http://dx.doi.org/10.4135/9781446273050145295 28

Meta-Analysis: Why Do Citizens Vote (or Abstain) in National Elections?

Link to the Research Output

- Smets, K. & Van Ham, C. (2013). *The embarrassment of riches? A meta-analysis of individual level research on voter turnout*. *Electoral Studies*, 32, 344–359. doi: http://dx.doi.org/10.1016/j.electstud.2012.12.006 http://dx.doi.org/10.1016/j.electstud.2012.12.006

The question why people choose to vote or abstain in national elections has been extensively researched in the past decades. Yet, disagreement over what drives citizens to the polls persists. Literally, over a 100 different explanatory factors have been linked to the individual's decision to vote or abstain: ranging from individual characteristics like education, age and political interest, to socialization by friends and parents, to characteristics of the election-like political competition, and many more. Slowly, but surely, it has become difficult to see the wood for the trees. In the research project that we describe in this article, we wanted to take a step back and summarize where we stand and what we know about turnout. To this end, we carried out a meta-analysis of 90 empirical studies of individual-level voter turnout in national elections between 2000 and 2010. This allowed us to identify which variables are consistently linked to turnout, and which are not. In this article, we describe what meta-analysis is and give an example of how to carry out meta-analysis by describing our research project on turnout.

Learning Outcomes

By the end of this case study, you should

- Be able to understand the difference between a literature review and a meta-analysis
- Be aware of the various types of meta-analysis that exist and make an informed decision as to which kind of analysis is most suitable under which circumstances
- Comprehend the various steps of the meta-analysis research design
- Be mindful of the caveats of meta-analyses, especially of meta-analyses that are based on a selection of the literature rather than everything that has ever been published on a given topic

Introduction

Elections are events that attract a great deal of interest from citizens, politicians and the media alike. It is not surprising that elections create a lot of buzz as they play a crucial role in democracies. Elections empower citizens with the choice to vote for different candidates or parties who will represent them during the next 4–5 years. Elections also make those in power accountable. At every election, voters pass judgement on how well they think the government has done and choose between keeping a government in office and replacing it.

Precisely, because elections play such a central role in democracies, many scholars have sought to understand who votes and who abstains in elections, and more importantly, why. Answering the latter question became possible with the event of survey research in the 1940s. Surveys are based on representative samples of the population and collect information about individuals, for example, about political attitudes and (self-reported) political behaviour.

Often, research will accumulate around a certain research question, and different empirical studies will come up with different answers. Literature review articles provide a descriptive summary of what different studies have found and evaluate what we know and don't know yet about a topic. As researchers were interested in turnout, we were both searching for an overview study that summarized the most important findings of research on individual-level voter turnout. While we found an overview study on

aggregate turnout by Benny Geys, published in 2006, we were unable to find a similar study on individual-level turnout. We then decided to do an overview study ourselves.

We chose to carry out meta-analysis rather than a literature review, because a meta-analysis allowed us to systematically summarize and empirically evaluate the robustness of empirical findings. In the next section, we will explain what meta-analysis is, and in the subsequent section, we give an example of a meta-analysis based on our overview study of turnout research.

What Is Meta-Analysis?

Meta-Analysis literally is an analysis of analyses (Glass, 1976). Meta-analysis involves analysing the results of multiple empirical studies, so as to (a) summarize empirical findings and (b) evaluate the robustness of empirical findings. For example, if we want to know why people vote in elections, we can collect a number of empirical studies on turnout and evaluate to what extent these studies have similar findings. Imagine all studies test the effect of education (the independent variable) on turnout (the dependent variable) and a majority of studies find a positive and significant effect on turnout: we would be more confident that there is indeed an effect of education on turnout than if we had considered a single study. In a literature review, this is where the summary would stop. In a meta-analysis, researchers *test* whether the number of times the effect of education was found to be significant and in the expected direction significantly deviates from zero. Moreover, in more advanced meta-analyses, researchers can also analyse the strength of the effect of the independent variable on the dependent variable. Meta-analysis thus allows researchers to statistically test (a) *whether there is a consistent effect* of an independent variable on the dependent variable and (b) *what the strength of the effect* is.

