An assessment of the accuracy of survey estimates of the prevalence of problem gambling in the United Kingdom

Author: Patrick Sturgis, Department of Methodology, London School of Economics

Date: March 2020

Funding Statement

GambleAware is a wholly independent charity and has a framework agreement with the Gambling Commission to deliver the National Strategy to Reduce Gambling Harms within the context of arrangements based on voluntary donations from the gambling industry. GambleAware commissions research and evaluation to build knowledge of what works in prevention and reduction of gambling harms that is independent of industry, government and the regulator.

The author alone is responsible for the views expressed in this article, which do not necessarily represent the views, decisions or policies of the institutions with which they are affiliated.

Executive Summary

The 2016 combined health surveys find 4.2% of the adult population experience gambling harm[1] compared with 13.2% in the 2019 YouGov survey. This is a substantial discrepancy that cannot be explained by sampling variability.

This report provides a description of the range of errors that may be present in the different survey estimates and assesses which of the two surveys is likely to be most affected by these errors and in which ways.

The YouGov survey is a non-probability sample in which respondents complete the questionnaire online. The combined health surveys use probability sampling and face-to-face interviewing, although the measure of gambling harm is completed by the respondents via a paper questionnaire.

Statistical theory should lead us to prefer surveys that use probability sampling over designs which use non-probability sampling. A large and growing body of evidence supports this expectation empirically – probability samples tend to produce more accurate estimates of population characteristics than non-probability samples across a range of topic domains.

Although probability sampling should be preferred in principle, there are a range of systematic and random errors that affect both probability and non-probability designs as they are actually implemented and these are likely to be the cause of the differences between the estimates from the two surveys. The most relevant sources of error here are coverage error, nonresponse error, and measurement error.

Both surveys are likely to be affected by under-coverage, that is the exclusion of some population sub-groups from the possibility of being included in the survey. For the combined health surveys, this includes people who do not live in private residential addresses which is likely to produce a small *under*-estimate of gambling prevalence. The

[1] Gambling harm refers to a Problem Gambling Severity Index (PGSI) score of 1 or above and 'problem gambling' to a PGSI score of 8 or above.

YouGov survey has under-coverage of the offline population, a group which is probably less likely to have gambling problems leading to an *over*-estimate of gambling prevalence. Under-coverage likely explains some but by no means all of the difference between the two surveys.

Nonresponse error, or selection bias, is when people with certain characteristics are more likely to agree to (or volunteer to) take part in a survey. It is plausible that the combined health surveys *under*-estimate gambling harms due to those experiencing gambling problems being less likely to be contacted or to agree to take part in the survey once contacted. Correspondingly, the YouGov survey may *over*-represent gambling prevalence in its panel of respondents with the quota and weighting procedures employed providing little or no mitigation of this over-representation.

Measurement error is when there is a difference between the true value on the characteristic of interest and the measured value in the survey, which arises due to the procedures used to make the measurement. Three forms of measurement error are relevant here: socially desirable responding; context effects; and survey satisficing.

While we might expect the YouGov survey to have lower levels of socially desirable responding due to the absence of an interviewer, any difference from the combined health surveys is likely to be largely or wholly removed by the use of self-completion procedures in the combined health surveys.

Part of the discrepancy in the estimates might have arisen due to differences in the subject content of the two surveys. However, while this is a theoretical possibility, the fact that the YouGov survey placed the PGSI questions very early in the survey means that context effects are unlikely to be a major driver of the difference between the estimates.

Survey satisficing, or careless responding is likely to be more prevalent amongst respondents in the YouGov survey and may account for some of the difference between estimates. However, the mechanisms that might lead this kind of responding to result in higher prevalence estimates in the YouGov survey are not obvious.

Because these combined errors mostly influence the estimates from the two surveys in *opposite* directions, it seems credible that the true level of gambling harm prevalence lies somewhere in between their two bounds. However, my assessment is that the true value probably lies closer to the combined health surveys than to the YouGov survey.

Even if the YouGov survey *does* over-estimate the true level of gambling harm prevalence the UK, which seems likely, this does not mean that it is not of potentially high value in estimating the distribution of gambling harm prevalence across demographic groups in the population, its degree of stability and change over time, or of understanding the attitudes and experiences of individuals with gambling problems.

