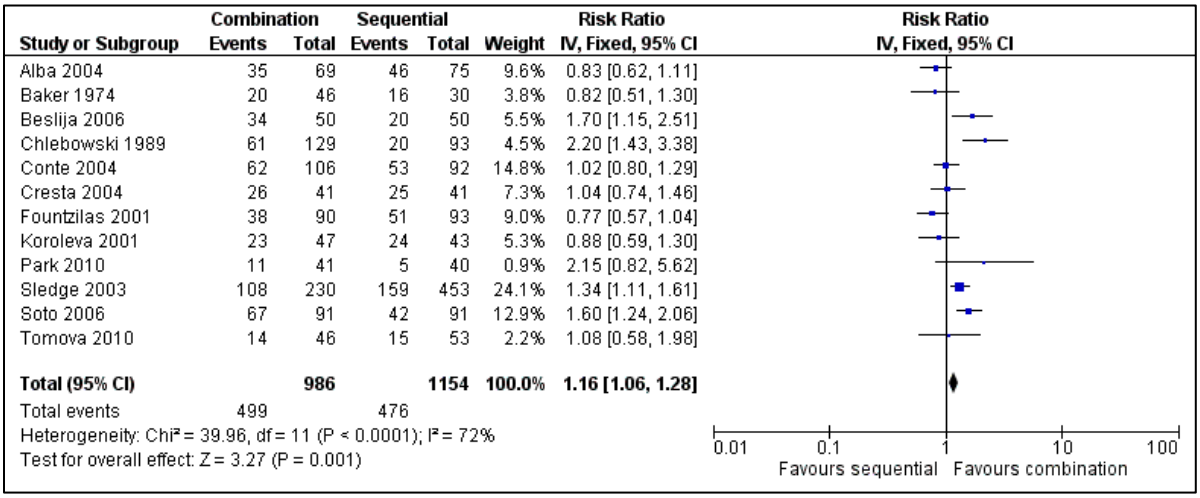
Only reviews with the meta-analyses containing at least more than three studies were selected because we only have the definitive conclusions about the claims when a review examines equal to or more than three studies. Reviews in the CDSR with the meta-analyses can either consider fixed effects or random effects: the fixed effects models assume that all the studies in the review have a common parameter of the effect size, only allowing the random errors within studies while the random effects model allows between-study variances. Each review provides the meta-analysis *p*-values and confidence intervals based on the methodology proposed by Higgins et al., (2011), combining the statistical information from the studies that a review attempts to examine. For instance, review CD008792 published in 2013 evaluates the

[1] From 1995 to 2009, CDSR was published quarterly; since then, it has been published in a monthly basis.

[2] The full list of the review groups can be found in https://www.cochranelibrary.com/about/cochrane-review-groups. Note that some topics such as influenza virus can be address by several groups including Acute Respiratory Infections Group, Childhood Cancer Group, etc.

efficacy of the combination chemotherapy in women with metastatic breast cancer comparing to the sequential application of the same drugs, using 12 studies. Figure 1 shows the overall response rates of the two different therapies from those complied by the authors of the review.

Figure 1. An Example of a Report from the CDSR [3]



Study or Subgroup	Combination Events	Combination Total	Sequential Events	Sequential Total	Weight	Risk Ratio IV, Fixed, 95% CI
Alba 2004	35	69	46	75	9.6%	0.83 [0.62, 1.11]
Baker 1974	20	46	16	30	3.8%	0.82 [0.51, 1.30]
Beslija 2006	34	50	20	50	5.5%	1.70 [1.15, 2.51]
Chlebowski 1989	61	129	20	93	4.5%	2.20 [1.43, 3.38]
Conte 2004	62	106	53	92	14.8%	1.02 [0.80, 1.29]
Cresta 2004	26	41	25	41	7.3%	1.04 [0.74, 1.46]
Fountzilas 2001	38	90	51	93	9.0%	0.77 [0.57, 1.04]
Koroleva 2001	23	47	24	43	5.3%	0.88 [0.59, 1.30]
Park 2010	11	41	5	40	0.9%	2.15 [0.82, 5.62]
Sledge 2003	108	230	159	453	24.1%	1.34 [1.11, 1.61]
Soto 2006	67	91	42	91	12.9%	1.60 [1.24, 2.06]
Tomova 2010	14	46	15	53	2.2%	1.08 [0.58, 1.98]
Total (95% CI)		**986**		**1154**	**100.0%**	**1.16 [1.06, 1.28]**
Total events	499		476			

