How to estimate the policy preferences of party supporters: disaggregating data from voting advice applications versus modeling survey responses

Toshkov, D.D.; Romeijn, J.

Electoral Studies 74 (2021) 102403

Contents lists available at ScienceDirect

Electoral Studies

journal homepage: www.elsevier.com/locate/electstud

How to estimate the policy preferences of party supporters: Disaggregating data from voting advice applications versus modeling survey responses

Dimiter Toshkov *, Jeroen Romeijn

Institute of Public Administration, Leiden University, the Netherlands

ARTICLE INFO

Keywords:
Measurement
Political parties
Preferences
Public opinion
Public policy
Voting advice applications

ABSTRACT

Valid and reliable estimates of the policy preferences of political parties' supporters are essential for the study of political representation. However, such estimates are not directly available from standard surveys of public opinion, which are typically representative by design only at the national level and rarely ask questions about public support for specific policies. In this article, we explore the possibility to use data from voting advice applications (VAA) to estimate the policy preferences of party supporters. To do that, first, we identify 10 questions on preferences towards issues of public policy that were asked around the same time and with similar wording in traditional surveys of public opinion and in VAAs fielded in Germany and in the Netherlands. Then we compare the VAA data disaggregated by political affiliation of the respondents to the survey data adjusted via multilevel regression modeling with poststratification (MRP). We find strong positive correlations between the estimates derived from both methods, especially after weighting the VAA data. Yet, point estimates are not always very close, and the match is sensitive to the treatment of neutral and 'don't know' answers. Overall, our results bode well for the validity of using VAA data in empirical research on political representation.

1. Introduction

Valid and reliable estimates of the policy preferences of the supporters of different political parties are essential for a considerable number of political science questions, and for the study of political representation in particular. For example, we need to know the policy preferences of party supporters in order to measure the effect of partisanship on societal polarization, to analyze how well parties represent their constituents, how much preference congruence influences electoral choices, and how parties adjust their policy positions in response to changing preferences of the(ir) voters.

However, measures of the policy preferences of party supporters are not directly available from traditional sources of political data, such as nationally-representative surveys of public opinion. First, such surveys are typically representative, by design, only for the population of a state as a whole, and not for sub-populations like the supporters of particular parties. In fact, current estimates of the opinions of the supporters of smaller parties are often based on no more than a dozen respondents. Second, surveys used to ask only occasionally questions about preferences on concrete issues of public policy and on attitudes towards specific policy changes, focusing instead of broader ideological self-placement or on questions of issue salience. Given the low correlations between policy preferences on specific issues and left-right ideology (Lesschaeve 2017, Toshkov and Krouwel, 2021), the study of party congruence on specific issues has been an important addition to the literature on representation (Romeijn 2020; Costello et al., 2021). The availability of representative survey data on public policy questions is increasing in several countries, but there is still scope for improvement of the breadth and depth of our knowledge about the policy preferences of citizens and voters, especially when it comes to the supporters of relatively small parties.

Recently, there has been considerable progress in deriving estimates valid for sub-populations from nationally-representative surveys, using techniques such as multilevel modeling with poststratification (MRP) (Park et al., 2006) or Bayesian additive regression trees (Bisbee 2019). Yet, even when feasible, such techniques require additional data, complex statistical models, do not work well for very small groups, and cannot compensate for the fact that large batteries of questions on policy preferences are still rarely asked.

In this context, voting advice applications (VAA) can provide a valuable additional source of data. VAA are online tools that allow users to express opinions on a variety of policy issues and, based on the

* Corresponding author.
E-mail address: d.d.toshkov@fgga.leidenuniv.nl (D. Toshkov).

https://doi.org/10.1016/j.electstud.2021.102403
Received 5 February 2021; Received in revised form 21 September 2021; Accepted 23 September 2021
Available online 6 October 2021

answers, provide information about the match of the user with different political parties (see Germann and Gemenis 2019 for a recent overview). Over the last two decades, VAA have spread rapidly in many countries and have been used for a large number of elections. Importantly, VAAs ask many questions on preferences towards concrete policies and policy changes. Furthermore, they typically have many more respondents than traditional surveys, in some cases reaching well into the hundreds of thousands. Importantly, they locate respondents and political parties in common (policy) space, which is necessary for valid comparisons between voters and parties. However, since VAA users do not constitute a (random) probability sample of the total population, it is unclear to what extent VAA data can provide valid estimates of public preferences: a limitation that has impeded the use of VAA data in empirical political science.

In this article, we explore the validity of VAA-derived estimates of the preferences of the supporters of different parties. To do that, first, we identify ten policy-related questions that were asked in two countries – Germany and the Netherlands – at approximately the same time and with similar wordings in the context of VAA applications *and* in traditional nationally representative surveys of public opinion. Importantly, we pool data from several survey waves to reconstruct the demographic profile of different party supporters. Then, we use multilevel regression modeling with poststratification (MRP) to reconstruct party-level preferences from the traditional surveys (with data pooled from several waves) and simple disaggregation to recover party-level preferences from the VAA data.

Comparing the estimates derived from these two methods and data sources, we find a high level of correspondence in Germany and a moderate one in the Netherlands. The correlations between the two sets of measures are consistently positive and rather high (0.63 in the Netherlands and 0.93 in Germany). However, despite the strong correlations, there is variation in the match between the point estimates. Occasionally, there can be significant bias in the estimates, even when the correlations remain high, and the point estimates based on the two approaches can be far apart. Our analysis also shows that data-coding decisions, such as including or excluding "don't know" responses and the treatment of "neutral" answers can influence the match between the two sets of estimates.

