

Earlier this month, YouGov conducted and published data from their survey to determine how most Americans like their breakfast foods. More specifically, the survey's goals were to find the preferences of the population of US adults (18 years and over) when eating the classic breakfast items of bacon, eggs, toast and coffee. With the questions on coffee and bread, respondents were shown pictures with varying levels of milk added and bread toasted to varying degrees (Ballard, 2019), accompanied by descriptive phrases like "with light milk, brown" or "somewhat lightly toasted" (Ballard, 2019). They were then asked to pick their preference on the spectrum. With the questions pertaining to eggs and bacon, no pictures were provided, and respondents were asked to pick a description best matching their preference; for example, "scrambled" or "somewhat crispy" (Ballard, 2019), respectively. All questions also contained the answer option of not eating that food at all.

Although this survey design is generally good, one main point of interest is that it is an opt-in survey, on two levels. This survey was conducted through an online interview that was administered to members of YouGov's panel through email. The YouGov "panel" is a list of adults who have previously signed up and elected to be sent various surveys by YouGov via email. Thus on the first level, the sample frame itself - the list of emails of the panel - was a list of individuals who opted in to be sent surveys to complete. On the second level, the frame members who then chose to open the email, click on the survey link and complete the survey opted in a second time.

The fact that this survey is opt-in may primarily have caused selection bias and thus sampling bias. Selection bias is one of the contributing factors to sampling bias, and occurs when some subgroups of the population have a disproportionately higher or lower chance of selection than others (Blair & Blair, 2015). In this case, certain groups of the population - US adults - may have been more likely to have opted in to YouGov generally - these include younger people, people who spend more time on the internet, or people with more time on their hands due to fewer study or work commitments. Furthermore, similar demographic characteristics (age, familiarity with the internet, school or career involvement)

mean that some members of the sample frame - the panel - potentially had a higher chance of seeing this particular email sent by YouGov and then choosing to fill out the survey. For these reasons, selection bias may have occurred in this survey, as some members of the population had a higher or lower probability of being sampled than others. This selection bias may have further caused sampling bias, as there is a possibility that the subgroups outlined were different in systematic ways, thus causing a “systematic failure to observe some elements” (Groves et al., 2006, p. 98) that could correlate to the research criteria of breakfast preferences. Therefore, while this survey attempts to generalise its data to American adults broadly, it may have produced “estimates that are too high or too low” (Groves et al., 2006, p. 98) in relation to the population’s true values.

Coverage error is another problematic aspect of this survey. The fact that it was a web survey means that coverage error most likely occurred, in relation to people without access to the internet. In fact, “only 75% of the US population currently uses the internet” (U.S. Census Bureau, 2012). There may be a systematic correlation between people who do not use or have access to the internet with socio-economic background (Anderson & Kumar, 2019), which may have a direct relationship with breakfast preferences. For example, some lower socio-economic groups may be busier due to the pressure to work more jobs to receive more disposable income; a busier subgroup may have different breakfast preferences (as expanded on below) or not eat breakfast at all. Thus, in internet surveys such as this one, “the theoretical potential for coverage bias is high” (Blair & Blair, 2015). That being said, selection and sampling bias are the focus of this critique.

Given that selection bias may have occurred, how could the characteristics of any systematically underrepresented groups correlate with and thus skew the results of the survey? Firstly, people who are more aware and up-to-date on online opportunities like YouGov, and who spend more time on the internet generally, are more likely to have heard of and then signed up for the YouGov panel. This characteristic could potentially align with younger, more tech-savvy individuals, as “while internet usage

has grown among all age groups and is virtually universal among younger Americans, older age groups lag behind” (Blair & Blair, 2015, p. 34). This means that the panel itself and thus the respondents sampled in this survey are more likely to be younger. Younger people may have different biases in breakfast preferences than older people; for example, younger people may enjoy black coffee less, or be more health conscious about processed meats like bacon due to recent revelations about them increasing cancer risk.

Moreover, as well as older adults, another potentially underrepresented subgroup could have been busier, more time-poor individuals. This also pertains to the potential for coverage error in this survey. People who know about YouGov but actively chose not to be a panelist at some point in the past, or who *are* panelists but were too busy to complete this particular survey, are likely more time-poor individuals, perhaps due to work or study commitments. Busier people may have skewed preferences about cooking times for foods, perhaps preferring more lightly toasted bread or less crispy bacon in the interest of time and efficiency in the mornings. It is important to note that the survey questions do not specify that respondents should choose their preferences on the basis of a particular characteristic, like taste or pleasantness. Different people prioritise different things when choosing their preferences for anything, and efficiency could be the priority in choosing food for a particularly busy demographic of individuals. Thus it is wholly possible that the opt-in group’s preferences were skewed towards more time-demanding breakfast options than the opt-out group’s choices would have been.