Heterogeneity: Chi² = 39.96, df = 11 (P < 0.0001); I² = 72%
Test for overall effect: Z = 3.27 (P = 0.001)

In total, the number of the IDs of the Cochrane articles for this study is 4,543, which were published before Nov 18, 2017. Each review has its history. Protocols in a preparation stage is first produced by the authors (which sometimes are not published) and they conduct a review after a set of original research articles are identified, which means that a Cochrane review article at least has two elements in its history. If a major updated is conducted later, a newer version will replace the last one, but it maintains the same review ID. Minor amendments such as changes in formatting and an addition of a summary in plain English can also take place, but it does not lead to an updated version of a review. For example, review CD001321 that investigates the efficacy

[3] Dear, Rachel F., Kevin McGeechan, Marisa C. Jenkins, Alexandra Barratt, Martin HN Tattersall, and Nicholas Wilcken. "Combination versus sequential single agent chemotherapy for metastatic breast cancer." *Cochrane Database of Systematic Reviews* 12 (2013).

of Cranberries in preventing urinary tract infections was first published in 2003 but updated in 2004, 2008, and 2012. Considering this, I developed a web-scraper to collect all the version histories of 4,543 reviews (which led to 8,212 different versions of the 4,543 reviews) and selected the most recent articles before Nov 18, 2017. The mean of the number of authors of the most recent reviews is 4.59 (std = 2.11, min = 1, max = 41).

The reference sections of each Cochrane review disclose both the included and excluded original studies.[4] The metadata of these original studies (e.g., the individual studies in Figure 1) in each review, including the names of the authors, the titles, the journals' names, and the publication year, and the PubMed IDs of original research articles (when it is possible) were also collected using a web-scraper. In total, we identified 128,563 original articles included in 4,543 original studies.

The PubMed ID is of importance here. Each PubMed ID can be used to identify a medical research article in the MEDLINE database, which is co-maintained by the National Library of Medicine (NLM) and the National Center for Biotechnology Information (NCBI) at the National Institutes of Health (NIH) of the United States. The PubMed IDs of each original research article examined by each Cochrane review allowed us to collect the additional data such

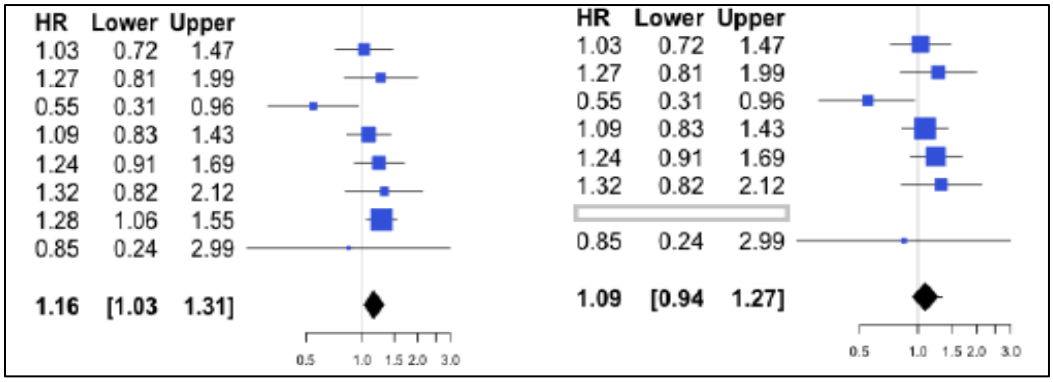
[4] The authors first collect the relevant published research and sort them with some evaluation criteria to safeguard the scientific validity or to guarantee fair comparisons. The authors decide to exclude studies when they do not meet the criteria or if the study designs are inconsistent with other studies. A reader can find detailed inclusion criteria in each review. In addition to this, note that each review only includes one of studies duplicated by the same authors. For example, the reference section of the review CD001321 includes the two studies conducted by Salo et al. (2010) and Salo et al. (2011). Since the two studies are conducted by the same group of the authors, the review CD001321 only cites Salo et al. (2010). But when the PubMed IDs were identified, the metadata of every original article in the reference sections was collected regardless of whether it is included or excluded, and whether it is directly cited.

