Who are the users of voting advice applications?

Ioannis Andreadis[*], *Matthew Wall*[†], *André Krouwel*[‡]

Introduction

Voting Advice Applications (VAAs) are web applications that have been used in many countries (most of them European) to help voters compare their positions on political issues with the positions of political parties and/or candidates on the same issues. VAA sites frequently attract millions of users, and are now a normal part of election campaigns in a growing number of established democracies. These sites can generate enormous tranches of public opinion data – containing the responses of individuals to policy questions, evaluations of political leaders and parties, demographic information, opinions on the functioning of democracy, personal political efficacy, and a range of country-specific items. These large datasets, which can be gathered cheaply and rapidly, allow us to explore public opinion, campaigns, and party-voter responsiveness in exciting new ways. Of particular interest is the fact that such sites collect their data throughout election campaign periods, with each observation being time stamped. Furthermore, the data on party/candidate issue stances is also highly rich, typically comprising 25-40 separate salient political issues. VAA data have been used lately by various researchers for many tasks: to explain the electoral behaviour of the voters, to study voter-party congruence, to position parties and their voters on political maps, etc. (Germann & Mendez, 2013; Germann et al, 2014; Jiglău et al, 2013).

In spite of these potential advantages, the confidence around the conclusions of studies using VAA data is consistently undermined by the fact that samples are not representative of the electorate. Online VAA opt-in surveys generate non-probability samples, the results of which cannot be straightforwardly generalized to the total population. Datasets collected online generally suffer from problems of under-coverage and self-selection that can potentially bias estimates (Bethlehem 2010; Hooghe and Teepe 2007). However, there are also benefits to opt-in web surveys. First, computerized self-administration reduces measurement error relative to other modes of data collection, increasing both the level of reporting and the report accuracy of opinions and attitudes compared with more "conventional" surveys (Kreuter et al. 2008; Sakshaug et al. 2010). Moreover, online survey questions are answered more truthfully and carefully, compared to interviewer-administered surveys (Olson 2006). This combination of self-selection and self-administration leads to a pool of respondents less likely to misreport their preferences and behaviour, and thus measurement errors should be smaller (Sakshaug et al. 2010).

The samples collected by VAAs have been found to be non-representative of the corresponding electorates, but the extent and nature of bias differs across countries. Knowing the source of bias is the first important step towards tackling this problem. In this paper we run a cross-national comparative analysis of VAA users in order to

[*] Aristotle University of Thessaloniki, Fulbright Visiting Scholar at the University of Michigan
[†] Swansea University
[‡] Free University of Amsterdam

examine if the observed differences of the sample bias can be explained by the differences of the Internet populations among countries.
For this purpose, this paper uses the Eurobarometer data 74.3 and compares it to VAA data in seven European countries. The countries were selected on the basis of the response rate of extra questions in the VAA tools. This leads us to an analysis of Western European, Eastern European, Southern European and a Scandinavian state: Greece, the Netherlands, Italy, France, Poland, Portugal and Sweden.

According to Andreadis (2012), many of the problems we meet with VAA data are common with the problems that appear in other web surveys (Couper, 2000, Couper 2008). According to Dillman (2007) the quality of a survey is affected by the overall survey error which consists of four components: coverage error, sampling error, nonresponse error, and measurement error. Coverage error is the error that occurs when some sectors of the population cannot be included in the sample. Sampling error is the error (inaccuracy) in estimating a quantity based on the sample instead of the whole population. Nonresponse error occurs when some people in the survey sample do not respond to the questionnaire and there is evidence that they differ significantly from those who respond. Measurement error occurs when answers to survey questions are inaccurate or wrong. In this paper we make an attempt to deal with the analogous of coverage error in VAAs, i.e. we deal with the problem that occurs by the fact that many people do not have the opportunity to use a VAA and we focus on Internet access limitations.

Several socio-demographic factors appear to be related with web survey participation, but this finding should be considered taking into account respondents' Internet resources and computer literacy (Diment & Garrett-Jones, 2007). Firstly, we should point out that not everyone is on the internet (at least not yet). Couper et al. (2007) using a panel study of people aged >= 50 years find significant demographic, financial, and health-related differences in Internet access and conclude that lack of access to the Internet appears to be of greater concern than unwillingness to participate, given access, for representation in web surveys (at least for people of older age).

In this paper we try to provide some answers to the question of whether the samples generated by VAA websites are representative of the total population of *web users* in the countries where they are deployed. We do this using data from the Greek Voting Advice Application HelpMeVote (Andreadis, 2013), Kieskompas (Krouwel et al, 2012) and EU profiler (Krouwel et al, 2012; Trechsel, 2010). Voting Advice Applications (VAAs) are web applications that enable voters to compare their political views with the positions of the political parties. VAAs have been used successfully in many European countries for more than a decade. Before analysing data of any European VAA, we first need to understand the profile of the Internet users in the European Union.

Internet users in the European Union

In this section we analyse the profile of the Internet user in the European Union and we focus on Greece and the Netherlands using data from the Eurobarometer 74.3 (European Commission, 2013): it includes a lot of useful variables with regard to the frequency of Internet use and the type of use. Secondly, the findings presented in the following sections of this paper are based on the analysis of VAA data. Since most of

the successful VAAs have appeared in EU countries (a significant exception being Smartvote in Switzerland), using the aforementioned Eurobarometer appears as a reasonable choice.

