Perspectives of paid panel survey research in cybercrime victimization and offending: Validity of global online market research sampling and data collection

Parti, K. (corresponding author), Virginia Tech

Dearden, T., Virginia Tech

Hawdon, J., Virginia Tech

Acknowledgment: This paper was sponsored by the Institute of Society, Culture and Environment at Virginia Tech and the Center for Peace Studies and Violence Prevention at Virginia Tech.

Abstract

Surveys are common tools for gathering data on a myriad of topics, including the study of online deviance and cybercrime. Topics such as teen cybervictimization (Wright et al., 2021), differential association and self-control applied to cybercrime (Dearden & Parti, 2021), cyber hatred (Reichelmann & Costello, 2021), cybercrime and institutional anomie theory (Dearden et al., 2021), and cybercrime and COVID (Hawdon et al., 2020; Kemp et al., 2021) have been studied using such non-probability data collected from online panels and tools such as Mechanical Turk. However, the validity of these data is of concern. This chapter considers the validity of using online polling for cybercrime data. First, we compare data from two of the most prominent global data vendors. Next, we use a control question about the 2020 presidential election. This item allows us to consider the validity of both internet-polling services by comparing our online surveys with data from FiveThirtyEight's "polls-only" and actual election results. Finally, we provide estimates of cybercrime offending and victimization from our data, and we discuss the differences in the estimates from the two samples.

Introduction

Surveys are standard tools used to gather representative data on a myriad of topics, including the study of online deviance and cybercrime. However, the increased use of the internet has led to the increased use of non-probability samples by researchers. Topics such as teen cybervictimization (Wright et al., 2021), problematic social media use and cybervictimization (Marttila et al., 2021), differential association and self-control and cybercrime (Dearden & Parti, 2021), cyber hatred (Reichelmann & Costello, 2021), cybercrime and institutional anomie theory (Dearden et al., 2021), and cybercrime and COVID (Hawdon et al., 2020; Kemp et al., 2021) have been studied using such non-probability data collected from online panels and tools such as Mechanical Turk.

Criminologists have long expressed concerns about measuring crime (Quetelet, 1839/1969). Since crime rates fluctuate, insufficient policies exist, and official statistics are limited, the use of victim surveys is common and provide a needed alternative to official data (Decker, 1977; Lynch 2014). Typically, victim surveys are used to inform criminological theories and to shape crime policies, particularly in Anglo-Saxon countries. These data are valuable as a counter to official data because they allow citizens direct participation in the documenting of the "crime problem." Moreover, since the official agencies providing protective services have a direct interest in how the "crime problem" is presented, using self-reported assessments of victimization protects against this bias by allowing the collection of data concerning crime and public safety to be separated from those agencies charged with providing protective services. However, despite their importance and development over time, victim surveys still present

limitations that make measuring crime difficult. Nevertheless, these surveys can provide valuable data for analyzing victims' behaviors and needs, but their value is a direct function of the accuracy with which they measure the concepts they aim to measure.

This chapter aims to evaluate online surveys as sources for estimating rates of online offending. We do this by comparing data collected by two leading industry panel research companies, juxtaposing the results with FiveThirtyEight, an American company focusing on opinion poll analysis. We use polls from the 2020 US presidential election as a baseline of analysis.

Literature review – Characteristics of non-probability sampling

Although most studies are performed using samples, census-based estimates are preferred whenever possible (Suresh et al., 2011; Rothman et al., 2013), independent of whether samples represent any target population. It would be nice to know how many people actually engage in cybercrimes and what is the actual rate of victimization; however, there are several theoretical and practical reasons that prevent researchers from carrying out census-based surveys. Barriers to census-based data include ethical issues (it may be unethical for researchers to include a greater number of individuals than is required to answer the research question), budgetary limitations and the costs of census surveys, logistical challenges with the required staff and equipment, the amount of time required of respondents to complete a census-based survey may be excessive, and unknown target population size (the actual pool of actors engaging in cybercrime is difficult to estimate). These reasons explain why samples are far more likely to be used than are census-based studies.

When it comes to sampling practices, there are two major sampling types: probabilistic and non-probabilistic. In probabilistic samples, every individual in the population has an equal likelihood of being selected into the sample. In non-probabilistic samples, the likelihood of selecting every individual from the target population to the sample is unknown or zero, which results in coverage error. Therefore, the results are usually not generalizable to the target population. Cybercrime researchers often use convenience samples such as classes of students (Back et al., 2018; Meldrum et al., 2019; Ngo & Paternoster, 2011; Bossler, 2021; Holt & Bossler, 2008), purposive sampling such as asking the opinion of experts in a particular field (Priya, 2021; Montoh et al., 2021; Correia, 2022; Ncubukezi et al., 2020), snowball sampling where current research participants help recruit future participants (Holt, 2007; Holt et al., 2017; Kinkade et al., 2013; Li & Chen, 2014; Steinmetz, 2015), or quota sampling (Reichelmann et al., 2021; Costello et al., 2021; Dearden et al., 2021; Parti et al., 2022). In quota sampling, the population is first classified by age, gender, race/ethnicity, etc., then sampling units are selected to complete each quota (For an overview, see Martinez-Mesa et al., 2016.)

Panel vendors use non-probabilistic quota samples. These vendors are organizations that sample and recruit potential respondents who have agreed to complete surveys for various forms of compensation. Once contracted by a researcher, the vendor then identifies a specified target audience of potential participants and invites them to complete a survey. Questions typically include survey items tapping participants' opinions, preferences, and behaviors. These vendors allow researchers to accrue survey responses while protecting respondents' anonymity as the vendors manage the sampling, selecting, and inviting panelists while verifying identities and compensating participating panelists for completing the survey. Once the data is collected, the

vendor delivers the survey data to the researchers. Researchers can get data expeditiously from respondents who previously signed up for survey participation. Offering data in a relatively short time at reasonable prices while managing the complexities of sampling and data collection, the use of panel vendors is quite attractive in social research. Compared to college samples, online panels provide access to participants with a wide range of demographics given how widespread digital devices are connected to the internet (e.g., smartphones, tablets, laptops, etc.). Even difficult-to-reach populations, such as marginalized communities and specific age groups, can be available via panel vendors for researchers, albeit not without challenges (Ibarra et al., 2018). Of course, one of the greatest attractions of online panel data is the relatively low expense of collecting it. Participants who sign up to be included in the online panels can be paid between $2 and $10 for completing a survey (Craig et al., 2013) or receive other rewards such as donations to charities in their name, "status points" from the vendor, or other inexpensive gifts.

