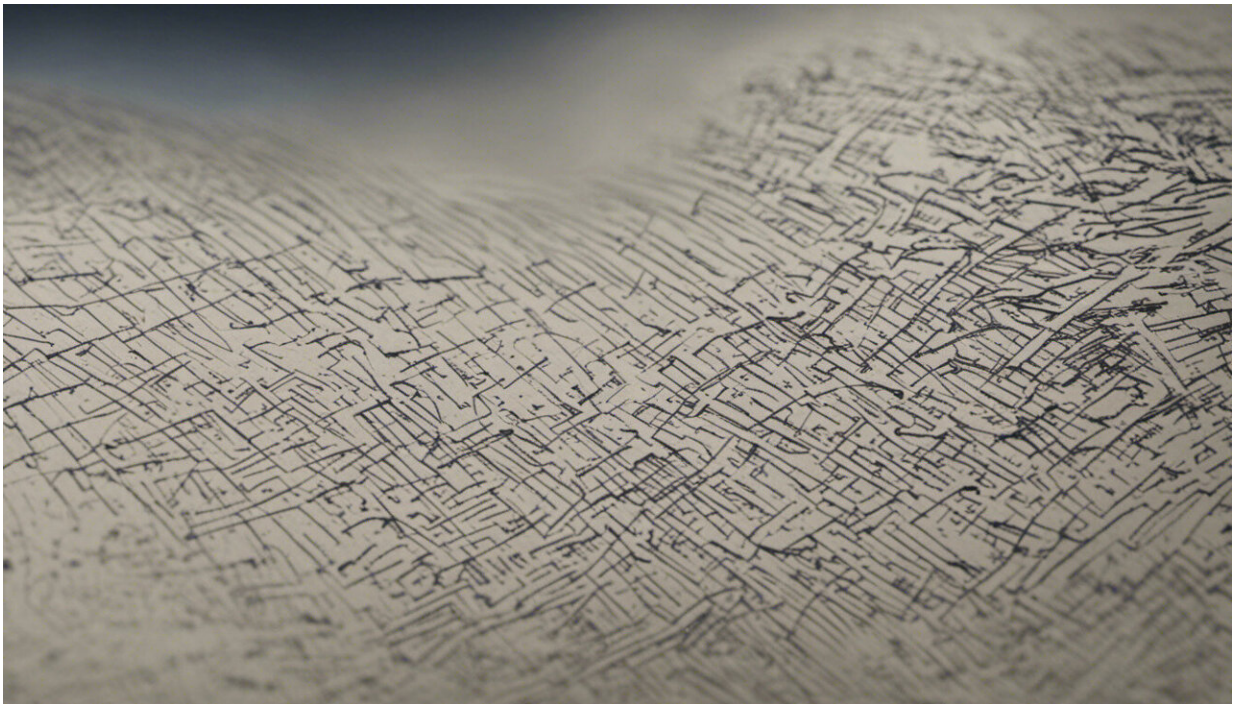


How to boil down a pile of diverse research papers into one cohesive picture

May 2 2017, by Mohammad S. Jalali



Credit: AI-generated image ([disclaimer](#))

From social to natural and applied sciences, overall scientific output has been growing worldwide – it [doubles every nine years](#).

Traditionally, researchers solve a problem by conducting new experiments. With the ever-growing body of scientific literature, though,

it is becoming more common to make a discovery based on the vast number of already-published journal articles. Researchers synthesize the findings from previous studies to develop a more complete understanding of a phenomenon. Making sense of this explosion of studies is critical for scientists not only to build on previous work but also to push research fields forward.

My colleagues [Hazhir Rahmandad](#) and [Kamran Paynabar](#) and I have developed a new, more robust way to pull together all the prior research on a particular topic. In a five-year joint [project](#) between MIT and Georgia Tech, we worked to create a new technique for research aggregation. Our recently published paper in PLOS ONE introduces a flexible method that [helps synthesize findings from prior studies](#), even potentially those with diverse methods and diverging results. We call it [generalized model aggregation](#), or GMA.

Pulling it all together

[Narrative reviews](#) of the literature have long been a key component of scientific publications. The need for more comprehensive approaches has led to the emergence of two other very useful methods: [systematic review and meta-analysis](#).

