

**Mode (still) matters: Evidence of social desirability bias in concern for poverty and public support for overseas aid**

Jennifer vanHeerde-Hudson\*  
Associate Professor of Political Behaviour  
University College London  
[jennifer.hudson@ucl.ac.uk](mailto:jennifer.hudson@ucl.ac.uk)

David Hudson  
Professor of Politics and Development  
University of Birmingham  
[d.e.hudson@bham.ac.uk](mailto:d.e.hudson@bham.ac.uk)

Paolo Morini  
Post-Doctoral Researcher  
University College London  
[p.morini.11@ucl.ac.uk](mailto:p.morini.11@ucl.ac.uk)

Paper prepared for delivery at the Western Political Science Association, San Francisco, California, March 2018. \*Corresponding author.

FIRST DRAFT: please do not cite without permission. Comments and suggestions are most welcome. The authors thank Agnes Magyar and Madeleine Leftwich for their research assistance.

## **Abstract**

In this paper we present the results of a field experiment measuring concern for poverty in poor countries and support for overseas aid in Britain. Surveys have shown levels of concern to be high – averaging above 70% and stable since the late 1990s – however, recent evidence has shown a significant decline. We argue that the decline is not real, but rather the result of social desirability bias stemming from changes in survey mode. We ask two questions: first, do online and FTF samples produce and distributions on items measuring public support for overseas aid and concern for poverty; and second, whether differences are consistent across types of questions (i.e. those measuring attitudes vs. knowledge). To test the mode effects hypothesis, we commissioned two identical surveys using an online (non-probability) and FTF (quota sample). Our findings show that the FTF sample provides consistently higher estimates for not just the item measuring concern for poverty in poor countries, but for items measuring support for overseas aid spending as well, subject to a number of robustness tests. We find no difference in estimates for the knowledge item. Our findings suggest that some questions are better suited to online formats.

Key words: field experiment; mode effects; social desirability bias; concern for poverty; overseas aid

## **Introduction**

In this paper, we present the findings of a field experiment designed to test whether online (non-probability) and face-to-face (quota) samples produce significantly different point estimates on questions measuring support for overseas aid and concern for poverty in poor countries, and whether differences are consistent across types of questions (i.e. those measuring attitudes vs. knowledge). Scholarly interest in online surveys has coincided with their increased use in social and political science research in the past decade (Ansolabehere & Schaffner 2014), in particular, in US and UK election studies (Karp & Lüthiste 2016; Chang & Krosnick 2009; Vavrek & Rivers 2008; Sanders et al. 2006).

A key question of interest has been the degree to which online surveys produce valid and reliable estimates, compared to the industry standard probability sample. Existing evidence is mixed. Using the 2005 British Election Study (BES), Sanders et al. (2006) compared responses from online (opt-in) and traditional face-to-face (probability) samples and find, that while the two samples produce ‘small variation in the distributions of key variables’, they also generate ‘remarkably similar results when it comes to estimating parameters in voting behaviour models’ (2006, 279). In contrast, Yeager et al. (2011) and Malhotra and Krosnick (2007) have shown significant variation in the distributions and association between variables generated by online and probability samples. Differences in estimates have also been found in results comparing online probability and opt-in samples. Karp and Lüthiste (2016, 687) demonstrate that data gathered from opt-in online panels produce different theoretical interpretations, and consequently, ‘while online panels have many advantages, they may not be suited for addressing certain types of questions’.

Research has shown that among the many advantages of online surveys over traditional face-to-face (FTF) techniques is to reduce social desirability bias (Powell 2013; Chang & Krosnick 2009). Social desirability bias (SDB) results when respondents answer survey questions that makes them look more favourable to interviewers, rather than revealing their real preferences. SDB has been shown to be particularly acute for sensitive questions, e.g. same-sex marriage, alcohol/drug use, abortion rights (Lax, Phillips & Stollwerk 2016), but can be present whenever expectations around social norms influence survey responses. A classic example is turnout; survey respondents may be more likely to say they voted or to ‘lie’ to interviewers given the ‘civic duty’ norm underpinning political participation. However, biases can exist across a range of survey questions, particularly where items are salient and anchored in perceived social norms – for example – on climate change or global poverty. At the extreme, SDB can produce estimates that fail to represent the ‘true’ nature of public opinion, which can be consequential for policymakers and other stakeholders who design policies around them.

We ask two questions: first, do online and FTF samples produce and distributions on items measuring public support for overseas aid and concern for poverty; and second, whether differences are consistent across types of questions (i.e. those measuring attitudes vs. knowledge). Using a field experiment, we test for evidence of mode effects for a small number of question items on public attitudes towards overseas aid and concern for poverty in poor countries. Because concern for global poverty is underpinned by norms of a ‘moral obligation’ to help people in need, we expect to find evidence of SDB resulting from differences in survey mode. We commissioned two identical surveys, the first using YouGov’s online, opt-in panel and the second using a random probability sample with face-to-face interviewing conducted by TNS Opinion & Social.<sup>1</sup>

Our findings contribute to previous literature on mode effects in two ways. First, our data show significant mode effects: questions measuring respondents’ attitudinal preferences on overseas aid and global poverty have significantly different means and distributions, however, the ‘fact-based’ item measuring respondents’ understanding of current spending on overseas aid, showed no significant difference. Second, the estimates produced from each sample lead us to different substantive conclusions in understanding the British public’s concern for global poverty.

#### *Wither concern?*

Our motivation for this paper results from an observed decline in concern for poverty in poor countries among the British public. Public support has long been regarded by policymakers, NGOs and international organizations as essential in legitimizing government expenditure, efforts to reform aid programs, and as noted in the 2009 International Development Committee’s inquiry into ‘Aid under Pressure’, is ‘essential to effective development policy’. This view has not come to pass. In February of this year, the Secretary for International Development, Penny Mordaunt, noted the support of the British public as a ‘necessary condition’ of DFID’s work.

