Behavioural science

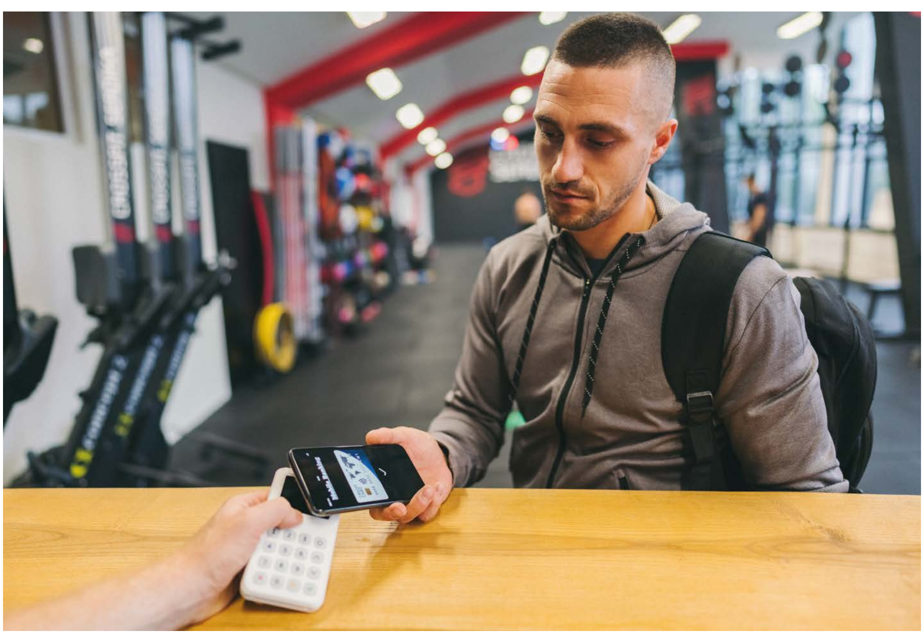
Benefits of a megastudy approach

Heather Royer

Trials of behavioural interventions are hard to compare, hampering policy decision-making. The effects of more than 50 interventions on exercise behaviour have been compared using an experimental design called a megastudy. See p.478

Over the past few decades, empirical work in the behavioural and social sciences has undergone a credibility revolution — a movement to use data to distinguish causality from correlation. Not surprisingly, one of the leading tools in this revolution is the randomized controlled trial (RCT). In these trials, participants are randomly allocated to groups that receive, or do not receive, an intervention. However, the conclusions that can be drawn from RCTs of behavioural interventions can be limited, and the results of different RCTs are difficult to compare. Milkman *et al.*[1] show on page 478 how a 'megastudy' — a collection of simultaneous RCTs — could overcome some of the limitations of individual RCTs. The authors' megastudy tested the effectiveness of 54 behavioural interventions on gym attendance.

Although RCTs are often described as the gold standard for estimating the causal effects of interventions, critics say that their usefulness for guiding future research and policymaking is limited, because they usually test only a narrow set of hypotheses. Therefore, the evidence often comes from a large collection of research, including separate RCTs and other study types, that are inherently difficult to compare. For example, one team of researchers might study how giving membership subsidies increases gym attendance at a private gym in Boston, Massachusetts, whereas another research group might investigate the effect of offering gym-attendance incentives to students visiting a university gym in southern California. Because the circumstances — including the location and participants — of these two studies are not the same, it is impossible to conclude whether the results from these studies diverge because of differences in the populations studied or in the efficacy of the interventions. This ambiguity makes policy decisions difficult.



GETTY

Figure 1 | A gym user making a digital payment. The effects of interventions to increase exercise behaviour can be relatively straightforward to track — for example, using electronic records of gym attendance or of digital payments for gym visits.

Milkman and co-workers propose the use of megastudies to address these challenges. Their team of researchers from various academic disciplines simultaneously compared and contrasted a control treatment and 53 interventions aimed at bolstering the exercise habits of 61,293 members of a chain of gyms in the United States. The interventions involved various combinations of prompts to encourage the participants to plan gym visits; text-message reminders of scheduled plans; and micro-incentives (points redeemable on an e-commerce website). Interventions also included other features, such as bonus micro-incentives when returning to the gym after missing a planned workout. The authors found that 45% of the tested interventions had a positive effect on gym attendance.