as the abstracts of the original articles, the references of the original articles, and the list of other original studies published by the authors with the same name, all of which we used to compute the social and intellectual dependencies of the original studies within review articles.

Dependent variable: The r-value as a measure replicability

In this paper, the *r*-value (Shenhav et al., 2015) is used as a measure of replicability, which can be computed based on a meta-analysis. The computation of the *r*-value basically follows the logic of sensitivity analysis in which whether the significance of a claim holds when samples of studies in a systemic review are excluded. By measuring the impact of each study or sets of studies on the overall conclusion of a review, the *r*-value can provide more rigorous evidence of the replicability of scientific claims by considering that one study can drive the significance of a meta-analysis with the increased power. Figure 2 graphically shows how the logic can be applied using an example from a review in the CDSR: when study 7 is excluded, the overall conclusion becomes insignificant.

Figure 2. The Forest Plot from Review CD008792 (Left) and Excluding Study 7 (Right)



Like *p*-value, the significance level of the *r*-value depends on a researcher; but the smaller the *r*-value, it is less likely that the null hypothesis (i.e., no actual effect of treatment based on a meta-analysis) is true. Thus, the smaller the *r*-value, it is more likely that a conclusion from a meta-analysis is not driven by a specific single study. Conceptually speaking, the *r*-value is the largest *p*-value among the *p*-values from the *N* meta-analysis when the basic Leave-Out-Out procedure is applied but other frameworks such False Discovery Rate, which controls multiplicity of end points. The computational details as following.

Let $p^L_{i_1,\ldots,i_k}$ and $p^R_{i_1,\ldots,i_k}$ be the left and right *p*-values from a meta-analysis on the subset $S = (i_1, \ldots, i_k) \subset 1, \ldots, N$, where *N* = the full number of studies in the meta-analysis and $k < N$. And let $\prod(k)$ be the set of all possible subsets of size *k*. Each of subsets from (*N - u + 1*) studies is used to compute the *r*-value of a systematic meta-analysis. The value of *u* is subject to a choice, but we can see that $\binom{N}{u-1}$ subsets of $(N - u + 1)$ studies are used to calculate the r-value for a given value of *u*. For example, if *u* = 2, a meta-analysis of each of the *N* subsets of *N* - 1 studies will be employed.

For a left-sided alternative, the *r*-value is computed as

$$r^L = max_{\{i_1,\ldots i_{(N-u+1)}\}\in\prod(N-u+1)} p^L_{i_1,\ldots,i_{(N-u+1)}}$$

For a right-sided alternative, the r-value is computed as:

$$r^R = max_{\{i_1,\ldots i_{(N-u+1)}\}\in\prod(N-u+1)} p^R_{i_1,\ldots,i_{(N-u+1)}}$$