The Department for International Development (DFID) is the agency responsible for delivering the UK Government’s policies on overseas and development. From 1999-2011, DFID monitored public support for overseas aid and development through the use of national opinion surveys. Figure 1 shows the distribution of one of DFID’s standard support items, ‘Which best describes how you feel about levels of poverty in poor countries’?<sup>2</sup> Evidence from

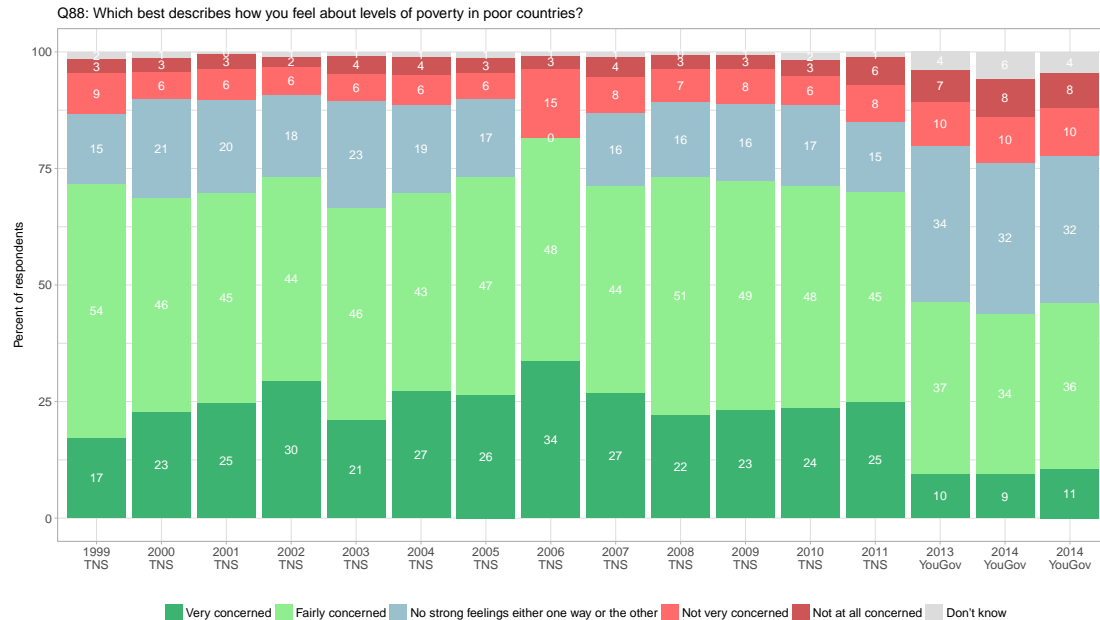
---

<sup>1</sup> TNS Opinion & Social is part of WPP plc holdings and from September 2016 is Kantar TNS.

<sup>2</sup> The concern measure has been subjected to some criticism (Hudson & vanHeerde-Hudson 2012), primarily in terms of its face validity as a measure of support for ODA, for which it has been both implicitly and explicitly used. However, the concern measure is useful in that it is the longest running time-series measure of public support and it has shown itself to be a good overall proxy measure for support for development related programmes. The item uses a 5-point scale, ranging from ‘very concerned’ to ‘not at all concerned’.

these surveys show that from 1999 – 2011, concern for poverty averages above 70% and stable over the time-series. However, in 2013 we see a significant decline in the percentage of respondents saying they are (very) concerned from, falling from 70% to 48%. Nearly all the collapse stems from respondents moving from ‘very concerned’ or ‘concerned’ into the middle category, ‘No strong feelings’. What explains the shift in attitudes?

**Figure 1. Concern for poverty in poor countries time-series, 1999-2014**



One potential explanation for the decline in the concern for poverty was the economic downturn and global recession. In addition to the Coalition Government’s ‘austerity’ agenda at home, it is plausible that as the recession began to bite, ‘charity begins at home’ sentiment shifted the focus away from poverty overseas to growing poverty at home. While this view is certainly plausible – it suggests that the British public took a long time to feel the pain – more than seven years from the onset of the recession in 2007 with the collapse in Lehman Brothers.

We do not test the economic crisis explanation here, instead focusing on a more likely explanation, social desirability bias resulting from changes in survey mode. From 1999-2011, DFID’s annual survey was conducted using a probability sample, with face-to-face interviews, first by COI (1999-05) and then by TNS Opinion & Social (2006-11). DFID stopped all public monitoring and engagement work in 2011 as part of the Coalition’s cuts to public engagement work. In 2013, parts of the DFID survey were resurrected through the Aid Attitudes Tracker (AAT), a cross-national survey of public attitudes and behaviour with overseas aid and development using YouGov’s online, opt-in panel.<sup>3</sup> We hypothesize that the difference in concern is a result of social desirability bias resulting from mode effects: in the face-to-face

<sup>3</sup> The Aid Attitudes Tracker tracks public attitudes and behaviour towards aid and development in Great Britain, France, Germany and the U.S. and is funded by the Bill & Melinda Gates Foundation. One aim in funding the study was to continue the valuable DFID time series.

format, respondents may be more willing to ‘lie’ about their attitudes towards people living in poor countries.

### **Empirical strategy**

In this section we describe our experimental design, data collection and weighting procedures employed by the two survey houses. To test the mode effects hypothesis we commissioned two identical, short surveys measuring concern for poverty in poor countries. The target population for our study is British adults. To identify whether the effects of the fall in concern shown, in Figure 1 are due to social desirability bias stemming from the different modes, we used the same two survey houses previously involved, YouGov UK and TNS Opinion & Social.

