

Review of ‘A Proposal for Replicating Evanschitzky, Baumgarth, Hubbard, and Armstrong’s (2007) “Replication Research’s Disturbing Trend”

In his proposal, Prof. Hubbard (H from hereon) describes how to replicate one of his earlier papers which assesses how often replication papers are published in highly regarded marketing journals. While I fully agree it is important to regularly measure how often replication studies get published, I think that H’s proposal can benefit from more discussion and clarification on two issues, both related to the definition of various types of replication.

First, H proposes a plan to ‘exactly replicate’ the paper by EBHA but it is not very clear what definition of replication will be used to judge whether other articles are replications or not

*“In determining whether an article qualifies as a replication, we will each read and classify them all independently.... To qualify as a “replication” an article has to contain an explicit citation of the original work. This does not mean that the replicating author(s) must identify their own research as such. This will be our responsibility. If doubt arises concerning whether the article is a replication, I will err on the liberal side and include it. If anything, then, the estimate of replication research will be exaggerated.”*

It would be good to make it more clear what kind of replications are considered here – only exact replication or any of the types of replications mentioned in table A? More general, if one would want to make the proposed replication replicable one would need to have a very precise and operational definition of what replication means. A person looking for type (1) replications is likely to find few but a person looking for type (4,5 or 6) replications might well find many if the definition of ‘different population’ and ‘different measurement or analysis’ are made sufficiently wide.

Related, H proposes to use the following population

*“Since EBHA used a census of articles appearing in the JCR, JM, and JMR for 1990-2004, I will do likewise for the 16-year period 2005-2020. I will also add 3 more journals to those above, namely, the Journal of the Academy of Marketing Science (JAMS), Marketing Letters (ML), and Marketing Science (MS), because of their general nature and wide readership among marketing academics.*

*So this will make my study as close as it can get to being an exact replication—faithfully repeating the methods used in the earlier work on a new sample (census in this case) from the same population (highly regarded marketing journals).”*

But is this really the same population as EBHA? More general, how does one determine whether the population is the same?

In some cases this is easy to determine. From H’s proposal:

*“EBHA replicated HA’s study which examined how often replications were published in three leading marketing journals—Journal of Consumer Research (JCR), Journal of Marketing (JM), and the Journal of Marketing Research (JMR)—over the 16-year period 1974-1989. They did so based on a content-analysis of 31 randomly selected annual issues from each journal. This 50% sampling of all JCR, JM, and JMR issues yielded a total of 835 empirical articles.”*

In this example, sampling from the same population would have been clear: take the other 50% of articles, those that were not evaluated by HA. But this is not what EBHA did: they ‘used a census of articles appearing in the *JCR*, *JM*, and *JMR* for 1990-2004’.

So how can one take another sample from this population which already has been exhaustively covered? H proposes “*Since EBHA used a census of articles appearing in the JCR, JM, and JMR for 1990-2004, I will do likewise for the 16-year period 2005-2020.*”

Could one not argue that this is sampling from a different population (and hence the proposed replication is not really an exact replication). A different time period means different editors, different submitting authors, different papers submitted – all things one could argue that lead to a different population. And hence a non-successful replication would not mean the original study is not trustworthy, it just would mean that there is heterogeneity. Some truths just turn out to be time-dependent.

If one would want to stick to the population of EBHA, one could evaluate the same time period as EBHA but use the alternative journals suggested by H (the *Journal of the Academy of Marketing Science (JAMS)*, *Marketing Letters (ML)*, and *Marketing Science (MS)*) – that is one would sample different ‘highly regarded marketing journals’. But even then could argue, that a non-successful replication would again just mean heterogeneous effects, that what was true for those journals highly regarded by EBHA is not necessarily true for the journals that H considers as highly regarded.

Now suppose H finds that in his replication he finds a precisely estimated 20% replication frequency – this would lead H to conclude that the replication of EBHA and HA are not successful as confidence intervals are clearly not overlapping. But what would be the meaning of this? That EBHA and HA did do something wrong? Or just that over time things have changed? Given that the implications of a ‘non-successful’ replication are so different for the different types of replications in table A one should be very careful in explaining to the reader what one can conclude from the replication – is the replication checking whether there are heterogeneous effects ( like for replications of type (5) and (6)) or checking whether the initial study is in one way or another flawed ( for replications of type (1) from table A) or (possibly un)lucky (for replications of type (3)).