For a two-sided alternative,

$$r = 2\min(r^L, r^R).$$

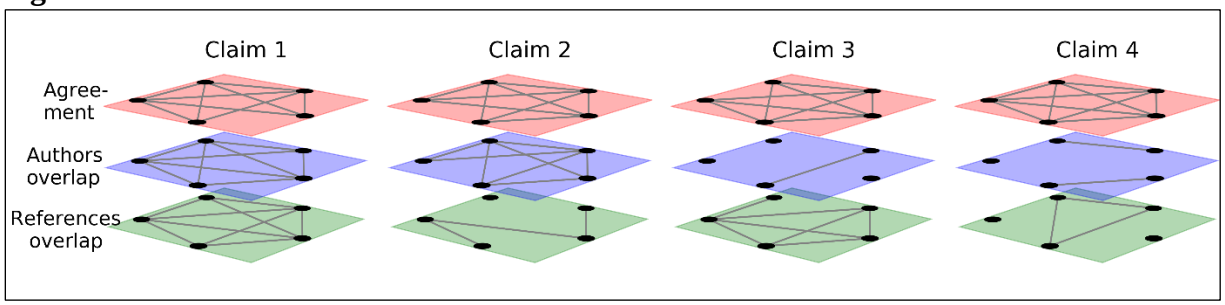
The sensitivity of the confidence interval from a meta-analysis is also can be calculated using the similar logic. Basically, the sensitivity interval can be defined as the union of the confidence intervals from the meta-analysis using the $\binom{N}{u-1}$ subsets of $(N - u + 1)$ studies. With the α level significance level, the upper limit of the (1 - α) sensitivity interval equals to the upper limit of the (1- α) confidence interval from the meta-analysis on $L = \{i_1^L, \ldots, i_{(N-u+1)}^L\}$, where L is the subset resulting in the largest *p*-value for the left-side *r*-value. The lower limit of the (1 - α) is computed using the same logic. This implies that a meta-analysis is robust (with a given *u*) if and only if the sensitivity interval does not contain the value proposed in the null hypothesis (e.g., an odds-ratio equals to 1).

Furthermore, when multiple primary endpoints need to be considered, Family-Wise Error Rate (FWER) or False Discovery Rate (FDR) controlling procedures can also be employed to lower the significance threshold. For example, review CD005211, which examines the effectiveness of exercise interventions to prevent/minimize the upper-limb dysfunction for those who receive breast cancer treatment, investigates 4 different types of study designs that can be grouped according to the timing of the introduction of the exercise intervention programs. In this case, we have multiple point estimates of which the relatedness should be controlled. Using FDR framework, we can apply the Bejamini-Hochberg (BH) adjustment[5] (Benjamini & Hochberg, 1995) to achieve this.

[5] The BH-adjusted *r*-values for a sorted list of *r*-values: $r_{(1)} \leq \ldots \leq r_{(M)}$ for $min_{i \geq j} \frac{M * r_{(i)}}{i}$, $j = 1, \ldots, M$. For instance, if we have S= (0.1231, 0.0017, 0.0167, 0.1176), the sorted list is (0.0017, 0.0167, 0.1231, 0.1776) and the adjusted r-values with BH procedure is (0.0068, 0.0334, 0.1641, 0.1776).

Independent Variables

Figure 3



Quantifiying Author/Reference Overlap

In this paper, the Jaccard Coefficient (JC) is used to quantify the degree of overlap for authors and references between the original papers in each Cochrane review. The JC is defined as the proprotion of the intersection of two sets to the union of them: $JC\,(A_i, A_j) = \frac{|A_i \cap A_j|}{|A_i| + |A_j|\,|A_i \cap A_j|}$. For example, if paper 1 and paper 2 were written by author A, B and author B, C, D, E, the authorship JC between the two papers is computed as 0.2. The resulting values for authors and referecnes were employed to construct the edge weights between pairs of articels in the network layers represented in Figure 3.

Indepedence scores

The indepedency score (*IND score)* for authorship and references of each Cochrane review is defined as following: $IND = \frac{E_{max} - W}{E_{max}}$, where $E_{max} = \frac{E}{n(n-1)\,/2}$ and W is the sum of all edges in a review article's each layer of authorship and references. (Note that E corresponds to the JC between a pair of pair in terms of authorship or reference). Social and knolwedge

independence are computed as on the JC of authors overalp and referecnes overlaps, respectively. For example, Claim 1 in Figure 3 has low social and knowledge indepedence scores where as Claim 4 has high depepdence scores for the two dimensions. Claim 2 scores high in knolwedge indepedence but low for social indepenence whereas Claim 3 has the opposite pattern.

