Published in final edited form as:
Nutr Rev. 2014 December ; 72(12): 737–740. doi:10.1111/nure.12156.

Big data and systematic reviews in nutritional epidemiology

Ambika Satija and **Frank B. Hu**
Departments of Nutrition and Epidemiology, Harvard School of Public Health, Boston, MA 02115

Nutritional epidemiology today is characterized by "big data." The last 50 years have seen an accumulation of numerous large scale prospective studies, as well as countless smaller epidemiologic studies of varying quality, providing us with a wealth of information on the dietary underpinnings of a wide range of health outcomes, in particular chronic diseases.

Such a profusion of data necessitates the ability to summarize it into a cohesive body of evidence. Systematic reviews, carried out by following a pre-defined, systematic, reproducible methodology, constitute the preferred method of summarizing literature in nutritional epidemiology. These are usually divided into two categories, based on how data from individual studies are summarized. If the review provides a qualitative summary of evidence, it is simply called a systematic review (SR). If on the other hand a review quantitatively summarizes individual study results into one effect estimate, it is called a meta-analysis (MA). Several such summaries of evidence have shown us that diets characterized by a high intake of plant-source foods such as whole grains, fruits, vegetables, nuts, and tea & coffee, as well as of fish & low-fat dairy, along with low intakes of certain animal-source foods such as red and processed meats, as well as sugar-sweetened beverages and refined grains, are associated with reduced risk of several chronic diseases, including cardiovascular disease (CVD), diabetes, obesity, and certain cancers.(1-4)

While the concept of systematically reviewing & summarizing scientific evidence dates back to the early 20th century, it was introduced into the field of epidemiology relatively recently.(5) Despite this, developments in this methodology within the field of epidemiology have advanced rapidly, with the number of systematic reviews and meta-analyses published increasing at an astronomical rate; a PubMed search with the search strategy ["meta-analysis" or "systematic reviews"] results in 12,403 citations in 2013 alone, compared with just 1 citation in 1971. Narrowing this search further and looking for meta-analyses and systematic reviews within the field of nutritional epidemiology shows a similar rising trend, with 523 citations showing up in 2013 relative to just 1 in 1985. Keeping pace with this trajectory, overviews of systematic reviews and meta-analyses are becoming increasingly common in epidemiology – reviews of literature with systematic reviews and/or meta-analyses as their units of inclusion, as opposed to individual studies. Analogously to systematic reviews of studies, overviews of SRs and MAs can be classified as either 'qualitative overviews', which qualitatively summarize key findings of included reviews

Correspondence: Dr. Frank Hu, Department of Nutrition, Harvard School of Public Health, 665 Huntington Avenue, Boston, MA 02115, frank.hu@channing.harvard.edu.

Disclosures:
None

using text, tables and/or figures, or as ‘quantitative overviews’, which summarize the effect estimates of included reviews into a single summary effect estimate. In the social sciences, specifically education and psychology, conducting quantitative overviews of meta-analyses, called “meta-meta-analyses”, is a fairly common practice.(6, 7) In the field of epidemiology however, summarizing systematic reviews is a relatively new development, and has most commonly been done using the qualitative approach.(8-13)

In this issue of *Nutrition Reviews*, Fardet and Boirie(14) present a qualitative overview of pooled/meta-analyses and systematic reviews (PMASR) examining the dietary determinants of chronic diseases. Specifically, they systematically searched for and qualitatively summarized pooled/meta-analyses and systematic reviews published since 1950, that had examined the association between any one of 17 food/beverage groups, and any one of 10 chronic diseases, including overweight/obesity, type 2 diabetes, cardiovascular disease (CVD), cancers, mental and neurological disorders, musculoskeletal diseases, digestive tract diseases, liver diseases, and kidney diseases. What they found was not unexpected; plant-source foods, especially whole grains, are more convincingly associated with reduced risk of most diseases relative to animal-source foods, and red meat consumption is consistently associated with increased risk of various diseases, as is sugar-sweetened beverage intake. This study provides a holistic, comprehensive overview of the relationship between several dietary factors and key chronic diseases. An overarching study such as this one is also useful in identifying important gaps in knowledge. For instance, the review found that a majority of the PMASRs so identified had studied obesity, type 2 diabetes, CVD and cancers as outcomes, but relatively few had examined the dietary determinants of mental/neurological, musculoskeletal, digestive, liver and kidney diseases. Given that these diseases together account for 8% of deaths and 20% of DALYs,(15) and have certain biological pathways in common with the other more commonly studied diet related chronic diseases,(16) more research is needed to identify their dietary determinants.