Introduction

Gambling policy in the UK, as in many other countries, is a contentious issue. Following the loosening of controls on the licencing of gambling premises and on maximum levels of stakes and rewards in the Gambling Act 2005, there has been a particular concern regarding the effects of this less restrictive policy environment on the prevalence of problem gambling.[2] Problem gambling covers a wide range of social, psychological, and behavioural factors which make it difficult to define and measure at the individual and population levels. Due to the primarily psychological and behavioural nature of the concept, it is not realistic to measure its prevalence through administrative records or commercial databases, such as the volume of betting or contacts made to gambling helplines. The only realistic means of assessing the extent of gambling harm in the general population, and of detecting change over time, is through representative sample surveys.

It is well known, however, that survey estimates are subject to a wide range of random and systematic errors relating to population coverage, sample representativeness, and measurement quality (Groves and Lyberg, 2010). This means that when surveys produce divergent estimates of the same quantity, it is usually difficult to determine which is the most accurate. In an ideal situation, there is a 'gold standard' external criterion against which different survey estimates can be compared. For example, the validity of different methods of opinion polling can be calibrated periodically against actual election results, because vote shares are the primary quantity that polls seek to estimate and these population quantities are known once the election has taken place. However, in most cases no such gold standard exists. Indeed, if a gold standard does exist then one might question the wisdom of expending scarce resources obtaining error-prone survey estimates of the same quantity.

This is the context of this report; surveys conducted at approximately the same time and using the same measure of problem gambling have produced widely divergent estimates of its prevalence in the general adult population. Specifically, the combined estimates from the

[2] https://www.gov.uk/government/publications/gambling-related-harms-evidence-review/gambling-related-harms-evidence-review-scope

2016 English, Scottish, and Welsh Health Surveys find just 0.7% problem gamblers compared to 2.7% in a 2019 YouGov survey commissioned by GambleAware. Additionally, the combined health surveys find 4.2% of the adult population to experience gambling harms with 13.2% in the YouGov survey. A third survey, undertaken using telephone interviewing for the Gambling Commission in 2019, produced prevalence estimates of 0.5% (problem gambling) and 5.3% (any level).

The purpose of this report is to provide a description of the range of errors that may be present in the different survey estimates and to assess which of the surveys is likely to be most affected by these errors and in which ways. I begin with a brief discussion of the theoretical underpinnings of survey inference using probability and non-probability[3] sampling methods and provide an overview of the findings of studies that have empirically compared the accuracy of estimates from each type of sample design. This is important because the YouGov survey uses non-probability sampling while the Health surveys use probability sampling. I then present the prevalence estimates from the surveys, alongside equivalent estimates from other recent surveys of gambling harm prevalence carried out in the UK and asses the likely errors affecting the different designs. I conclude with a judgement on which of the two surveys of primary consideration is likely to be the most accurate.

Accuracy of probability and non-probability samples

It has long been accepted that random probability sampling is the best method for achieving accurate population inference, in principle at least. This is because random sampling is based clearly and explicitly on mathematical probability theory. It is, therefore, straightforward to demonstrate, both theoretically and empirically, that statistics (e.g. means, proportions, ratios) calculated on sample data that were drawn using random procedures of selection will, in expectation (i.e. over many repeated samples), and assuming no nonresponse or measurement error, converge on the true population value.

[3] There is a wide variety of sampling approaches that fall under the heading of non-probability sampling. In this report I generally mean the method of quota sampling when I refer to non-probability sampling.

This can be contrasted with non-probability sampling, for which the underlying theory is much weaker. It is often said that there is *no* theoretical basis for non-probability sampling but this is not correct. Inference for non-probability sampling is based on statistical modelling; in effect, the analyst uses a statistical model to predict the outcome of interest and, if the model is correctly specified, inference to the population will be accurate (Sturgis et al 2017). What *can* be said of the theoretical basis of non-probability sampling is that it is a) not general in that different models will usually be required for each different outcome and b) it requires strong and largely untestable assumptions about model specification. Purely in theoretical terms, then, there is little question that random probability sampling should be preferred, if the goal is to make accurate inferences to a broader population. Indeed, this position was codified in a 2010 American Association of Public Opinion Research (AAPOR) report on online non-probability panels which concluded that, "Researchers should avoid nonprobability online panels when one of the research objectives is to accurately estimate population values".[4]

Surveys are not, however, conducted in *theory* but in practice and there are a number of additional factors beyond sampling procedure that must be taken into account when assessing the likely accuracy of different survey designs. A particular practical problem for random probability designs is the low and declining response rates that tend to characterise contemporary surveys. Nonresponse will lead to biased estimates of population parameters to the extent that non-respondents differ from respondents on the quantity the survey is seeking to measure. When response rates are low, there is a high (though generally unquantifiable) risk that survey estimates will be biased due to differential nonresponse. Weighting adjustments can and usually are implemented to mitigate this risk, although their effectiveness is rarely easy to assess. Indeed, it is true to say that for low response rate surveys, the statistical methods and assumptions required for valid inference are not very far removed from those used in non-probability sampling. I will return to the issue on nonresponse bias later in the report, for now it is useful to note that the theoretical benefits of random sampling may not always accrue fully in the real world.