Table 1 Internet use frequency in EU27

	Frequency	Percent	Cumulative Percent
Everyday/Almost everyday	12815	48,2	48,2
Two or three times a week	2699	10,2	58,4
About once a week	1007	3,8	62,2
Two or three times a month	411	1,5	63,7
Less often	588	2,2	65,9
Never	6813	25,6	91,6
No Internet access (SPONT.)	2241	8,4	100,0

Source: Own analysis of Eurobarometer 74.3 (using W22 WEIGHT EU 27 which provides adjustments for each national sample in proportion to its share in the total EU population aged 15 and over). Internet use frequency is created using three variables in the dataset: D62 INTERNET USE FREQ: AT HOME, D62 INTERNET USE FREQ: AT PLACE OF WORK, and D62 INTERNET USE FREQ: SOMEWHERE ELSE by taking the minimum value of them i.e. if someone has responded "Never" at home, "about once a week" at work and "Less often" somewhere else, I have kept the answer: "about once a week".

As Table 1 shows, not all EU citizens use the internet. Approximately one out of three EU citizens is either without Internet access or has Internet access but never uses it. In addition, not every internet user uses the Internet with the same frequency. In fact, less than half of the EU population uses the Internet daily or almost daily. This should be kept in mind when we try to assess the popularity of a website or a web application. Simply put, if people do not have Internet access or do not use the Internet at all, they will not have any chance to visit a VAA website.

Table 2 shows that the highest rate of no access/use of the Internet is observed in Portugal (almost two out of three respondents), followed by Romania, Greece, Bulgaria, Cyprus. In all these countries more than half of the respondents were found with no Internet use. Denmark and Sweden are at the bottom of the list (about 13 per cent of no Internet use), while the last row is occupied by the Netherlands (5.6 per cent of no Internet use).

Nevertheless, there are other differences than the obvious ones between people who use and do not use the Internet. Firstly, there are differences between Internet users regarding the frequency of use. For instance, Germany does not have many people who do not use Internet at all, (it is among the five nations with the lowest rate) but on the other hand there are a lot of Internet users in Germany who do not use the Internet daily or almost daily, thus Germany ranks below the top ten according to the percentage of frequent (everyday/almost everyday) users.

The large differences across countries should be taken into account when we try to estimate the ratio of the total population of a country that has "selected" to visit a website of national interest. For instance, if a Portuguese website was used by one third of the total Portuguese population, we could argue that practically, the website was visited by everyone who was able to access it. If a similar ratio was observed for

a website in the Netherlands (where almost everyone has Internet access) the conclusion would be totally different, since the website would have been visited only by a small part of the group of Dutch people who were able to do so.

Table 2 Internet use frequency per nation in EU27

Nation	Everyday/Almost everyday	Two or three times a week	Less often	No access or never use
Portugal	26,9%	6,4%	2,8%	63,8%
Romania	28,8%	6,6%	7,5%	57,1%
Greece	25,0%	12,1%	6,5%	56,3%
Bulgaria	32,9%	9,3%	3,3%	54,5%
Cyprus (Republic)	34,9%	7,0%	4,7%	53,5%
Hungary	34,1%	11,2%	8,3%	46,4%
Spain	40,2%	9,1%	7,3%	43,3%
Italy	35,6%	11,3%	9,8%	43,3%
Poland	41,8%	8,8%	6,4%	43,0%
Lithuania	43,9%	7,0%	8,0%	41,2%
Malta	45,5%	9,1%	4,5%	40,9%
Austria	43,6%	14,0%	7,9%	34,4%
Slovenia	50,4%	8,8%	7,1%	33,6%
Czech Republic	40,7%	14,0%	11,8%	33,6%
Slovakia	43,6%	14,4%	10,4%	31,5%
Ireland	46,8%	14,5%	7,3%	31,4%
Belgium	54,1%	10,7%	6,4%	28,8%
Latvia	55,3%	7,4%	8,5%	28,7%
France	60,0%	6,7%	5,5%	27,8%
Estonia	61,7%	5,0%	6,7%	26,7%
Germany	49,5%	16,3%	9,2%	25,0%
UK	59,2%	8,6%	8,4%	23,7%
Luxembourg	59,3%	11,1%	7,4%	22,2%
Finland	68,2%	6,2%	5,5%	20,1%
Denmark	81,4%	4,4%	4,4%	9,8%
Sweden	81,4%	5,3%	4,6%	8,7%
The Netherlands	84,1%	6,5%	3,8%	5,6%

Source: Own analysis on Eurobarometer 74.3 data (using W22 WEIGHT EU 27)

This fact should be kept in mind when analysing VAA use. VAAs offer political information but we should not conclude that people who have not used a VAA are not necessarily interested in gathering political information. In order to talk about self selection of not using a VAA, we need to assume first that i) the voter has internet access and ii) the voter was informed about the existence of the VAA. The second condition is sometimes neglected, but similarly to web surveys (Fan and Yan, 2010), it means that everyone who has Internet access does not necessarily have an equal chance to visit a VAA website.