The domain of exercise is well suited for such a megastudy approach, because changes in exercise behaviour in response to intervention can occur rapidly and be tracked through objective, high-frequency outcome measures — for example, number of gym visits (Fig. 1). Moreover, there are sizeable margins for improvement in behaviours, as reflected by the gap between exercise intentions and actual exercise habits[2]. According to the authors' definitions of a megastudy, large numbers of participants from a common pool were allocated to each of the interventions; the interventions had the same 'treatment' and 'follow-up' periods (the periods when and after the intervention was applied), and a standard set of metrics was used to evaluate the effectiveness of the interventions.

If researchers could accurately predict which interventions would be most effective when designing RCTs, there would be less need for a megastudy. However, Milkman and colleagues' data suggest that the efficacy of interventions is difficult to predict. They asked behavioural scientists, public-health researchers and non-expert, online-survey respondents to predict which interventions would be effective, and found a surprising lack of correlation between the predicted and actual effects of the interventions. This finding highlights one of the main benefits of a megastudy: it is challenging to identify interventions that are likely to be successful before an RCT and, therefore, trying many interventions at once — as was done in the megastudy — could accelerate scientific discovery.

Unfortunately, however, megastudies might be too costly to study the effects of larger incentives. The micro-incentives offered in Milkman and colleagues' study are between one-fifteenth and one-fiftieth the monetary value of incentives used in other work (see, for example, refs 3–7) and are given as points for use on an e-commerce website that might have a lower value than their cash equivalent. Yet, offering larger incentives could make a megastudy approach unfeasible. For example, replicating the size of the incentives used in a previous study of the effect of incentivizing exercise[7], but using the sample size of Milkman and co-workers' study, would require about US$2.6 million for the incentives alone.

The multi-armed nature of megastudies might influence the types of intervention selected for testing. During the early stages of creating an RCT proposal, a researcher often drafts and deliberates over a long list of potential interventions, before choosing a few to test, on the basis of previous results or hypotheses. The researcher might pursue an intervention that they are confident will work, in the hope of obtaining a positive effect, while excluding untested interventions that are very unlikely to have an effect, but that still have a small chance of being highly successful. When selecting interventions to include in a megastudy, however, the researcher might decide to include these more-risky interventions. The researcher might reason that the likely failure of a riskier intervention could be worthwhile because of the likelihood that at least one of the other tested interventions has a positive effect.

It is unclear whether such a change in intervention-selection strategy would be better or worse for the advancement of science. In extreme cases, it could lead researchers to pursue outlandish interventions that have little hope of success, leaving the less-risky interventions under-studied. Alternatively, megastudies could encourage the investigation of more-innovative treatments.

Although megastudies push the frontier of RCTs, caution must be taken when interpreting their findings. Assuming that the best-performing intervention in one megastudy is necessarily the most promising of the tested interventions is ill-advised. First, this conclusion can suffer from what is known as a 'winner's curse' bias. That is, as demonstrated in statistics[8], the true effect of the best-performing intervention is likely to be smaller than the effect measured in one study — which means that, when the study is replicated, the 'best' treatments are likely to differ. The winner's-curse bias becomes more acute with the number of treatments being studied and therefore could be particularly amplified in a megastudy setting[8].

Second, megastudies do not overcome the challenge of achieving external validity, in that the findings of one megastudy might not generalize to other scenarios that involve other places and people. Whereas the present megastudy draws participants from the same participant pool, it is exciting to imagine that the next generation of megastudies might not only expand the number of interventions being studied, but also widen the pool of participants across different geographies, and, in doing so, might address the frequent concerns about the external validity of RCTs.

Heather Royer is in the Department of Economics, University of California, Santa Barbara, Santa Barbara, California 93106, USA.
e-mail: royer@econ.ucsb.edu

1. Milkman, K. L. *et al. Nature* **600**, 478–483 (2021).
2. Carrera, M., Royer, H., Stehr, M., Sydnor, J. & Taubinsky, D. *J. Health Econ.* **62**, 95–104 (2018).
3. Acland, D. & Levy, M. R. *Mgmt Sci.* **61**, 146–160 (2015).
4. Beshears, J., Lee, H. N., Milkman, K. L., Mislavsky, R. & Wisdom, J. *Mgmt Sci.* **67**, 4139–4171 (2021).
5. Carrera, M., Royer, H., Stehr, M. & Sydnor, J. *Mgmt Sci.* **66**, 1890–1908 (2020).
6. Charness, G. & Gneezy, G. *Econometrica* **77**, 909–931 (2009).
7. Royer, H., Stehr, M. & Sydnor, J. *Am. Econ. J. Appl. Econ.* **7**, 51–84 (2015).
8. Andrews, I., Kitagawa, T. & McCloskey, A. *NBER Working Pap.* No. 25456 (Natl Bur. Econ. Res., 2019).

The author declares no competing interests.
This article was published online on 8 December 2021.