Importantly, once we weight the VAA data to represent better the demographic profiles of the supports of different parties, the match between the VAA-based and MRP-ed survey estimates in the Netherlands improves further, reaching a correlation of 0.86.

Overall, our results provide strong evidence that VAA data can be used in empirical research that correlates the preferences of party supporters to other variables of interest, but that care is needed when interpreting estimates of absolute levels of support for a policy among party supporters based on (unweighted) VAA data. Our analysis underlines the high convergence validity (Adcock and Collier 2001) of VAA-based estimates of party supporter preferences, which expands the sources of data that can be used in the study of party representation, especially on specific policy issues. There are already examples that use such data to study congruence between parties and voters (e.g. Romeijn 2020; Costello et al., 2021), as well as to map the structure of policy positions of voters (e.g. Toshkov and Krouwel, 2021; Garry et al., 2017). We conclude by recommending practices for VAA developers, such as making the data openly available and encouraging the collection of socio-demographic data for the respondents, that can make this a valuable and widely-used data source for political analysis.

2. The problem

Researchers often use survey data to estimate the policy preferences of the supporters of a political party. This is prone to difficulties, however. Especially in political systems with many parties, estimates of the preferences of supporters for the smaller political parties can be based on as few as a dozen respondents. Currently, scholars relying on survey data address this problem by ignoring it or by excluding parties for which they have too few respondents, but the chosen thresholds are typically rather low (cf. Dalton 2017). To indicate the scale of the problem, consider the following example: if a party has 5% support in the population, a typical survey with 1000 respondents will produce an estimate of the proportion from its supporters in favor of a policy coming with a 90% confidence interval that can be 12 percentage points wide (see Part 1 of the Supplementary Material for details).

Moreover, the estimates of the preferences of subgroups, such as party supporters, can be biased even when the number of respondents per subgroup is large, because the surveys are not designed to be representative for subgroups, but only for the population as a whole.

Data from VAAs could help address this problem. As online tools that help evaluate the match of users' policy preferences against the positions of political parties, VAA are usually designed with the help of political scientists in cooperation with media and other organizations. This ensures that features of VAA, such as question wording and item selection, follow professional standards, while the involvement of societal organizations means that the tools can reach large audiences and record many responses.

Scholars have explored the applicability of VAA data for studying representation by comparing estimates derived from VAA to survey data (for recent extensive overviews of VAAs and their use, see Garzia and Marschall, 2016 and Rosema et al., 2014). Usually, this has not been done with the goal of verifying estimates of sub-groups in the population, but either to establish the representativeness of the overall sample, or for specific purposes like studying the effects of VAAs on turnout or vote choice (e.g. Marschall and Schultze 2012; Gemenis and Rosema 2014; Dinas et al. 2014; Garzia et al. 2017; Kleinnijenhuis et al., 2019; Germann and Gemenis 2019). Jackman et al. (2019) apply model-assisted post-stratification procedures to VAA data from Australia and validate it against small-area opinion estimates from a 2017 plebiscite on same-sex marriage. Our study is also related to the study by Popp et al. (2016), who use MPR to model VAA responses in order to forecast election outcomes. VAA data has also been extensively used to map party positions in multidimensional space (Mendez and Wheatley 2014; Germann et al., 2015; Wheatley 2015; Garry et al., 2017; Wheatley and Mendez 2021), but rarely for subpopulations such as groups of party supporters (Garry et al., 2017, Toshkov and Krouwel, 2021). König and Nyhuis (2020) study the applicability of vote advice applications for estimating the party positions of the parties themselves. In light of the many studies that already use VAA data to infer ideological positions and measure policy preferences, a validation study – such as the one offered here – is relevant and valuable, as it speaks to the broader generalizability and validity of these published results.

Unlike survey data, VAA data is not based on a probability sample and is not representative by design for any population. Indeed, existing evidence suggests that the typical VAA user is more likely to be male, younger, more highly educated and, more politically interested, and more likely to engage in other forms of political communication (see the review of evidence provided in Van de Pol et al., 2014 and Part 4 of the Supplementary Material). Yet, Wheatley et al. (2014) forcefully argue that VAA data can be used to study the preferences of party supporters. They also suggest that for the Scottish data they analyze, the VAA users, and especially those with an explicit party affiliation, were relatively politically committed but more polarized and more coherent in their views compared to a sample from the nationally representative election survey. Hence, the lack of representativeness of VAA data does not preclude the possibility to study the issue positions of party supporters.

The fact that the number of party supporters in VVA data is generally much higher than in existing surveys may help to facilitate more detailed analyses of these subgroups than existing surveys. In addition, VAA data include many more questions on public preferences on concrete policy issues than most existing surveys, which would be a further advantage of demonstrating the validity of VAA-based estimates of subgroup preferences.

3. Our approach

3.1. Matching VAA data and survey responses

In order to assess the suitability of using VAA data for estimating the preferences of the supporter of a political party, first, we compare unmodeled estimates from VAA data to modeled survey data. Finding an overlap between the two sets of estimates would provide strong evidence for the convergent validity of VAA-derived estimates (Adcock and Collier 2001). In the absence of an objective benchmark against which to compare the estimates, convergence between the results of these two rather different approaches is an important criterion for their validity. Note that validating against the official party positions is not an option, as we cannot assume that the party and its voters agree (see Costello et al., 2021).