To test whether there is a consistent effect of an independent variable on the dependent variable, the vote-counting procedure can be used. In the vote-counting procedure, each time the effect of an independent variable on the dependent variable is found to be statistically significant and in the hypothesized direction, it is considered a ‘success’. If the effect is not significant, it is considered a ‘failure’, and if the effect is significant but in the opposite direction as expected, it is considered an ‘anomaly’. Considering all

tests together for each independent variable, the modal category gives an estimate of the most common relationship between the independent variable and the dependent variable, and dividing the number of ‘successes’ by the total number of tests provides a measure of the success rate (see equation 1 below)

$$\text{Success rate} = (\text{number of successes}/\text{number of tests}) \times 100 \quad (1)$$

The higher the success rate, the more confident we are that an independent variable has the hypothesized effect on the dependent variable, in terms of both direction and significance. Note that because some studies include more tests than others (e.g. the same hypothesis can be assessed in multiple models within in a given study), looking at the separate test results may lead to biased results when the distribution of tests over studies is highly skewed (some studies including many tests and other including few). Moreover, various test results from a single study are not independent from one another as they often use the same data. Hence, in general, it is more reliable to calculate the success rate per study rather than per test, for example, by taking the modal outcome of all tests in one study as the outcome of that study.

If your studies are sufficiently comparable, it is advisable to use the vote-counting procedure only as a descriptive first step in your meta-analysis. The problem with vote counting is that it does not take into account the information provided by the confidence intervals (i.e. an effect that is highly significant is equated with an effect that is only marginally significant: they are both coded as ‘successes’) (Doucouliagos & Uluba#o#lu, 2008). Also, depending on the statistical power of studies, vote-counting procedures may increase the likelihood of type II errors: that is, finding there is no effect when actually there is one in the population (Borenstein, Hedges, Higgins, & Rothstein, 2009).[1] Hence, if your studies are sufficiently comparable, it is preferable to use the combined test procedure. In the combined test procedure, you calculate the average *effect size* of the effect of the independent variable on the dependent variable in the different studies and calculate a confidence interval around the average effect size to evaluate whether the effect of all studies taken together is significant. For that, you need to *standardize* the effect sizes so that they become comparable across studies, for example, by using the partial correlation. To subsequently calculate confidence intervals, you can use a

bootstrap procedure. [2] The advantage of this procedure is that it also allows you to evaluate the strength of the effect. In addition, it is possible to analyse the source of variation in effect sizes between studies using meta-regression analysis, for example, evaluating how the use of different statistical techniques influences the effect sizes found (see Doucouliagos & Uluba#o#lu, 2008 for a very nice example of how to do this).

However, studies often will not be sufficiently comparable, and one might still want to move beyond a simple vote-counting procedure. In that case, using a simplified version of the combined test procedure is possible (Geys, 2006). The simplified version of the combined test procedure uses not the real effect sizes, but approximations of effect sizes based on whether the effects were ‘successes’, ‘failures’ or ‘anomalies’. To do this, successes (significant and in hypothesized direction) are assigned a weight of 1, failures (not statistically significant) a weight of 0 and anomalies (significant but not in hypothesized direction) a weight of -1. A proxy of the effect size, called r, at the level of tests can then be calculated with the formula in equation 2

$$\text{Effect size } r = (\text{successes} - \text{anomalies})/\text{number of tests} \quad (2)$$

The average effect size of a given independent variable across all studies (r *av*

) is subsequently given by the mean effect size, as in equation 3

$$\text{Average effect size } r_{av} = \Sigma r_i/\text{number of studies} \quad (3)$$

The average effect size has a theoretical lower bound of -1 and an upper bound of +1, and gives the number of standard deviation units with which the dependent variable changes if the independent variable changes by one standard deviation. By calculating a confidence interval around this statistic, we can judge whether or not there is a statistically significant effect on the dependent variable (testing the null hypothesis that the mean effect across all studies is zero) (see Smets & Van Ham, 2013, p 4, for a calculation example).