In this paper we argue that the differences we observe between VAA users and the general population (i.e. more male, younger, more educated, etc.) are very similar to the differences we observe between Internet users and the total population. Thus, a significant factor for VAA use is Internet use (in fact, it is a necessary condition: people who do not use the Internet are unable to use a VAA).

Age

Table 3 clearly shows that there is a strong negative correlation between age and Internet use. Less than one out of four European citizens over the age of 64 uses the Internet. Moreover, only 13.6 per cent of this age group use the Internet daily or almost daily.

Table 3 Internet use frequency per age group in EU27

	Everyday/Almost everyday	Two or three times a week	Less often	No access or never use
18-24	75,8%	11,2%	5,8%	7,2%
25-34	69,3%	10,9%	7,1%	12,8%
35-49	54,3%	14,1%	9,3%	22,3%
50-64	39,3%	9,4%	9,0%	42,3%
>64	13,6%	5,0%	5,5%	75,9%

Source: Own analysis of Eurobarometer 74.3 (using W22 WEIGHT EU 27 which provides adjustments for each national sample in proportion to its share in the total EU population aged 15 and over). Cases with respondent's age less than 18 years old were filtered out.

What are the implications of the age distribution of Internet users for the analysis of the traffic of a European website? If we assume that a website is equally appealing to everyone regardless of the age group he/she belongs to, we should not expect to find the website visitors with an age distribution similar to the age distribution of the entire population; it should look more like the age distribution of the population of Internet users. To make the differences between the two distributions clear on Diagram 1 I present a comparison of age distributions between total sample (using W22 WEIGHT EU 27 of Eurobarometer 74.3) and the Internet users of the same dataset. It becomes obvious that people aged 50 or more are under-represented in the Internet population.

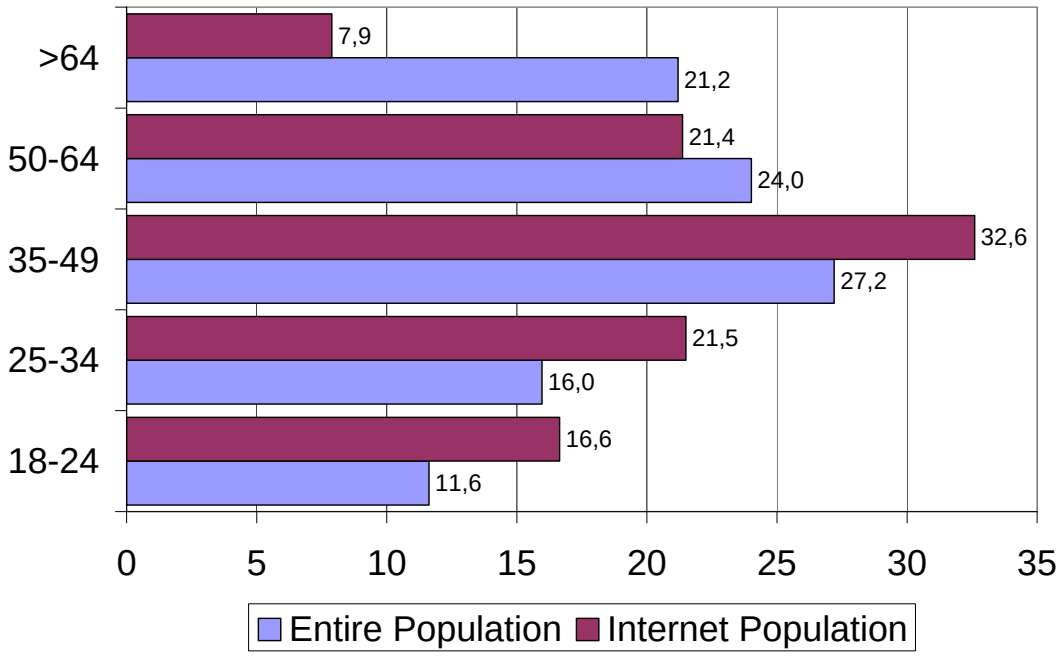

Diagram 1 Comparison of age distributions between total sample and Internet users

Our main point here is that if we observe that the age group of 65+ years old corresponds only to 7.9% of the total population of the users of a European VAA (instead of 21.2% which is the share of this age group in the total EU27 population), we should not be surprised. On the contrary, if the website is equally appealing to all age groups, then the age distribution of its visitors will be similar to the age distribution of the Internet population.