Nevertheless, there are also drawbacks of online surveys. First, panel vendors cannot guarantee sample representativeness. Panel samples tend to over-represent higher educated, white, non-Hispanic, and younger populations, whose income exceeds the national median. For example, Craig et al. (2013) found that panel vendor samples underrepresent low socioeconomic respondents. Higher-income populations have access to better (faster and more reliable) internet connections at home, which probably explains their higher participation rate. This non-coverage bias affects quota sampling practices since only those with internet access can opt-in to online survey participation. Approximately 89.8% of the US population access internet in 2022 (InternetWorldStats, 2022); however, younger, more educated, and high-income people the most likely to have internet access. According to the Pew Research Center (2021), 96% to 99% of

adults under 65 maintain daily access to broadband internet, whereas only 75% of people 65 and above do so. While the once pronounced, gender and racial/ethnicity gaps in internet access are no longer significant (e.g., 94% of men and 93% of women; 93% of white, 91% of black, and 95% of Hispanics have access to the internet), only 57% of households with less than $30,000 income have broadband internet compared to 92% of households with $75,000 or higher income (Pew Research Center, 2021). Education also patterns one's chances of connecting to the World Wide Web: while 98% of college graduates are connected, only 71% of those without a high school diploma are (Perrin & Atske, 2021). Furthermore, online surveys show biases towards English-speaking, literate, non-visually impaired, and persons with low time costs (Eysenbach & Wyatt, 2022; Liu et al., 2010; Craig et al., 2013). When these demographic and socioeconomic characteristics cumulate in intersectional disadvantages, underprivileged communities and societal groups will be left out of panel vendor samples (Perrin & Atske, 2021). These biases may affect generalizability.

Characteristics of Dynata and Qualtrics (Prolific or Middleman Services)

In research examining data quality and representativity juxtaposing selected platforms (e.g., Amazon Mechanical Turk or MTurk, CloudResearch, Prolific), online platforms such as MTurk, where research participants use the platform as their primary source of income but spend just a few hours per week with surveys, did not provide quality data. These studies concluded that MTurk and other online platforms offer convenience samples at a reasonable price, but the quality of survey responses is low (Porter et al., 2019; Thompson & Pickett, 2019; Wang et al., 2015; Yeager et al., 2011). In addition, MTurk users can be highly active to the point that they are

considered professional survey takers, which could further hamper data quality (Chandler et al., 2014; Keith et al., 2017).

Survey platforms such as MTurk, CloudResearch, Prolific, and SurveyMonkey are *self-service platforms* that provide researchers complete control over the sampling and administering their study. These platforms apply river sampling, routing, or banners to attract survey participants. Panel vendors such as Qualtrics and Dynata are *middleman services* (Eyal et al., 2021) that handle sampling and administration of the study on behalf of the researchers. Prolific panel vendors can also run the study independently through common industry software platforms (such as Qualtrics and QuestionPro), and they invite panel members who have previously agreed to take online surveys based on set population characteristics (e.g., age, sex, education, etc.). Among other advantages, middleman services score higher in honesty, attention checks, and internal reliability (Cronbach's alpha between different sites; Eyal et al., 2021). Moreover, many participants (41-42%) of self-service platforms said they use the site as their primary source of income, but only 7% of Qualtrics users did so (Eyal et al., 2021). Studies suggest an inverse relationship between hours spent on the site and data quality, with self-service platform users showing relative high frequency in usage (eight hours a week or more) and providing lower quality data than middleman services with less than two hours a week (Eyal et al., 2021). Overall, Qualtrics and Dynata provide higher data quality than self-service platforms, yet many of their participants fail attention-check questions and cheat when given the opportunity. Furthermore, these panels are more expensive than self-service panels such as MTurk (Eyal et al., 2021). Still, Qualtrics and Dynata have the highest data quality among these types of online survey providers (Eyal et al., 2021).

Studies have also tested if probability-based selection occurs with panel vendors. Kimball (2019) found that prolific panel vendors such as Qualtrics and Dynata provide much more reliable data than self-service panels. Kimball (2019) tested sample composition in eight US states as well as on a nationwide sample. A demographic survey was fielded by prolific panels and MTurk. Demographic composition showed significant differences in some states compared to the US census data. Although basic demographics such as gender, race/ethnicity, and party affiliation only showed significant differences from census data in one or two in nine studies, educational attainment was significantly different from census data in half of the eight states. The age distribution showed the highest variability and diversion: MTurk was skewed towards a younger demographic because it was low in older age cohorts, whereas the prolific panels had a more normal distribution by age (Kimball, 2019). The study shows that panel vendors provide a demographic composition closer to the US census than self-service platforms.

In the current analysis, we compare two samples from prolific panel vendors, Dynata (formerly SSI Research Now) and Qualtrics, with data from the US Census and FiveThirtyEight, a company that reports on results from opinion polls while adjusting for the quality of the survey. Dynata and Qualtrics are commercial, prolific panelists operating globally with locations in the Americas, Europe, and Asia-Pacific, and are recognized as the leaders in the market research industry. After the panel companies are contacted by researchers/clients, they recruit participants and implement the survey compiled and designed by the clients.

Prolific commercial panel vendors use sampling quotas to obtain panel diversity. Prolific panelists may deliver samples from their own panel or leverage various technologies and platforms to aggregate/blend participants from a combination of sample sources (ESOMAR,

2021). These sources include opt-in databases of individuals who agreed to complete research projects and undertake other non-market research activities (such as watching ads, downloading an app, completing marketing offers, also known as loyalty programs, or getting rewards on "Get Paid To" sites). The "opt-in for market research" process requires respondents to submit an initial registration form requesting participation in market research studies. Potential participants build their profiles from a standardized list of questions. Panel vendors then select from the profiles that best fit the case specifications. There is a double opt-in requirement, meaning that those who do not reconfirm will not be contacted to participate in a survey.

To ensure data validation, Qualtrics replaces respondents who finish in less than half the median survey completion length (ESOMAR 2012). Then, within seven days of completion of sample quotas, the company sends the database to the client to check data quality and if any additional data replacement is needed to address any sampling biases and improve quality. The company and the client can also work together as the data are being collected to monitor any potential issues. For example, if the sample appears to be over-representing females, the vendor can increase its efforts to recruit males.

Characteristics of FiveThirtyEight

FiveThirtyEight is an American company that focuses on opinion poll analysis. Taking its name from the number of electors in the United States Electoral College (538), it was founded as a polling aggregation website in 2008. During the presidential primaries and general election of 2008, the site compiled polling data through a unique methodology called sabermetrics to "balance out the polls with comparative demographic data" (Romano, 2008a; Romano, 2008b).

Each poll is weighted based on the pollster's historical track record, sample size, and recentness of the poll (Silver, 2008). Since the 2008 election, the site has published articles on various topics in current politics and political news, typically creating or analyzing statistical information. These included monthly updates on numerous topics, including the prospects for turnover in the Senate; federal economic policies; Congressional support for legislation; public support for health care reform; global warming legislation; LGBT rights; elections around the world; marijuana legalization; and numerous other topics. The site and its founder are best known for election forecasts, including the 2012 presidential election in which FiveThirtyEight correctly predicted the vote winner of all 50 states and the District of Columbia.

Research questions

We aim to answer the research questions:

RQ1) Do panel sample data accurately reflect the nation in terms of demographic characters when compared to US Census data?

RQ2) How reliably did each panel and FiveThirtyEight predict the results of the 2020 US presidential election?

By answering these research questions, we seek to highlight the generalizability of panel surveys and the reliability of panel vendors for cybercrime researchers.