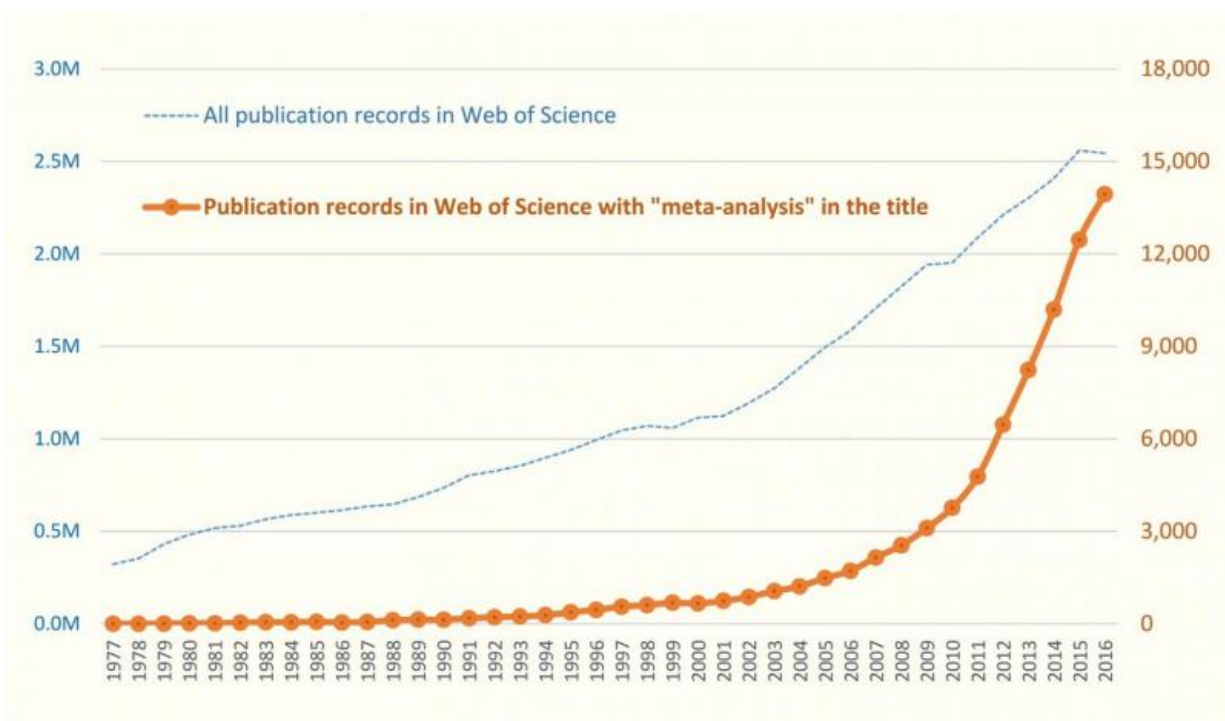
In a systematic review, an author finds and critiques all prior studies around a similar research question. The idea is to bring a reader up to speed on the current state of affairs around a particular research topic.

In a meta-analysis, researchers go one step further and synthesize the findings quantitatively. Essentially, it takes a weighted average of the findings of several studies on one topic. Pooling results from multiple studies is meant to generate a more reliable finding than that of any single study. This is crucially helpful when prior studies reported diverging findings and conclusions. And the rise in the publications of

meta-analysis has shot up over the last decade, underscoring their importance across research communities.

Meta-analysis has been helpful in increasing our understanding of many scientific problems. But it has some challenges. [A typical meta-analysis](#) combines just one explanatory variable (that is, a treatment controlled by the experimenter) and one response variable (for instance, a health outcome). Also, a researcher has to be very careful not to lump apples and oranges together in the meta-analysis. She must be selective and make sure to include only previous work that shared a very similar study design.

Here is where our simple and flexible generalized model aggregation method comes in. Using GMA, the prior studies do not necessarily need to have the same study design or method. They can also have different explanatory variables. As long as they are all answering a similar research question, GMA can synthesize them.



Publications of meta-analyses are on the rise, based on Web of Science search results for articles that included the term ‘meta-analysis’ in their title. Credit: Mohammad S. Jalali, CC BY-ND

Pooling findings from across a field

Consider an example from the health literature. Obesity and nutrition researchers need reliable equations that estimate basal metabolic rate (BMR) – the amount of energy the human body spends at complete rest. Understanding BMR has big implications for real-world questions of weight management.

Researchers often estimate BMR as a function of different attributes: age, height, weight, fat mass and fat-free mass. The challenge is that current publications in research journals [provide over 200 such equations](#) estimated for different samples and age groups. These equations also include different subsets of those attributes.

For example, one of these equations included weight and age, but another included only fat-free mass. Another equation considered the impact of all these attributes, but the sample size was too small to make it reliable. More interestingly, and confusingly, there have been several studies with similar samples and variables but they have reported very different equations to explain the relationships.