To summarize, while testing whether there is an effect is generally possible in meta-analysis, assessing the average strength of effects is not always possible. When making the choice what type of meta-analysis to do and to be aware of the limitations of your meta-analysis, three considerations are important: (a) What do you want to know? (b) What universe of studies do you include? (c) How comparable are the studies you include?

What Do You Want to Know?

Meta-analyses can focus either on multiple independent variables (i.e. seeking to establish which independent variables are consistently linked to a dependent variable) or on a single independent variable (i.e. seeking to establish whether a particular independent variable has an effect on a dependent variable, and if so, how strong the effect is). Often, meta-analyses of the first type only evaluate whether independent variables have an effect on the dependent variable of interest. Meta-analyses of the second type often also evaluate the strength of the effect of the independent variable on the dependent variable.

What Universe of Studies Do You Include?

Since the objective of meta-analysis is to summarize and evaluate the robustness of empirical research on a certain topic, the selection of studies shapes the results and conclusions of the meta-analysis. Four considerations are important for the selection of studies: should the meta-analysis encompass (a) all studies, (b) a representative sample of studies, (c) only studies of high quality or (d) only the most comparable studies, or a combination of these?

Include All Studies or a Representative Sample?

While including all available studies in a meta-analysis is preferable, with many topics, this is not possible for practical reasons. If including all available studies is impossible, it is important to consider how representative your selection of empirical studies is for the entire body of research on the topic. Moreover, it is crucial that you think carefully about how your sample selection might affect your results.

The most well-known source of bias in the selection of studies for meta-analysis is the so-called file-drawer problem. The file-drawer problem results from the fact that significant results are more likely to get published than non-significant results. If a meta-analysis is only based on published studies, this may inflate the number of studies with significant results found in those studies. The researcher may subsequently conclude that there is a significant effect of an independent variable, whereas in fact there is not. As it is often difficult to gather non-published studies (let alone a representative sample of non-published studies), in practice, most meta-analysis do include only published studies.

In meta-analyses that evaluate the effect of multiple independent variables, the file-drawer problem can be mitigated by checking the results when only considering the independent variables that were not of theoretical interest, that is, those included as control variables. The file-drawer problem is likely to occur only for those independent variables that are crucial for the theoretical argument made in a study. Hence, only including control variables in the meta-analysis decreases this source of sample bias.

Only Include Studies of High Quality?

Some researchers argue that including only published work in a meta-analysis is actually good. Published work will have gone through the process of peer review and is therefore thought to be of high quality and report more reliable results than unpublished

work. In this case, the selection of studies is not aimed to include all studies, or a representative selection of all studies, but rather the ‘best’ studies.

Include the Most Comparable Studies?

The sample of studies included in meta-analysis may also be chosen to increase the comparability of studies (see point 3 below), for example, studies may be limited to a specific country (e.g. the United States) or set of countries (e.g. only established democracies), or studies may be limited to a sub-set of cases (e.g. national elections). Finally, sample selection may also partly be based on pragmatic reasons. For example, when the body of research is too large to evaluate, studies may be limited to a specific time period. In all these cases, it is important to realize that your findings are only applicable to that particular country or set of countries, time period or sub-set of cases.

How Comparable Are the Studies You Include?

As in all research, when doing meta-analysis, it is important to prevent comparing apples to oranges, so to speak. There are three questions to consider before choosing how advanced the meta-analysis can be. First, is the dependent variable measured in the same way in all the studies included? Second, how comparable are the model specifications of the different studies? Do studies include the same or at least similar independent variables or do they include widely varying independent variables? Third, what kind of statistical models are used to analyse the results?