Sample and methods

We requested samples from two prominent panel vendors, Dynata and Qualtrics, two weeks before the 2020 US presidential elections. We requested that the sample quotas meet the

following requirements: 1,300 adult respondents representing the American population by age, sex, race, and ethnicity. In addition to questions about typical demographics such as age, sex, race, ethnicity, and education, the survey included items about participants' online presence, online activities, the use of computers in terms of both frequency and activities, familiarity with computers, cybercrime victimization, and cybercrime offending. We also asked about the respondents' voting preferences in the upcoming presidential elections. In the following analysis, we concentrate on common demographic information and the preferences in the presidential election.

The data from Dynata was collected between Oct 19, 2020, and Oct 27, 2020. Overall, 1,717 participants started the survey. One-hundred-twenty individuals who started the survey were dropped because they did not consent to participate as required by standard Institutional Review Board (IRB) ethical considerations. In addition, 197 participants were dropped from the sample because they completed the survey in less than 180 seconds and were considered "speeders." Dropping these respondents resulted in a final sample of 1,400.

Qualtrics data was collected between Oct 19, 2020, and Oct 26, 2020, and included 1,490 participants. Like the Dynata data, some participants were initially dropped from the data, including 114 who did not consent to the research per IRB requirements. In addition, 105 respondents were dropped because they completed the survey faster than 180 seconds (see Table 1). The final Qualtrics sample had 1,271 respondents.

Table 1: Sample Sizes for Total and Remaining Sample after Dropping Respondents

	# Participants starting the survey	# Participants not accepting IRB consent	Dropped # speeders (under 180s)	Remaining sample
Dynata	1,717	120	197	1,400
Qualtrics	1,490	114	105	1,271
Total	3,207	234	302	2,671

Results: Evaluating the Quality of the Data

To judge the accuracy of the two samples, we begin by comparing the general demographic information of our samples with US Census data. To do so, we use Pearson's chi-square tests, and the results can be found in Table 2. Based on these results, both Dynata and Qualtrics do reasonably well returning a sample that reflects the nation's demographic profile. First, both samples are well within expected margins of error in terms of sex and most racial/ethnic categories. The Qualtrics sample significantly under-represented the Hispanic or Latino, and it slightly over-represented the Black or African American population. Both samples significantly over-represented people of American Indian and Alaska Native origins.

Next, we compare age categories to the US census. Because our samples were limited to those 18 or over, the census figures reported below are derived from the US census instead of those reported by the census. The numbers below are the people in the respective age category as a percentage of the total US population 18 and over rather than the people in the category as a percentage of the total US population. Given this adjustment, age categories were reasonably well represented in both samples, at least for those between the ages of 35 and 74. Both samples over-sampled those ages 20 to 34 while under-sampling those 75 years old or over. Dynata performed slightly better reflecting those 75 and over while Qualtrics provided a slightly better sample for those between 20 and 34. This under-sampling of the older population and over-

sampling of the younger population is unsurprising given the relationship between age and internet use. These biases, slight as they may be, should be taken into consideration when trying to predict social phenomena related to age.

In terms of education, both samples return somewhat biased samples, but in different ways. The Dynata sample significantly under-represented those with a High School or equivalent degree, and substantially over-represented those with a Bachelor's degree and those with a graduate or professional degree. The only educational grouping the Dynata sample accurately reflected the nation's population was for those with some college. Qualtrics did not perform much better in terms of education, as it only accurately reflected the nation for those with a High School degree or equivalent, although it only slightly over-represented those with a Graduate or professional degree. It significantly over-represented those with some college and those with a Bachelor's degree. It is interesting to note that the samples did not differ in terms of the percentage of respondents with a Bachelor's degree. This educational category is clearly over-represented in both samples.

The Dynata sample more accurately represented the employment status of the nation. It did not significantly differ from the nation in terms of employment (59.3% for the sample, 59.6% for the nation). The Qualtrics sample, however, under-represented the employed, as only 53.9% of that sample reported being employed. This under-representation may reflect that in previous studies 7% of Qualtrics users said completing surveys was their primary source of income (Eyal et al., 2021); however, we lack the data to verify this possible explanation.

Finally, in terms of income, it appears that both samples did reasonably well. It is difficult to accurately test the representativeness of income since our survey collected data in income

ranges instead of actual dollar amounts. Collecting income in ranges is a common practice in survey research as people frequently will skip the income question unless ranges are provided. Nevertheless, the fact that the nation's median household income is approximately the mid-point of the median-range of the samples is encouraging.

Overall, the two samples appear to reasonably reflect the nation's demographics, although there are important shortcomings in terms of education and employment. Given the likely correlation between these factors and important topical concerns—voter preference and cybercrime in our case—this is worrisome and casts concerns about the true representativeness of the samples. It is however, possible to perfectly balance any desired demographic variable through both vendors. This aim can be achieved by not interviewing individuals whose demographic profile match those of the established quotas that have been filled. Of course, this approach changes the nature of the survey design as demographic questions would need to be at the beginning of the survey. It also can significantly affect the price of the survey. Weights to correct for over- or under-sampling certain groups can also be constructed to correct for sampling biases. Yet, the data here suggest that setting basic quotas on sex and race/ethnicity can result in a sample that reasonably reflects the general population.

Table 2: Sample and US Census Comparisons on Key Demographic Variables

	Dynata	**Qualtrics**	**US Census (Percent of Population over 18)**
Median age	41 (n/a)	42 (n/a)	38.2 (+/-0.1)
Persons 20 to 34, percent	30.3**	28.5*	26.5
Persons 35 to 54 years, percent	35.0	32.7	33.1
Persons 55 to 74 years, percent	26.2	29.8	28.4
Persons 75 years and older, percent	5.1**	3.3***	8.3

Female persons, percent	51.3	51.6	50.5
White alone, percent	73.7	75.0	75.8
Black or African American, percent	14.3	15.5*	13.6
Asian alone, percent	5.8	5.6	6.1
American Indian, Alaska Native	3.3***	2.5***	1.3
Hawaiian and Pacific Islander, percent	1.1***	0.3	0.3
Hispanic or Latino, percent	18.5	13.1***	18.9
High school or equivalent degree, percent	19.5***	28.5	26.7
Some college, no degree, percent	22.7	24.8***	20.3
Bachelor's degree, percent	29.2***	29.8***	20.2
Graduate or professional degree, percent	25.5***	14.8*	12.7
Employed, percent	59.3*	53.9***	59.6 (+/-0.1)
Median household income (in 2020 dollars)	$50-75 000 (n/a)	$50-75 000 (n/a)	$64,995 (+/-$128)

***p<.000; **p<.010; *p<.050; n/a: no p values

Next, we consider the accuracy of the polling services by comparing a question asking who planned to vote for in the 2020 presidential election to the election's final results. Respondents were asked the question, "If the 2020 election were held today, would you vote for: a) Donald Trump, b) Joe Biden, c) some other candidate, or d) will not vote/not sure." We only analyze here those who specified a candidate. We do this since most polls reported are of likely voters, and the fact that are data were collected between October 19, 2020 and October 26, 2020, within weeks of the actual election. We also compare data from FiveThirtyEight's collected polls. We use the reported predictions from the end of our data collection period, October 26, 2020 (FiveThirtyEigh 2020). As reported by FiveThirtyEight's "Who's ahead in the national polls," 52.3% of likely voters were predicted to vote for Joseph Biden, 42.8% for Donald Trump, and 3.1% for other candidates. Because FiveThirtyEight does not report undecided and not all polls report other candidates, we adjusted the reported percentages to reflect 100% of the total votes.