These potential areas of bias mean that there is a much higher chance that the inferences and means drawn from this survey may be far from accurate in depicting the true mean of the population (Groves et al., 2006). Thus, the survey results may not be an accurate depiction of breakfast preferences of all American adults generally; insofar as this is true, YouGov should not present them as such.

In terms of some recommendations to improve this survey, YouGov has already attempted to rectify some of the issues by weighting the data, stating that the “figures have been weighted and are representative of all US adults” (Ballard, 2019). It is unspecified whether US Census data or another source was used for this weighting process. Although weighting may help decrease biases, it is key to remark that it is a limited strategy for improvement. This is because weighting on “any characteristic will ensure that the sample is representative with respect to that characteristic..., but it will not ensure that the sample is representative with respect to characteristics that are not correlated with the characteristic used for weighting” (Dillman, Smyth, & Christian, 2014, p. 89). This means that weighting has the potential to have no impact on estimates at all, or can even make some less accurate - it all depends on how the characteristics used for weighting are related to the variables of interest: breakfast preferences (Dillman et al., 2014). Further, if an entire group of the population was missing from the sample in this survey - for example, if no extremely busy people with two jobs completed it at all - their responses will not exist, and thus they cannot be upweighted. No amount of upweighting could balance out the responses of subgroups who were nonexistent altogether in this survey.

One recommendation is to apply a quota approach, wherein the sample is forced to include certain subgroups that cater for factors like gender, race and particularly in this case, age. This improvement would mean that YouGov can continue using its panel with adjustments, but it requires knowledge of the key attributes of the frame population and their correlation with key survey criteria and values. With this knowledge, the survey can be properly balanced and it can be ensured that certain demographic characteristics and their relevant impacts on answers are not neglected to be measured due to the opt-in nature of the survey. If done correctly with full knowledge about the pertinent attributes, this survey could yield results that closely resemble the actual population (Groves et al., 2006). However, this knowledge is “usually very difficult to obtain” and be sure of (Groves et al., 2006, p. 410). Although YouGov has been known to apply quotas to some of their surveys, it is unspecified whether or not that approach was used for *this* survey. Another recommendation would be to introduce mixed modes into this survey; for

example, sending out the original email link first and then following up in a second round with telephone calls or mail surveys, depending on how much contact information is available from panelists. This mixed mode approach would improve coverage and biases, reaching systematically underrepresented groups who do not have access to the internet or email; for example, lower socio-economic subpopulations. Also, the alternative of a second or third mode with which to complete the survey may allow an individual to respond to it when they are unable to respond by a different mode (Dillman et al., 2014). An example is that “individuals who cannot respond on a computer because of not having developed those skills may be quite comfortable responding by paper or by telephone” (Dillman et al., 2014, p. 13). When used in conjunction, the quota and mixed mode methods may improve the accuracy of this survey’s estimates on the variables of US adults’ breakfast preferences.

Works Cited

- Anderson, M., & Kumar, M. (2019, May 7). Digital divide persists even as lower-income Americans make gains in tech adoption. Retrieved October 14, 2019, from <https://www.pewresearch.org/fact-tank/2019/05/07/digital-divide-persists-even-as-lower-income-americans-make-gains-in-tech-adoption/>.
- Ballard, J. (2019, October 2). How do Americans like their breakfast foods? Retrieved October 14, 2019, from <https://today.yougov.com/topics/food/articles-reports/2019/10/03/how-americans-eggs-coffee-toast-bacon-breakfast>.
- Blair, E., & Blair, J. (2015). *Applied survey sampling* [Yuzu]. Retrieved from <https://reader.yuzu.com/#/books/9781483346915/>
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, Phone, Mail, and Mixed-Mode Surveys : The Tailored Design Method* (Vol. Fourth edition). Hoboken: Wiley. Retrieved from <http://search.ebscohost.com.ezp-prod1.hul.harvard.edu/login.aspx?direct=true&db=nlebk&AN=827492&site=ehost-live&scope=site>
- Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology* (Vol. 2nd ed). Hoboken, N.J.: Wiley. Retrieved from <http://search.ebscohost.com.ezp-prod1.hul.harvard.edu/login.aspx?direct=true&db=nlebk&AN=587983&site=ehost-live&scope=site>
- Keeble, C., Law, G. R., Barber, S., & Baxter, P. D. (2015). Choosing a Method to Reduce Selection Bias: A Tool for Researchers. *Open Journal of Epidemiology*, 5(3), 155–162. doi: 10.4236/ojepi.2015.53020