[4] https://www.aapor.org/Education-Resources/Reports/Report-on-Online-Panels.aspx

What then of empirical comparisons between probability and non-probability surveys? Until comparatively recently, assessments of the relative accuracy of probability and non-probability sampling were prohibitively expensive. With the advent and proliferation of cheap online non-probability surveys in the 2000s, however, this situation has changed and there is now a large and growing body of evidence comparing estimates from the two sampling approaches to one another and to external gold standards. A recent paper by Cornesse et al (2019) synthesises the findings of this research. Probability samples, they find, consistently produce more accurate estimates than non-probability samples across topics such as voting, health, consumption behaviour, and sexual attitudes and lifestyles. This leads these authors to conclude that “even in the age of declining response rates, accuracy in probability sample surveys is generally higher than in nonprobability sample surveys” (ibid, p22).

In addition to finding that non-probability surveys tend to be less accurate than probability surveys, this body of empirical research has also revealed that non-probability samples often produce widely varying estimates of the same quantity (Callegaro et al 2014). For example, a study by the Pew Research Foundation[5] in the US found that estimates of knowledge of which party controlled the House and Senate covered a range of 26 percentage points. Several other variables also showed a very wide range of estimates. This suggests that variation in recruitment strategies and model specification in non-probability surveys can have a large impact on the accuracy of estimates.

Sources of error in survey estimates

Probability samples *should*, then, produce more accurate estimates than non-probability samples according to statistical theory, an expectation that is confirmed by actual empirical comparisons. However, while this is a reasonable expectation on average, it cannot be taken as a general rule. Indeed, across the many comparisons considered by Cornesse et al (2019), a minority showed non-probability survey estimates to be more accurate than the corresponding probability survey. Another recent example of non-probability samples out-performing probability samples comes from the 2016 EU referendum. The NatCen

[5] https://www.pewresearch.org/methods/2016/05/02/overall-variability-in-estimates-across-samples/

probability panel found a 6 percentage point lead for Remain[6] in the penultimate week of the campaign, while YouGov polls were regularly showing a narrow lead for Leave[7]. For a proper assessment of the likely accuracy of the prevalence estimates, it is therefore necessary to focus on the designs and errors of the specific surveys in question.
Table 1 shows the estimates of low-risk, moderate-risk and problem gambling behaviour according to the Problem Gambling Severity Index (PGSI) measure of gambling harm from the 2016 combined health survey, 2019 Gambling Commission telephone survey, and the YouGov survey. The PGSI is based on respondents' answers to 9 questions, each with 4 response alternatives ranging from 0= never, 1=sometimes, 2=most of the time, and 3=almost always. The total PGSI score is the sum of the scores for each individual items and the low-risk, moderate-risk, and problem gambling categories are for respondents with total scores of 1 to 2, 3 to 7, and 8 or above, respectively. I do not discuss the measurement properties of the PGSI further here but see Ferris and Wynne (2001) for more detail. An important caveat is that the Gambling Commission telephone survey uses a subset of just 3 items from the full PGSI, with thresholds set at 1=low-risk, 2-3=moderate-risk, 4+=problem gambler. The prevalence estimates for this survey are not, therefore, directly comparable to those of the other surveys which use the full 9 item PGSI.