The online, opt-in sample for our experiment comes from YouGov UK, a global market leader in online survey research. Respondents are recruited to the UK panel (c. 750,000) through a variety means and are incentivized for participating in surveys.<sup>4</sup> The sample was generated by placing our instrument on YouGov’s Daily Polling surveys. Respondents are selected from the main panel using quota sampling. Individual samples from the daily polls (target n= 1,500-2,000 British respondents aged 18+) were then merged (duplicate respondents were removed) and then re-weighted according to the targets for Daily Polling by age and gender (interlocked), social grade, region, party identification, newspaper readership demographics. The survey opened on 23 November and closed on 27 November 2014, generating a sample of n= 6,998.<sup>5</sup>

The FTF sample comes from TNS Opinion & Social, now Kantar TNS. In this case, our questions were added to TNS’s Weekly Omnibus, a subscription-based survey where clients buy space. Our questions featured among a range of social, political and commercial topics. The survey is administered by computer-assisted personal interview (CAPI) in respondent’s households and data are weighted according to a range of socio-demographic variables to produce a representative sample of British adults. The survey fielded on 21<sup>st</sup> – 25<sup>th</sup> November 2014, generating a sample of n= 1,948.

Our design differs from previous studies in two ways. First, whereas previous studies have used question items from existing, large-scale instruments, for example the British or American Election Studies (Karp & Lüthiste 2016; Sanders et al. 2006) or National Health Indicators Survey (Ansolabehere & Schaffner 2014), we use a small number of items embedded within larger omnibus/daily surveys. For example, in the online sample respondents were asked

---

<sup>4</sup> The vast majority of panelists are recruited directly by YouGov contact. Points are allocated according to the size and complexity of the instrument, which can be accrued and redeemed for payment.

<sup>5</sup> Responses were distributed over the five days the survey was open: 23 Nov. N= 913 (13.05%); 24 Nov. n= 1,199 (17.13%); 25 Nov. n= 2,250 (32.15%); 26 Nov. n= 1,527 (21.82%); 27 Nov. n= 1,109 (15.85%).

questions on a range of topics – Ebola, European issues and Brexit, FIFA corruption, airport expansion and our topic of interest – poverty in poor countries. Because YouGov has a number of surveys in the field at any one time, respondents are sent an invite to participate, but do not know the content until they open the link. One advantage of this approach is to reduce selection effects based on interest in or knowledge of our topic. Of course, people who sign up to do online surveys may be different from the general population (more interested in politics, better educated) however, participation in the survey may not be driven by substantive interest in global poverty.

Second, unlike election studies or long-standing questions on social practices (i.e. church attendance), there are no population estimates of public attitudes on concern for poverty and overseas aid to benchmark either the YouGov or TNS sample estimates against. For example, in Karp & Lühiste (2016) and Sanders et al. (2006) are able to compare self-reported turnout against validated measures to determine how well their sample estimates reflect actual turnout. Similarly, Chang and Krosnick (2009) and Ansolabehere and Schaffner (2014) are able to compare their samples' representativeness against benchmarks from the Current Population Survey (CPS). Consequently, we cannot say which of the two sample estimates are more accurate estimates of the population parameter, only the extent to which the two sample estimates differ and in which direction.

Table A1 in the Appendix shows the question wording and scaling for the survey items. In addition to the concern item, we measure five indicators of support for UK overseas aid/development spending. We include the aid support/effort measures for two principal reasons. First, similar to the concern measure we hypothesize that these measures also suffer from SDB as support for overseas aid and government efforts to address global poverty are also underpinned by the moral obligation social norm. Attitudes towards government provision of overseas aid are also highly correlated with concern for global poverty and so provide additional evidence of whether these types of items are susceptible to social desirability biases. Second, the AAT routinely measures knowledge of government spend on overseas aid, which has been shown to be wildly overestimated by donor publics (vanHeerde-Hudson 2014; Pew Research Centre 2014). We include a single knowledge item to see if SDB extends to what respondents 'know' about overseas aid spending; our contention is that because the item is not measuring an attitudes or preference for government aid effort, the samples should not produce significantly different estimates.

Before moving to our analysis, we compare the demographic composition of the two samples across five key variables: gender, education, social grade (class), age and party identification. Looking at the unweighted data, there are significant differences between the samples across each category, although the direction is not consistent. In comparing mean estimates for the weighted data, there is no longer a significant effect for gender, however, all

other differences remain. With regard to education, the TNS sample over estimates respondents in the lower educational bands (no award, GCSE) and underestimates those in the higher bands (university degrees) compared to the YouGov sample. And there are significantly fewer partisans in the TNS compared to YouGov sample. However, absent population statistics, we cannot determine which of these are more or less representative of the population. The TNS weighted sample is representative of the larger population with respect to social grade, whereas YouGov overestimates those in the higher social classes (ABC1). Finally, there are significant differences across the different age ranges between the two samples: compared to the population, YouGov underestimates the percentage of young people (18-29) and overestimates the percentage of 50-64s. Thus, between the two weighted samples there are significant differences across education, age, social grade and party. Where we can compare the sample profiles to population estimates, the TNS/quota sample is more representative on balance.

