Internet Penetration and Voting Advice Applications

Andreadis, Ioannis; Wall, Matthew; Krouwel, André

2015

document version
Other version

Internet Penetration and Voting Advice Applications

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

Abstract
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. In this paper we try to provide some answers to the question of the representativeness of a VAA sample. We do this using data from the Greek Voting Advice Application HelpMeVote/Votemach 2014 and Euvox which was used in most of the EU member states. We run a cross-national comparative analysis of VAA users in order to examine if the observed differences of the sample bias can be explained by the differences of the Internet populations among countries.

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 and to explore the dimensionality of the political space (Germann & Mendez 2013; Germann et al. 2014; Jiglău et al. 2013, Wheatley et al., 2012).

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

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

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 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 80.1 and compares it to VAA data in 25 European countries: FR - France
NL - The Netherlands, DE - Germany, IT - Italy, DK - Denmark, IE - Ireland, GB - Great Britain, GR - Greece, ES -Spain, PT - Portugal, FI - Finland, SE - Sweden, AT - Austria, CY - Cyprus, CZ - Czech Republic, EE - Estonia, HU - Hungary, LV - Latvia, LT - Lithuania, PL - Poland, SK - Slovakia, SI - Slovenia, BG - Bulgaria, RO - Romania, and HR – Croatia. There were not enough Euvox users in Belgium, Malta and Luxembourg, so the data for these countries are not available. Euvox had dual websites for the Netherlands, France and Sweden and the data were stored on different databases. In order to use the maximum available information we have merged these datasets. For the same reason, we have also used the Greek Voting Advice Application: HelpMeVote/VoteMatch 2014 which includes circa 80000 cases (Andreadis and Chadjipadelis, 2014).

According to Andreadis (2012; 2014), 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 issue of coverage error in VAAs, i.e., we deal with the problem that occurs due to 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 to 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. 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, 2013b), and Euvox, partner of Kieskompas (Krouwel et al, 2012). 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 using data from the Eurobarometer 80.1 (European Commission, 2014) which includes fresh data (Date of Collection: 02.11.2013 - 17.11.2013) on the frequency of Internet use by the EU citizens. We use Eurobarometer because the findings presented in the following sections of this paper are based on the analysis of European VAA data.

Table 1 INTERNET USE (INDEX) frequency in EU28

	Frequency	Percent	Cumulative Percent
Everyday/Almost everyday	15677	57,7	57,7
Two or three times a week	2725	10,0	67,7
About once a week	790	2,9	70,6
Two or three times a month	367	1,4	72,0
Less often	771	2,8	74,8
Never/No access	5065	18,6	93,4
No Internet access at all	1781	6,6	100,0

Source: Eurobarometer 80.1 (using W23 WEIGHT EU 28 which provides adjustments for each national sample in proportion to its share in the total EU population aged 15 and over)".

As Table 1 shows, not all EU citizens use the internet. Approximately one out of four EU citizens is either without Internet access or has Internet access but never uses it. In addition, not every internet user uses the web with the same frequency. In fact, just under than 6 out of 10 EU citizens use 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, Bulgaria, Greece, Croatia and Cyprus. In all these countries more than 35% of the respondents were found with no Internet use. The Netherlands and Sweden are at the bottom of the list (less than 5 per cent of no Internet use).

Nevertheless, there are other differences than the obvious (i.e. between people who use and do not use the Internet). There are differences between Internet users regarding the frequency of use. For instance, Ireland 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 Ireland who do not use the Internet daily or almost daily, thus Ireland ranks tenth according to the percentage of 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 half 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	Less often	No access or never use
PT - Portugal	38,6%	12,7%	48,7%
RO - Romania	37,9%	18,3%	43,8%
BG - Bulgaria	41,7%	17,0%	41,3%
GR - Greece	44,9%	14,7%	40,4%
HR - Croatia	47,8%	15,0%	37,2%
CY - Cyprus	52,3%	11,4%	36,4%
HU - Hungary	41,5%	24,4%	34,0%
PL - Poland	45,8%	21,6%	32,6%
ES -Spain	54,0%	15,2%	30,8%
LT - Lithuania	53,9%	15,7%	30,4%
IT - Italy	50,6%	19,4%	30,0%
SK - Slovakia	56,9%	18,6%	24,5%
SI - Slovenia	61,3%	14,3%	24,4%
CZ - Czech Republic	51,2%	26,3%	22,5%
AT - Austria	54,1%	23,9%	22,0%
DE - Germany	57,0%	21,5%	21,6%
EE - Estonia	69,2%	9,2%	21,5%
LV - Latvia	67,3%	13,3%	19,4%
FR - France	68,0%	14,0%	18,0%
GB - Great Britain	69,0%	13,1%	17,9%
IE - Ireland	60,5%	24,4%	15,1%
FI - Finland	72,9%	13,0%	14,0%
DK - Denmark	84,0%	8,1%	7,8%
NL - The Netherlands	87,7%	8,0%	4,3%
SE - Sweden	88,4%	7,4%	4,2%