Doing this resulted in predictions of Biden receiving 53.3%, Trump 43.5%, and other candidates 3.2% of the vote. The predictions from the two samples and FiveThirtyEight are reported in Table 3 along with the results of the actual election as reported by the Federal Election Commission (FEC, 2020).

Table 3: Comparison of Dynata, Qualtrics, FiveThirtyEight and Actual Election Results

Source	***Joe Biden***	***Difference from Biden Actual***	***Donald Trump***	***Difference from Trump Actual***	***Other candidate***	***Difference from Other Actual***
Actual Results	51.3%		46.9%		1.8%	
Dynata	50.4%	-0.9	46.5%	-0.4	3.1%	+1.3
Qualtrics	53.8%	+2.5	42.1%	-4.8	4.1%	+2.3
FiveThirtyEight	53.3%	+2.0	43.5%	-3.4	3.2%	+1.4

The numbers in Table 3 indicate that all three predictions were reasonably accurate, and all three sources accurately predicted the eventual winner. Yet, it is clear from the data the Dynata sample outperformed both the Qualtrics sample and FiveThirtyEight's polls-only data. Dynata was within 1.0% of the actual results for both President Biden and former President Trump, and it only over-predicted votes for other candidates by 1.3%. It was the only source of data that did not over-estimate the share of the vote that would eventually go to President Biden, and it nearly perfectly predicted the vote for former President Trump. Both Qualtrics and FiveThirtyEight underestimated the vote that would go to then-president Trump.

Results: Estimating Cybercrime

Now that we are confident that our samples are reasonably reflective of the general population, we can ask what percentage of Americans are victimized by or engage in cybercrime. To measure cybervictimization, respondents were asked if they had been a victim of seven different types of cybercrime in the 12 months preceding the survey. In this chapter, we operate with a reactivist or relativist definition of deviance (Meier, 2007) when defining cybercrime. That said, to investigate the prevalence and patterns of cybervictimization and offending, we included behaviors that are clearly normatively "deviant" or illegal such as "hacking," identity theft, or publishing someone's intimate images without permission, but we also considered behaviors that are likely to elicit negative social reactions that can be considered "mildly deviant." An example of such an act would be excluding someone from an online community. Therefore, all the included "offending" types will likely provoke normative response if someone is victimized. As such, our definition of cybercrime is a compound of normative and reactivist definitions of deviant activities. Hence, to measure cyber-offending, respondents were asked if they engaged in any of 10 behaviors in the 12 months preceding the survey. Table 4 reports the number and percentages of respondents who were victimized or engaged in cybercrime for each survey.

Looking at Table 4, it becomes obvious that the two samples produce different estimates of involvement with cybercrime. Respondents in the Dynata sample consistently report less victimization and less cyber-offending than do the respondents in the Qualtrics sample. In fact, rates of victimization and offending in the Qualtrics sample are approximately twice those in the Dynata sample, on average. The differences between the two samples are more pronounced for offending than for victimization. For offending, respondents in the Qualtrics sample are approximately 2.2 times more likely to report committing the offense than are those in the

Dynata sample, and at times the estimate is 2.4 times higher in the Qualtrics sample than in the Dynata sample. This is the case for three offenses: hacking into an unauthorized area of the Internet, distributing malicious software, and posting nude photos without the person's permission. The two samples' estimates of victimization are relatively closer, with Qualtrics respondents being approximately 1.6 times more likely than Dynata respondents to report being a victim of a cybercrime, on average.

Table 4: Sample-based Cybervictimization and Cyber-offending Data

	Dynata		**Qualtrics**	
	Count	Percent	Count	Percent
Cybervictimization				
Lost money due to an email, website or other computer scam	162	13.0%	324	23.4%
Had your identity used by someone else to start a bank account, credit card or loan	125	10.0%	230	16.6%
Had unknown transactions in your bank/investment account, credit card, or other online payment system	208	16.6%	332	24.0%
Received notification from a company or organization that your private information, such as name, social security, credit card or password, has been stolen or posted publicly	192	15.1%	317	22.9%
Experienced hurtful comments, pictures or videos about you about posted online	168	13.5%	301	21.6%
Experienced unwanted sexual comments or advances online	182	14.6%	276	19.8%
Had a computer virus or malware that affected how your computer operated	154	12.3%	300	21.7%
Cyber-offending				
Posted hurtful Information about Someone on the Internet	118	9.4%	306	22.0%
Threatened or insulated others through email or instant messaging	97	7.8%	214	15.4%
Excluded someone from an online community	132	10.5%	299	21.5%
Hacked into an unauthorized area of the internet	74	5.9%	192	13.9%
Distributed malicious software	80	6.4%	214	15.4%
Illegally downloaded copyrighted files or programs	117	9.3%	247	17.8%
Illegally uploaded copyrighted files or programs	80	6.4%	193	13.9%
Used someone else's personal information on the internet without their permission	91	7.2%	213	15.4%
Bought prescriptions (without a prescription) or other drugs on online pharmacies or websites	97	7.7%	219	15.8%
Posted nude photos of someone else without his/her permission	81	6.5%	221	15.9%

Finally, to provide global estimates of the extent to which Americans are victimized by cybercrimes, we code respondents as either experiencing any of these cybercrime victimizations or not. We also code respondents as either engaging in any of the cybercrimes or not. Calculating estimates from each sample using these global measures of victimization and offending generate the estimates presented in Table 5. However, realizing that our samples over-represent young people and under-represent older people and knowing that age is related to cyber-offending and cybervictimization, we also re-estimate these figures after weighting the data so that it accurately reflects the US population in terms of age. These estimates are also presented in Table 5.

Table 5: Estimates of Global Cybervictimization and Cyber-Offending from Two Samples Using Unweighted and Weighted Data

	Victimization		**Offending**	
	Unweighted Data	Data Weighted on Age	Unweighted Data	Data Weighted on Age
Dynata	37.2%	35.8%	18.4%	17.1%
Qualtrics	46.5%	45.4%	31.6%	30.8%

The estimates in Table 5 reflect the same pattern as those in Table 4 in that the respondents in the Qualtrics sample were more likely to report being a victim of or engaging in cybercrime than were the respondents in the Dynata sample. However, the discrepancy between the two samples is smaller using this global measure than when individual crimes were reported. The Qualtrics estimate of victimization was only 1.3 times higher than the Dynata estimate, and the estimate of offending was only approximately 1.7 times higher in the Qualtrics sample as compared to the Dynata. Unsurprisingly, the data that is weighted by age provide lower estimates of cybercrime involvement than do the unweighted data. This is

expected since both samples under-represented youth and over-represented the older population. The correlation between age and cybervictimization and cyber-offending in both samples is negative, so using the unweighted data would over-estimate cybercrime involvement because it includes a disproportionate number of younger respondents who are disproportionately more likely to be both victims and offenders.