**Table 1. Demographic composition of the samples: Weighted and unweighted data**

<i>Variables</i>	<i>Unweighted data</i>			<i>Weighted data</i>			<i>Pop.<sup>6</sup></i>
	<i>YouGov</i>	<i>TNS</i>	<i>Diff.</i>	<i>YouGov</i>	<i>TNS</i>	<i>Diff.</i>	
<b>Gender</b>							
Male %	50.04	46.61	-3.43***	48.35	48.25	-0.1	49.32
Female %	49.96	53.39	3.43***	51.65	51.75	0.1	50.68
<b>Education</b>							
No award %	5.06	20.74	15.68***	6.04	15.68	9.64***	--
GCSE %	13.92	19.51	5.59***	16.28	18.94	2.66***	--
A-level %	18.53	20.33	1.80 <sup>+</sup>	20.43	20.33	-0.1	--
Undergrad %	25.45	14.78	-10.67**	23.09	17.3	-5.79***	--
Postgrad %	12.32	6.47	-5.85***	9.74	8.49	-1.25 <sup>+</sup>	--
<b>Social grade</b>							
ABC1 %	69.61	41.58	-28.03***	58.54	54.86	-3.68***	55%
C2DE %	30.39	58.42	28.03***	41.46	45.14	3.68***	45%
<b>Age</b>							
18-29 %	15.1	24.33	9.23***	18.68	22.06	3.38***	23.24%
30-39 %	15.2	15.25	0.05	18.71	16.92	-1.79 <sup>+</sup>	15.91%
40-49 %	18.58	13.91	-4.67***	16.2	18.12	1.92**	16.29%
50-64 %	32.05	20.23	-11.82***	28.79	20.71	-8.08***	22.67%
65+ %	19.06	26.28	7.22***	17.62	22.19	4.57***	21.89%
Mean	48.68	47.97	-0.71	46.83	46.97	-0.14	
<b>Party id</b>							

<sup>6</sup> Population statistics are from the Office of National Statistics: <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalescotlandandnorthernireland>. Statistics for education levels are protected data and we are currently awaiting approval for access.)



Con %	25.43	13.98	-11.45***	26.25	15.68	-10.57***	--
Labour %	28.62	26.97	-1.65	27.84	25.82	-2.02 <sup>+</sup>	--
Lib-Dem %	6.8	3.16	-3.64***	6.9	3.26	-3.64***	--
No party %	23.8	41.07	17.27***	26	40.4	14.4***	--

(<sup>+</sup> p < .10; \*\* p < .05; \*\*\* p < .01)

## Results

Our analysis proceeds as follows. First, we compare the weighted and unweighted differences in sample estimates. Second, we look to see whether observable differences in the estimates are robust across different socio-demographic groups. Third, we the confounding effect of the demographic variables in three ways, all of which consistently show that the FTF mean estimates for attitudinal measures are higher than the ones obtained using the online sample.

Table 2 compares the mean estimates produced by the two samples. For each of the six items, the samples produced significantly different mean estimates, for both the weighted and unweighted data (with the exception of knowledge in the weighted data). Moreover, the FTF sample produced higher average estimates than the online sample, suggesting respondents are more supportive of aid efforts and more concerned for global poverty, on average. For four of items – *concern*, *generosity*, *aid spend* and *government effort* – the weighting strategies applied by the two houses *increases* the mean differences. The magnitude of the differences varies from just over 1% (knowledge) to 12% (aid spend).

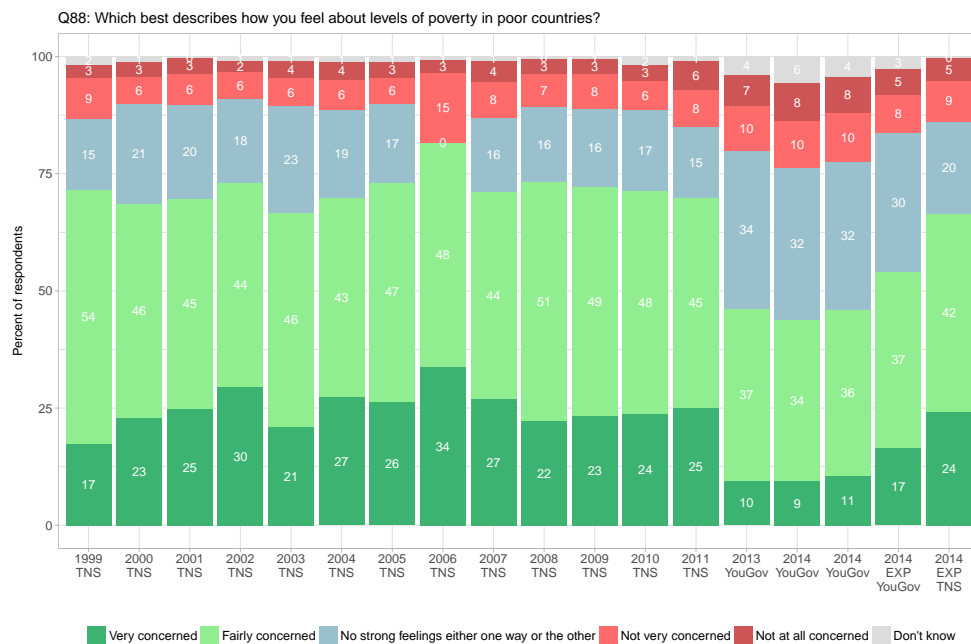
**Table 2. Comparison of mean estimates from the online and FTF samples**

<i>Variables</i>	<i>Unweighted data</i>			<i>Weighted data</i>		
	<i>YouGov</i>	<i>TNS</i>	<i>Difference</i>	<i>YouGov</i>	<i>TNS</i>	<i>Difference</i>
<b>Concern</b>	3.59 (1.04)	3.69 (1.10)	0.10***	3.53 (1.04)	3.72 (1.07)	0.19***
<b>Knowledge</b>	15.70 (19.76)	20.38 (20.70)	4.68***	18.37 (21.63)	19.48 (19.99)	1.11 <sup>+</sup>
<b>Preferred</b>	9.35 (13.26)	16.45 (18.02)	7.10***	10.39 (14.65)	15.79 (17.21)	5.40***
<b>Generosity</b>	5.05 (2.95)	5.94 (2.69)	0.90***	4.85 (2.87)	6.00 (2.67)	1.15***
<b>Aid spend</b>	2.51 (1.19)	2.97 (1.12)	0.46***	2.43 (1.16)	3.02 (1.11)	0.59***
<b>Govt effort</b>	2.70 (1.28)	3.02 (1.18)	0.32***	2.61 (1.25)	3.08 (1.17)	0.47***