For additional context, Table 1 also presents estimates from five other recent UK surveys of gambling harm prevalence, the 2015 and 2012 Gambling Prevalence surveys based on the Health Surveys, the standalone 2010 Gambling Prevalence Survey, the 2016 Northern Ireland Gambling Prevalence survey, and the age 24 (2016) Avon Longitudinal Study of Parents and Children (ALSPAC) cohort survey. The combined health surveys are all conducted as household surveys with the sample drawn randomly from the Postcode Address File (PAF), with a single randomly selected individual interviewed using face-to-face interviewing and the PGSI completed via a paper questionnaire during the interview. The Northern Ireland Gambling Prevalence Survey and the 2010 Gambling Prevalence Survey also use this sampling method but ask respondents to enter their answers for the PGSI into the interviewer's laptop directly. ALSPAC used 'opportunistic' non-probability sampling of

[6] http://natcen.ac.uk/our-research/research/public-opinion-on-the-eu-referendum-question/
[7] https://www.bbc.co.uk/news/uk-politics-eu-referendum-36648769

pregnant mothers and have followed up the children of the mothers who joined the survey periodically since the 1990s. It employed a combination of online and paper self-completion for the 2017-18 wave, although the technical report of the survey does not state in what proportions. ALSPAC has high rates of dropout with only 1921 respondents completing the PGSI in the 2017-18 wave from a total of 20,248 eligible pregnancies in 1990-1992. For comparability with the ALSPAC estimate, Table 1 also shows estimates for the 20 to 29 age group from the 2016 combined health surveys and the 18 to 34 estimate from the YouGov survey. I will focus mostly on the overall prevalence (i.e. scores of 1 or above) rate given the small cell sizes in the three categories and the somewhat arbitrary nature of the placement of the thresholds which produce the categories.

The estimated total prevalence rate is three times as high (13.2%) in the YouGov survey compared to the 2016 combined health surveys (4.2%) and the Gambling Commission showing approximately the same estimates of gambling harm prevalence as the combined health surveys (5.3%). The 2012 and 2015 combined health surveys are very similar to the 2016 estimates at 4.6% and 4.5% respectively, while the 2010 prevalence survey shows a higher overall rate of 8.1%, though this is mostly concentrated in the low prevalence category. The Northern Ireland survey shows a distribution quite similar to the YouGov survey, with an overall rate of 13.9%. Finally, the ALSPAC survey shows that at age 24, the cohort in this region had an estimated prevalence rate of 21.8% compared to 9.5% in the 2016 combined health surveys for the 20 to 29 age group and 18% for the 18 to 34 age group in the YouGov survey.

Table 1 Prevalence estimates of problem gambling from different UK surveys 2010 to 2019

Survey	Year	Mode	Sampling	sample size	response rate	PGSI Score Low-risk	Moderate-risk	Problem Gambler	Total
Gambling Commission	2018	phone/RDD	quota	4000	-	3.3%	1.5%	0.5%	5.3%
GambleAware/YouGov	2019	online	quota	16000	-	7.2%	3.3%	2.7%	13.2%
Combined health surveys	2016	f-t-f /paper	random	12161	55%	2.4%	1.1%	0.7%	4.2%
Combined health surveys*	2015	f-t-f /paper	random	8034	57%	2.8%	1.1%	0.6%	4.5%
Combined health surveys*	2012	f-t-f /paper	random	7676	56%	3.2%	1.0%	0.4%	4.6%
Gambling Prevalence Survey	2010	f-t-f /CASI	random	7748	47%	5.6%	1.8%	0.7%	8.1%
NI Gambling Prevalence Survey	2016	f-t-f /CASI	random	1003	53%	6.7%	4.9%	2.3%	13.9%
ALSPAC (aged 24)	*2017-18*	*online/paper*	*Non-random*	*1921*	*(<20%)*	*15.9%*	*4.4%*	*1.5%*	*21.8%*
Health Survey for England (20 to 29)	*2016*	*f* f-t-f /paper	*random*	*972*	*55%*	*6.5%*	*2.0%*	*1.0%*	*9.5%*
GambleAware/YouGov (18 to 34)	*2016*	online			-	*9.0%*	*4.0%*	*5.0%*	*18.0%*

*England and Scotland Only

Several points can be gleaned from Table 1. First, there are marked differences in magnitudes of overall gambling prevalence, ranging from a low of 4.2% to a high of 13.9%. Second, the differences between the YouGov estimates and those from the random samples drawn from PAF cannot be attributed to sampling variance. The large sample sizes mean that these differences are highly statistically significant.[8] The size of the estimates is related to mode and sampling method, with non-probability online surveys showing higher prevalence than face-to-face, although this is not uniform; the 2010 prevalence survey and 2016 Northern Irish survey both show markedly higher estimates than the combined health surveys. This may be partly due to higher actual prevalence in Northern Ireland and change over time in the rest of the UK, although it is worth noting that these surveys both use Computer Assisted Self Interview, where the respondent completes the PGSI on the interviewer's laptop. This may afford more privacy than completing a paper questionnaire, as is done in the combined health surveys.