Discussion

In this chapter, we considered the validity of panel data used for cybercrime offending and victimization analysis. Analyzing data from two of the leading prolific panel vendors, Dynata and Qualtrics, we compared essential demographic characteristics of the samples to current US Census data. In addition, we utilized a control question about the likelihood of voting for the presidential candidates in the 2020 election, considering FiveThirtyEight, an American company rating polls by quality. Furthermore, we consider potential differences and future implications of utilizing panel vendors' non-probability samples for measuring cybercrime.

Samples collected by the two prolific panel vendors, Dynata and Qualtrics, did not show significant differences: the 1,400 US adult population samples were completed within a week, and the demographic variables showed little difference from actual US Census data, partly confirming previous studies. The differences observed in the panels' age, racial/ethnicity, education, and employment components indicated a slightly higher rate of 65-74 years of age individuals but a lower representation of 75+ individuals in the panels compared to then current US Census data. This result is similar to previous research scrutinizing quota sampling (Eyal et al., 2021). Next, both panels performed well representing sex. In contrast to Kimball (2019), our

samples did not under-represent racial and ethnic minorities. Comparing our sample to the US census, both samples accurately reflected the White and Asian populations, and Dynata also accurately reflected the Hispanic and Black or African American populations. Hispanic or Latinx individuals were under-represented in the Qualtrics sample. Both panel samples over-represented highly educated individuals, similar to previous research findings (Craig, 2013; Szolnoki & Hoffmann, 2013). The most striking difference was the rate of overrepresentation of people with college and professional degrees in the panels. Apparently, individuals with higher socioeconomic status (Craig, 2013; Szolnoki & Hoffmann, 2013) – a composition of better education and older age in our sample -- are indeed more likely to participate in panel surveys. Employment status was also slightly under-represented in the Qualtrics sample as it showed a 5.7% difference over-representing unemployed participants. Regarding income differences, the median household income of panel participants was very similar to the US Census median.

There are several issues to consider when trying to answer why specific demographics are over-represented or under-represented in prolific panels. First, interest in participating in surveys might differ by the survey topic, and this selection bias can dramatically influence survey results (see Shropshire et al., 2009). For example, women are more concerned about social and health care-related issues than are men (Bidmon & Terlutter, 2015; Rainie & Fox, 2000). Second, broadband internet access, an inherent necessity of taking part in online surveys, is not equally available for everyone. In the United States and elsewhere, internet access is positively associated with being male, being younger, having higher income and more education, being White or Asian, and not being Hispanic (Horrigan, 2010; Bethlehem & Biffignardi, 2012). However, measuring internet penetration has become more challenging since people can get

reliable internet over their phones (Horrigan, 2010), and many survey platforms such as Qualtrics now design their surveys to be visually appealing and well-functioning on mobile phones. Still, out of all the above differentiating characteristics, our samples seriously over-represented only higher educated individuals. Panel vendors typically field a survey until the requested quotas are filled, and we cannot tell why participants took the survey at an earlier or later stage. As such, we cannot tell if early and later survey-takers differ in their motives for participating in the survey. Future research should consider asking participants directly about why they completed a particular survey, including cybercrime-related social surveys. Fourth, when turning to panel vendors, researchers should also consider that recruitment rates for panel survey participation are 33% or less (nonresponse bias; Callegaro & DiSogra, 2009), which undoubtedly creates bias in the samples and decreases their representativeness.

The study's second goal was to determine how reliably each panel predicted the result of the US presidential election. Our samples were also compared to the predictions reported by FiveThirtyEight. All three sources accurately predicted Joseph Biden as the winner, but Dynata provided the most accurate results. Qualtrics and FiveThirtyEight both under-estimated the share of the vote that President Trump would eventually receive, possibly indicating a bias in these data sources. It is possible that both sources under-sampled the groups most likely to vote for then-president Trump. This potential problem notwithstanding, our result indicate that panel vendor surveys can predict the outcomes of an upcoming presidential election with similar levels of accuracy as more traditional methods of polling potential voters. While no polling will achieve 100% accuracy, all these sources accurately reflected the final results within expected margins of error.

Regarding data quality and representativeness, we can conclude that the panel samples' demographic composition is slightly different from that of the US Census; however, the samples overall provided reasonable reflections of the nation's demographic profile. Nevertheless, survey participants' drives and motivations are unknown, and these could include money (or the small rewards offered by panel vendors), general interest in participating in social surveys, or specific interest in participating in a survey about cybercrime victimization and offending. Since the channels of circulating the polls gathered by FiveThirtyEight and our panel vendors were slightly different (with FiveThirtyEight polls being circulated not exclusively online but also over dial-up services), we can expect more accuracy in FiveThirtyEight polls. However, panel surveys still offer a great deal of advantages to researchers, including quota sampling for a reasonable price, which produces more reliable results compared to other non-probability sampling techniques such as convenience sampling on college campuses or snowball sampling.

Although our samples from these non-probabilistic panel vendors reasonably reflected US demographics and presidential election results, there are issues to consider when selecting panel vendors. Although prolific panels utilize a sample larger and more aligned with the US census than other non-probabilistic samples, realistically, every participant has the potential to provide poor quality data if they are incorrectly screened. Panel vendors admittedly (Fawson & Lorch, 2021) try to build in quality checks to the process, such as fraud or cheating prevention, by employing machine learning or AI, reasonable rewards, accuracy screeners, and mobile-friendly surveys. According to ESOMAR (2015) recommendations for enhancing data quality, agencies must validate each research participant's claimed identity, and providers must ensure that research participants complete the same survey only once. However, researchers could also

take actions to enhance data quality as well. For example, applying a good online questionnaire design can provide clarity and motivation to complete the survey once it is started. The longer the survey, the more likely research participants will disengage and potentially drop out before finishing the survey. This process is known as "survey fatigue" (Shropshire, et al., 2009). Repeated matrix or grid-style questions can result in straight-lining and other patterned responses. Furthermore, a phenomenon known as primacy, where questions with a large number of answer categories can result in research participants choosing responses from the top of the list more often than the bottom, can also be problematic. Conversely, the use of images can increase respondent interest and data quality (Witte et al., 2004). In addition to the application of good survey design, building speed checks and attention check questions in the survey can identify a large number of problematic participants and enhance response validity (for standardized reporting checklists, see Cheung et al., 2017; DeSimone et al., 2015; Keith et al., 2017).

Cybercrime researchers face additional challenges compared to social researchers with other interests. Studies of cybercrime found that participants with the highest income level were largely unavailable, and the recruitment of older people took longer than for the rest of the sample (Miller et al., 2020). Moreover, gathering data about people's engagement in crimes and deviant activities present further challenges. According to Hutchings and Holt (2018), researchers tend to conceptualize cybercrime and deviant cyber activities such as hacking, cyberbullying, and cyberstalking, differently. This can complicate the collection of data on these topics and make comparisons across samples difficult or impossible. In addition, there are ethical issues associated with researching cybercrime, as recruiting offenders and convincing them to

participate in research can be challenging (Hutchings & Holt, 2018). Considering that participant honesty is compromised in asking about illegal and ambiguous activities, we recommend using triangulation (multimode design) to ensure data reliability. We believe that a mix of cross-sectional surveys or longitudinal panel measurements, supported by qualitative research techniques such as interviews and focus groups, would provide a good balance for studying cybercrime and its impact on society.