(Note: Means with standard deviation in parenthesis; p<sup>+</sup> = .10; \*\* p < .05; \*\*\* p < .01)

Consistent with our expectations, mean estimate for *concern* is produced by the FTF sample (3.72) is significantly higher than that produced by the online sample (3.53) ( $p < .01$ ), with similar standard errors. Although the magnitude of the difference is not large at 4% (.19 on a 1-5 scale), it is not insignificant. We illustrate the result of our experimental test in Figure 2, showing the distribution of responses on levels of concern. The patterns shown in the test generally conform to the data produced by the two methods previously: in the FTF format, respondents are less likely to say, ‘no strong feelings’, and instead opt for ‘very’ or ‘fairly concerned’.

**Figure 2. Concern for poverty in poor countries, with experimental test**



Turning next to the aid spending and support items, we again observe statistically significant differences between the mean estimates produced by the FTF and online samples. For *aid generosity*, where 0= ‘The UK Government should not give aid at all’ and 10= ‘The UK Government should give aid generously, the mean estimate for the FTF sample (6.0) is 1.15 points higher than that for the online sample (4.85) ( $p < .01$ ), again with fairly similar standard errors. For *aid spend*, where 1= ‘Government should decrease aid spending a great deal’ and 5= ‘Government should increase aid spending a great deal’, the mean estimate for the FTF sample is 3.02 compared to 2.43 for the online sample ( $p < .01$ ). Finally, for *government effort*, which asks about respondents’ preferences for the UK Government’s role in reducing poverty overseas (1= Government is doing far too much; 5= Government should do a lot more), the mean estimate from the FTF sample is 3.08 compared to 2.61 for the online sample ( $p < .01$ ). On balance then, respondents in the FTF sample appear to be warmer towards spending on overseas aid and the Government’s efforts to reduce global poverty. We infer these differences to be the result of SDB produced by the FTF format.

The final two items included in our experiment are related to respondents' knowledge of the percent of the UK national budget spent annually on overseas aid (*knowledge*) and their preferences for how much should be spent (*preferred*). These items have been used to determine the distance between public preferences for, and actual, aid spending as well as to understand whether preferences are sensitive to judgments about how much is spent (Scotto et al. 2017). As set out previously, if social desirability bias is present in the estimates produced in the FTF sample, then an additional test is provided by these two items. The former – we argue – is a simple measure respondents knowledge of aid levels, while the latter is a question measuring preferences, although we recognise that some may interpret this question normatively.<sup>7</sup>

If, as we have suggested, respondents in the FTF format are more likely to lie so as to not contravene the morality and guilt norms perceived to operate for attitudes towards global poverty, then we should not observe meaningful differences on *knowledge* but should on *preferred*. The results are consistent with our expectations. For the knowledge item, the average estimate is 19.48 in the FTF sample compared to 18.37 in the online sample, but the difference is not statistically significant ( $p = .10$ ). For the preferred spend item, on average, respondents in the FTF sample prefer government to spend 15.79% of the national budget on overseas aid, compared to 10.39% for the online sample. The difference is significant at the  $p < .01$  level.

We next consider how robust these differences are when looking at the socio-demographic characteristics in our two samples. One of the main confounding effects when comparing the online and FTF samples is the different distributions for the demographic characteristics of respondents. This is to be expected, as existing literature shows that online samples can differ from the overall population when looking at education, income levels, age and social class (Callegaro et al. 2014). As we have shown in Table 1, our two samples show significant differences for education, social grade, age and party id.

In Table 3 we show the results of a regression analysis for each of the six items looking at the impact of social profile. The crux of this analysis helps us to answer the question of whether respondents who participate in online or FTF samples are inherently different. In other words, we do not expect – a priori – that women in the online panel are different to women in the FTF sample. Put another way, we do not think different types of women participate in different survey activities. If gender is constant, then we would also expect roughly similar attitudes by women and men in the respective samples. As we can see there are differences across the difference social profile variables, but these are inconsistent. First, we do observe a significant base house effect: mean estimates are likely to be higher for each of the items we measure, however, they are only significantly higher for *generosity*, *aid spend* and *government*

---

<sup>7</sup> The precise question wording is 'What per cent of the national budget do you think the UK Government *should* spend each year on overseas aid, if any?'

effort. Second, the remaining items do not tell a consistent story. We look to investigate these results in the next section.

**Table 3. The impact of social profile on measures of concern & support for overseas aid**

Interactions	Concern	Knowledge	Preferred	Generosity	Aid spend	Govt effort
<b>Base house effect</b>	ns (+)	ns (+)	ns (+)	+	+	+
<b>Education</b>	+	+	ns	-	ns	ns
<b>Women</b>	ns	-	ns	ns	ns	ns
<b>Soc grade C2DE</b>	-	ns	+	ns	+	ns
<b>Age</b>	ns	ns	-	ns	ns	ns
<b>Conservative</b>	+	ns	+	+	+	ns
<b>Labour</b>	ns	ns	ns	-	ns	-

*Are mode effects really mode effects?*

We address the confounding effect of the demographic variables in three ways, seeking to show that FTF sample estimates for our variables are significantly higher than for our online sample estimates. First, we adjust the simple mean comparisons we presented above in two ways, by employing demographic variables as controls and to build survey weights. Second, we use the variables in a propensity score matching model. Third, we “block” these variables by creating a set of 24 subsamples<sup>8</sup> which we use to compare mean estimates for our six items; we then use frequency distributions and confidence intervals for the estimated average mode effect. We describe the results of each of these three approaches in the following sections. Overall, we find that results from all three tests consistently show that the FTF mean estimates for attitudinal measures – but not the knowledge item – are higher than those obtained from the online sample.