The highest prevalence estimate comes from ALSPAC, although this is to be expected because this survey comprises only 24 year olds and this is a peak age to experience gambling problems. For a more appropriate comparison with ALSPAC, I have also presented estimates for the 20 to 29 age group in the combined health surveys and the 18 to 34 age group for YouGov. This shows that for the overall prevalence rate, ALSPAC is much closer to YouGov than it is to the combined health surveys, though the majority of this difference comes from the Low prevalence group. For the problem gambling (PGSI 8+) group, ALSPAC is much closer to the combined health surveys than to the YouGov estimate. There are many differences in the designs of these surveys[9], so it would not be sensible to draw strong conclusions about the reasons for these differences.

[8] The Prevalence survey and YouGov reports present 95% confidence intervals for the PGSI estimates which show this to be the case, although it is not clear how the YouGov intervals were calculated, given it uses non-probability sampling.

[9] Note also that the target population for ALSPAC is 24 year olds in the south west region of the UK so part of the difference in estimates may be due to actual differences in prevalence in this region compared to the rest of Great Britain.

Coverage error

A source of error which is likely to affect these surveys differentially is coverage error, particularly under-coverage. Coverage error is the discrepancy between the defined target population, such as all adults in Great Britain, and the population elements listed on the sampling frame. If some elements do not appear on the sampling frame, they have a zero probability of being included in the sample. To the extent that the under-covered elements are different on the population characteristic of interest, the survey estimates will be biased. Because non-probability sampling does not use a sampling frame, the definition of coverage error for probability sampling does not read across directly, although the notion of sub-groups of the target population are entirely excluded from potential inclusion in the survey does still apply. Coverage error is likely to affect the PAF based probability surveys because PAF excludes people who do not have a residential address, such as homeless people. PAF also excludes people who live in institutional addresses such as student halls of residence, hospitals, prisons, and military barracks.

To the extent that these groups have higher rates of problem gambling than the general population, which seems a reasonable assumption for some of them, their under-coverage will result in under-estimation of problem gambling in these surveys. That said, these groups are of very low incidence in the general population so their exclusion will likely have only a very modest impact on estimates of the prevalence rate. It may also be noted in this regard that the Gambling Commission telephone survey uses Random Digit Dialling with a 50/50 mix of landline and mobile phones and so would not have the same under-coverage of these groups, because anyone with a landline or mobile phone is, in principle covered. Yet it has approximately the same prevalence estimates as the combined health surveys[10].

The YouGov survey will also likely have under-coverage of these groups and so suffer from the same biases, although it is likely to have considerably better coverage of people living in student residences, military accommodation, and even homeless people because these groups can all, in principle at least, sign up to take YouGov surveys. A likely more important

[10] Although note the caveat about the comparability of the PGSI measure in this survey, as it is based on a subset of 3 items.

source of under-coverage in the YouGov survey derives from the exclusion of the off-line population as well as of very infrequent users of the internet. The 2018 British Social Attitudes survey estimated the offline population in Great Britain at 13.5%. This group cannot, by definition be included in the YouGov survey and also cannot engage in online gambling, a large and growing proportion of the gambling in the UK. Other demographic features of this group, i.e. being older and living in more rural areas, also suggest they are likely to have lower rates of gambling compared to the general population. By way of illustration, if we make the not unreasonable assumptions that the under-coverage rate is 15%, the rate of problem gambling is 10% in the covered population and 1% in the uncovered population, the survey would *over-estimate* the true population prevalence of problem gambling by 1.1 percentage points. This example shows that, while under-coverage of the offline population is likely to make some contribution to the high rate of problem gambling prevalence in the YouGov survey, it is a long way from being able to explain all of the difference from the combined health surveys.