If the studies you aim to include in a meta-analysis are highly comparable, using similar dependent variables, similar independent variables and similar statistical models, then more advanced meta-analysis techniques are possible. If the studies included in your meta-analysis are less comparable, for example, using differently measured dependent variables, different sets of independent variables and different statistical models, the possibilities of meta-analysis are often limited to the vote counting and simplified combined test procedures.

Example: Why Do Citizens Vote?

As explained above, the fact that we could not find a good overview study of research on individual-level turnout triggered the idea of carrying out a meta-analysis. In our article, ‘The embarrassment of riches? A meta-analysis of individual-level research on voter turnout’ (Smets & Van Ham, 2013), we hoped to make a contribution to the literature by (a) providing a systematic overview of factors linked to individual-level voter turnout in the literature and (b) evaluating which of these factors were consistently found to explain voter turnout in empirical research. We decided to complement Benny Geys' study from 2006 in which he sought to explain what drives aggregate-level turnout and conduct a similar study focussing on research explaining why individuals vote or not.

As research on individual-level turnout has been carried out since the 1940s, it was clear from the start that we would have to restrict the time period of our investigation. Taking into consideration all research published on individual voter turnout would simply be impossible. We therefore decided to focus on the most recent decade as we thought that the outcomes of more recent research would be of most interest to both scholars and policymakers. However, even restricting our time period from 2000 to 2010, we knew that it would be impossible to look at every conference paper, working paper, journal article, book chapter and book addressing individual-level turnout. For pragmatic reasons, we therefore chose to restrict our sample to peer-reviewed journal articles from 10 top journals in political science. Our goal was to include general political science journals as well as journals specializing in political behaviour and to strike a balance between European and American journals. To make sure our selection of journals was representative, we benchmarked our selection against the journals cited by the Social Science Citation Index as having published the most articles on turnout in the period between 2000 and 2010. This made us more confident in our selection.

There were, however, more ways in which we restricted the sample of articles we included in our study. We decided to focus exclusively on the decision of individuals to vote in *national* elections. This implies that we excluded studies investigating why people participate in – for example – primary, municipal, regional and European elections. We felt that including research looking at such ‘second-order’ elections might distort the results of our meta-analysis, as academics do not seem convinced that what

drives turnout in these elections is the same as what makes people decide to vote in a country's most important election. Moreover, we felt more confident looking at a homogenous group of elections knowing that our results would at least be applicable to this sub-set of elections, namely, national elections, and therefore boost our confidence in the findings.

A last restriction we applied was to look at research on individual-level turnout in established democracies only. We thus excluded studies on turnout in, for example, new democracies. This decision was prompted by the fact that the literature on individual-level turnout in new democracies has only recently started to emerge. While certainly finding similarities with established democracies, turnout in new democracies seems to be affected by a number of important factors that are not found to affect electoral participation in established democracies. In this instance too, we thought it would be better to restrict our sample to a smaller sub-set, but to have more confidence in the results.

The fact that we restricted our sample of studies led our research to suffer from a number of caveats that needed to be taken into consideration when interpreting the results. Apart from the limitations on the generalizability of our results due to the restricted time period and exclusive focus on national elections in established democracies, our focus on peer-reviewed journal articles made our analyses potentially suffer from the file-drawer problem discussed above. Upon advice from one of the anonymous reviewers of our manuscript, we performed robustness checks to see whether our results were influenced by the file-drawer problem. This appeared not to be the case.

With a clear idea of the type of articles we wanted to include in our study and aware of the limitations of our study, the next step in the research process was to identify the exact articles to include. To this extent, we read the abstracts of every article published between 2000 and 2010 in the 10 journals that we selected. We downloaded all articles that we thought might have to be included in our study, making sure to err on the side of inclusion rather than exclusion. To be extra safe that we did not exclude any articles at this stage, we double-blinded the selection procedure meaning that we each went through all abstracts without knowing which articles the other person had selected. In this process, we gathered 190 articles to be investigated in more detail. Subsequently,

we checked whether the articles met the criteria of our sampling frame and documented why they did or did not meet these criteria. This meant that all articles that investigated turnout at the aggregate level, turnout in new democracies and turnout in elections other than in national elections were left out of our final meta-analysis. This left us with 90 articles to code.