First, we estimate a set of six linear regression models using survey weights and heteroscedastic-robust standard errors. Each of the models estimates the difference between the online and FTF samples while controlling for age, gender, education levels, social class, and party identification. Table 4 reports the differences between the FTF and online samples (with the latter used as a baseline). With exception for the *knowledge* item, the mean estimates obtained through FTF surveys are significantly higher than their online counterparts, and all results are significant at the 1% level. The “% effect” shows that the magnitude of effects varies quite substantially across questions. While *concern* for global poverty and *preferred* levels of aid expenditure show differences of approximately 4% – we observe much bigger effects for generosity, aid spend and government effort – all of which are all at or above 10% in size. This evidence suggests that although mode differences are present across all five of the attitudinal

<sup>8</sup> We obtain 24 samples by calculating each possible combination of three factors: education (whether a respondent has no education, whether they hold a university undergraduate degree, or neither of the two), party identification (whether a respondent identifies as a Conservative, a Liberal-Democrat, or as not being a supporter of any political party).

items, there magnitude of differences are much larger for the aid spending and government effort. We consider reasons why this may be in the case in the final section.

**Table 4. Difference between FTF and online samples**

<b>Profile</b>	<b>Effect</b>	<b>95% low</b>	<b>95% high</b>	<b>% effect</b>	<b>Sign</b>
<b>Concern</b>	0.22	0.15	0.28	4.4%	1%
<b>Knowledge</b>	-1.08	-2.34	0.18	1.1%	ns
<b>Preferred</b>	4.49	3.49	5.49	4.5%	1%
<b>Generosity</b>	1.25	1.10	1.42	11.4%	1%
<b>Aid spend</b>	0.65	0.59	0.71	13%	1%
<b>Govt effort</b>	0.51	0.45	0.58	10%	1%

(Note: the online sample is the baseline)

Next, we estimate six propensity-score matched treatment effects models. We model the propensity score using a logistic regression model using age, gender, social class, education, and party identification. We calculate the effects using one-to-one matches across the two samples using the propensity score. As Table 5 shows, results are consistent with what we obtained with linear regression models. Once again, the FTF sample produces significantly higher estimates for all five attitudinal measures. The effect size is smaller than estimated in the linear models but shows a similar pattern. The effects size estimated for *concern* in the linear models is 4.4%, compared to 4% in the propensity score matched models. The biggest differences occur for the aid level/effort variables with generosity (-1.6%), aid spend (-1.8%) and government effort (-1.4%). Again, the differences in the *knowledge* item are non-significant.

**Table 5. Propensity score matched models**

<b>Profile</b>	<b>Effect</b>	<b>95% low</b>	<b>95% high</b>	<b>% effect</b>	<b>Sign</b>
<b>Concern</b>	0.20	0.13	0.27	4%	1%
<b>Knowledge</b>	0.74	-0.47	1.96	0.7%	ns
<b>Preferred</b>	4.48	3.47	5.48	4.5%	1%
<b>Generosity</b>	1.08	0.89	1.26	9.8%	1%
<b>Aid spend</b>	0.56	0.48	0.64	11.2%	1%
<b>Govt effort</b>	0.43	0.35	0.51	8.6%	1%

Finally, we estimate ‘profiles models’ where we create 24 samples using combinations of our socio-demographic variables. Table 6 shows the mean mode effects obtained through our profile regressions. As both medians and means with confidence intervals show, these results are consistent with those we obtained with descriptive statistics and results from the multivariate regression and matching approaches. In all cases FTF survey estimates are significantly higher than those obtained through the online sample. We fail to find any significant differences controlling for the *knowledge* question in any of our analyses.

**Table 6. Profiles models**

<b>Profile</b>	<b>Median effect</b>	<b>Mean effect</b>	<b>95% low</b>	<b>95% high</b>
<b>Concern</b>	0.21	0.19	0.06	0.33
<b>Knowledge</b>	0.20	-0.51	-3.01	1.98
<b>Preferred</b>	4.33	4.12	2.08	6.15
<b>Generosity</b>	1.25	1.27	0.90	1.64
<b>Aid spend</b>	0.64	0.64	0.51	0.78
<b>Govt effort</b>	0.50	0.41	0.23	0.60

## Discussion

Our motivation in this paper was to assess the extent to which an observed decline in concern for poverty in poor countries was the result of a change of survey mode. We argue that the underlying mechanism for the change is social desirability bias – the tendency for respondents to give interviewers answers to survey questions that conform to perceived social norms. Using a field experiment, we show that with respect to the concern item, respondents in online samples are more likely to move from saying they are ‘very’ or ‘fairly concerned’ to ‘no strong feelings’. Under the veil of anonymity, respondents are simply less likely to ‘lie’ about levels of concern for poor people in poor countries.

We also find that mode effects are present in other questions measuring support for the British Government’s efforts on overseas aid and spending levels. In fact, the size of the effects here are much larger than found for concern. While we cannot test different explanations for the effects sizes here, one explanation for the difference in magnitude may lie in the nature of the questions which ask respondents to make decisions about public spending. For example, the aid spending question presents information to respondents about the volume and percent of the UK budget that is given towards overseas aid annually. In terms of volume, the UK Government gave £11.5 billion in aid in 2014, or just over 1% of the total budget. Elsewhere, research has shown that regardless the way in which aid spending is presented, the British public want to cut aid, even if they support higher aid spending in the first instance (Scotto et al. 2017; vanHeerde-Hudson 2014). We suspect that in the presence of an interviewer respondents moderate their preferences – again social desirability bias – but in online samples, respondents are free to express preferences to cut aid spending with impunity.