Source: Eurobarometer 80.1 data (using W23 WEIGHT EU 28)

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.) can be at least partly explained by 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 35% of European citizens over the age of 64 use the Internet. Moreover, only 21.4 per cent of this age group use the Internet daily.

Table 3 Internet use frequency per age group in EU28

	Everyday	Less often	Never
18-24	88,0%	9,1%	2,9%
25-34	79,9%	14,8%	5,2%
35-49	66,6%	21,9%	11,6%
50-64	46,2%	21,8%	32,0%
>64	21,4%	13,3%	65,4%

Source: Eurobarometer 80.1 (using W23 WEIGHT EU 28 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. Our main point here is that if we observe that the age group of 65+ years old corresponds only to a small fraction of the total population of the users of a European VAA, 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.

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 25 EU countries we can observe that this age group is seriously under-represented in Eastern and Southern Europe, but only slightly under-represented in Sweden and the Netherlands. As a result if people aged over 64 represent only 1,2% of the visitors of a Greek website and 15,4% of the visitors of a Dutch website, this difference can be largely explained by the differences of the age distributions between the corresponding Internet populations.

Table 4. Age Distribution in General Population and Internet users

	General Population					Internet Users				
	18-24	25-34	35-49	50-64	65+	18-24	25-34	35-49	50-64	65+
FR	12,9%	14,7%	25,9%	24,9%	21,5%	15,5%	17,9%	30,3%	24,4%	11,9%
NL	11,4%	15,6%	27,0%	27,7%	18,2%	12,0%	16,3%	28,1%	28,3%	15,4%
DE	9,1%	14,0%	25,3%	25,5%	26,1%	11,4%	17,7%	30,4%	27,0%	13,6%
IT	10,2%	14,7%	29,1%	22,0%	24,0%	14,4%	20,4%	37,4%	18,5%	9,4%
DK	11,2%	14,9%	25,1%	27,5%	21,4%	12,2%	16,3%	27,0%	28,5%	15,9%
IE	13,5%	20,4%	29,1%	22,2%	14,8%	15,5%	23,8%	32,1%	20,7%	7,8%
GB	11,6%	16,8%	26,2%	24,1%	21,2%	13,6%	20,1%	29,8%	23,9%	12,7%
GR	14,8%	19,4%	24,1%	21,6%	20,1%	25,0%	29,5%	31,0%	13,4%	1,2%
ES	9,5%	17,9%	31,1%	20,8%	20,7%	13,6%	25,1%	39,6%	16,3%	5,4%
PT	10,1%	11,6%	28,3%	26,8%	23,2%	18,8%	20,2%	38,6%	18,4%	4,0%
FI	12,6%	14,0%	24,9%	27,6%	20,8%	14,7%	15,9%	27,9%	28,3%	13,1%
SE	7,4%	12,7%	30,7%	25,4%	23,9%	7,7%	13,3%	31,9%	26,1%	21,0%
AT	13,7%	15,2%	28,8%	22,2%	20,1%	17,6%	18,5%	35,0%	21,5%	7,4%
CY	16,3%	23,3%	25,6%	18,6%	16,3%	22,2%	37,0%	29,6%	7,4%	3,7%
CZ	13,1%	19,3%	26,1%	23,5%	18,1%	16,5%	23,7%	31,1%	21,3%	7,4%
EE	13,3%	18,3%	25,0%	23,3%	20,0%	17,0%	23,4%	31,9%	21,3%	6,4%
HU	13,1%	16,3%	29,9%	21,8%	18,9%	18,5%	22,3%	34,7%	17,1%	7,4%
LV	17,2%	20,4%	31,2%	19,4%	11,8%	21,6%	24,3%	33,8%	16,2%	4,1%
LT	14,8%	15,8%	24,6%	25,1%	19,7%	21,4%	22,2%	30,2%	21,4%	4,8%
PL	13,4%	19,4%	24,7%	25,8%	16,6%	19,8%	26,2%	32,0%	18,2%	3,8%
SK	15,2%	17,9%	29,7%	22,3%	14,9%	20,0%	22,7%	34,5%	18,2%	4,5%
SI	10,5%	17,5%	27,2%	24,6%	20,2%	14,1%	23,5%	32,9%	23,5%	5,9%
BG	13,1%	16,8%	23,9%	25,7%	20,5%	21,4%	25,4%	31,3%	18,3%	3,6%
RO	12,8%	18,8%	28,7%	21,1%	18,6%	21,7%	26,1%	31,8%	15,8%	4,6%
HR	11,5%	18,4%	24,2%	24,6%	21,3%	18,1%	28,2%	30,2%	16,8%	6,7%