While selecting the articles, we also started to think about the information that we would need to document about each of the articles that met our selection criteria. We developed a codebook specifying the characteristics of the data, the type of dependent variable used, the independent variables modelled, the statistical techniques used and of course the study results: direction of the relationship (positive or negative), statistical significance, size of the coefficient, and so forth. Once again, we made sure to be as inclusive as possible. We thought it would be better to include too much information and not use everything than to have to go back to all the articles because we forgot to code something important.

We now felt ready to start coding all of our articles. Even though we had help from a research assistant and were aware of the sheer volume of the task ahead of us, this stage of the research process was very, very time-consuming. We ensured inter-coder reliability by test-coding a substantial sub-set of the articles at the beginning, meaning that all three of us coded the same articles and then checked whether we coded consistently. Also, once the first version of the dataset was complete, all data entries were double-checked by one of us.

We still were not ready to start our analyses, however. During the coding stage, some issues arose, and we took the decisions on how to deal with these together. We were careful to document all of these decisions so that we could review them at a later stage if necessary. For example, in some instances, independent variables were measured through several dichotomous variables rather than through a single ordinal variable. So instead of including a variable for educational level with, say, seven categories, we found articles in which authors included six dichotomous variables for each of the categories minus one (the reference category). Should we treat these dummy variables as six separate variables or as one because they measure a single underlying concept? We decided to go with the last option, coding the variable to be a ‘success’ if the majority of dummy variables were significant and in the expected direction.

Also, how to deal with authors researching the same concept but using different operationalizations? The degree of competition between parties, for example, can be operationalized and measured in many different ways. Should we treat all of these different measurements as different explanatory factors, or should we group them together because they all measured the same concept? We used common sense as well as our expertise of the topic in deciding when to group variables together and when it was better to keep them separate. Once again, we documented carefully how we regrouped variables.

During the coding stage, we also noted that because of the completely different nature of the statistical analyses performed in the studies, we would not be able to carry out the more sophisticated types of meta-analysis described in the previous section. Also, different authors measured individual-level turnout – the outcome variable that we were interested in – in very different ways. Some used self-reported turnout (citizens reporting whether or not they voted in surveys), others used validated turnout (self-reported turnout that is checked against official records), and we also encountered some articles that measured turnout intentions in future elections. Ideally, we would have performed robustness checks to assure that our findings held for all three types of dependent variables, but this turned out difficult because only a small number of studies used validated turnout and turnout intention, resulting in too small sample sizes. Vote intention, for example, is used so sporadically that we did not have enough observations to calculate the effect size.

The small *n* problem occurred for other reasons as well. In the 90 articles that we ended up coding to be included in our database, we found no fewer than 176 different explanatory factors linked to voter turnout! (Hence, the choice to entitle our article ‘The embarrassment of riches’.) Of all of those 176 variables, only eight (less than 5%) were included in more than 25% of the studies. Even the two most common independent variables – age and education – were included in only 72% and 74% of studies, respectively. This means that for most variables we did not have many observations. We decided not to report variables that were included in only one or two studies, since this would not allow us to carry out t-tests to assess the significance of the average effect size.

Carrying out the analyses and writing up the results was a breeze compared to selecting the sample, selecting the cases, developing a codebook, coding the articles, constructing a dataset and dealing with the various issues that arose during the coding stage. However, all the preparatory work was extremely important in safeguarding the validity and robustness of our results. Looking back, we would not have done it any other way.

Conclusion

Meta-Analysis is literally ‘an analysis of analyses’. Meta-analysis involves analysing the results of multiple empirical studies together, so as to (a) summarize empirical findings and (b) evaluate the robustness of empirical findings. Meta-analysis is a much more powerful summary tool than a literature review, because it collects the information from empirical studies in a more systematic way, and it allows the researcher to test empirically whether there is an effect of independent variables on a dependent variable. When studies are highly comparable, meta-analysis even enables researchers to assess the average strength of the effect of an independent variable on a dependent variable, as well as to analyse the source of variation in results between studies.