We can use survey data as a benchmark against which to compare the VAA estimates only because we (a) pool several survey waves together to build demographic profiles of the supporters of different parties, and (b) we model the sub-population estimates using MRP, which requires complex modeling and external data. This approach to survey data is of limited general applicability, because it requires a relatively large survey to begin with (or several waves pooled together), relevant policy questions being included in the survey, external (census) data for the MRP, and the use of a relatively complex modeling procedure. Hence, if we show that VAA data provides a reliable, alternative, easier to use data source for deriving estimates of the policy positions of party supporters, this would significantly extend the availability of such estimates for political analysis.

In addition to the comparisons between disaggregated VAA data and MRP-ed survey data, we match the disaggregated VAA data against disaggregated data from the pooled survey in order to make sure that the results are not driven by the modeling choices part of the MRP procedure. Even though we believe that MRP is an improvement over the simple disaggregation in the case of surveys, it will be reassuring to see that the VAA estimates can reasonably approximate the disaggregated survey estimates as well. We also check whether it is possible to further improve the performance of the VAA estimates based on weighting the data using the demographic profiles of the groups of party supporters in order to address possible bias due to self-selection.

For the empirical part of this study, we focus on the 2013 German general election and the 2012 Dutch general election. To select the questions included in the study, we started from a list of policy-related public opinion items asked in nationally-representative surveys in Germany and in the Netherlands collected in the period 2011–2013. This list was collected in the framework of the GOVLIS project (Rasmussen et al., 2019). Then we restricted the list to questions that were asked repeatedly, so that the data could be pooled to obtain a larger number of observations per group of party supporters. Finally, we searched for close matches to these questions in two VAAs: *Kieskompas* (in the Netherlands) (Krouwel et al., 2012) and *Bundeswahlomat* (Germany) (Krouwel et al., 2013).

In Germany, we matched five VAA questions to survey items in the *Politbarometer* (Jung et al., 2015) that was run throughout the election year. In the Netherlands, we matched five VAA questions to survey items from the *National Election Survey* (Nationaal Kiezersonderzoek: Stichting Kiezersonderzoek Nederland – SKON et al., 2012), which was held after the election. The questions concern opinions on public policy issues (e.g. whether a general speed limit should be introduced on German motorways or the pension age in the Netherlands should stay at 65) of varying salience and generality. Table A2 in Part 2 of the Supplementary Material contains the exact question wordings.

Although we took care to ensure that the match was as close as possible, three factors remain that may inflate the differences in estimates between the surveys and the VAA data. Firstly, VAAs could be filled out in the months running up to the national elections, while the surveys were typically fielded for a shorter period of time. This could increase differences if the supporters of a party shifted position on an issue in the (short) period between the VAA and the survey. Secondly, although we identified survey items that matched rather closely with the wording of the questions as asked in the VAA, the question formulations were not always identical. Specifically, the answers to some questions had to be recoded, as the VAA questions sometimes included a negation and the survey data did not (and vice versa), and we know from previous studies that this can affect estimates of public opinion (Holleman et al., 2016). Thirdly, there are several studies of the design of VAAs that point to the importance of the response scales (Gemenis, 2013; Rosema and Louwerse, 2016). The VAA items always included a neutral category, where the survey questions only provided occasionally (see also, Gemenis 2013). Taken together, these considerations all bias our approach somewhat *against* finding correspondence between estimates from different sources.

3.2. Disaggregating VAA data

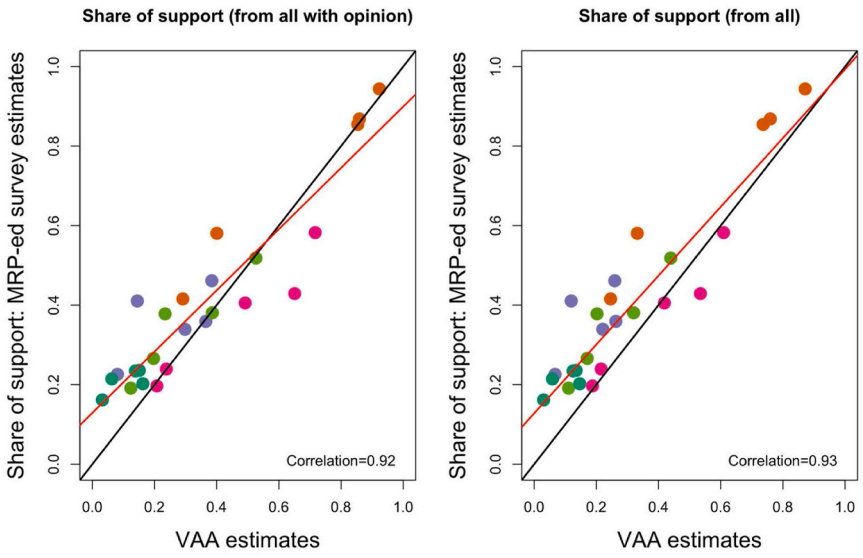
To estimate the policy positions of party supporters from the VAA data we need first to identify party supporters. In the case of Germany, we use the information for the vote in the previous 2009 general elections. In the Dutch case, we use the question asked to the VAA users about the party they voted for at the previous 2010 general election. Table 1 shows the number of VAA users per party that we identify using this approach. In Part 6 the Supplementary Material we report results based on more strict definitions of party supporters that take into account vote intention as well, in addition to prior votes: the results remain substantively the same when we use the stricter operationalization of party supporters.