Source: Eurobarometer 80.1 (using W23 WEIGHT EU 28 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.

Gender

Table 5 shows the relationship between gender and frequency of Internet use. 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 5 Internet use frequency per gender

	Everyday	Less often	Never
Male	62,3%	15,7%	22,0%
Female	53,4%	18,4%	28,2%

Source: Eurobarometer 80.1 (using W23 WEIGHT EU 28 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.

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. This is the reason that in Table 6 in addition to the gender distributions in the general population, and in the group of Internet users we present the gender distribution in the group of frequent Internet users.

It is clear that in Germany, Greece, Austria, Portugal, Spain and Slovakia women are underrepresented by 5-7 more percentage units when comparing the total population with the population of frequent (everyday) internet users.

Table 6. Gender Distribution in General Population and Internet users and Everyday Internet users

	General Population		Internet Users		Everyday Internet Users	
	male	female	Male	female	male	female
FR	48%	52%	49%	51%	50%	50%
NL	49%	51%	50%	50%	50%	50%
DE	49%	51%	52%	48%	56%	44%
IT	48%	52%	50%	50%	52%	48%
DK	49%	51%	49%	51%	50%	50%
IE	49%	51%	49%	51%	47%	53%
GB	49%	51%	51%	49%	52%	48%
GR	49%	51%	55%	45%	56%	44%
ES	49%	51%	52%	48%	55%	45%
PT	47%	53%	52%	48%	53%	47%
FI	49%	51%	50%	50%	50%	50%
SE	50%	50%	50%	50%	50%	50%
AT	48%	52%	51%	49%	55%	45%
CY	49%	51%	52%	48%	52%	48%
CZ	49%	51%	52%	48%	51%	49%
EE	44%	56%	45%	55%	45%	55%
HU	47%	53%	46%	54%	49%	51%
LV	46%	54%	47%	53%	45%	55%
LT	46%	54%	49%	51%	49%	51%
PL	48%	52%	49%	51%	51%	49%
SK	48%	52%	50%	50%	53%	47%
SI	49%	51%	52%	48%	52%	48%
BG	48%	52%	47%	53%	48%	52%
RO	48%	52%	49%	51%	50%	50%
HR	48%	52%	50%	50%	50%	50%

Digital divide and VAAs

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.

Of course, 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. This can be verified by comparing the Internet use tables presented in the previous sections of this article with the tables presented in previous versions of this article (see Andreadis, Wall and Krouwel, 2014), in which we were using data from: Eurobarometer 74.3 (European Commission, 2013) (see also Andreadis, 2013a). In about three years (Collection period of EB 74.3: 25.11.2010 - 17.12.2010) the percentage of EU citizens who never use the Internet dropped from about 1 out of 3 to about one out of 4.

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 with no differences in Internet use the main factor for VAA use should be political interest.

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, as the literature shows, are not independent of each other.

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.

Although there is a relationship between Internet use and political interest, we are dealing only with Internet use because the primary focus of this paper is to display the impact of Internet use on VAA use.

VAA users, General and Internet populations

Age

First, we test the hypothesis that the differences in the representation of older people (65+) in VAA samples can be attributed to differences in the general populations between the European Countries. It is obvious from Diagram 1 that there is not a significant correlation coefficient between the percentage of older people (65+) in the general population in each country and the percentage of older people (65+) in the group of VAA users in each country. The Pearson correlation coefficient is 0.066 (p=0.756).