For researchers wishing to carry out meta-analysis, it is important to think carefully: (a) whether they want to evaluate the effect of a single or multiple independent variable(s) on the dependent variable of interest, (b) what universe of studies to include (all studies, a representative sample, a selection of the ‘best’ studies or the most comparable studies) and (c) how comparable the studies included in the meta-analysis are. The universe of studies included in a meta-analysis is important because it delimits the boundaries of generalization of the meta-analysis. The comparability of the studies is important because it determines whether the researcher can use only simple meta-analysis procedures (such as vote counting and the simplified combined test procedure) or more advanced meta-analysis procedures (such as the combined test procedure and meta-regression analysis). When discussing the results of their meta-analysis, researchers should be careful to document these – often inevitable – choices and evaluate the ways in which these choices may have affected their results.

Meta-Analysis is a crucial tool for mapping the state of the art of a certain research question, systematically evaluating what we know and do not yet know. Meta-analyses contribute to the accumulation of scientific knowledge and provide crucial information for further research. In our case, meta-analysis allowed us to evaluate which explanatory factors were consistently linked to individual-level turnout in national elections, which we hope has contributed to a better understanding about what drives citizens to the polls.

Notes

1. The latter issue is less problematic in studies that tend to have high statistical power, such as individual-level turnout studies. However, it is an issue to consider for meta-analysis of studies with lower levels of statistical power (*link to chapter on statistical power*).

2. Good guides on how to use these more advanced meta-analysis techniques can be found in Borenstein et al. (2009) and on the website: http://www.deakin.edu.au/buslaw/aef/meta-analysis/articles/ajpsapril2007.pdf

Exercises and Discussion Questions

- 1. What is the difference between a literature review and meta-analysis?
- 2. What procedures can be followed to evaluate *whether there is an effect* of an independent variable on a dependent variable, and which one is preferable?
- 3. What procedures can be followed to evaluate *the strength of an effect* of an independent variable on a dependent variable, and which one is preferable?
- 4. What considerations are important when choosing whether to use simple meta-analysis procedures (like vote counting and the simplified combined test procedure) or advanced meta-analysis procedures (like the combined test procedure and meta-regression analysis)? Name two.
- 5. What is the ‘file-drawer’ problem?

Further Reading

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis* (Vol. 49, *Applied social research methods series). London*, England: SAGE.

Littell, J. H., Corcoran, J., & Pillai, V. (2008). *Systematic reviews and meta-analysis*. Oxford, UK: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195326543.001.0001

Wolf, F. M. (1986). *Meta-analysis: Quantitative methods for research synthesis (Quantitative applications in the social sciences 07–059)*. Beverly Hills, CA: SAGE.

References

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, UK: John Wiley & Sons. http://dx.doi.org/10.1002/9780470743386

Doucouliagos, H., & Uluba#o#lu, M. A. (2008). *Democracy and economic growth: A meta-analysis*. *American Journal of Political Science*, 52, 61–83. doi: http://dx.doi.org/10.1111/j.1540-5907.2007.00299.x http://dx.doi.org/10.1111/j.1540-5907.2007.00299.x

Geys, B. (2006). *Explaining voter turnout: A review of aggregate-level research*. *Electoral Studies*, 25, 637–663. doi: http://dx.doi.org/10.1016/j.electstud.2005.09.002 http://dx.doi.org/10.1016/j.electstud.2005.09.002

Glass, G. V. (1976). *Primary, secondary, and meta-analysis of research*. *Educational Researcher*, 5, 3–8. doi: http://dx.doi.org/10.3102/0013189X005010003 http://dx.doi.org/10.3102/0013189X005010003

http://dx.doi.org/10.4135/9781446273050145295 28