Table 4. Age Distribution. Population, Internet users and average difference

	Population					Internet users				
Nationality	18-24	25-34	35-49	50-64	>64	18-24	25-34	35-49	50-64	>64
Greek	13,8%	17,9%	27,4%	20,6%	20,3%	26,5%	29,9%	33,1%	9,4%	1,1%
Dutch	9,5%	14,9%	30,3%	27,0%	18,3%	10,1%	15,8%	31,9%	27,8%	14,4%
French	11,6%	14,6%	25,7%	26,3%	21,8%	15,9%	19,3%	31,0%	24,1%	9,7%
Italian	9,5%	15,1%	29,1%	22,3%	23,9%	15,0%	21,9%	38,6%	18,8%	5,6%
Polish	14,6%	19,2%	24,3%	25,4%	16,5%	24,3%	27,2%	28,9%	16,5%	3,1%
Swedish	9,5%	12,4%	29,1%	26,0%	23,1%	10,5%	13,5%	31,6%	27,0%	17,3%
Portuguese	14,9%	5,2%	26,7%	33,2%	20,0%	35,0%	10,8%	34,2%	17,4%	2,6%

Source: Eurobarometer 74.3 (using W22 WEIGHT EU 27 which provides adjustments for each national sample in proportion to its share in the total EU population aged 15 and over). Cases with respondent's age less than 18 years old were filtered out.

We should point out that the differences of the age distributions between the entire population and the Internet populations are not the same in each EU country. Table 4 provides adequate evidence to support this argument: by comparing the shares of people aged over 64 between the national populations and the corresponding Internet populations in seven EU countries we can observe that this age group is seriously under-represented in Eastern and Southern Europe and slightly under-represented in Sweden and the Netherlands. As a result if people aged over 64 represent only 1% of the visitors of a Greek website and 14,4% of the visitors of a Dutch website, this difference can be totally explained by the differences of the age distributions between the corresponding Internet populations.

Education

Table 5. Internet use frequency per age when stopped full-time education

	Everyday/Almost everyday	Two or three times a week	Less often	No access or never use
No full-time education	10,3%	,6%	,6%	88,5%
Up to 15 years	14,5%	5,2%	5,9%	74,4%
16-18 years	41,9%	13,4%	10,9%	33,8%
19 - 23 years	61,8%	11,7%	7,8%	18,7%
24 years and older	77,8%	9,3%	3,9%	9,1%
Still studying	87,6%	7,3%	3,4%	1,7%

Source: Eurobarometer 74.3 (using W22 WEIGHT EU 27 which provides adjustments for each national sample in proportion to its share in the total EU population aged 15 and over). Cases with respondent's age less than 18 years old were filtered out

.

Table 5 shows the relationship between age of respondents when they stopped full-time education and frequency of Internet use (Denissen, Neumann, and van Zalk, 2010). The range of values for the percentage of the people who do not have Internet access or they never use it, is even larger than the corresponding range observed for the age groups, i.e. from 7.2% for the younger group to 75.9% for the older group. For the education level groups these values range from 1.7% for people who are still studying up to 88.5% for people who have not had full time education

Table 6 Comparison of education level distributions between total sample and Internet users

	Entire Population	Internet Population
No full-time education	0,6%	0,1%
Up to 15 years	22,1%	8,6%
16-18 years	34,0%	34,1%
19 - 23 years	23,1%	28,4%
24 years and older	10,9%	15,0%
Still studying	9,2%	13,8%

Source: Eurobarometer 74.3 (using W22 WEIGHT EU 27 which provides adjustments for each national sample in proportion to its share in the total EU population aged 15 and over). Cases with respondent's age less than 18 years old were filtered out.

Similarly to age distribution of the visitors of a EU website, we should expect that the education level distribution should look more like the education level distribution of the population of Internet users. To make the differences between the two distributions clear, we present on Table 6 a comparison of education level distributions between the total sample (using W22 WEIGHT EU 27 of Eurobarometer 74.3) and the Internet users of the same dataset. It is obvious that the group of people who stopped full time education when they were 15 years old or younger is under-represented in the Internet population, and the group of people who continued after the age of 18 is over-represented.

We use data from the Eurobarometer 74.3, and we present the differences of education level distributions between the populations and the Internet users of seven European countries. It is obvious that the group of people who stopped full time education when they were 15 years old or younger is far more in all countries, except for the Netherlands and Sweden. In fact the education distribution of the Dutch and Swedish Internet populations is very similar to the education distribution of the general Dutch population. At this point, we should mention that according to Andreadis (2013a) the education level is also associated with the frequency and the type of internet use: i) the less educated are people the less likely they are to use Internet, ii) among Internet users: higher education corresponds to more frequent internet use and iii) among Internet users: the higher the education level of an Internet user, the more probable it is that he/she will visit a website that provides useful information (e.g. government services).

TABLE 7 ABOUT HERE

Gender

Table 8 shows the relationship between gender and frequency of Internet use. The percentages of the two middle columns ("Two or three times a week" and "Less often") are very similar for both men and women. The difference is observed on the daily and almost daily Internet use category(the male group is 9% larger than the female group), and on the last column showing that women are more likely than men to have no Internet access or to never use it.

Table 8 Internet use frequency per gender

	Everyday/Almost everyday	Two or three times a week	Less often	No access or never use
Male	52,9%	10,1%	7,2%	29,8%
Female	43,9%	10,2%	7,9%	38,1%

Source: Eurobarometer 74.3 (using W22 WEIGHT EU 27 which provides adjustments for each national sample in proportion to its share in the total EU population aged 15 and over). Cases with respondent's age less than 18 years old were filtered out.

In order to be able to compare these findings with the VAA data presented in the following sections, we present in Table 9 the differences of gender distributions between the general population and Internet users in seven European countries. It is clear that Greece, Italy, Poland and Portugal women are underrepresented when comparing the total population with the population of internet users.