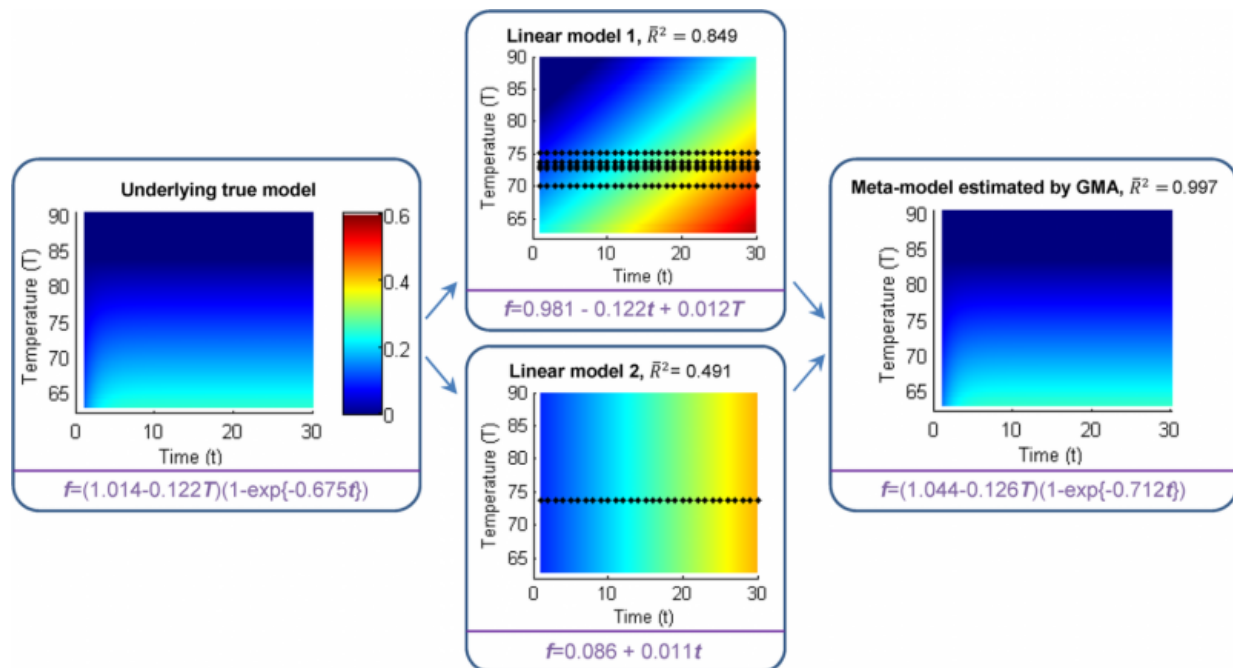
So which equations are you going to choose to accurately estimate BMR? How do you ensure that your selected equation is more reliable than the rest?

In order to address these questions, [we identified 27 published BMR equations](#) for white males from published studies. Then we used GMA to aggregate them into a single [equation](#), which we called a meta-model.

Through validation tests, we showed that our meta-model is more precise than any of the prior equations for estimating BMR. It also can deal with a logarithmic relationship between two variables – something not captured by any of the original 27 linear equations.

We tested our method by putting it up against more complex situations. What if all the equations we aggregate using GMA are actually off the mark? Would GMA still get close to what is really going on?

To investigate, we imagined two researchers coming up with two different linear equations to describe what they did not realize is actually a nonlinear phenomenon. The findings of the two researchers are far from reality. But again, our meta-model provided an extremely close estimate of reality – even when aggregating these two incorrect and biased models.



The meta-model (on the right) relies only on reported information from the two incorrect models in the middle – not their observed data or the true data. And it is much closer to reality (on the left) than either incorrect model. Credit: Rahmandad et al, DOI: 10.1371/journal.pone.0175111, CC BY

How GMA gets at the truth

So how does it all work? There is no magic here. In fact, the [intuition behind GMA is simple](#), which lets researchers with no extensive statistical background use it.

Broadly, each previous empirical study is an attempt to estimate an underlying reality. Let's call this the "true model." And it is unknown to us; whatever is actually driving the phenomenon under investigation is nature's secret. The empirical studies report relevant information about the true model, even if they are biased or incomplete.

Generalized model aggregation uses computer simulations to replicate prior studies. This time, though, the simulated studies attempt to estimate a meta-model instead of the true model (that is, reality).

We feed the empirical studies' reported estimates into the simulation. The flexibility of the GMA allows us to also use any other additional information about the underlying true model, too – such as the relationships among the variables or the quality of empirical studies' estimates. This extra information helps increase the reliability of GMA estimates.

The GMA algorithm carefully applies the same sample characteristics to each previous study and replicates their same method. Then it compares the outcomes of the simulated studies with the actual results of the [empirical studies](#), trying to find the closest match. Through this matching process, GMA estimates the meta-model.

If the simulated and actual outputs match, the meta-model may be a good representation of the true model – that is, by running a bunch of studies through the GMA algorithm, we are able to tease out a closer approximation of how the phenomenon in question actually works.

Wide range of applications for GMA

In our paper, we [discussed a wide range of examples](#), from health to climate change and environmental sciences, that can benefit from generalized model aggregation. Using GMA to synthesize prior findings into a coherent meta-model can increase the accuracy of aggregation.

In the current replicability crisis, GMA can help not only identify studies that are reproducible, but also distinguish reliable findings from less robust ones.

We reported [all the steps of our analysis](#) for further replication. A recipe for using GMA and its codes, along with instructions, is also [publicly available](#).

We hope that GMA can extend the reach of current research synthesis efforts to many new problems. GMA can help us understand the bigger picture of phenomena by aggregating their parts. Consider a puzzle with its pieces scattered about; the overall picture is revealed only when the pieces have been put together.

This article was originally published on [The Conversation](#). Read the [original article](#).

Provided by The Conversation

Citation: How to boil down a pile of diverse research papers into one cohesive picture (2017, May 2) retrieved 14 August 2024 from <https://phys.org/news/2017-05-pile-diverse-papers-cohesive-picture.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--