Commentary: Towards a definite coherent heterogeneity in meta-analyses

Jeanine M Genkinger1,2* and Mary Beth Terry1,2

1Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, USA and 2Herbert Irving Comprehensive Cancer Center, New York, USA

*Corresponding author. Department of Epidemiology, Mailman School of Public Health, 722 W 168th St, New York, NY, 10019, USA. E-mail: jg3081@columbia.edu

‘Meta-analysis’, coined by Gene Glass in 1976, is the ‘analysis of analyses’, denoting a systematic approach to summarizing the available peer-reviewed literature using statistical techniques.1 Most readers view a meta-analysis as a definitive summary of the state of the evidence, and meta-analyses have become an essential element in evidence-based decision making. Because meta-analysis involves a quantitative summary of the published data, it is sometimes interpreted as less biased and more objective than qualitative reviews of the literature or even individual studies. Over the past two decades the use of systematic reviews has exploded, with over 4000 published on human data in 2013 alone (as inventoried by PubMed).

The power of meta-analyses is realized when they provide clarity on the exposure and outcome association not revealed through individual studies. For example, in a meta-analysis on therapeutic trials for myocardial infarction, treatment with intravenous streptokinase reduced mortality from acute myocardial infarction by approximately 20% (P-value < 0.001).2 In contrast to the overall finding from the meta-analysis, 28 of the 33 clinical trials did not observe a statistically significant finding between treatment with intravenous streptokinase for acute myocardial infarction; the individual study results were heterogeneous, with suggestions of inverse, null and positive findings, thus illustrating how a meta-analysis clarified an important causal factor advancing clinical care.

Nevertheless, meta-analyses, just like their individual parts, are only as valid as the data and methods used (including design aspects, eligibility criteria and analyses). Thus, it is entirely possible to see multiple meta-analyses on the same topic with different conclusions, if there are different data and methods represented in each individual meta-analysis. Fortunately, most epidemiologists embrace heterogeneity and its determinants in evaluating and synthesizing individual studies. The proliferation of meta-analyses, however, requires constant vigilance and recognition of the fact that methodological critique extends to describing and understanding heterogeneity in meta-analyses. Without such a critical evaluation to explain the heterogeneity across meta-analyses, public health and clinical care decisions will not be informed by the best empirical evidence.

As an illustration of the task we face, in this issue of IJE we have an example of a well-conducted meta-analysis on adiponectin concentrations and breast cancer risk by Macis and colleagues.3 As two other meta-analyses on this same topic have been published recently in other journals4,5 with some differing conclusions, particularly for the selected subgroups, we compare these three meta-analyses as an illustration of heterogeneity in meta-analyses, and we articulate and distil the differences across them for a more coherent understanding of the evidence.

© The Author 2014; all rights reserved. Published by Oxford University Press on behalf of the International Epidemiological Association
Macis and colleagues\textsuperscript{3} report on meta-analysis of 15 studies on the association between adiponectin concentrations and breast cancer risk. Specifically, they observed a 34\% reduction in breast cancer risk [95\% confidence interval (CI) = 13-50\%] for those with the highest compared with the lowest circulating adiponectin concentrations. Compared with the overall results, risk estimates for both premenopausal women (0.70, 95\% CI = 0.28–1.76), based on four studies, and postmenopausal women (odds ratio (OR) = 0.80, 95\% CI = 0.63–1.01), based on six studies, were similar in effect size to the overall findings but were not statistically significant. The similar effect sizes across menopausal groups suggest that mechanisms linking adiponectin levels to breast cancer are similar for all women regardless of menopausal status.

This meta-analysis by Macis and colleagues\textsuperscript{3} was well designed and conducted and followed the guidelines set forth for conducting and reporting meta-analyses.\textsuperscript{6} If this were the only meta-analysis on the topic, we might be tempted to call the matter to a close and conclude that adiponectin concentrations play a similar role in the aetiology of pre- and postmenopausal breast cancer, given the similar effect sizes, but recognizing the limitations of the small sample size to examine subgroups even in meta-analyses. This conclusion could be justified as these results are supported by a body of evidence that has shown that low adiponectin concentrations are related to insulin resistance, as well as other properties or mechanisms that may predispose to breast cancer (e.g. obesity, hormone regulation, inflammation, angiogenic).\textsuperscript{7–14}