TABLE 9 ABOUT HERE

Internet Populations and VAA users

The digital divide is not permanent. As technology costs decline, differences in all aspects of living standards decrease. Initially, all new technologies are available at a high price and are used only by wealthier people, but as the volume of use increases, prices decrease, and the new technology products become available to almost everybody.

Selwyn (2004) suggests that the digital divide does not simply mean the binary distinction of ICT access or not, e.g. accessing online information from a home-based computer is different from accessing the same materials in a public library regarding time, privacy and ease of use.

Norris (2001, p.4) describes three different dimensions of the digital divide: i) the global divide that refers to the lower rates of Internet penetration in the developing societies, ii) the social divide that refers to the information gap between the rich and poor in each nation and iii) the democratic divide, which concerns differences within the online community between those who actively use the Web for political information and those who do not. Norris (p.12) seems to agree with the idea that even if Internet penetration rates gradually widen throughout society, a substantial democratic divide may remain in place. She offers several possible explanations: i) the increase of Internet access will lead towards the opposite direction by reinforcing the divisions between the information rich and the information poor, thus further facilitating the participation of the activists while leaving the disengaged from the

politics of the real world further behind in the digital world or ii) the Internet will not have a significant impact, the situation will remain the same, i.e. "Politics as usual" and/or the traditional interests and established authorities will reassert their control on the virtual political sphere.

Min, S.J. (2010) analysing data from the 2004 US General Social Survey shows that political Internet users are individuals with high Internet skills and political interest. According to Min as the Internet penetration rates increase in all sectors of society, the importance of Internet skills and political interest will matter even more, and in the same time, we may observe a decrease in the importance of socioeconomic and demographic factors as the Internet is more and more evenly accessed and used across the population.

According to Johnson and Kaye (2003) the respondents to their web survey report that their involvement in politics has increased or greatly increased since they first became online users. Lupia & Philpot (2005) conclude that the Internet can increase young adults' interest in politics, but this impact depends on the web sites visited because some sites are more likely than others to affect political interest. Strömbäck and Shehata (2010) found that attention to political news and political interest are related, and this relation is both causal and reciprocal.

In countries where almost everyone is online daily or almost daily, it is meaningless to attribute differences of VAA use to differences of frequency or type of Internet use, because practically there are no such differences. Thus, in these countries the main variable that discriminates between VAA users and non users is political interest. Thus if we examine two subgroups that display different levels of political interest (such as the gender subgroups in the Netherlands) we should expect differences of VAA use between these two groups. In fact, since the literature indicates that political interest is lower among women than among men (Banwart 2007; Verba et al. 1997), we should expect that in countries

From the aforementioned literature we can conclude that a user of a VAA can be described by two significant characteristics: i) Use of Internet (including the frequency and the type of use) ii) political interest. Of course these two characteristics are not independent of each other, but the primary focus of this paper is to display the impact of Internet use on VAA use.

TABLE 10 ABOUT HERE

Although in Greece the gender distribution of HelpMeVote is very similar to the gender distribution of the general Greek Internet population, in the Netherlands male users are over-represented in the group of Kieskompas users by about nine percentage points. Moreover, in all other countries, female VAA users are heavily underrepresented.

The difference may be explained by the differences observed in political interest. Using the NES from Greece 2012 and the Netherlands 2010 we observe that in both countries the percentage of men who are very interested in politics is larger than the percentage of women who are very interested in politics, but the difference between men and women is different between the two countries. The gap in Greece is 6.4%

(men 19.1%, women 12.8%) while in the Netherlands is 11.2% (men 19.8%, women 8.5%). A similar difference is observed if we use the European Social Survey wave 5 2010 data: NL: male: 18,2%, female: 9,5% Greece: male: 9,6%, female: 5,2%.

Another reason is the difference of missing values between HelpMeVote, Kieskompas and EU profiler datasets. In HelpMeVote, before giving the output, users are asked to fill-in a short form with their personal information (mostly demographics, i.e. Sex, Age group, Education Level, but also related to their voting behaviour, i.e. Vote Choice, Confidence to their vote intention). Although it is not mandatory (users can click "continue" and move on to the output without answering) the vast majority responds to these questions, probably, because they are in responsive mood, or because they consider this form as part of the procedure. As a result, the gender information is available for almost all HelpMeVote users and the missing values correspond to 6.7% of the total number of users. On the other hand, 23.8% of the Kieskompas users have not indicated their gender. Thus, if more female than male Kieskompas users avoid to provide their gender information, then the aforementioned difference could be a result of non-response bias. In the EU profiler datasets the demographic questions are available for only a small subset of the sample (the percentage varies between countries but it is less than 5%).

Table 11 Comparison of education level distributions between VAA and Internet users in Greece and the Netherlands

	HelpMeVote users	Greek Internet users	Kieskompas users	Dutch Internet users
Compulsory (up to 15 years)	4,6%	2,7%	7,7%	7,1%
Secondary (16-18 years)	27,9%	32,3%	33,7%	24,3%
Tertiary (>18 years)	67,5%	65,0%	58,6%	70,6%

Table 11 displays the comparison of education level distributions between HelpMeVote population and Greek Internet users and between Kieskompas users and Dutch Internet users. The distributions are very similar.