Controls: Journal Prominence as EigenFactor Score

To control prominence of journals that each original study was published, we employed the journal eigenfactor score (Bergstrom et al., 2008). The calculation of journal EigenFactor score basically follows the logic of the Google's Page-Rank algorithm which exploits the global structure of citation network in a recursive way.[6] The Comparing to some basic centrality measures such as a degree centrality in which only local influence is measured, it gives more weights to citations from prominent journals than less prominent ones. However, at the same time, it assumes that each subfield can have different standards for citation. In other words, it accounts for the fact 10 citations over 3 threes years in field A is good as much as 100 citations over a year in field B. The formulation can be expressed as following:

$$PR(p_i) = \left(\frac{1 - d}{N}\right) + d \sum_{p_j \in M(p_i)} \frac{PR(p_j)}{L(p_j)},$$

$$where,$$

[6] Page-Rank is one variant of a family of eigenvector centralities. Comparing to the most basic eigenvector centrality, Page-Rank and EigenFactor accounts for the weights and direction of edges. For more sociological interpretation of a family eigenvector centralities, please refer to Bonacich (1987) and Bonacich (2007).

$$d = a\ dampening\ constant,$$

$$p_j = a\ paper\ j\ citing\ paper\ p_i,$$

$$L(p_j) = the\ outbound\ edge\ of\ p_j,$$

$$M(p_i) = \{p_1, \ldots, p_j\}.$$

For each claim in a Cochrane review, we calculated the mean journal EigenFactor by taking average of the EigenFactor of the journals that studies included in a Cochrane review were published.

Final Model

Model 1

$$r_i = \beta_0 + \beta_1 * S_{IND} + \beta_2 * K_{IND} + \beta_3 * J + e_i,$$

$$where,$$
$$S_{IND} = social\ independence,$$
$$K_{IND} = Knowledge\ indepdence,$$
$$J = Journal\ EigenFactor$$

Model 2

$$\text{logit P}(R_i = 1) = \beta_0 + \beta_1 * S_{IND} + \beta_2 * K_{IND} + \beta_3 * J$$

$$where,$$
$$R_i = 1\ if\ r_i < 0.05\ otherwise\ R_i = 0,$$
$$S_{IND} = social\ independence,$$
$$K_{IND} = Knowledge\ indepdence,$$
$$J = Journal\ EigenFactor$$

References

Benjamini, Yoav, and Yosef Hochberg. "Controlling the false discovery rate: a practical and powerful approach to multiple testing." *Journal of the Royal statistical society: series B (Methodological)* 57, no. 1 (1995): 289-300.

Büchter, Roland Brian, and Dawid Pieper. "Most overviews of Cochrane reviews neglected potential biases from dual authorship." *Journal of clinical epidemiology* 77 (2016): 91-94.

Lindson, Nicola, Samantha C. Chepkin, Weiyu Ye, Thomas R. Fanshawe, Chris Bullen, and Jamie Hartmann-Boyce. "Different doses, durations and modes of delivery of nicotine replacement therapy for smoking cessation." *Cochrane Database of Systematic Reviews* 4 (2019).

Handoll, Helen H., William J. Gillespie, Lesley D. Gillespie, and Rajan Madhok. "The Cochrane Collaboration: a leading role in producing reliable evidence to inform healthcare decisions in musculoskeletal trauma and disorders." *Indian journal of orthopaedics* 42, no. 3 (2008): 247.

Hoff, Peter D., Adrian E. Raftery, and Mark S. Handcock. "Latent space approaches to social network analysis." *Journal of the american Statistical association* 97, no. 460 (2002): 1090-1098.

Higgins, Julian PT, Douglas G. Altman, Peter C. Gøtzsche, Peter Jüni, David Moher, Andrew D. Oxman, Jelena Savović, Kenneth F. Schulz, Laura Weeks, and Jonathan AC Sterne. "The Cochrane Collaboration's tool for assessing risk of bias in randomised trials." *Bmj* 343 (2011): d5928.

Kivelä, Mikko. "Multilayer networks library for python (pymnet)." (2017).

Olsen, Ole, Philippa Middleton, Jeanette Ezzo, Peter C. Gøtzsche, Victoria Hadhazy, Andrew Herxheimer, Jos Kleijnen, and Heather McIntosh. "Quality of Cochrane reviews: assessment of sample from 1998." *Bmj* 323, no. 7317 (2001): 829-832.