For the main analyses, we use simple disaggregation to obtain the party supporters' estimates from the total pool of responses. This approach is the easiest to implement, so if we discover that it provides reliable estimates, this will speak in favor of the utility of using VAA data to estimate the policy preferences of party supporters. The original VAA responses are recorded on a 5-point Likert scales from "Strongly agree" to "Strongly disagree" with a "Neutral" midpoint and an additional "Don't know"/"No opinion" option provided. We estimate policy support in two ways: First, we calculate the share of "Strongly agree" and "Agree" responses from all responses expressing an opinion ("Strongly agree", "Agree", "Disagree", "Strongly disagree"), per group of party supporters. Second, we calculate the share of "Strongly agree" and "Agree" responses from all responses, including "Neutral" and "No opinion", per group of party supporters. We report the results with respect to both ways of measuring policy support.

While we use simple disaggregation for the VAA data for the main analyses – since such estimates are most directly available to researchers, without the need for complex modeling and post-stratification that requires external data – we also apply weighting of the VAA data based on information about the demographic profile of party supporters, for the case of the Netherlands. This approach gives more weight to observations that are relatively underrepresented in the

Table 1
Number of VAA records identified as party supporters per party.

a. Germany (2013)			
Party	N	Party	N
CDU/CSU	2515	FDP	1391
Die Grünen	3202	SPD	3526
Die Linke	1230		
b. The Netherlands (2012)			
Party	N	Party	N
CDA	3832	PvdD	664
CU	1581	PVV	3511
D66	6434	SGP	330
GL	5905	SP	4356
PvdA	8522	VVD	10524



Notes: VAA-derived estimates on the X-axis and the MRP-derived estimates on the Y-axis. Black lines show the diagonal (perfect fit). Red lines show the OLS fits. Dots are colored by policy issue. *Data from Germany, 2013: 5 policy issues and 5 parties.*

Fig. 1. Share of party supporters who support a particular policy position from those with opinion (left) or from all (right) in Germany. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 2
Comparing two sets of party support estimates per policy issue & per party group in Germany.

Policy issue/ Party group	COR	LCC	MAE	COR *(with DK)*	LCC *(with DK)*	MAE *(with DK)*
adoption	0.94	0.83	6	0.94	0.72	9
euro	0.69	0.15	10	0.71	0.12	9
speed	0.96	0.79	9	0.97	0.94	4
taxes	0.99	0.92	7	1	0.85	12
Turkey (DE)	0.63	0.38	11	0.6	0.19	16
CDU/CSU	0.89	0.79	9	0.89	0.7	10
Die Gruenen	0.97	0.95	8	0.97	0.91	10
Die Linke	0.93	0.92	8	0.96	0.95	8
FDP	0.41	0.29	9	0.34	0.18	11
SPD	0.95	0.92	9	0.97	0.89	10

Notes: COR = Pearson's correlation coefficient; LCC = Lin's concordance coefficient; MAE = Mean absolute error, in percentage points. 'with DK' = incl. respondents who answered "Neutral"/"Don't know".

VAA (e.g. old females with high education within the group of CU supporters) and less weight to observations that are overrepresented (e.g. young males with high education within the group of PVV supporters). Hence, it is supposed to result in more representative samples at the sub-population level. To apply this approach, we need demographic data for the VAA users (which we have for the case of the Netherlands, but not Germany) and demographic profiles of the groups of supporters of different parties, which we obtain from a big panel survey (LISS, see also the section of MRP and the Supplementary Material, Part 3)[1].

The two VAAs we used collected a large number of responses: more than 650,000 in the Netherland and more than 66,000 in Germany. However, the number of records with complete information of the voting and party affiliation module is much smaller: 45,659 in the Netherlands and 12,737 in Germany. While these numbers are still high compared to traditional surveys, they are by no means unprecedented for VAAs, whose popularity is growing over time (see Fig. 2 on page 152 in Germann and Gemenis 2019).

3.3. Modeling survey responses

To increase the precision of survey estimates of the preferences of political parties we rely on an adaptation of multilevel regression with post-stratification (MRP). MRP is a method developed to estimate the preferences of subgroups in a population of interest (Park et al., 2006; Kastellec et al., 2010). It has been shown to provide accurate point estimates of public preferences in US states (Lax et al., 2019) that correlate highly with the true values, even if MRP performs less well producing point estimates and rankings (Toshkov, 2015). The method starts by running multilevel regressions with random intercepts for a number of demographic characteristics to get a prediction of public opinion for each intersection in the data. For each demographic subgroup, MRP then uses census data to weight these predictions for the relevant geographic subunit. The advantage of this approach is that it uses more data than disaggregation to estimate levels of support for subgroups (Lax et al., 2019).

When expanding the approach to political party supporters instead of geographical units, one would ideally have a demographic profile of all supporters of the political party in an election year in a country. Since such data is not directly available, we pool data from existing surveys to obtain a demographic profile of party supporters. By pooling several surveys, we increase the number of observations available per group of supporters of political parties, so that there is coverage for more cells defined by the intersection of the values of the socio-demographic variables we use for poststratification. We then use this data to run MRP and poststratify by age (ten categories), education (four categories), and gender (two categories) to predict support for our policy issues amongst supporters of political parties. Part 3 of the Supplementary Material contain details on the exact approach followed in each of our countries.

In order to make sure that the results from comparing the VAA estimates to those derived from the surveys are not driven too much by the application of the MRP method, we also match the VAA ones to survey estimates derived from simple disaggregation.

[1] Instead of weighting, one could apply other related forms of adjustment, such as MRP, other model-assisted procedures with post-stratification (cf. Jackman et al., 2019) or raking to the VAA data.

4. Results: comparing VAA and survey estimates

In this part of the article we present the results from comparing the different sets of estimates: those derived from disaggregating the VAA data and those obtained from applying MRP to survey data.