The challenge in interpretation, however, is that it is not the only meta-analysis on the topic; two other meta-analyses on the same topic were published in 2013, with different conclusions about premenopausal breast cancer. One of these other meta-analyses, by Liu and colleagues,\textsuperscript{4} concluded that although high adiponectin levels were associated with reduced postmenopausal breast cancer risk (OR = 0.75, 95\% CI = 0.60–0.94), based on 13 studies, there was no association with premenopausal breast cancer (OR = 0.90, 95\% CI = 0.64–1.26). The other meta-analysis, by Ye and colleagues,\textsuperscript{5} a meta-analysis of eight studies, similarly reported that risk was limited to postmenopausal women [standard mean difference (SMD) between cases and controls = −0.39 µg/ml, 95\% CI = −0.619 to −0.161], with no association among premenopausal women (SMD = 0.02 µg/ml, 95\% CI = −0.164 to 0.204). Placing the Macis finding within the context of these two existing meta-analyses, without any comparison of the drivers of the heterogeneity, would mean that we might be tempted to conclude that adiponectin concentrations are truly not related to premenopausal breast cancer, as two of the three meta-analyses made this conclusion, suggesting a different mechanism for adiposity and premenopausal breast cancer risk.

These heterogeneous conclusions beg us to dig deeper to understand the drivers of the inconsistencies across these meta-analyses. The main differences can arise in the following: (i) search strategy and eligibility criteria applied for inclusion; (ii) exposure constructs and outcome measures assessed (e.g., absolute vs relative effects); (iii) statistical methods employed (e.g. random vs fixed effects); and (iv) subgroup assessment.

Search strategy

Although each of the three meta-analyses assessed the association between adiponectin and breast cancer risk during a similar time frame, each meta-analysis identified and included different studies. Specifically, 17 unique studies were included in at least one of the meta-analyses; only seven studies were in common across all three. Whereas all three used PubMed as a search engine, they also all used other search engines to identify studies for inclusion (e.g. EMBASE, Medline). All three articles also included ‘adiponectin’ and ‘breast cancer’ as search terms, but Macis et al.\textsuperscript{3} and Liu et al.\textsuperscript{4} also included terms such as ‘ACDC’, ‘ADPN’, ‘APM1’, ‘APM-1’ that relate to single nucleotide polymorphisms (SNPs) associated with adiponectin concentrations. Further, Ye et al.\textsuperscript{5} also included eight studies that assessed standardized mean differences, and included studies only presented in abstract form. The included studies for Ye et al. mostly focused on studies prior to 2008, except for one later study published in 2010;\textsuperscript{15} that of Ye and colleagues was the only meta-analysis to include a publication which only reported mean values.\textsuperscript{16} Thus, Macis et al.\textsuperscript{3} (15 studies for total breast cancer) and Liu et al.\textsuperscript{4} (13 studies for total breast cancer) were the most consistent with regard to their study inclusion and had 12 overlapping studies. These slight differences in search engines and search terms, along with variations in eligibility criteria, resulted in the inclusion of different studies in the three meta-analyses. Thus, based on this criterion alone, both the Macis and the Liu meta-analyses\textsuperscript{3,4} seem more complete and argue for giving more weight to these two meta-analyses in our overall interpretation. However, these two meta-analyses also need to meet high quality standards on other essential data and design features.

Exposure constructs and outcome assessments

The different inclusion criteria resulted in the inclusion of different risk estimates in each meta-analysis. However, even for studies that were included in all three, the different meta-analyses sometimes extracted different
information on exposure constructs (e.g. means vs quantiles) and used different measures of association to evaluate differences (e.g. mean differences vs relative ratios). The most distinct difference across the three meta-analyses was that Ye et al. included standardized mean differences as their measure of effect, whereas Macis et al. and Liu et al. used relative measures of differences (e.g. odds ratios, hazard ratios). However, even when comparing Macis et al. and Liu et al., who had similar inclusion criteria and study inclusion, different estimates of risk were used in five of the 12 overlapping studies. The differences in risk estimates that are included in each of these meta-analyses contributed to the differences in summary risk estimates observed. This is further compounded by the fact that meta-analyses rely on the results as presented in the publication. Thus, heterogeneity in how the exposures, covariates and outcomes are defined and modelled in each individual publication can occur and can lead to the combining of estimates that are quite different in their exposure construct and result in differences in summary risk estimates. The only way to overcome these differences in exposure constructs and outcome assessment may ultimately be through pooled analyses of individual data. Unlike meta-analyses of published data which rely on the results as presented in the publication, in pooled analyses of individual data, potential sources of heterogeneity are removed across the studies included by standardizing how the exposures, covariates and outcomes are defined and modelled. For example, this pooling approach allows flexibility to uniformly categorize the exposures and covariates across studies into quantiles, by common absolute cut-points, or as continuous variables. Pooling of individual data also ensures that subgroup analyses can be evaluated with all of the studies that collected the data rather than just the studies that report on the subgroups in their publication.