In conceptualizing "cybercrime," we followed a broader approach of deviance that included normative illegal as well as relative or reactionist deviant activities. Other research may utilize a more constructivist definition of deviance and ask why certain offenses are defined as crimes while others that are equally or even more harmful are not, whose interests are served by defining some but not other behaviors as crimes, whether the act is harmful, and what reactions they provoke. Yet, we maintain that even constructivist studies of crime and deviance require accurate data. For example, studies concerning the creation and sustaining of moral panics (e.g., Hawdon, 2001; Bonn, 2010) need good estimates of the "problem" to establish if a constructed moral panic was a disproportionate response to the potential harm posed by the targeted behavior and group. Therefore, the focus on respondents' perceptions of what they considered to be "crimes," "deviant," and "harmful" is important.

Another limitation to the current research is that we only included US samples in the analysis. First, the commercial survey platforms can only be utilized in countries where these platforms are available (i.e., in countries with developed economies). In contrast, countries with lower GDPs and non-capitalist cultures are typically not represented; hence, part of the global population is excluded from the possibility of professional panel-based data collection. This is

clearly an issue of coverage bias related to the use of these vendors. Yet, it should be noted that collecting data using other methods would also be difficult in many of the nations that are not covered by online panels. Moreover, a significant number of US users are registered in the commercial platforms under scrutiny, which increases reliability and cost effectiveness of panel research. Furthermore, the presence and availability of fairly accurate census polls in the US provided opportunity for comparison. In addition, the fact that our data was collected at the same time of the presidential election allowed us to directly compare our prediction of that event with the actual results provided an excellent test of the data's accuracy and utility. Still, professional panels have limitations. In addition to the coverage bias mentioned above, there is a likely selection bias due to marginalized populations not being able to reach online platforms and consequently being under-represented. Coverage biases may also be introduced by educational level, race and ethnicity, income, and other factors. While these biases can be estimated and steps to correct them can be taken, they are nevertheless issues that must always be considered when using survey research data.

Finally, it is important to note that our two samples provide very different estimates of cybercrime involvement. Despite both being reflecting the demographic characteristics of the country reasonably well and providing relatively accurate predictions of the 2020 presidential election, respondents from the Qualtrics sample were significantly more likely to report being victims of cybercrime and engaging in cyber-offenses than were the Dynata respondents. This difference is likely a function of who participated in the survey. As noted above, topical interest is an important determinant of survey participation (Shropshire et al., 2009), and we suspect that interest in cybercrime may have varied between the two samples. Given that the surveys used

to collect the data were identical, other survey design factors could not have influenced these estimates. Unfortunately, we are unable to definitively know why the estimates varied so much, but the fact that they did should give researchers pause when trying to use surveys to estimate rates of cybercrime involvement. This problem dramatically highlights the need for the collection of high-quality data on cybercrime using techniques such as those used to collect the National Crime Victimization Survey (NCVS). While collecting these data are extremely expensive, our estimates of the extent to which Americans experience cybercrime victimization or participate in cybercrime will be questionable and open to the errors that plague all survey research.

Conclusion

The issues to consider when applying panel samples are responsiveness, loss of attention, attrition due to survey length and ambiguities, and failure to maintain sociodemographic diversity. However, some studies do not require representativeness to ensure internal validity. For example, experimental designs that test theory-driven hypotheses within a defined population sampled by a panel vendor can legitimately test the hypotheses. Before deciding on prolific panel vendors, we recommend that researchers proactively ask questions about panels and carefully consider the strengths and drawbacks of online survey features, including quota sampling.

References

Back, S., Soor, S., & LaPrade, J. (2018). Juvenile hackers: An empirical test of self-control theory and social bonding theory. *International Journal of Cybersecurity Intelligence & Cybercrime, 1*(1), 40-55. https://www.doi.org/10.52306/01010518VMDC9371

Bethlehem, J., & Biffignandi, S. (2012). *Handbook of web surveys.* Wiley & Sons

Bidmon, S., & Terlutter, R. (2015). Gender differences in searching for health information on the internet and the virtual patient-physician relationship in Germany: Exploratory results on how men and women differ and why. *Journal of Medical Internet Research, 17*(6), e156. https://doi.org/10.2196/jmir.4127.

Bonn, S. A. (2010). *Mass deception. Moral Panic and the US war on Iraq.* Rutgers University of Press.

Bossler, A.M. (2021). Perceived formal and informal sanctions in deterring cybercrime in a college sample. *Journal of Contemporary Criminal Justice, 37*(3), 452-470. https://doi.org/10.1177/10439862211001630

Callegaro, M., & DiSogra, C. (2009). Computing response metrics for online panels. *Public Opinion Quarterly, 72*(5), 1008-1032. https://doi.org/10.1093/poq/nfn065

Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods, 46*(1), 112–130. https://doi.org/10.3758/s13428-013-0365-7

Cheung, J.H., Burns, D.K., Sinclair, R.R., & Sliter, M. (2017). Amazon Mechanical Turk in organizational psychology: An evaluation and practical recommendations. *Journal of Business and Psychology, 32*(4), 347–361. https://doi.org/10.1007/s10869-016-9458-5

Craig, B.M., Hays, R.D., Pickard, A.S., Cella, D., Revicki, D.A., & Reeve, B.B. (2013). Comparison of US panel vendors to online surveys. *Journal of Medical Internet Research, 15*(11), e260. http://www.jmir.org/2013/11/e260/

Costello, M., Restifo, S.J., & Hawdon, J. (2021). Viewing anti-immigrant hate online: An application of routine activity and social structure-social learning theory. *Computers in Human Behavior, 124*. https://doi.org/10.1016/j.chb.2021.106927

Correia, S.G. (2022). Making the most of cybercrime and fraud crime report data: A case study of UK Action Fraud. *International Journal of Population Data Science, 7*(1), 9-17. https://doi.org/10.23889/ijpds.v7i1.1721

Dearden, T.E., & Parti, K. (2021). Cybercrime, differential association, and self-control: Knowledge transmission through online social learning. *American Journal of Criminal Justice, 46,* 935–955. https://doi.org/10.1007/s12103-021-09655-4

Dearden, T., Parti, K., & Hawdon, J. (2021). Institutional anomie theory and cybercrime—Cybercrime and the American dream, now available online. *Journal of Contemporary Criminal Justice, 37*(3), 311-332. https://doi.org/10.1177/10439862211001590

Decker, S.H. (1977). Official crime rates and victim surveys: An empirical comparison. *Journal of Criminal Justice, 5,* 47-54.