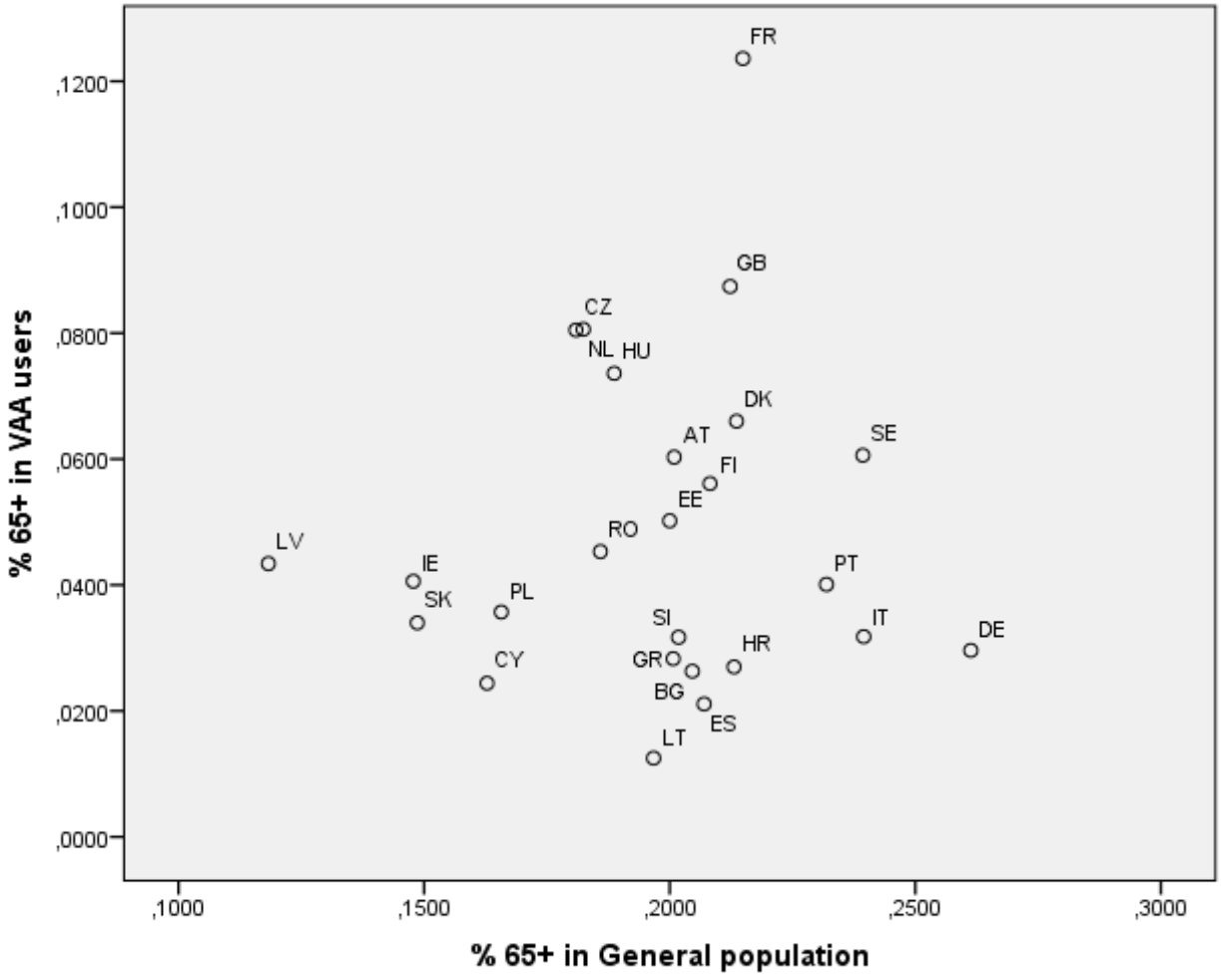


Diagram 1 Percentage of older people (65+) in the General population and in the group of VAA users.

Then, we test the hypothesis that the differences in the representation of older people (65+) in VAA samples can be attributed to differences in the Internet populations between the European Countries. Diagram 2 shows a significant positive correlation

coefficient between the percentage of older people (65+) in the Internet population in each country and the percentage of older people (65+) in the group of VAA users in each country. The Pearson correlation coefficient is 0.544 (p=0.005).