The focus on food groups is also a key strength. This focus represents a shift in nutritional epidemiology away from the study of single nutrients towards the study of foods and food groups. Overviews of evidence such as the one performed by Fardet and Boirie form the basis of dietary recommendations, which would be easier to follow when framed in terms of food groups such as nuts and vegetable oils, than in terms of nutrients such as individual vitamins and minerals. A more recent development in nutritional epidemiology however, goes further beyond the study of isolated nutrients and foods towards the study of comprehensive dietary patterns – combinations of foods and beverages consumed together.(17) Such an approach overcomes several methodological shortcomings inherent in the single nutrient/food approach, and also has the above mentioned advantage of ease of translatability into dietary recommendations. The dietary pattern approach is not without its limitations, and hence cannot replace the single nutrient/food approach. Nevertheless, it serves to compliment the latter, and hence deserves its own summary of evidence.

The vastly accumulating body of literature in nutritional epidemiology makes apparent the need for overviews of meta-analyses and systematic reviews, like the one undertaken by Fardet and Boirie. However, such an undertaking comes with numerous challenges, especially when the literature is so expansive; hence it is important to understand the

potential methodological limitations of such overviews. A key drawback of a qualitative overview, such as the present one, is that it makes it difficult to compare effect estimates across multiple exposures, in this case food/beverage groups. A qualitative overview also has the disadvantage of requiring the use of methods of summarizing data which are potentially problematic. An example of such a method is 'vote counting', in which the number of positive studies is tallied up against the number of negative studies. This could lead to biased conclusions as there could be a larger number of positive studies merely due to publication bias – the tendency of statistically significant results to get published more often than non-significant findings. Another method of summarizing data is reporting the range of effect estimates (e.g. Odds Ratios) found for a given exposure-disease association, which the authors did in this overview. While this is a slight improvement on the vote counting method, it shares other limitations with it. Specifically, both methods do not give consideration to confidence intervals. Confidence intervals give us an idea of the extent of uncertainty surrounding effect estimates, which tend to be wider in smaller studies, and narrower in larger studies. Thus, these methods don't account for the varying sample sizes of individual studies included in the constituent reviews. Both methods also fail to give adequate consideration to the underlying heterogeneity in effect estimates, which might have systematic mechanisms driving it. A quantitative overview could resolve these issues, by providing a single effect estimate summarizing all included reviews, along with a confidence interval surrounding it. In a quantitative approach, formal statistical tools could also be used to gauge the extent of heterogeneity and publication bias in the included reviews.

Given the clear advantages of a quantitative approach, why haven't more overviews of reviews in epidemiology adopted this approach? Nutritional epidemiologic studies are usually characterized by a diversity of exposures, outcomes, analytic methods, and study populations. This heterogeneity across studies and perhaps meta-analyses makes the pooling of effect estimates difficult. Thus, while statistical tools for evaluating heterogeneity can be considered an advantage of quantitative overviews, such an overview would be inappropriate when there is considerable heterogeneity in individual reviews and their constituent studies. Heterogeneity can also be caused by varying study quality and differing magnitudes of biases such as confounding and selection bias. While the draw of a quantitative overview is the ability to summarize a vast range of literature into one single effect estimate, this estimate can be very misleading if the individual reviews included in the overview, and their constituent studies, are of low methodological quality. Conducting a meta-analysis of nutritional epidemiologic studies requires a clear understanding of methodological considerations unique to the field, deep content knowledge of the research question at hand, and a commitment to not let preconceived notions influence one's approach. Without these, the pooled effect estimate so obtained can be very misleading, as is evident from the contrary findings of recent meta-analyses of hitherto well characterized associations.(18, 19) Clearly, pooling together the effect estimates of such methodologically compromised meta-analyses in a quantitative overview can lead to erroneous conclusions that have the false credibility of 'big data' backing them. The perils of not considering study quality while formulating one's conclusions applies to qualitative overviews as well. In their review, Fardet and Boirie, erroneously concluded that coffee intake is positively associated with body mass index (BMI). This direction of association is usually observed in cross-

sectional studies, with well carried out prospective studies showing coffee to be inversely associated with weight gain.(20) Similarly, their conclusion that coffee is positively associated with CVD and certain cancers could be attributable to inadequate control for confounding by smoking, which is a strong, and often imperfectly measured confounder of this relationship.

There are other considerations which might make a quantitative overview difficult to undertake. A formal statistical method of pooling summary estimates from meta-analyses hasn't been developed as yet in epidemiology. Overviews of epidemiologic studies that have adopted a quantitative approach have extracted and pooled individual study data from each included meta-analysis.(21, 22) While this is a reasonable approach when there are a small number of meta-analyses with few studies included in each, it can become a time-consuming task as the scope of the literature expands. For the goal set out by Fardet and Boirie, it is nearly impossible. Even if a statistical method of directly pooling summary estimates did exist, there would still be the problem of double-counting – using data from individual studies more than once, for instance when one study contributes to the respective summary estimates of two or more meta-analyses. This would lead to an overly precise, and hence misleading summary effect estimate. It should be noted that double-counting could be an issue in qualitative overviews as well. However, the reliance on a single summary effect estimate makes this especially problematic for quantitative overviews. The only solution for this problem would be to comb through the included studies in each meta-analysis to assess the extent of double-counting, and then to extract and pool study-specific data. This brings us back to the original practical constraint with the existing methodology for conducting quantitative overviews.