DeSimone, J.A., Harms, P.D., & DeSimone, A.J. (2015). Best practice recommendations for data screening. *Journal of Organizational Behavior, 36*(2), 171–181. https://doi.org/10.1002/job.1962

ESOMAR (2012). *28 questions to help buyers of online samples.* ESOMAR World Research. Retrieved July 24, 2022 from https://swiss-insights.ch/wp-content/uploads/2020/05/ESOMAR-28-Questions-to-Help-Buyers-of-Online-Samples-September-2012.pdf

ESOMAR (2015). *ESOMAR/GRBN Guideline for Online Sample Quality.* Global Research Business Network. Retrieved July 24, 2022 from https://grbn.org/wp-content/uploads/2016/12/Online_Sample_Quality_Guideline.pdf

ESOMAR (2021). *Questions to help buyers of online samples.* ESOMAR Global Research. Retrieved July 24, 2022 from https://esomar.org/uploads/attachments/ckqqecpst00gw9dtrl32xetli-questions-to-help-buyers-of-online-samples-2021.pdf

Eyal, P., Rothschild, D., Gordon, A., Evrden, Z., & Damer, E. (2021). Data quality for platforms and panels for online behavioral research. *Behavioral Research Methods, 54,* 1643–1662. https://doi.org/10.3758/s13428-021-01694-3

Eysenbach G, & Wyatt J. (2002). Using the Internet for surveys and health research. *Journal of Medical Internet Research, 4*(2), e13. https://doi.org/10.2196/jmir.4.2.e13

Fawson, B., & Lorch, J. (March 29, 2021). *The new dynamics of online sample quality.* Greenbook. Retrieved July 24, 2022 from https://www.greenbook.org/mr/market-research-trends/the-new-dynamics-of-online-sample-quality/

FEC (2020). *Official 2020 presidential general election results.* Federal Election Commission. Retrieved August 26, 2022 from https://www.fec.gov/resources/cms-content/documents/2020presgeresults.pdf

FiveThirtyEight (2020). *Who's ahead in the national polls?* FiveThirtyEight. Retrieved August 26, 2022 from https://projects.fivethirtyeight.com/polls/president-general/2020/national/

Hawdon, J. (2001). The role of presidential rhetoric in the creation of a moral panic: Reagan, Bush, and the war on drugs. *Deviant behavior*, *22*(5), 419-445.

Hawdon, J., Parti, K. & Dearden, T.E. (2020). Cybercrime in America amid COVID-19: The initial results from a natural experiment. *American Journal of Criminal Justice, 45,* 546–562. https://doi.org/10.1007/s12103-020-09534-4

Holt, T.J. (2007). Subcultural evolution? Examining the influence of on- and off-line experiences on deviant subcultures. *Deviant Behavior, 28*(2), 171-198. https://doi.org/10.1080/01639620601131065

Holt, T.J., & Bossler, A.M. (2008). Examining the applicability of lifestyle-routine activities theory for cybercrime victimization, Deviant Behavior, *30*(1), 1-25. https://doi.org/10.1080/01639620701876577

Holt, T.J., Freilich, J.D., & Chermak, S.M. (2017). Exploring the subculture of ideologically motivated cyber-attackers. *Journal of Contemporary Criminal Justice, 33*(3), 212-233. https://doi.org/10.1177/1043986217699100

Horrigan, J.B. (2010). *Broadband adoption and use in America. Results from an FCC survey.* Federal Communications Commission. Retrieved July 24, 2022 from https://transition.fcc.gov/DiversityFAC/032410/consumer-survey-horrigan.pdf

Hutchings, A., & Holt, T.J. (2018). Interviewing Cybercrime Offenders. *Journal of Qualitative Criminal Justice & Criminology* [Preprint]. https://doi.org/10.21428/88de04a1.1fdab531

Ibarra, J.L., Agas, J.M., Lee, M., Pan, J.L., & Buttenheim, A.M. (2018). Comparison of online survey recruitment platforms for hard-to-reach pregnant smoking populations: Feasibility study. *Journal of Medical Internet Research Research Protocols, 7.*

Internet World Stats (2022). *Internet users and 2020 population in North America.* InternetWorldStats. Retrieved July 24, 2022 from https://www.internetworldstats.com/stats14.htm#north

Keith, M. G., Tay, L., & Harms, P. D. (2017). Systems perspective of Amazon Mechanical Turk for organizational research: Review and recommendations. *Frontiers in Psychology, 8,* 1359. https://doi.org/10.3389/fpsyg.2017.01359

Kemp, S., Buil-Gil, D., Moneva, A., Miro-Llinares, F., & Diaz-Castano, D. (2021). Empty streets, busy internet: A time-series analysis of cybercrime and fraud trends during COVID-19. *Journal of Contemporary Criminal Justice, 37*(4): 480-501. https://doi.org/10.1177/10439862211027986

Kimball, S.H. (2019). Survey data collection; Online panel efficacy. A comparative study of Amazon MTurk and Research Now SSI/ Survey Monkey/ Opinion Access. *Journal of Business Diversity, 19*(2), https://doi.org/10.33423/jbd.v19i2.2054

Kinkade, P., Bachmann, M., & Bachmann, B. (2013). Hacker Woodstock: Observations of an off-line cyber culture at the Chaos Communication Camp 2011. In T.J. Holt (Ed.), *Crime on-line: Correlates, causes, and context* (2nd ed., pp. 27-53). Carolina Academic Press.

Li, W. & Chen, H. (2014). Identifying top sellers in underground economy using deep learning-based sentiment analysis, *2014 IEEE Joint Intelligence and Security Informatics Conference* (pp. 64-67) https://doi.org/10.1109/JISIC.2014.19

Liu, H., Cella, D., Gershon, R., Shen, J., Morales, L.S., Riley, W., & Hays, R.D. (2010). Representativeness of the patient-reported outcomes measurement information system Internet panel. *Journal of Clinical Epidemiology, 63*(11), 1169-1178. https://doi.org/10.1016/j.jclinepi.2009.11.021

Lynch, J. (2014). The evolving role of self-report surveys of criminal victimization in a system of statistics on crime and the administration of justice. *Statistical Journal of the IAOS*, *30*(3), 165-169.

MacInnis, B., Krosnick, J.A., Ho, A.S., & Cho, M-J. (2018). The accuracy of measurements with probability and nonprobability survey samples: replication and extension. *Public Opinion Quarterly*, *82*(4), 707–744, https://doi.org/10.1093/poq/nfy038

Martinez-Mesa, J., Gonzalez-Chica, D.A., Duquia, R.P., & Bastos, J.L. (2016). Sampling: How to select participants in my research study? *Anais Brasileiros de Dermatologia, 91*(3):326-330. http://dx.doi.org/10.1590/abd1806-4841.20165254

Marttila, E., Koivula, A. & Räsänen, P. (2021). Cybercrime victimization and problematic social media use: findings from a nationally representative panel study. *American Journal of Criminal Justice, 46,* 862–881. https://doi.org/10.1007/s12103-021-09665-2

Meier, R.F. (2007). Deviance, normative definitions of. In R.F. Meier & C.M. Cain: *The Blackwell Encyclopedia of Sociology.* (pp. 1116-1117) Wiley. https://doi.org/10.1002/9781405165518

Meldrum, R.C., Boman, J.H. & Back, S. (2019). Low self-control, social learning, and texting while driving. *American Journal of Criminal Justice, 44,* 191–210. https://doi.org/10.1007/s12103-018-9448-4

Miller, C.A., Guidry, J.P.D., Dahman, B., Thomson, M.D. (2020). A tale of two diverse Qualtrics samples: Information for online survey researchers. *Cancer Epidemiology Biomarkers Preview, 29*(4): 731-735. https://doi.org/10.1158/1055-9965.