TABLE 12 ABOUT HERE

We believe that Table 12 is very interesting because VAA researchers usually find that older people are under-represented and younger people are over-represented in VAAs. For instance, in Greece this is only partially true: it is true if one compares VAA users with the total electorate, but not all of them are able to use the VAA. The eligible population is the group of Internet users. And if we compare HelpMeVote users with the corresponding eligible population (Greek Internet users) we will observe that older people are in fact over-represented. The 65 years and older age group is overrepresented in Italy and Portugal and underrepresented in the rest of the countries under analysis.

On the other hand, in Greece younger people (aged 18-24) are under-represented. Of course this can be easily explained in terms of both type of use (younger people tend

to spend more time online on non political activities) and political interest which is lower in the age group 18-24. As Table 12 indicates in addition to Greece, in most other countries the age group 18-24 is slightly under-represented (France, Italy, Sweden) or heavily under-represented (Sweden) and only in the Netherlands and Poland seems to be evidence supporting over-representation of this age group.

Conclusions

In spite of this emerging interest in VAAs among academic political scientists, research thus far has been slow to realise the potential of VAAs as a source of voluminous and dynamic public opinion data. While techniques such as panel studies and rolling cross section survey design (Brady and Jonhston, 2009) have been used to capture campaign dynamics in elections in the United States and Canada, such designs have been less common elsewhere. van der Eijk (2002) was an early proponent of the potential advantages of using online methodologies to collect bulk samples that would then be matched against smaller randomly collected samples which would capture information on 'core' variables (i.e., the variables that will be used to adjust the non-representative bulk sample). Thus, matched or weighted survey data that includes information on the timing of questionnaire completion could provide valuable new insights into the evolution of public opinion in a range of countries.

The research of Alvarez et al. (2011) seeks to unlock this potential by using three weighting and matching approaches for data generated by the 2009 EU Profiler VAA. As such, the approaches developed to treat VAA data could ultimately prove useful for all public opinion surveys in societies with sufficiently high levels of internet access. Alvarez et al. (2011) make some promising advances, most notably in using combinations of responses as merging variables. Emerging research from the USA (Vavreck and Rivers, 2008) indicate that 'sample matching' can alleviate biases in such samples, although this is still a matter of contention in the literature (Bethlehem, 2008). However, the potential of VAA generated data to improve our understanding of politics in established democracies remains significantly underexplored, and the proposed project seeks to begin to address this lacuna in the literature.

The findings presented in this paper represent an empirically-oriented effort to evaluate the sources and nature of VAA bias, by relating it to the factors that drive internet use more generally. To the extent that these factors overlap, survey data on internet use can be an effective source for correcting VAA data and unlocking its potential as a source of political analysis.

References

Andreadis, I. (2012) To Clean or not to Clean? Improving the Quality of VAA Data, at the XXIInd World Congress of Political Science, (IPSA), Madrid http://www.polres.gr/en/sites/default/files/IPSA12.pdf

Andreadis, I. (2013a) Who responds to website visitor satisfaction surveys? General Online Research Conference GOR 13 March 04-06, 2013, DHBW Mannheim, Germany http://www.polres.gr/en/sites/default/files/GOR13.pdf

Andreadis, I. (2013b) Voting Advice Applications: a successful nexus between informatics and political science. *BCI '13*, September 19 - 21 2013, Thessaloniki, Greece http://www.polres.gr/en/sites/default/files/BCI-2013.pdf

Alvarez, M, Franklin, M., del Sio, L, Trechsel, A. (2011) 'Weighting and Matching as Methods for Combining Disparate Survey Data – The European Election Study 2009 and the EU Profiler' paper presented at: Midwest Political Science Association Annual Conference, Chicago.

Banwart, M.C. (2007). Gender and young voters in 2004 the influence of perceived knowledge and interest. American Behavioral Scientist 50 (9): 1152-68.

Bethlehem, J. (2008). Can we make official statistics with self-selection web surveys?. In Proceedings of Statistics Canada Symposium.

Bethlehem, Jelke. (2010). Selection bias in web surveys. International Statistical Review 78 (2): 161-188.

Brady, H., Johnston R. (2009) 'The Rolling Cross-Section and Causal Attribution' in H. Brady & R. Johnston (Eds.) *Capturing Campaign Effects.* Ann Arbor: the University of Michigan Press pp. 164-199.

Chiranov, M. (2011) Applying pop-up survey software to incorporate users' feedback into public library computing service management. *Performance Measurement and Metrics*, Vol. 12 Iss: 1, pp.50 - 65

Couper, M.P. (2000) Web surveys – A review of issues and approaches, *Public Opinion Quarterly*, 64, pp. 464–494

Couper, M.P. (2008). *Designing Effective Web Surveys*. New York: Cambridge University Press

Couper, Mick P., A. Kapteyn , M. Schonlau, and J. Winter. (2007). Noncoverage and nonresponse in an Internet survey. *Social Science Research*, 36(1): 131-148.