Nonresponse bias

A second source of error to be considered is nonresponse error, or selection bias. In probability sampling, this results from failure to contact sampled elements or from their refusal to take part in the survey once contacted. Because in non-probability sampling, people choose to opt in to the sample rather than being randomly selected, the same idea of nonresponse bias does not really exist. However, the tendency for the kinds of people who sign up to the YouGov panel to be different from the general population has the same ultimate result of making the sample unrepresentative on the characteristic the survey is seeking to measure. For random surveys, protection against nonresponse bias comes from the randomisation in the sample selection, bolstered by the application of nonresponse and post-stratification weights. For the non-probability sample, claims to representativeness are based entirely on the variables used to specify the sampling quotas and the calibration weights. As noted earlier, this means that non-probability sampling requires stronger assumptions than probability sampling for accurate inference. It seems plausible that the kinds of people who are most likely to experience gambling problems might also be difficult to contact and persuade to take part in surveys. This would result in under-estimation of problem gambling in the probability surveys. It is though, by definition, impossible to know the extent of any such nonresponse bias without an external criterion by which accuracy can be assessed. It is worth noting, however, that recent studies of the relationship between response rates and nonresponse bias have found only a weak correlation (for example see Groves 2006; Sturgis et al 2015). There is therefore no reason to assume that a low response rate will automatically equate to nonresponse bias and, in any event, the response rates for the combined health surveys are in the mid-fifties, which are reasonably high by contemporary standards.

Non-probability samples are also prone to this type of selection bias, although in ways that are more difficult to assess, not least because a response rate cannot be calculated for non-probability samples. If -gamblers are more likely to sign up to be members of the YouGov panel, then estimates will be biased upwards, unless the variables used to set the sampling quotas and calibration weights mitigate this wholly, or in part. It is certainly possible that gamblers might be more likely to sign up to be YouGov panellists, though it is difficult to know how such an assessment might be made empirically without knowing the PGSI distribution for the entire YouGov panel. It is worth noting that the YouGov survey uses a quite limited range of calibration variables: sex, age, social grade, and region. If the unweighted data does over-represent problem gambling due to selection bias into the full panel, it seems unlikely that this set of variables would be successful in eliminating it entirely.

While both survey designs are likely affected by nonresponse bias, at least to some extent, it would seem that the YouGov survey probably has the greater risk. A plausible pattern of nonresponse bias is that the combined health surveys *under*-estimate problem gambling prevalence while the YouGov survey *over*-estimates it, with the true value lying somewhere in-between.

Measurement Error

Last, I consider the potential impact for measurement error to have affected the distributions on PGSI across the different surveys. Measurement error pertains when there is a difference between the true value and the measured value on the characteristic of interest, which arises due to the procedures used to make the measurement. The primary forms of measurement error that are relevant in the current case are socially desirable responding, context effects, and survey satisficing. Socially desirable responding (SD) is when a respondent over-reports socially desirable attitudes and behaviours or under-reports socially undesirable behaviours. The survey methodological literature is replete with examples of SD responding (Tourangeau, Rips and Rasinski, 2000). Problem gambling is clearly an example of a socially undesirable behaviour, so it seems reasonable to expect that some people will under-report its true extent in surveys. Context effects are when the responses obtained to a question are influenced by the content of the questions that have preceded it (Schuman et al, 1981). Survey satisficing is when respondents do not put much or any effort into selecting a response alternative (Krosnick, 1991). Instead, they select a response alternative that is easy and convenient rather than the one which most accurately represents their true position. Examples of satisficing response-styles in surveys include selecting the mid-point, selecting Don't Know alternatives, non-differentiation of responses between adjacent items, and agreeing with statements (acquiescence).

Turning first to SD responding, it is known that its influence is heightened in the context of social presence, which in the context of a survey interview, generally means the presence of an interviewer. People are less willing to admit to socially undesirable attitudes and behaviours in the presence of another individual. For this reason, we might expect the YouGov estimates to represent better measures of problem gambling prevalence because no interviewer is present when the respondent completes the PGSI. Indeed, studies have found that self-administration reduces the frequency of SD responding (Tourangeau and Yan, 2000). However, because the social presence effect has long been known in survey research, face-to-face interview surveys generally use self-administration procedures to measure sensitive attitudes and behaviours. This is done either by requiring the respondent to read and input the answers into the interviewer's laptop themselves (sometimes this is done with the additional security of the respondent wearing headphones and listening to a recording of the questions being read aloud) or by placing the sensitive questions in a pencil-and-paper self-completion questionnaire. This is completed during the interview, or left for the respondent to finish after the interviewer has left and to post back once completed. Table 1 shows that all the face-to-face interview surveys use one of these self-completion methods. It is, of course, possible that some respondents may still be influenced in the direction of SD responding by the presence of an interviewer during self-completion. There is some evidence that the presence of an interviewer does have such an effect for self-completion schedules (West and Peytcheva, 2014). However, these interviewer effects tend to be very small or zero in magnitude, so cannot be expected to have a substantial influence on SD responding when the questionnaire is self-completed by the respondent. In short, while we might expect an online survey to have lower levels of SD responding compared to an interviewer administered survey, any such difference is likely to be largely or wholly removed by the use of self-completion procedures in interviewer surveys.