4.1. Germany

For the German items, we observe a very close relationship between the VAA and the MRP-adjusted survey estimates of the policy position of party supporters. For the five parties and five issues in our dataset, the correlation between the two sets of estimates is 0.92 (when neutral answers and "no opinion" are excluded) and 0.93 (when all response categories are considered). Given the limitations mentioned above, these correlations are striking and similar to correlations found in other studies validating the general use of MRP (e.g. Lax et al., 2019; Toshkov, 2015). Fig. 1 demonstrates these rather high correlations by showing two scatterplots of the estimates, with the VAA ones on x-axis and the MRP-ed survey estimates on the y-axis.

When we look at the correlations between the two sets of estimates within each issue (for all parties) and within each party (for all issues), we also discover very high correlations: see Table 2 and Part 5 of the Supplementary Material. The correlations are lowest for the FDP (by some distance the smallest party in the sample) and the issues of Turkey's accession to the EU and leaving the Eurozone. The mean absolute errors between the two sets of estimates per group are between 6 and 11 percentage points when we exclude 'neutrals' and 'don't knows' and between 4 and 16 when these are included. From this, one can see that although most point estimates are rather close, larger distances between the different estimates do occasionally occur (see Table 3).

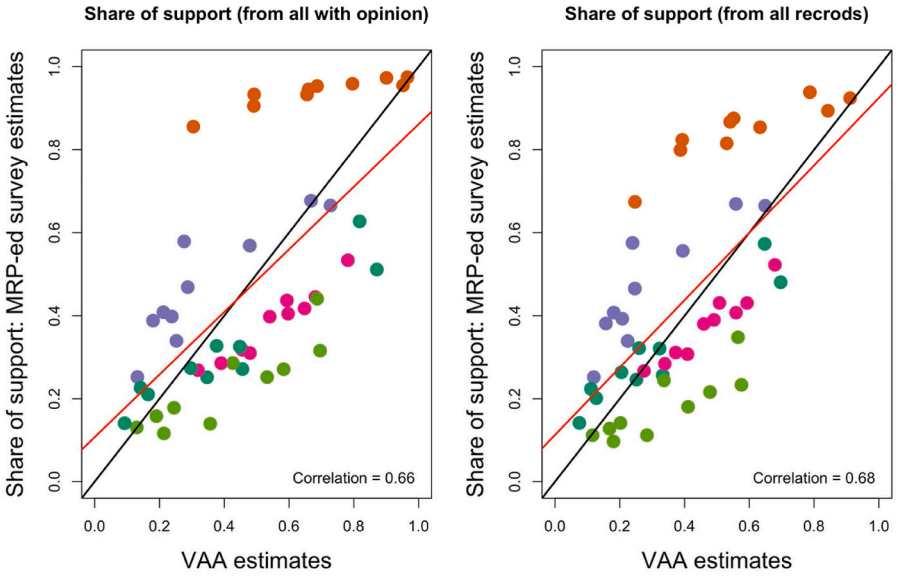
While the Pearson's correlations are very high, occasionally there is considerable bias. Looking at the Lin's concordance coefficients[2] (Lin 1989) reported in Table 2, we can see that for the issues on the retention of the euro and Turkey's membership in the EU, the concordance coefficients are rather low, and much lower than the Pearson's correlation coefficients. This implies that while correlated, the VAA estimates can be systematically 'shifted' or biased from the survey benchmark.

There is a significant correlation of 0.58 between the size of the party (the number of party supporters identified in the data) and the match between the two sets of estimates. The extremity of the party position (calculated as the absolute deviation of the mean of the positions of the supporters of party X from the mean of the positions of all respondents on an issue.) is correlated at −0.40 with the match, meaning that the estimates of more extreme positions diverge to a greater extent.

The results are very similar when we consider unmodeled, disaggregated survey estimates instead of the MRP-ed ones (see Part 8 of the Supplementary Material) and for some issues and parties the match with the VAA data is actually better. Using a stricter definition of party supporters (Part 7 of the Supplementary Material) leads to a very small drop in the overall correlations (to 0.89/0.91).

4.2. The Netherlands

Estimating the policy preferences of party supporters with traditional survey data (and MRP) should be harder in the Netherlands: the higher number of parties means that data needs to be disaggregated into more categories that contain fewer respondents. Compared to Germany, results for the Netherlands indeed look less promising. The overall correlation between the VAA and the MRP-adjusted survey estimates of the policy positions of party supporters for the 10 parties and 5 issues in our dataset is 0.66 (when neutral answers and "no opinion" are excluded) and 0.68 (when all response categories are considered). These lower



Notes: VAA-derived estimates on the X-axis and the MRP-derived estimates on the Y-axis. Black lines show the diagonal (perfect fit). Red lines show the OLS fits. Dots are colored by policy issue. *Data from the Netherlands, 2012: 5 policy issues and 10 parties.*

Fig. 2. Share of party supporters who support a particular policy position from those with opinion (left) or from all respondents (right) in the Netherlands. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

[2] Lin's concordance coefficient adds a measure of systematic bias to Pearson's correlation coefficient. The two measures can be seen as complementary.

Table 3
Comparing the two sets of party support estimates per policy issue and per party group in the Netherlands.