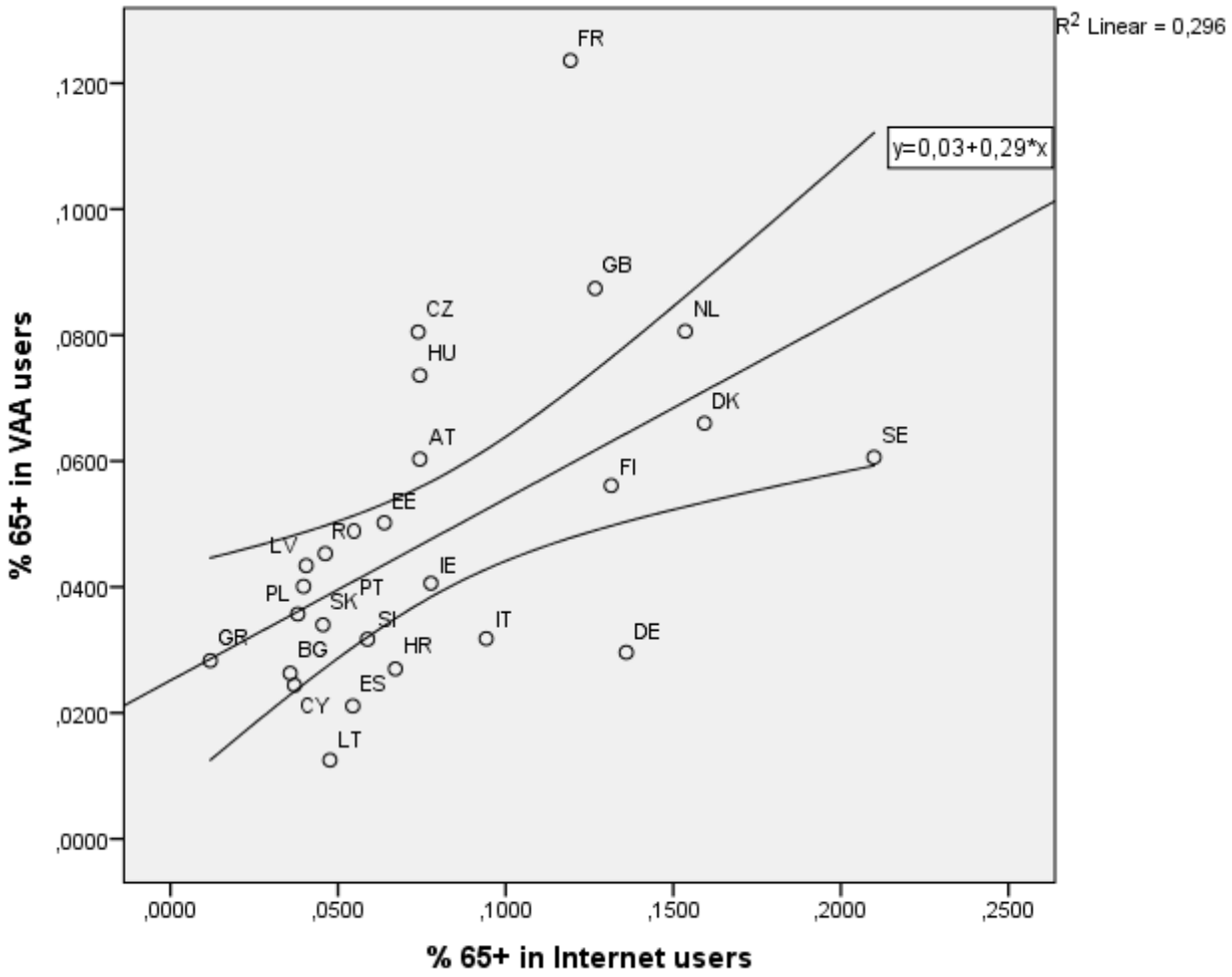


Diagram 2 Percentage of older people (65+) in the Internet population and in the group of VAA users.

We believe that Diagram 2 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 Greek VAA users with the corresponding eligible population (Greek Internet users) we will observe that older people are in fact over-represented.

Gender

First, we should note that the differences between the group of Internet users and the general population are smaller for gender than for age. For instance, the average difference between the percentage of women in the general population and the Internet population is about 2% and the corresponding difference for the group of older (65+) people is about 12%. Since the variability of the gender distributions in the Internet populations is smaller, we expect their explanatory power to be smaller

too. Thus we expect the variability of female percentages in VAA samples to have some relationship with the female percentages in the general population.