Montoh, A., Mirahyanti, G.A., & Rantung, T. (2021). Cyber security analysis on small medium enterprise's online shop. *International Journal of Information Technology and Education, 1*(1), https://ijite.jredu.id/index.php/ijite/article/view/14

Ncubukezi, T., Mwansa, N., & Rocaries, F. (2020). A review of the current cyber hygiene in small and medium-sized businesses. *15th International Conference for Internet Technology and Secured Transactions (ICITST)* (pp. 1-6) https://doi.org/10.23919/ICITST51030.2020.9351339

Ngo, F.T., & Paternoster, R. (2011). Cybercrime victimization: An examination of individual and situational level factors. *International Journal of Cyber Criminology*, *5*(1), 773-793.

Parti, K., Dearden, T., & Hawdon, J. (2022). Understanding the overlap of online offending and victimization: using cluster analysis to examine group differences. *Victims & Offenders, 17*(5), 712-734, https://doi.org/10.1080/15564886.2022.2036655

Perrin, A. & Atske, S. (April 2, 2021). *7% of Americans don't use the internet. Who are they?* Pew Research Center. Retrieved July 25, 2022 from https://www.pewresearch.org/fact-tank/2021/04/02/7-of-americans-dont-use-the-internet-who-are-they/

Pew Research Center (April 7, 2021). *Internet/Broadband Fact Sheet.* Retrieved July 24, 2022 from https://www.pewresearch.org/internet/fact-sheet/internet-broadband/

Porter, C. O., Outlaw, R., Gale, J. P., & Cho, T. S. (2019). The use of online panel data in management research: A review and recommendations. *Journal of Management, 45*(1), 319–344. https://doi.org/10.1177/0149206318811569

Priya, P. (2021). Effectiveness of planned teaching programme on knowledge regarding cybercrime among higher secondary students in selected school of Indore district in the year 2015-16. *GFNPSS Global Nursing Journal of India, 4*(2), Retrieved August 26, 2022 from https://ssrn.com/abstract=3930173

Quetelet, L.A.J. (1839/1969). *A treatise on man and the development of his faculties.* Gainesville, FL: Scholars Facsimilies and Reprints.

Rainie, L. & Fox, S. (November 26, 2000). *The online health care revolution.* Pew Research Center. Retrieved August 26, 2022 from https://www.pewresearch.org/internet/2000/11/26/the-online-health-care-revolution/

Reichelmann, A.V. & Costello, M. (2021). When patriot becomes hate-triot: The relationship between American identity and the production of cyberhate. *American Journal of Criminal Justice, 46,* 956–979. https://doi.org/10.1007/s12103-021-09659-0

Reichelmann, A.V., Hawdon, J., Costello, M., Ryan, J., Blaya, C., Llorent, V., Oksanen, A., Räsänen, P. & Zych, I. (2021). Hate knows no boundaries: Online hate in six nations. *Deviant Behavior, 42*(9), 1100-1111. https://doi.org/10.1080/01639625.2020.1722337

Romano, A. (2008a). *Where we were on June 16, 2004--and what it means for Nov. 4, 2008.* Newsweek. Retrieved July 24, 2022 from

https://web.archive.org/web/20080617214439/http:/www.blog.newsweek.com/blogs/stumper/archive/2008/06/16/the-electoral-map-is-obama-outperforming-kerry-or-is-mccain-outperforming-bush.aspx

Romano, A. (2008b). *Making his pitches: Nate Silver, an all-star in the world of baseball stats, may be the political arena's next big draw.* Newsweek. Retrieved July 25, 2022 from https://web.archive.org/web/20080619111251/http:/www.newsweek.com/id/140469

Rothman, K.J., Gallacher, J.E., & Hatch, E.E. (2013). Why representativeness should be avoided. *International Journal of Epidemiology, 42*(4), 1012-1014. https://doi.org/10.1093/ije/dys223

Shropshire, K., Hawdon, J., & Witte, J.C. (2009). Web-survey design: Balancing measurement, response, and topical interest. *Sociological Methods and Research, 37*(3), 344–370. https://doi.org/10.1177/0049124108327130

Silver, N. (March 1, 2008). *Frequently asked questions, last revised 8/7/08.* FiveThirtyEight. Retrieved July 24, 2022 from https://fivethirtyeight.com/features/frequently-asked-questions-last-revised/

Simmons, A.D. & Bobo, L.D. (2015). Can non-full-probability internet surveys yield useful data? A comparison with full-probability face-to-face surveys in the domain of race and social inequality attitudes. *Sociological Methodology, 45*(1), 357–387. https://doi.org/10.1177/0081175015570096

Steinmetz, K. F. (2015). Craft(y)ness: An ethnographic study of hacking. *The British Journal of Criminology, 55*(1), 125-145. https://doi.org/10.1093/bjc/azu061

Suresh, K., Thomas, S.V., & Suresh, G. (2011). Design, data analysis and sampling techniques for clinical research. *Annals of Indian Academy of Neurology, 14*(4), 287–290. https://doi.org/10.4103/0972-2327.91951

Szolnoki, G. & Hoffmann, D. (2013). Online, face-to-face and telephone surveys—Comparing different sampling methods in wine consumer research. *Wine Economics and Policy, 2*(2), 57-66. https://doi.org/10.1016/j.wep.2013.10.001.

Thompson, A. J., & Pickett, J. T. (2020). Are relational inferences from crowdsourced and opt-in samples generalizable? Comparing criminal justice attitudes in the GSS and five online samples. *Journal of Quantitative Criminology, 36*(4), 907–932. https://doi.org/10.1007/s10940-019-09436-7

Wang, W., Rothschild, D., Goel, S., & Gelman, A. (2015). Forecasting elections with non-representative polls. *International Journal of Forecasting, 31*(3), 980–991. https://doi.org/10.1016/j.ijforecast.2014.06.001

Weinberg, JD, Freese J, McElhattan D (2014). Comparing data characteristics and results of an online factorial survey between a population-based and a crowdsource-recruited sample. *Sociological Science, 1*(19), 292–310. http://dx.doi.org/10.15195/v1.a19

Witte, J.C., Pargas, R., Mobley, C., & Hawdon, J. (2004). The effects of images in web surveys: A research note. *Social Science Computer Review, 22*(3), 363-369. https://doi.org/10.1177/0894439304264531

Wright, M.F., Wachs, S., Yanagida, T., Sevcikova, A., Dedkova, L., Bayraktar, F., Aoyama, I., Kamble, S.V., Machackova, H., Li, Z., Soudi, S., Lei, L., & Shu, C. (2021). Associations between Severity and Attributions: Differences for Public and Private Face-to-face and

Cyber Victimization. *American Journal of Criminal Justice,* 46, 843–861 (2021). https://doi.org/10.1007/s12103-021-09660-7

Yeager, D. S., Krosnick, J. A., Chang, L., Javitz, H. S., Levendusky, M. S., 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*(4), 709–747. https://doi.org/10.1093/poq/nfr020