Faculty Op-ed

Less is not more when it comes to evidence-based policy - To solve policy issues, we need the right studies

All hail the systematic review! The gold standard of science! The top of the evidence pyramid!

Anyone that has worked in – or adjacent to – any public policy domain is likely familiar with the idea of a systematic review: a study of studies in which evidence from many different sources is synthesized into uniform comparisons. One common approach for these papers to directly compare many treatments for the same condition (e.g., "Which treatment has the greatest effect on X?"). Another common one is to look at studies testing the same treatment or observing the same concept to look for an overall average (e.g., "Across 25 studies, what was the average result for X?").

Seems like a great idea, right? It is only natural that, with greater global resources and understanding for conducting science, we should aim to learn from as many studies as possible in a way that allows for direct comparison. Accordingly, this is also how many policy researchers and advisers have approached systematic reviews (and their primary comparisons, known as "meta-analyses"). Entire networks have been established to deliver on this promise, and in many cases, have delivered. Unfortunately, the promise of systematic reviews also comes with concerns.

Early in my career, I was asked to coordinate the systematic review of treatments for relapsed or refractory multiple myeloma. In going through hundreds of papers, it became clear that a single comparison was never going to be appropriate. This was no fault of any researchers whose work we were reviewing; each of them had their own set of objectives and typically a small group of patients to work with. Each of those patients had very complex care needs. Yet we were asked to make a direct comparison. [We shared our concerns(link is external and opens in a new window)], and [similar critiques(link is external and opens in a new window)] have been raised in recent years, ranging from how some reviews [may simply compile a lot of bad studies(link is external and opens in a new window)] into a single analysis, or that [multiple small studies(link is external and opens in a new window)] do not automatically add up to one valuable one.

But there is actually an even broader problem with treating systematic reviews as the pinnacle of evidence in policy: does the evidence even help policymakers? Sure, if the FDA needs to review a drug for a specific disease or the NHTSA needs to know how many crashes are caused by distracted drivers each year, those very specific questions could be

informed by systematic reviews. However, such simplicity in questions is more likely the exception than the norm.

In working with policymakers, I increasingly heard feedback about studies not recognizing what information was most valuable to them. To a researcher, a systematic review of randomized controlled trials seems ideal. To a policymaker, it may be more useful to know how the public might react to a new approach or just existing reports from government agencies that have tried the same approach before.

[So last year, we decided to take a more formal approach to finding out what evidence truly helps policymakers(link is external and opens in a new window)](). We invited thousands of policymakers from all levels of government in the US, Canada, and Europe to rate 30 different types of evidence that may get considered for policy. What we found in doing that was quite remarkable and confirmed those suspicions: what counts as the most useful type of evidence hugely depends on the policy domain and policymaker.

Across those 30 types of evidence, systematic reviews were only the fifth most highly rated type – and even quite a few policymakers indicated that systematic reviews are of no value to the work they do. Furthermore, things like focus groups, business cases, and public opinion polls were all rated as typically important across policy domains. Yet such types of evidence are rarely considered as valuable as meta-analyses of systematic reviews within science.

To be sure, systematic reviews are valuable and should absolutely be considered when appropriate. The point of this article is not to say otherwise, but to encourage broader thinking about ways we synthesize evidence. To demonstrate this, we followed up the survey (in the same paper) with a review of policy outcomes based on the types of evidence that informed them originally. We found that policy outcomes were much more likely to be predicted when the scale of evidence was greater prior to the policy. In some cases, that did mean systematic reviews had been available in advance; in others, there had been sufficiently large or otherwise highly powered studies available.

The lesson we took from all of this was that there is no single type of evidence that is inherently best-suited for policy. Systematic reviews – [when appropriate(link is external and opens in a new window)]() – can be great for informing policy decisions. However, what really matters most is the scale and quality of evidence – and having more of that evidence will go further in setting realistic expectations for what policies can deliver to benefit the well-being of populations.

– Kai Ruggeri, June 2024