Whether qualitative or quantitative, however, an overview of reviews should go beyond just a descriptive summary of evidence, especially in the face of significant heterogeneity. In fact, it is when the conclusions of several systematic reviews and meta-analyses differ, that an overview of reviews can be most meaningful. Such overviews can provide an insight into factors driving the heterogeneity in reviews.(23) These factors could be study-specific. For instance, 'meta-epidemiological' studies of meta-analyses have shown that unpublished trials tend to show a less beneficial treatment effect than published trials, while trials published in other languages relative to in English, trials not indexed in MEDLINE relative to those indexed in MEDLINE, and trials of poor quality (e.g. not double blinded versus double blinded) tend to show more beneficial treatment effects.(5) Factors driving heterogeneity could be review-specific as well. For instance, Katerndahl and Lawler (1999) (24) found that among meta-analyses summarizing the benefits of cholesterol reduction interventions, those that were methodologically better tended to report more beneficial effects of cholesterol reduction than methodologically inferior meta-analyses. Knowledge of the determinants of inconsistent findings can advance the field by providing a deeper understanding of the totality of evidence, and highlighting areas that need further study.

Given this background, what is the role of overviews of reviews, whether qualitative or quantitative, in nutritional epidemiology? The field of nutritional epidemiology has seen a substantial increase in systematic reviews and meta-analyses over the past 50 decades, in part due to the growing accumulation of nutritional epidemiologic studies during this time.

And the amount of data on the dietary determinants of chronic diseases will continue to increase at a very rapid pace due to existing and future large cohorts and modern technologies that will be increasingly used in collecting diet and nutrition data. A method of summarizing this vastly accumulating pool of data is crucially needed, so that clinicians, policy makers, and the academic community can better access an expanding evidence base. The overview of reviews by Fardet and Boirie represents an important contribution to the field. However, in order for nutritional epidemiology to progress with respect to overviews of reviews, formal guidelines detailing the steps needed to create a methodologically sound overview are needed. Some important steps in this direction include the Cochrane overview of reviews protocol,(25) and the guidelines set forth by Smith et al (2011).(26) These represent the beginnings of what will hopefully become a standardized methodology for summarizing reviews of evidence in nutritional epidemiology, and a step towards resolving the challenges faced by the discipline in managing big data.

Acknowledgments

Dr. Hu's research is supported by NIH grants HL60712, DK58845, and P30 DK46200.

REFERENCES

1. Esposito K, Kastorini CM, Panagiotakos DB, Giugliano D. Mediterranean diet and weight loss: meta-analysis of randomized controlled trials. Metabolic syndrome and related disorders. 2011; 9(1):1–12. Epub 2010/10/27. [PubMed: 20973675]
2. Hu FB. Plant-based foods and prevention of cardiovascular disease: an overview. The American journal of clinical nutrition. 2003; 78(3 Suppl):544S–51S. Epub 2003/08/26. [PubMed: 12936948]
3. Ley SH, Hamdy O, Mohan V, Hu FB. Prevention and management of type 2 diabetes: dietary components and nutritional strategies. Lancet. 2014; 383(9933):1999–2007. Epub 2014/06/10. [PubMed: 24910231]
4. Jenkins DJ, Kendall CW, Marchie A, Jenkins AL, Augustin LS, Ludwig DS, et al. Type 2 diabetes and the vegetarian diet. The American journal of clinical nutrition. 2003; 78(3 Suppl):610S–6S. Epub 2003/08/26. [PubMed: 12936955]
5. Egger M, Ebrahim S, Smith GD. Where now for meta-analysis? International journal of epidemiology. 2002; 31(1):1–5. Epub 2002/03/27. [PubMed: 11914281]
6. Kazrin A, Durac J, Agteros T. Meta-meta analysis: a new method for evaluating therapy outcome. Behav Res Ther. 1979; 17(4):397–9. Epub 1979/01/01. [PubMed: 486043]
7. Hattie, J. Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement. 1st ed.. Routledge; New York: 2009.
8. Clarke M. Systematic review of reviews of risk factors for intracranial aneurysms. Neuroradiology. 2008; 50(8):653–64. Epub 2008/06/19. [PubMed: 18560819]
9. Flodgren G, Eccles MP, Shepperd S, Scott A, Parmelli E, Beyer FR. An overview of reviews evaluating the effectiveness of financial incentives in changing healthcare professional behaviours and patient outcomes. The Cochrane database of systematic reviews. 2011; (7):CD009255. Epub 2011/07/08. [PubMed: 21735443]
10. Jones L, Othman M, Dowswell T, Alfirevic Z, Gates S, Newburn M, et al. Pain management for women in labour: an overview of systematic reviews. The Cochrane database of systematic reviews. 2012; 3:CD009234. Epub 2012/03/16. [PubMed: 22419342]
11. Keus F, Gooszen HG, van Laarhoven CJ. Open, small-incision, or laparoscopic cholecystectomy for patients with symptomatic cholecystolithiasis. An overview of Cochrane Hepato-Biliary Group reviews. The Cochrane database of systematic reviews. 2010; (1):CD008318. Epub 2010/01/22. [PubMed: 20091665]