Context effects may be relevant to understanding the causes of the difference between survey estimates because the content of the combined health surveys and the YouGov surveys was different. In particular, while the combined health surveys contained a wide variety of questions about health and lifestyle, the YouGov survey focused almost entirely on gambling. If the respondent has already been asked a lot of questions about their gambling behaviour before they complete the PGSI, it is possible that their responses to the PGSI questions may be pushed upward due to priming and memory effects. That is to say, the preceding questions about gambling might stimulate respondents to think about and recall gambling episodes and experiences that would not otherwise have come to mind and these memories then influence respondents to have greater recall of problem gambling experiences. However, the PGSI was administered very early in the YouGov survey (Question 5), so there is little scope for this kind of effect to have exerted a notable upward influence on the PGSI scores for these respondents. The combined health surveys, on the other hand, included the PGSI toward the end of the survey following a wide range of questions on matters unrelated to gambling. It is possible that this sequence resulted in an under-estimation of problem gambling prevalence, although the mechanism that would cause such an effect is not obvious. Nonetheless, as an evident discrepancy between the surveys, the difference in the content of the questions that preceded the PGSI in the combined health surveys cannot be ruled out as having made some contribution to the difference in prevalence estimates.

With regard to survey satisficing, we should anticipate that the YouGov survey might have higher rates of this type of responding compared to the combined health surveys. This is because a key driver of satisficing is respondent motivation. The less motivated a respondent is to complete a survey, the more likely it is that he or she will use a satisficing response strategy. Because YouGov panellists have signed up to take surveys for financial incentives, it follows that some (though by no means all of them) will be motivated to complete the survey in as short a time and with as little effort as possible. Non-probability panels do take action to mitigate this type of responding and sometimes remove respondents who display signs of having taken a satisficing approach to the survey, such as very undifferentiated responses, or extremely fast response times. It is not clear from the YouGov survey report if any such action was taken in the YouGov gambling survey.

The implications of survey satisficing for the prevalence estimates is not as clear as it is for SD responding, which is to say that there is no clear expectation for the *direction* of any such effects. A recent study by the Pew Research Center found evidence of a minority of between 4% to 7% of what they refer to as 'bogus respondents' in online non-probability panels who use extreme satisficing response styles.[11] These respondents do not choose response alternatives at random but tend to select positive answer choices, presumably because they think that this is what the survey organisation wants from 'good respondents'. Because the PGSI items have a positive response alternative as the highest category 'almost always', it is possible that these estimates may be pushed upward by this kind of 'bogus respondent'.

A related possibility is that opt-in panellists become aware that answers to questions in a survey may result in invitations to participate in additional surveys. Because panellists are taking surveys primarily for pay, they select response alternatives in order to maximise the number of surveys they are invited to take. This was a possibility for the YouGov gambling survey, where respondents who reported some level of gambling harm were invited to participate in a follow-up survey. This filtering approach might have increased the level of reporting of problem gambling in the YouGov survey. However, while this is a theoretical possibility, it is not clear how respondents would have known that their responses to the PGSI items would result in an invitation to participate in a follow-up survey. So, as for the

[11] https://www.pewresearch.org/methods/2020/02/18/assessing-the-risks-to-online-polls-from-bogus-respondents/

other forms of survey satisficing, while it seems reasonable to expect the prevalence of this type of responding to be higher in the YouGov survey, there is no direct evidence that it was.

Conclusion

In this report, I have considered theories of inference for probability and non-probability sampling and empirical evidence on their relative accuracy. I have also assessed the various sources of error that are likely to affect recent survey estimates of problem gambling prevalence in the UK. These are coverage, nonresponse, and measurement error. My assessment of this evidence leads me to conclude that the 2016 combined health surveys may somewhat under-estimate the true prevalence of problem gambling as a result of under-coverage and nonresponse amongst groups with higher rates of problem gambling compared to the general population. Additionally, it seems likely that there may be some degree of downward bias in the combined health survey estimates due to measurement error, with socially desirable responding and the content of the questions preceding the PGSI having a downward effect on the prevalence reports amongst some respondents.