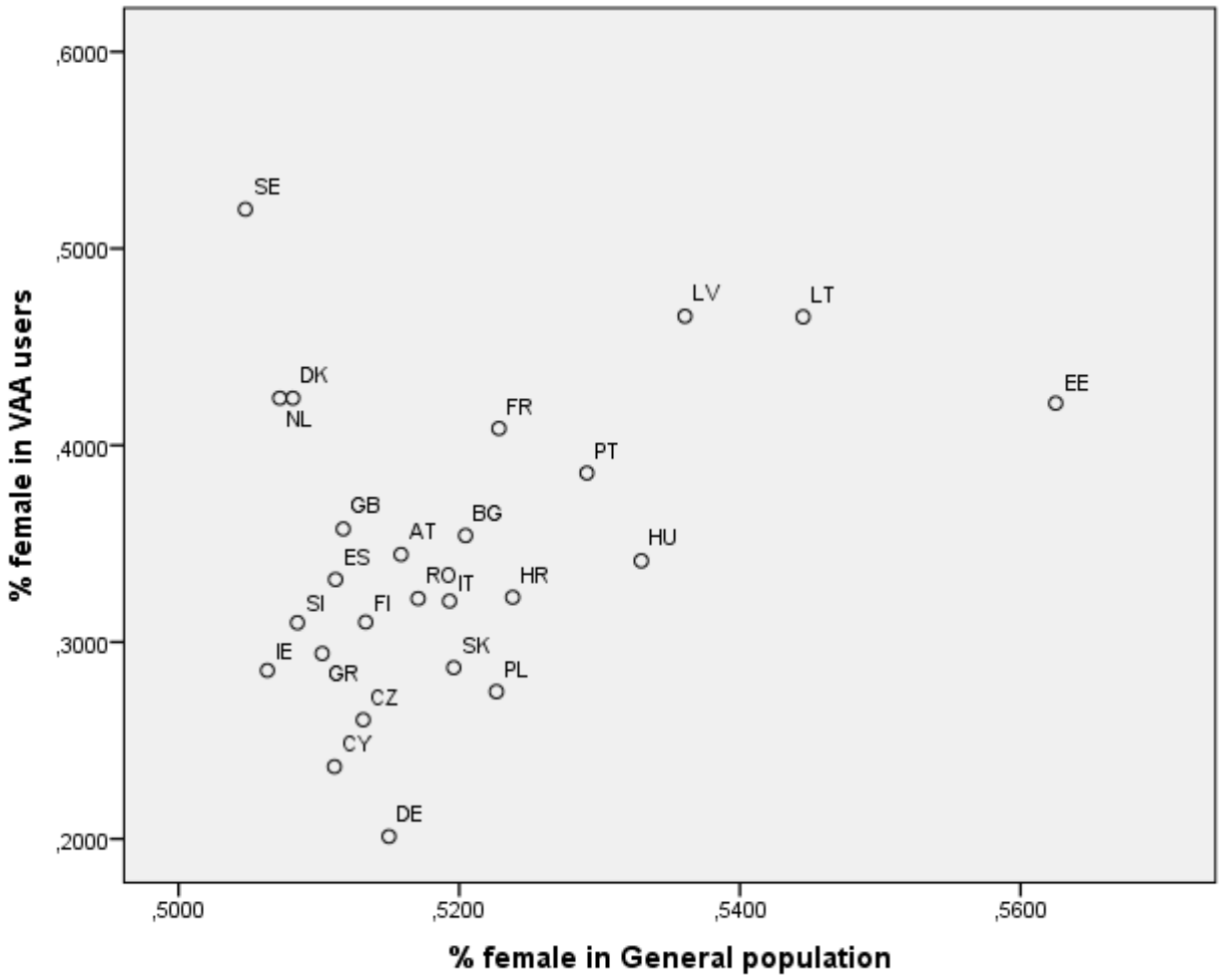


Diagram 3 Percentage of females in the General population and in the group of VAA users.

In Diagram 3 we test the hypothesis that the differences in the representation of females in VAA samples can be attributed to differences in the general populations between the European Countries. There is a correlation coefficient between female percentage in the general population in each country and female percentage in the group of VAA users in each country, but it is not statistically significant. The Pearson correlation coefficient is 0.330 (p=0.107).

In Diagram 4 we test the hypothesis that the differences in the representation of females in VAA samples can be attributed to differences in the Internet populations between the European Countries. The Pearson correlation coefficient is 0.465 (p=0.019).