Policy issue/Party group	COR	LCC	MAE	COR *(with DK)*	LCC *(with DK)*	MAE *(with DK)*
mortgages	0.97	0.40	17	0.97	0.60	10
nuclear	0.88	0.43	18	0.84	0.38	15
pensions	0.89	0.64	14	0.88	0.52	17
punishments	0.90	0.12	25	0.88	0.22	26
Turkey (NL)	0.94	0.73	12	0.95	0.84	7
CDA	0.73	0.68	16	0.67	0.61	16
CU	0.67	0.59	20	0.65	0.53	18
D66	0.53	0.50	20	0.50	0.46	18
GL	0.49	0.45	21	0.58	0.51	18
PvdA	0.55	0.49	19	0.53	0.42	19
PvdD	0.83	0.74	14	0.86	0.70	12
PVV	0.90	0.79	13	0.92	0.87	7
SGP	0.46	0.42	24	0.46	0.45	20
SP	0.90	0.82	11	0.90	0.76	13
VVD	0.87	0.84	13	0.85	0.83	11

Notes: COR = Pearson's correlation coefficient; LCC = Lin's concordance coefficient; MAE = Mean absolute error, in percentage points. 'with DK' = incl. respondents who answered "Neutral"/"Don't know".

correlations between the estimates are also visible in Fig. 2, which shows two scatterplots comparing the VAA and survey-based estimates. The match improves slightly (to correlations of 0.68/0.70)) if we exclude the three smallest parties (CU, SGP, PvdD), but we find no strong correlation between the size of a political party and the size of the differences between the estimates from the two sources (0.03). As in Germany, there is a moderately strong *negative* correlation between the extremity of the positions of the party supporters and the match between the two sets of opinion estimates.

However, when we look at the relationship between the two sets of measures for each policy (Figure A6.3), the correlations are actually very high, ranging from 0.88 to 0.97. When we look at the correlations per party, these are rather strong as well, with the lowest ones recorded for the Green-Left (GL) and the small protestant party SGP. The mean absolute differences within issues and parties are somewhat greater than in the German case, ranging between 11 and 25 percentage points (excluding 'neutrals' and 'don't knows'). When we examine the patterns more closely (see Part 6 of the Supplementary Material), it looks like the MRP might have pulled the party-level estimates too much towards the grand mean, reducing variation that is captured by the VAA estimates (see especially the 'punishment' issue). Also, the discrepancies are greatest where the question wording is not exactly the same in both data sources ('punishment', 'pensions').

When we compare the match of the VAA disaggregation to simple disaggregation of the survey data, instead of the MRP-derived estimates, there are no huge differences. In fact, in some cases the correlations are higher and the errors are smaller (for details, see Part 8 of the Supplementary Material). The results also do not differ much when we use a stricter definition of party supporters (Part 7 of the Supplementary Material), although the overall correlation drops to 0.58.

Importantly, however, weighting the VAA data leads to a significant improvement in the match with the survey estimates (for details, see Part 9 of the Supplementary Material). Overall, the correlations between the weighted VAA estimates and the MRP-ed survey ones increase to 0.86 (0.89 with 'neutrals' and 'don't' know' responses considered) and the LCC rises to 0.83/0.88). These increases are not an artefact of using very similar demographic profiles to weight the VAA data and to apply the MRP to the survey data, as the correlations of the weighted VAA estimates with the unmodeled, disaggregated survey estimates are very high as well: 0.83/0.88, with LCCs of 0.81/0.89. At the party level, the improvements come from some of the smaller parties (CU, SGP), but also the PvdA. At the issue level, the improvement of the correlation is greatest for the issue of nuclear energy, of the LCC for the pension age, and of the MAE for the stricter punishment issue.

Overall, we can conclude that the Dutch VAA and MRP-derived estimates are strongly correlated in our dataset for most issues and most parties, especially after we weight the VAA data to improve its representativeness. Like in the German case, this provides evidence that both can be used in studies interested in relating the relative preferences of party supporters to outcome variables of interest. However, we demonstrate that estimated *levels* of support can vary significantly across estimation methods (see also Toshkov, 2015), so more care is necessary when using this data to estimate absolute levels of public support for policies. Moreover, the results serve to remind us that researchers using survey data should be very careful in the interpretation of results and keep in mind that even questions that are seemingly about the same issue can yield widely varying estimates.

5. Conclusion

Across the two countries and ten policy issues included in this analysis, we find strong correlations between estimates based on MRP-adjusted survey data and data from voting advice applications. Given the ample evidence that question formulations can influence estimates and advice from VAAs (Gemenis, 2013; Holleman et al., 2016; Rosema and Louwerse, 2016) and the fact that the formulations of the policy issues we compared were not entirely identical, we take this as strong evidence of the applicability of both sets of estimates in studies of representation that relate them to some outcome of interest.

The need for estimating of the preferences of party supporters based on different data sources for empirical research in political science is only likely to grow in the future. Two trends contribute to this need. The first is the increasing fragmentation of the party systems in many European democracies. This leads to preponderance of smaller parties, the supporters of which are difficult to describe with traditional survey methods. The second trend is the decreasing strength of general ideological dimensions in structuring preferences over individual policy issues. This makes it difficult to approximate preferences on concrete policies from general ideological predispositions, such as left-right of conservative-liberal self-placements, which are typically the estimates available from traditional surveys. It also makes it harder to maintain only left-right ideological congruence as the only test of democratic representation (Lax et al., 2019). If, as our results suggest, VAA-based measures are valid, this opens up new opportunities to address old questions about congruence between parties and voters (Costello et al., 2021; Romeijn 2020), differential policy responsiveness (Rasmussen et al., 2019), and differences between the representation style of different parties. The fact that VAAs contain more questions about policy preferences on concrete policy issues makes them especially valuable for future research in this vein. One limitation for individual-level analysis using VAA data is that socio-demographic and other variables of potential interest are typically included only in opt-in modules in existing VAA implementations.

We should note that we find that there can be larger differences in the *point* estimates that are derived using different methods. But this does not imply that VAA-based estimates could not be used for this purpose. The fact that the estimates do vary suggests that users of VAA data and of traditional nationally-representative survey data should be careful when interpreting levels of support amongst subgroups of the public. Our results especially in the Dutch case underline the importance of the interpretation of survey responses and different answer categories, including neutrals and "don't knows".