Denissen, Jaap J.A. Neumann, Linus and van Zalk, Maarten (2010) How the internet is changing the implementation of traditional research methods, people's daily lives, and the way in which developmental scientists conduct research?. *International Journal of Behavioral Development* 34: 564-575, doi:10.1177/0165025410383746

Dillman DA (2007). *Mail and internet surveys: the tailored design method* (2nd edition). New York, NY: John Wiley and Sons, Inc.

Diment, K. and S. Garrett-Jones (2007) How demographic characteristics affect mode preference in a postal/web mixed-mode survey of Australian researchers *Social Science Computer Review*, 25 pp. 510–517

EGOV 2007, LNCS 4656, pp. 293–304, 2007. Springer-Verlag Berlin Heidelberg 2007

European Commission (2013): Eurobarometer 74.3 (2010). TNS OPINION & SOCIAL, Brussels [Producer]. GESIS Data Archive, Cologne. ZA5450 Data file Version 5.2.0, doi:10.4232/1.11627

Fan, Weimiao and Yan, Zheng. (2010). Factors affecting response rates of the web survey: A systematic review, Computers in Human Behavior, Volume 26, Issue 2, March, Pages 132–139 http://dx.doi.org/10.1016/j.chb.2009.10.015

Germann, M., & Mendez, F. (2013). Dynamic Scale Validation Reloaded: Assessing the Psychometric Properties of Latent Measures of Ideology in VAA Spatial Maps. Paper presented at the 2013 ECPR General Conference; 4–7 September, Bordeaux, France.

Germann, M., Mendez, F., Wheatley, J., & Serdült, U. (2014). Spatial maps in voting advice applications: The case for dynamic scale validation. Acta Politica.

Hooghe, Marc, and Wouter Teepe. (2007). Party profiles on the web: an analysis of the logfiles of non-partisan interactive political internet sites in the 2003 and 2004 election campaigns in Belgium. *New Media & Society*, 9 (6): 965-985.

Jiglău, G., Burean, T., & Bădescu, G. The Ideological Mapping of Political Parties in Romania. The Relationship between Dimensions of Competition and Ideological Consistency. Paper presented at the 2013 ECPR General Conference; 4–7 September, Bordeaux, France.

Johnson, Thomas J., and Barbara K. Kaye. 2003. A boost or bust for democracy? How the web influenced political attitudes and behaviors in the 1996 and 2000 presidential elections. *The International Journal of Press/Politics* 8 (3): 9-34.

Kloss, L., Zhang, Y. (2003), An evaluative case study of a real-time online reference service, *The Electronic Library*, Vol. 21 No.6, pp.565-75.

Kreuter, Frauke, Stanley Presser, and Roger Tourangeau. (2008). Social desirability bias in CATI, IVR, and Web surveys: The effects of mode and question of sensitivity. *Public Opinion Quarterly*, 72 (5): 847–865.

Krouwel, André, Thomas Vitiello, and Matthew Wall. 2012. The practicalities of issuing vote advice: a new methodology for profiling and matching. *International Journal of Electronic Governance*, 5 (3): 223-243.

Lankes, R.D. (2005), Digital reference research: fusing research and practice, *Reference & User Services Quarterly*, Vol. 44 No.4, pp.320-6.

Lupia, Arthur, and Tasha S. Philpot. 2005. Views from inside the net: How websites affect young adults' political interest. *Journal of Politics* 67 (4): 1122-42.

Miller, T.I., M. Miller-Kobayashi, E. Caldwell, S. Thurston, B. Collett (2002). Citizen surveys on the web: General population surveys of community opinion. *Social Science Computer Review*, 20 (2), pp. 124–136

Min, S.J. (2010) From the Digital Divide to the Democratic Divide: Internet Skills, Political Interest, and the Second-Level Digital Divide in Political Internet Use, *Journal of Information Technology & Politics*, 7:1, 22-35, DOI: 10.1080/19331680903109402

Norris, P. (2001). *Digital divide: Civic engagement, information poverty, and the internet worldwide*. Cambridge University Press.

Olson, Kristen. (2006). Survey participation, nonresponse bias, measurement error bias, and total bias. *Public Opinion Quarterly*, 70 (5): 737-758.

Ruppel, M., Fagan, J.C. (2002), Instant messaging reference: users' evaluation of library chat, Reference Services Review, Vol. 30 No.3, pp.183-97. Chang, H.R., Holland, M.P. (2005), "User satisfaction survey of ask-a-question service at the internet public library", *Internet Reference Services Quarterly*, Vol. 10 No.2, pp.61-73.

Sakshaug, Joseph W., Ting Yan, and Roger Tourangeau. (2010). Nonrepsonse error, measurement error, and mode of data collection. *Public Opinion Quarterly*, 74(5): 907–933.

Selwyn, N. (2004). Reconsidering political and popular understandings of the digital divide. *New Media & Society* 6 (3): 341-62.

Trechsel, A (2010). EU-Profiler: user profiles in the 2009 European Elections, European University Institute [Distributor] V5 [Version]

van der Eijk, C. (2002) 'Design issues in electoral research: taking care of (core) business' *Electoral Studies* 21(2): 189-206.