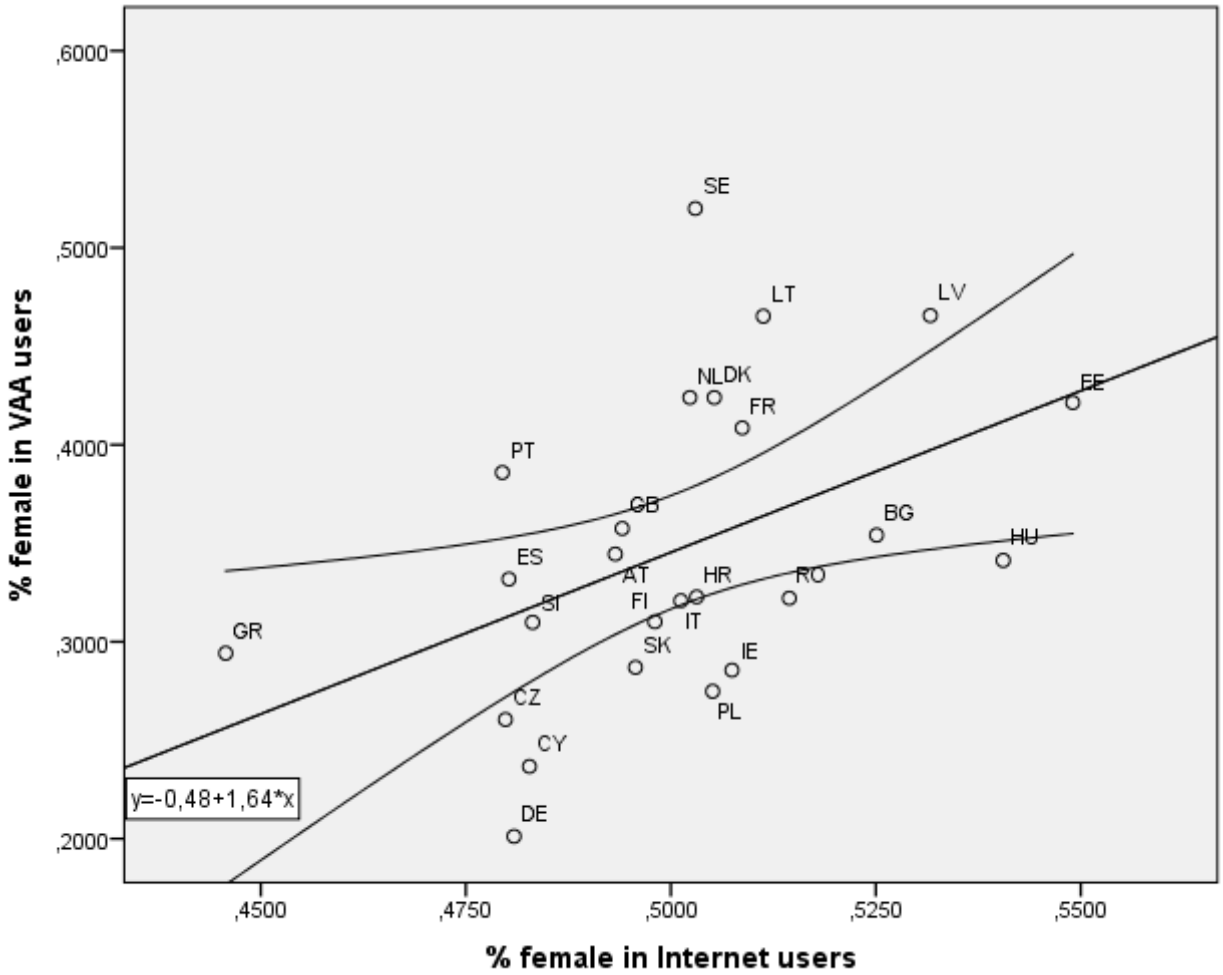
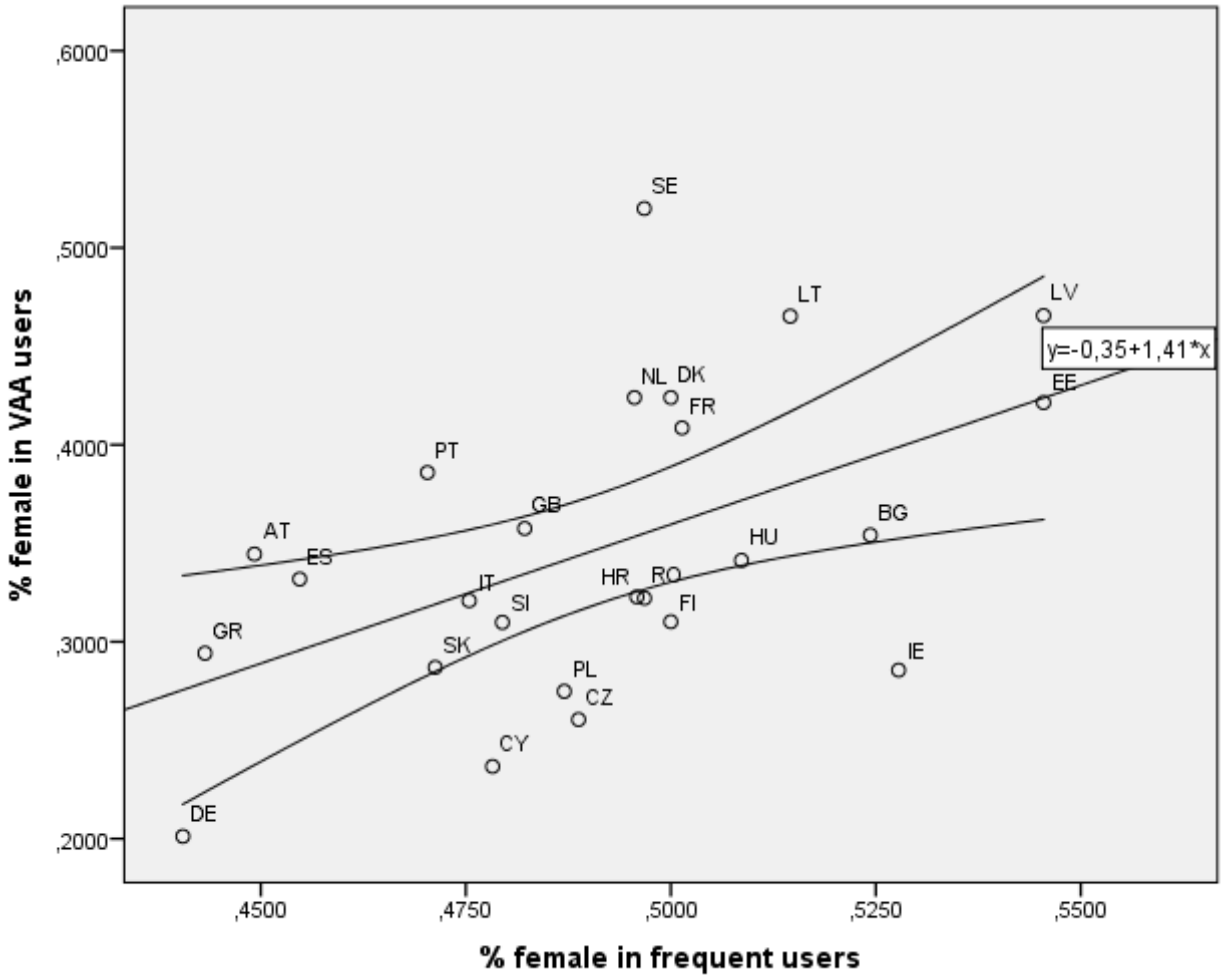