We present several tests to isolate and show the robustness of our mode effects thesis. First, we include an item measuring respondents’ knowledge of overseas aid spending and find no statistically significant difference between estimates from the two samples. Unlike the attitudinal measures, we do not find mode effects when the question measures respondents’ knowledge. Finally, we included a number of robustness test to attempt to isolate the effect of survey mode: across all of these controls, we find consistently significant differences between

the two samples. Of course, we cannot say conclusively from this analysis that the effects observed here are due strictly to mode, and not a consequence of the sampling frame.

We conclude with three points. First, while we have shown that the FTF sample produces higher average point estimates than the online sample, we have no way of knowing which of the two estimates better reflect the true population parameter. There are no comparable baselines to measure against similar to electoral turnout. Second, it is not yet clear whether these point estimates produce different substantive conclusions about concern for poverty or support for overseas aid. A preliminary analysis estimating concern for poverty in the two modes shows many consistencies – age, gender, Labour party identifier, greater support for government efforts and aid generosity – are positive predictors across both models. But there are also differences in sign and significance for education and social grade. In future work we turn our attention to understanding whether mode difference matter for understanding British attitudes towards global poverty.

Third and finally, our findings around concern may have important policy implications for those who work in the aid and development sectors. The DFID time-series provided an important public good; both the Government and NGOs used the DFID annual surveys to inform their communications and campaigns. If FTF samples produce artificially high(er) estimates, policies based on these data may fail to land with the public. To rephrase Karp and Lühiste (2016), online panels have many advantages, and addressing questions relating to global poverty and overseas aid may be one of them.

## References

- Berrens, R., Bohara, A., Jenkins-Smith, H., Silva, C., & Weimer, D. 2003. 'The advent of internet surveys for political research: A comparison of telephone and internet samples'. *Political Analysis* 11(1): 1-22.
- Callegaro, M., Baker, R., Bethlehem, J., Göritz, A., Krosnick, J. & Lavrakas, P., (eds). 2014. *Online Panel Research: A Data Quality Perspective*. Hoboken: John Wiley and Sons.
- Chang, L. & Krosnick, J. 2010. 'Comparing oral interviewing with self-administered computerized questionnaires'. *Public Opinion Quarterly* 74(1): 154-67.
- Chang, L. & Krosnick, J. 2009. 'National surveys via telephone interviewing versus the internet'. *Public Opinion Quarterly* 73(4): 641-678.
- Clarke, H., vanHeerde-Hudson, J., Hudson, D., Stewart, M. & Twyman, J. Aid Attitudes Tracker Survey, 2013-2018. Bill & Melinda Gates Foundation.
- Grönlund, K. & Strandberg, K. 2014. 'Online panels and validity: Representativeness and attrition in the Finnish eOpinion panel'. In *Online Panel Research: A Data Quality Perspective*, edited by M. Callegaro, R. Baker, J. Bethlehem, A. Göritz, J. Krosnick & P. Lavrakas. Hoboken: John Wiley and Sons.
- Hudson, D. & vanHeerde-Hudson, J. 2012. 'A Mile Wide and an Inch Deep': A review and critique of extant surveys on public opinion and development'. *International Journal of Development Education and Global Learning* 4(1): 5-23.
- Karp, J. & Lühiste, M. 2016. 'Explaining political engagement with online panels: Comparing the British and American Election Studies'. *Public Opinion Quarterly* 80(3): 666-693.
- Lax, J., Phillips, J. & Stollwerk, A. 2016 'Are survey respondents lying about their support for same sex-marriage? Lessons from a list experiment'. *Public Opinion Quarterly* 80(2): 510-533.
- Malhotra, N. & Krosnick, J. 2007. 'The effect of survey mode and sampling on inferences about political attitudes and behavior: Comparing the 2000 and 2004 ANES to internet surveys with nonprobability samples'. *Political Analysis* 15: 286–323.
- Pew Research Center. 2014 'From ISIS to Unemployment: What Do Americans Know'? <http://assets.pewresearch.org/wp-content/uploads/sites/5/2014/10/Final-Knowledge-Release.pdf> [Accessed 13 March 2018]
- Powell, R. 2013. 'Social desirability bias in polling on same-sex marriage ballot measures'. *American Politics Research* 41: 1052–1070.
- Sanders, D., Clarke, H., Stewart, M. & Whiteley, P. 2006. 'Does mode matter for modelling political choice? Evidence from the 2005 British Election Study'. *Political Analysis* 15: 257-285.
- Scotto, T., Reifler, J., Hudson, D. & vanHeerde-Hudson, J. 2017. 'We spend how much? Framing foreign aid in the United States and Great Britain'. *Journal of Experimental Political Science* 4: 119–128.
- Sparrow, N. & Curtice, J. 2004. Measuring the attitudes of the general public via Internet polls: An evaluation. *International Journal of Market Research* 46: 23–44.



Twyman, J. 2008. 'Getting it right: YouGov and online survey research in Britain'. *Journal of Elections, Public Opinion & Parties* 18:343–54.

vanHeerde-Hudson, J. 2014. 'Corruption, waste and information: Public opinion and overseas aid'. In P. Cowley and R. Ford (eds.), *Sex, Lies and the Ballot Box*. Biteback Publishers.

Vavrek, L. & Rivers, D. 2008. 'The 2006 Cooperative Congressional Election Study'. *Journal of Elections, Public Opinion & Parties* 18: 355-66.