Vavreck, L., Rivers, D.(2008) 'The 2006 Cooperative Congressional Election Study,' Journal of. Elections, *Public Opinion, and Policy*, 18 (4): 355-366.

Verba, S., Burns, N. and Schlozman K.L. (1997). Knowing and caring about politics: Gender and political engagement. *Journal of Politics* 59 : 1051-72.

Table 7 Comparison of education level distributions between total sample and Internet users in Seven European Countries

	Greek Population	Greek Internet users	Dutch Population	Dutch Internet users	French population	French Internet users	Italian population	Italian Internet users	Polish population	Polish Internet users	Swedish population	Swedish Internet users	Portuguese population	Portuguese Internet users
No full-time education	1,1%	0,6%	0,0%	0,0%	0,4%	0,0%	0,5%	0,0%	0,3%	0,3%	0,0%	0,0%	5,5%	0,0%
Up to 15 years	30,3%	2,1%	7,1%	5,1%	17,0%	6,8%	32,7%	10,3%	13,0%	2,4%	6,7%	4,1%	58,2%	22,4%
16-18 years	33,8%	32,3%	24,6%	24,3%	34,3%	33,1%	19,7%	22,8%	30,8%	18,7%	14,9%	14,0%	16,7%	32,5%
19 - 23 years	17,8%	29,3%	34,5%	35,4%	27,8%	33,0%	28,3%	38,2%	29,2%	36,4%	23,6%	24,3%	10,3%	23,4%
24 years and older	6,9%	13,4%	21,5%	22,2%	12%	15,4%	10,0%	13,5%	15,1%	22,1%	40,4%	41,7%	4,2%	7,3%
Still studying	10,2%	22,3%	12,3%	13,0%	8,5%	11,6%	8,9%	15,2%	11,6%	20,1%	14,4%	15,8%	5,2%	14,3%

Eurobarometer 74.3 (using W22 WEIGHT EU 27 which provides adjustments for each national sample in proportion to its share in the total EU population aged 15 and over). Cases with respondent's age less than 18 years old were filtered out.

Table 9 Comparison of gender distributions between total sample and Internet users in Seven European Countries

	Greek Population	Greek Internet users	Dutch Population	Dutch Internet users	French population	French Internet users	Italian population	Italian Internet users	Polish population	Polish Internet users	Swedish population	Swedish Internet users	Portuguese population	Portuguese Internet users
Male	49,0%	57,0%	49,3%	49,7%	47,7%	49,1%	48,1%	51,3%	47,7%	50,3%	49,4%	49,9%	47,7%	54,1%
Female	51,0%	43,0%	50,7%	50,3%	52,3%	50,9%	51,9%	48,7%	52,3%	49,7%	50,6%	50,1%	52,3%	45,9%

Eurobarometer 74.3 (using W22 WEIGHT EU 27 which provides adjustments for each national sample in proportion to its share in the total EU population aged 15 and over). Cases with respondent's age less than 18 years old were filtered out.

Table 10 Comparison of gender distributions between Internet Populations and VAA users

	HelpMeVote users	Greek Internet users	Kieskompas Users	Dutch Internet users	French EUProfiler users	French Internet users	Italian EUProfiler users	Italian Internet users	Polish EUProfiler users	Polish Internet users	Swedish EUProfiler users	Swedish Internet users	Portugese EUProfiler users	Portuguese Internet users
Male	57,8%	57,0%	58,9%	49,7%	72,03%	49,10 %	69,55%	51,30%	68,86%	50,30%	61,80%	49,90%	62,43%	54,10%
Female	42,2%	43,0%	41,1%	50,3%	27,97%	50,90 %	30,45%	48,70%	31,14%	49,70%	38,20%	50,10%	37,57%	45,90%

Source: HelpMeVote 2012; Kieskompas 2012; Euprofiler 2009; Eurobarometer 74.3

Table 12 Comparison of age distributions between VAA users and Internet population of Seven European Countries

	HelpMeVote users	Greek Internet users	Kieskompas Users	Dutch Internet users	French EUProfiler users	French Internet users	Italian EUProfiler users	Italian Internet users	Polish EUProfiler users	Polish Internet users	Swedish EUProfiler users	Swedish Internet users	Portugese EUProfiler users	Portuguese Internet users
18-24	16,3%	26,5%	17,9%	10,1%	13,4%	15,9%	13,6%	15,0%	40,7%	24,3%	8,1%	10,5%	11,4%	35,0%
25-34	31,6%	29,9%	24,4%	15,8%	28,1%	19,3%	23,9%	21,9%	30,9%	27,2%	23,6%	13,5%	26,4%	10,8%
35-49	35,0%	33,1%	32,1%	31,9%	29,2%	31,0%	27,6%	38,6%	14,0%	28,9%	36,2%	31,6%	32,5%	34,2%
50-64	14,4%	9,4%	19,5%	27,8%	21,8%	24,1%	23,5%	18,8%	11,3%	16,5%	24,3%	27,0%	23,0%	17,4%
>64	2,7%	1,1%	6,2%	14,4%	7,4%	9,7%	11,3%	5,6%	1,7%	3,1%	7,6%	17,3%	6,7%	2,6%

Source: HelpMeVote 2012; Kieskompas 2012; Euprofiler 2009