For the YouGov survey, my assessment is that the coverage and nonresponse errors are likely working in the *opposite* direction, with under-coverage of low prevalence gambling groups and higher rates of selection into the YouGov panel of higher gambling prevalence groups resulting in over-estimation of problem gambling prevalence in this survey. There are no strong reasons to expect that the combined health surveys are substantially more subject to socially desirable responding compared to the YouGov survey. This is because the combined health survey employs self-completion procedures for the gambling questions, which should be expected to mitigate interviewer influence on socially desirable responding about as effectively as online self-completion. Additionally, the YouGov estimates may be subject to other forms of measurement error relating to survey satisficing, a measurement error that is particularly prevalent in non-probability panels.

Because these combined errors influence the estimates from the two surveys in opposite directions, it seems credible that the true level of people experiencing gambling harms lies somewhere in between their two bounds. This is not to contend that we should simply 'split the difference' and take the mid-point as the most reasonable estimate. Rather, I would hazard that, for the reasons set out in this report, the true value lies rather closer to the combined health surveys than it does to the YouGov survey. However, without an external gold standard to which the estimates can be compared, such an assessment will always be somewhat speculative.

I finish by noting that, even if the YouGov survey *does* over-estimate the true level of problem gambling prevalence in the UK, this does not mean that it is not of potentially high value in estimating the distribution of problem gambling prevalence across demographic groups in the population, its degree of stability and change over time, or of understanding the attitudes and experiences of individuals with gambling problems. This is because estimates of the patterns of association between problem gambling and these demographic and attitudinal variables will be approximately accurate, even if the level of problem gambling prevalence is over-estimated (Pasek, 2016). Because of the substantially lower cost of online non-probability surveys, much valuable information of this kind can be obtained to inform policy that would otherwise not be available at all if there were an insistence on the use of expensive random probability sampling in all designs.

References

Callegaro, M., R. Baker, J. Bethlehem, A. S. Go¨ritz, J. A. Krosnick, and P. J. Lavrakas (2014), "Online Panel Research: History, Concepts, Applications and a Look at the Future," in *Online Panel Research: A Data Quality Perspective*, eds. M. Callegaro, R. Baker, J. Bethlehem, A. S. Goritz, J. A. Krosnick, and P. J. Lavrakas, pp. 1–22, UK: John Wiley and Sons.

Cornesse, C. Blom, A. Dutwin, D. Krosnick, J D De Leeuw, E. Legleye, S. Pasek, J. Pennay, D. Phillips, B. Sakshaug,J. Struminskaya, B. Wenz, A. (2020) A Review of Conceptual Approaches and Empirical Evidence on Probability and Nonprobability Sample Survey Research, *Journal of Survey Statistics and Methodology*, Volume 8, Issue 1, Pages 4-36.

Ferris, J., & Wynne, H. (2001). *The Canadian Problem Gambling Index user manual. Report to the Canadian Inter-Provincial Task Force on Problem Gambling*. Canada: The Canadian Centre on Substance Abuse.

Groves, Robert M. 2006. “Nonresponse Rates and Nonresponse Bias in Household Surveys.” *Public Opinion Quarterly* 70:646–75.

Groves, R. M., and L. Lyberg (2010), “Total Survey Error: Past, Present, and Future,” *Public Opinion Quarterly*, 74, 849–879.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5, 213-236.

Schuman, H., Presser, S., & Ludwig, J. (1981). Context effects on survey responses to questions about abortion. *Public Opinion Quarterly, 45*(2), 216–223.

Sturgis, P. Kuha, J. Jennings, W. Baker, N. Callegaro, M. Fisher, S. Green, J. Lauderdale, B. and Smith, P. (2017) An assessment of the causes of the errors in the 2015 UK General Election Polls. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 181(3), 757-781.

Sturgis, P, Williams, J, Brunton-Smith, I and Moore, J (2016) Fieldwork effort, response rate, and the distribution of survey outcomes: a multi-level meta-analysis, *Public Opinion Quarterly,* 81(2), 523-*542*.

Tourangeau, R., Rips, L. J., & Rasinski, K. (Eds.). (2000). *The psychology of survey response*. Cambridge University Press.

West, B, and Peytcheva E (2014) “Can Interviewer Behaviors During ACASI Affect Data Quality?” *Survey Practice* 7 (5).

Yan, T. and Tourangeau, R. (2008), Fast times and easy questions: the effects of age, experience and question complexity on web survey response times. *Applied Cognitive Psycholo*gy, 22: 51-68.