Importantly, we find that in some cases weighting or otherwise adjusting the VAA data can help a lot to improve the estimates (cf. Jackman et al., 2019). Weighting requires that demographic data is available for the VAA respondents and that the demographic profiles of supporters of different parties are possible to reconstruct. But when this is the case, weighting the VAA data is highly recommended, even if the performance of the simple disaggregation is great in the case of Germany

and acceptable in the case of the Netherlands.

Why does simple disaggregation of VAA data perform reasonably well even without modeling and adjusting the estimates, despite the fact that the VAA samples are opt-in and clearly unrepresentative? One potential explanation is that party-level differences dominate differences in policy opinions based on socio-demographic variables. Unfortunately, we cannot test this idea in the German case, where the fit between the disaggregated VAA data and the modeled survey data is excellent, but future research could address that.

Still, the fact that we find such high correlations between our sets of estimates will help move studies of representation forwards. One topic our study touches upon is the problem of political representation on concrete issues versus representation on general ideological left-right dimensions. In the past scholars have privileged the latter (e.g. Lefevere and Walgrave, 2014), but in normative terms it is not so clear that 'ideological' representation (which is not even ideological in the strict sense) is more desirable than representation on concrete policy issues (Rasmussen et al., 2019). The availability of VAA data has increased the number and range of concrete policy positions on which citizens, voters, party supporters, and political parties can be matched. Given that the VAAs can provide reliable data on public preferences, as our study suggests, assessments of responsiveness of political parties to different subsets of the general public will become more popular and more normatively consequential in the future.

An important limitation of this study relates to the choice of questions included in the analysis. While this choice has been constrained by availability of similarly-worded questions relevant to issues of public policy, it could be that an analysis of a different set of questions will lead to different results. Our sample includes questions of varying political polarization and salience – from introducing speed limits on highways to abandoning the euro as a national currency. Still, a more comprehensive study with a larger set of systematically-selected cases would be beneficial in probing the generality of our results. In this respect, VAA designers could include questions that are worded in the same way as in representative surveys to make comparisons easier.

We should point out that there are significant limitations to using VAA data as well. First, unlike comparative surveys, such as Eurobarometer or the European Social Survey, VAA tools deployed in different countries and elections are not always coordinated to ask the same questions in the same format (also, because the questions are supposed to track national political agendas). Still, sometimes there is significant overlap, and the EU Profiler (2009) and EUANDI (2014, 2019) were launched on a pan-European scale. VAA are also of limited use for historical work that looks before the 2000s and for comparisons over time, since not all policy questions are asked repeatedly.

Another limitation relates to the still limited public availability of VAA data for potential users. Since many VAA are developed in partnership with media and other commercial organizations, the data is sometimes kept proprietary. But even in such cases, VAA developers often make data available via individual user agreements. Other VAA datasets are already available under open access, including the ones used for the current study.[3]

To sum up, the utility of VAA data for political research will be enhanced even further if the data is made freely available, if some of the questions asked in the VAA are matched with questions in standard nationally-representative surveys and more questions are repeated over time, and if information on the socio-demographic background of respondents is systematically collected to allow filtering, weighting, modeling and poststratification of the data.

[3] For example, the EUANDI comparative data for 2019 is available via the EUI repository: https://cadmus.eui.eu/handle/1814/65484.

Data availability

The data used in this article, as well as the analysis scripts in R, are available at: https://doi.org/10.7910/DVN/WFCJHO.

Acknowledgements

We are grateful to the participants in the workshop '*Political Participation, Political Representation, and Voting Advice Applications*', ECPR Joint Sessions, University of Nicosia, Nicosia, 10-14 April 2018, for useful comments and suggestions. We would also like to thank André Krouwel for access to the VAA data. The authors acknowledge funding from the GOVLIS project (VIDI 452-12-008).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.electstud.2021.102403. Replication data and scripts in R are available at: https://doi.org/10.7910/DVN/WFCJHO.