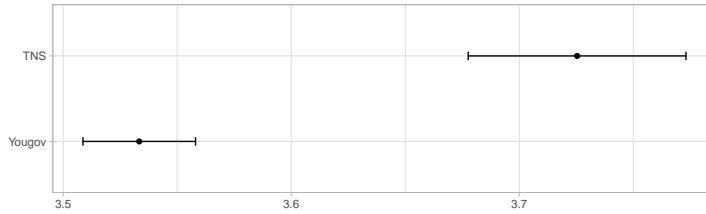
Yeager, D. Krosnick, J., Chang, L., Javitz, H., Levendusky, M., Simpser, A. & Wang, R. 2011. 'Comparing the accuracy of RDD telephone surveys and internet surveys conducted with probability and non-probability samples'. *Public Opinion Quarterly* 75: 709–47.

## Appendix

Figures A1 – A6: Point estimates and confidence intervals for comparison items

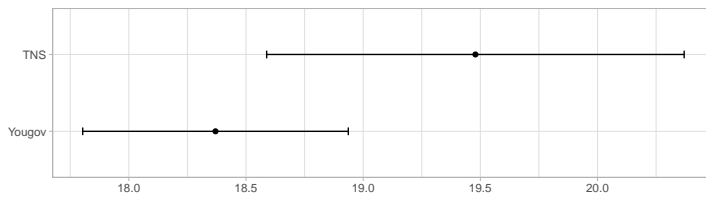
### A1. Concern for poverty in poor countries

Q1: Which best describes how you feel about levels of poverty in poor countries?



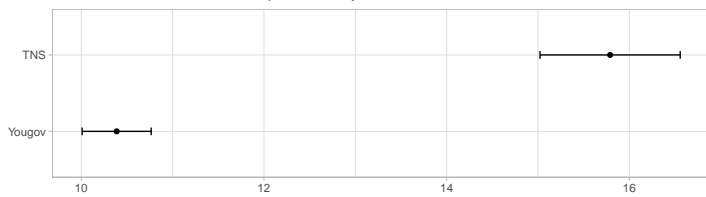
### A2. Knowledge

Q2: What per cent of the national budget do you think the UK Government currently spends each year on overseas aid?



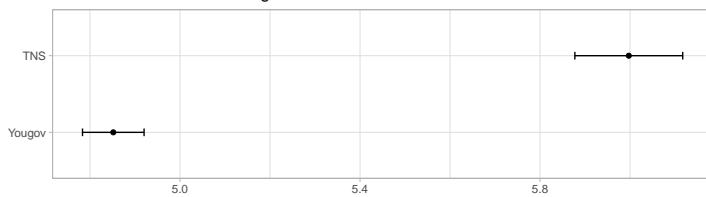
### A3. Preferred aid spend

Q3: What per cent of the national budget do you think the UK Government should spend each year on overseas aid?



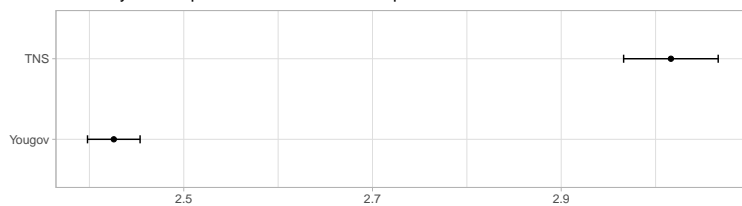
### A4. Aid generosity

Q4: Please indicate the extent to which you think that the UK Government should give overseas aid



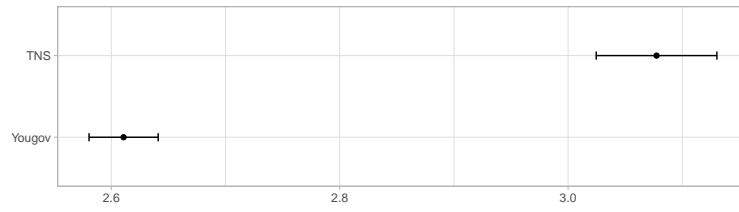
### A5. Aid spend

Q5: Do you think that the Government should increase or decrease the amount of money that it spends on overseas aid to poor countries?



## A6. Government effort

Q6: Thinking of the UK Government's role in reducing poverty in poor countries, which statement best describes how much the UK Government should do?



**Table A1. Question wording, scaling and descriptives**

Question wording	Scale	Mean YG	SD YG	Mean TNS	SD TNS
Which best describes how you feel about levels of poverty in poor countries?	1 = Not at all concerned 5 = Very concerned	3.53	1.04	3.72	1.07
What per cent of the national budget do you think the UK Government <i>currently</i> spends each year on overseas aid?	0 - 100%	18.37	21.63	19.48	19.99
What per cent of the national budget do you think the UK Government <i>should</i> spend each year on overseas aid, if any?	0 - 100%	10.39	14.65	15.79	17.21
Thinking about overseas aid to poor countries, please indicate the extent to which you think that the UK Government should give overseas aid, where a score of 0 means that it 'should not give aid at all' and a score of 10 means that it 'should give aid very generously'. Where would you place yourself on this scale?	0 = Should not give aid at all 10 = Should give aid very generously	4.85	2.87	6.00	2.67
Of its total budget of nearly £714 billion, the UK Government currently allocates 1.6 percent—£11.5 billion to overseas aid to poor countries. Do you think that the Government should increase or decrease the amount of money that it spends on overseas aid to poor countries?	1 = Decrease a great deal 5 = Increase a great deal	2.43	1.16	3.02	1.11
Thinking of the UK Government's role in reducing poverty in poor countries, which statement best describes how much the UK Government should do?	1 = The UK Government is doing far too much 5 = The UK Government should do a lot more	2.61	1.25	3.08	1.17