12. Smith V, Devane D, Begley CM, Clarke M, Higgins S. A systematic review and quality assessment of systematic reviews of fetal fibronectin and transvaginal length for predicting preterm birth. European journal of obstetrics, gynecology, and reproductive biology. 2007; 133(2): 134–42. Epub 2007/04/25.
13. Smith V, Devane D, Begley CM, Clarke M, Higgins S. A systematic review and quality assessment of systematic reviews of randomised trials of interventions for preventing and treating preterm birth. European journal of obstetrics, gynecology, and reproductive biology. 2009; 142(1): 3–11. Epub 2008/11/11.
14. Fardet A, Boirie Y. Food groups and beverages against major diet-related chronic diseases: a state of the art based on an exhaustive review of pooled/meta-analyses and systematic reviews. Nutrition Reviews.
15. Craig WJ, Mangels AR. Position of the American Dietetic Association: vegetarian diets. Journal of the American Dietetic Association. 2009; 109(7):1266–82. Epub 2009/07/01. [PubMed: 19562864]
16. Fardet A, Boirie Y. Associations between diet-related diseases and impaired physiological mechanisms: a holistic approach based on meta-analyses to identify targets for preventive nutrition. Nutr Rev. 2013; 71(10):643–56. Epub 2013/10/15. [PubMed: 24117841]
17. Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. Current opinion in lipidology. 2002; 13(1):3–9. Epub 2002/01/16. [PubMed: 11790957]
18. Chowdhury R, Warnakula S, Kunutsor S, Crowe F, Ward HA, Johnson L, et al. Association of Dietary, Circulating, and Supplement Fatty Acids With Coronary RiskA Systematic Review and Meta-analysis. Annals of Internal Medicine. 2014; 160(6):398–406. [PubMed: 24723079]
19. Flegal KM, Kit BK, Orpana H, Graubard BI. Association of all-cause mortality with overweight and obesity using standard body mass index categories: A systematic review and meta-analysis. JAMA. 2013; 309(1):71–82. [PubMed: 23280227]
20. Pan A, Malik VS, Hao T, Willett WC, Mozaffarian D, Hu FB. Changes in water and beverage intake and long-term weight changes: results from three prospective cohort studies. International journal of obesity. 2013; 37(10):1378–85. Epub 2013/01/16. [PubMed: 23318721]
21. O'Connell NE, Wand BM, McAuley J, Marston L, Moseley GL. Interventions for treating pain and disability in adults with complex regional pain syndrome. The Cochrane database of systematic reviews. 2013; 4:CD009416. Epub 2013/05/02. [PubMed: 23633371]
22. Singh JA, Christensen R, Wells GA, Suarez-Almazor ME, Buchbinder R, Lopez-Olivo MA, et al. Biologics for rheumatoid arthritis: an overview of Cochrane reviews. The Cochrane database of systematic reviews. 2009; (4):CD007848. Epub 2009/10/13. [PubMed: 19821440]
23. Delgado-Rodriguez M. Systematic reviews of meta-analyses: applications and limitations. Journal of epidemiology and community health. 2006; 60(2):90–2. Epub 2006/01/18. [PubMed: 16415253]
24. Katerndahl DA, Lawler WR. Variability in meta-analytic results concerning the value of cholesterol reduction in coronary heart disease: a meta-meta-analysis. American journal of epidemiology. 1999; 149(5):429–41. Epub 1999/03/06. [PubMed: 10067902]
25. Becker, LA.; AD, O. Chapter 22: Overviews of reviews. In: Higgins, JPT.; Green, S., editors. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.0.1 [updated September 2008]. The Cochrane Collaboration; 2008. Available from www.cochrane-handbook.org
26. Smith V, Devane D, Begley CM, Clarke M. Methodology in conducting a systematic review of systematic reviews of healthcare interventions. BMC medical research methodology. 2011; 11(1): 15. Epub 2011/02/05. [PubMed: 21291558]