Diagram 4 Percentage of females in the Internet population and in the group of VAA users.

Finally, in Diagram 5 we test the hypothesis that the correlation coefficient is stronger if we replace the Internet population with the group of frequent Internet users between the European Countries. The Pearson correlation coefficient is 0.516 (p=0.008).

Conclusions

In this paper we have shown that the variability of the age and gender distributions in VAA samples can be (at least) partially explained by the variability of the corresponding distributions in the Internet populations of the European countries. We have also shown that in the countries of Southern and Eastern where the Internet penetration is less balanced among the age groups, we should expect VAA samples that would be less representative of the general population and more similar to the population of Internet users in the country.



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

Andreadis, I. (2014) Data Quality and Data Cleaning in Garzia, D. Marschall, S. (eds) Matching Voters with Parties and Candidates. Voting Advice Applications in Comparative Perspective, ECPR Press, ISBN: 9781907301735

Andreadis, I. and Chadjipadelis, T. (2014). HelpMeVote-VoteMatch Greece: VAA for the elections for the European Parliament 2014: Votematch-2014-Greece. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2014-07-19. http://doi.org/10.3886/E11266V3

Andreadis, I., Wall, M. and Krouwel, A. (2014). Who are the users of voting advice applications? presented at 2014 MPSA conference, Chicago April 2-6, 2014

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
European Commission, Brussels (2014): Eurobarometer 80.1 (2013). TNS Opinion, Brussels [producer] . GESIS Data Archive, Cologne. ZA5876 Data file Version 1.0.0, doi:10.4232/1.11881
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.

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.

Wheatley, J., Carmann, C., Mendez, F. and Mitchell, J. (2102). The dimensionality of the Scottish political space: Results from an experiment on the 2011 Holyrood elections. *Party Politics*. published online 30 September 2012 DOI: 10.1177/1354068812458614