Statistical analyses

Across the three meta-analyses, different statistical analysis plans were employed for pooling the data for a summary estimate. Whereas Macis et al. pooled using a random effects model, both Liu et al. and Ye et al. employed a combination of random and fixed effect models as deemed appropriate or only employed the random effects model if statistically significant heterogeneity was observed. Fixed effect models assume that there is one true effect or a common effect across all studies; thus, only the variation within studies, and not between studies, is incorporated. In contrast, random effect models assume that each study’s true effect is randomly distributed around a central effect. To account for this, both the

between and the within study variance are incorporated into the statistical weights needed to construct a summary estimate. This inclusion results in more conservative estimates or wider confidence intervals in random effect models. As these meta-analyses include observational data from vastly distinct populations from different geographical regions with different age ranges and percentages of postmenopausal women, use of a random effects model, as in the Macis study, may have been a more appropriate statistical approach.

Subgroup analyses

Finally, all three studies assessed potential modification of the association between adiponectin concentrations and breast cancer risk by various reported factors. Ye et al. examined effect measure modification by fasting blood glucose, body mass index, physical activity, alcohol intake, geographical area and laboratory assays. Stronger inverse summary estimates were observed for studies where they did not adjust for alcohol intake compared with those that did, those that were conducted in Asia compared with Europe, and within studies that measured adiponectin using enzyme-linked immunosorbent assay (ELISA) compared with radioimmunoassay (RIA). Liu et al. only examined effect measure modification by geographical region; no differences were observed when examining Asia, Europe, American and Australasia. Macis et al. examined effect measure modification by study design and laboratory methods; although there was no difference observed by laboratory methods, they observed a stronger association among case-control studies in comparison with cohort and nested case-control studies.

Most importantly, the Macis et al. study suggested that the inverse association between adiponectin and premenopausal breast cancer risk exists, as the effect estimates for both the pre- and the postmenopausal groups were very similar, whereas Liu et al. and Ye et al. concluded that no association exists for premenopausal women. As Macis et al., compared with Liu et al., included three additional studies, it is important to note that two of the three studies were conducted in populations in which more of the women were premenopausal. Thus the Macis et al. meta-analysis provides more evidence to evaluate the association in premenopausal women.

Overall, individual studies do not have enough power and/or sample size to adequately address a number of research questions. Meta-analyses can overcome this limitation of individual studies, attaining higher statistical power by pooling the individual results. The combining of results also enables identification of heterogeneity between studies and the factors driving this heterogeneity.
However, results by subgroup need to be published within the original manuscripts in order to be appropriately assessed within the meta-analysis. What we observed across these three meta-analyses published within 1 year of each other is that they reported different summary estimates and corresponding conclusion statements, particularly with regard to risk in premenopausal women. After evaluating the three on search criteria, exposure constructs and outcomes assessed, statistical approach and subgroup analyses, we conclude that the Macis\(^3\) article provides the best evaluation of this rapidly evolving literature.

Moving forward, we should not rely on each individual study or meta-analysis as the ‘truth’, but instead we should evaluate and review each meta-analysis as we evaluate and review an original research article, with regard to its design, analysis and conclusion. As Herbert Spencer theorized that ‘evolution is an integration of matter and concomitant dissipation of motion during which the matter passes from an indefinite incoherent homogeneity to a definite coherent heterogeneity’,\(^{26}\) the differences between these meta-analyses have become more coherent from evaluating the heterogeneity in the results. This principle is demonstrated and tested when a meta-analysis is conducted to examine the causal relationship between an exposure and disease; the relationship becomes more coherent as we begin to understand the underlying heterogeneity between all the prior research studies that addressed this question. However, for this coherence to be realized, the meta-analysis needs to be well designed, conducted systematically and carried out with an understanding of the methodological concerns surrounding the topic, the individual study designs and the meta-analytical techniques.

**Conflict of interest:** None declared.

**References**