References

Adcock, R., Collier, D., 2001. Measurement validity: a shared standard for qualitative and quantitative research. Am. Polit. Sci. Rev. 95 (3), 529–546.
Bisbee, J., 2019. BARP: improving mister P using bayesian additive regression trees. Am. Polit. Sci. Rev. 113 (4), 1060–1065.
Costello, R., Toshkov, D., Bos, B., Krouwel, A., 2021. Congruence between voters and parties: the role of party-level issue salience. Eur. J. Polit. Res. 61 (1), 92–113.
Dalton, R., 2017. Party representation across multiple issue dimensions. Party Polit. 23 (6), 609–622.
Dinas, E., Trechsel, A., Vassil, K., 2014. A look into the mirror: preferences, representation and electoral participation. Elect. Stud. 36, 290–297.
Garry, J., Matthews, N., Wheatley, J., 2017. Dimensionality of policy space in consociational Northern Ireland. Polit. Stud. 65 (2), 493–511.
Garzia, D., Marschall, S., 2016. Research on voting advice applications: state of the art and future directions. Pol. Internet 8, 376–390.
Garzia, D., Trechsel, A., De Angelis, A., 2017. Voting advice applications and electoral participation: a multi-method study. Polit. Commun. 34 (3), 424–443.
Gemenis, K., 2013. Estimating parties' policy positions through voting advice applications: some methodological considerations. Acta Politic. 48 (3), 268–295.
Gemenis, K., Rosema, M., 2014. Voting advice applications and electoral turnout. Elect. Stud. 36, 281–289.
Germann, M., Gemenis, K., 2019. Getting out the vote with voting advice applications. Polit. Commun. 36 (1), 149–170.
Germann, M., Mendez, F., Wheatley, J., Serdült, U., 2015. Spatial maps in Voting Advice Applications: the case for dynamic scale validation. Acta Politic. 50 (2), 214–238.
Holleman, B., Kamoen, N., Krouwel, A., van de Pol, J., de Vreese, C., 2016. Positive vs. Negative: the impact of question polarity in voting advice applications. PLoS One 11 (10), e0164184.
Jackman, S., Ratcliff, S., Mansillo, L., 2019. Small Area Estimates of Public Opinion: Model-Assisted Post-stratification of Data from Voter Advice Applications (Unpublished manuscript, available at: this link).
Jung, M., Schroth, Y., Wolf, A., 2015. Politbarometer 2013 (GESIS Studiennr. ZA5677, v1.1.0. GESIS Datenarchiv, Köln. https://doi.org/10.4232/1.12171.
Kastellec, J.P., Lax, J.R., Phillips, J., 2010. Estimating State Public Opinion with Multi-Level Regression and Poststratification Using R (Unpublished manuscript).
Kleinnijenhuis, J., van de Pol, J., van Hoof, A.M., Krouwel, A.P., 2019. Genuine effects of vote advice applications on party choice: filtering out factors that affect both the advice obtained and the vote. Party Polit. 25 (3), 291–302.
König, P.D., Nyhuis, D., 2020. Assessing the applicability of vote advice applications for estimating party positions. Party Polit. 26 (4), 448–458.
Krouwel, A., Ceulemans, S., Koedam, J., van Heck, S., Boiten, M., van Dieren, L., 2012. Kieskompas (Election Compass) Voting Advice Application Data for the 2012 General Dutch Election. Kieskompas, Amsterdam.
Krouwel, A., Eckert, T., Fanto, S., 2013. Kieskompas (Election Compass) Voting Advice Application Data for the 2013 German Federal Election. Kieskompas, Amsterdam.
Lax, J.R., Phillips, J.H., Zelizer, A., 2019. The party or the purse? Unequal representation in the US senate. Am. Polit. Sci. Rev. 113 (4), 917–940.
Lefevere, J., Walgrave, S., 2014. A perfect match? The impact of statement selection on Voting Advice Applications' ability to match voters and parties. Elect. Stud. 36, 252–262.
Lesschaeve, C., 2017. The predictive power of the left-right self-placement scale for the policy positions of voters and parties. W. Eur. Polit. 40 (2), 357–377.
Lin, L., 1989. A concordance correlation coefficient to evaluate reproducibility. Biometrics 45, 255–268.
Marschall, S., Schultze, M., 2012. Voting advice applications and their effect on voter turnout: the case of the German Wahl-O-Mat. Int. J. Electron. Govern. 5, 349–366.

Mendez, F., Wheatley, J., 2014. Using VAA-generated data for mapping party supporters in the ideological space. In: Garzia, D., Marschall, S. (Eds.), Matching Voters with Parties and Candidates. ECPR Press, Colchester, pp. 161–174.
Park, D.K., Gelman, A., Bafumi, J., 2006. State-level opinions from national surveys: poststratification using multilevel logistic regression. Public opinion in state politics 209–228.
Popp, R., Horvath, L., Banducci, S., Coan, T., Krouwel, A., 2016. What Voting Advice Applications Can Teach Us about Voters and Elections. ECPR General Conference, Prague.
Rasmussen, A., Reher, S., Toshkov, D., 2019. The opinion-policy nexus in Europe and the role of political institutions. Eur. J. Polit. Res. 58 (2), 412–434.
Romeijn, J., 2020. Do political parties listen to the (ir) public? Public opinion–party linkage on specific policy issues. Party Polit. 26 (4), 426–436.
Rosema, M., Louwerse, T., 2016. Response scales in voting advice applications: do different designs produce different outcomes? Pol. Internet 8 (4), 431–456.
Rosema, M., Anderson, J., Walgrave, S., 2014. The design, purpose, and effects of voting advice applications. Elect. Stud. 36, 240–243.
Stichting Kiezersonderzoek Nederland – Skon, Centraal Bureau voor de Statistiek - Cbs, Kolk, H., Tillie, J., Erkel, P., Velden, M., Damstra, A., 2012. *Dutch Parliamentary Election Study 2012* - DPES 2012. DANS.
Toshkov, D., 2015. Exploring the performance of multilevel modeling and poststratification with Eurobarometer data. Polit. Anal. 23 (3), 445–460.
Toshkov, D., Krouwel, A., 2021. Beyond the U-Curve: citizen and party preferences on European integration in multidimensional political space. Unpublished manuscript. Submitted for publication.
Van de Pol, J., Holleman, B., Kamoen, N., Krouwel, A., De Vreese, C., 2014. Beyond young, highly educated males: a typology of VAA users. J. Inf. Technol. Polit. 11 (4), 397–411.
Wheatley, J., 2015. Identifying latent policy dimensions from public opinion data: an inductive approach. J. Elections, Public Opin. Parties 25 (2), 215–233.
Wheatley, J., Mendez, J., 2021. Reconceptualizing dimensions of political competition in Europe: a demand-side Approach. Br. J. Polit. Sci. 51 (1), 40–59.
Wheatley, J., Carman, C., Mendez, F., Mitchell, J., 2014. The dimensionality of the Scottish political space: results from an experiment on the 2011 Holyrood elections. Party Polit. 20 (6), 864–878.