

AAPOR Report on Online Panels

Prepared for the AAPOR Executive Council by a Task Force operating under the auspices of the AAPOR Standards Committee, with members including:

Reg Baker, Market Strategies International and Task Force Chair

Stephen Blumberg, U.S. Centers for Disease Control and Prevention

J. Michael Brick, Westat

Mick P. Couper, University of Michigan

Melanie Courtright, DMS Insights

Mike Dennis, Knowledge Networks

Don Dillman, Washington State University

Martin R. Frankel, Baruch College, CUNY

Philip Garland, Survey Sampling International

Robert M. Groves, University of Michigan

Courtney Kennedy, University of Michigan

Jon Krosnick, Stanford University

Sunghee Lee, UCLA

Paul J. Lavrakas, Independent Consultant

Michael Link, The Nielsen Company

Linda Piekarski, Survey Sampling International

Kumar Rao, Gallup

Douglas Rivers, Stanford University

Randall K. Thomas, ICF International

Dan Zahs, Market Strategies International

MARCH, 2010

CONTENTS

1. Executive Summary	3
2. Background and Purpose of this Report	7
3. An Overview of Online Panels	9
4. Errors of Nonobservation in Online Panel Surveys	18
5. Measurement Error in Online Panel Surveys.....	29
6. Sample Adjustments to Reduce Error and Bias	42
7. The Industry-wide Focus on Panel Data Quality	48
8. Conclusions/Recommendations	52
9. References and Additional Readings	55
Appendix A: Portion of the CASRO Code of Standards and Ethics dealing with Internet Research..	68
Appendix B: ESOMAR 26 Questions to Help Research Buyers of Online Samples	74
Appendix C: AAPOR Statement - Web Surveys Unlikely to Represent All Views	79
Appendix D: AAPOR Statement - Opt-in Surveys and Margin of Error	81

1. EXECUTIVE SUMMARY

In September, 2008, the AAPOR Executive Council established an Opt-In Online Panel Task Force and charged it with “reviewing the current empirical findings related to opt-in online panels utilized for data collection and developing recommendations for AAPOR members.” The Council further specified that the charge did not include development of best practices, but rather would “provide key information and recommendations about whether and when opt-in panels might be best utilized and how best to judge their quality.” The Task Force was formed in October, 2008. This is its report.

TYPES OF ONLINE PANELS

A key first step was to distinguish among the various kinds of online panels based on their recruitment methods. The vast majority are not constructed using probability-based recruitment. Rather, they use a broad range of methods, most of which are online, to place offers to join in front of prospective panelists. Those offers are generally presented as opportunities to earn money but also emphasize the chance to have a voice in new products and services and the fun of taking surveys. People join by going to the panel company’s Web site and providing varying amounts of personal and demographic information that is later used to select panelists for specific surveys.

A few panels recruit their members using traditional probability-based methods such as RDD sampling. In cases where a sampled person may not have Internet access, the panel company might choose to provide access as a benefit of joining. Probability-based panels generally have many fewer members than the nonprobability panels that dominate online research.

A third type of online sample source is generally referred to as *river sampling*. In this approach respondents are recruited directly to specific surveys using methods similar to the way in which nonprobability panels are built. Once a respondent agrees to do a survey he or she answers a few qualification questions and then is routed to a waiting survey. Sometimes, but not always, these respondents are offered the opportunity to join an online panel

Because nonprobability panels account for the largest share of online research and because they represent a substantial departure from traditional methods the report’s overriding focus is on nonprobability panels.

TOTAL SURVEY ERROR

Early on the Task Force decided to conduct its evaluation from a Total Survey Error perspective. Not surprisingly, coverage error is a major factor when the research goal is to represent the general population. The best estimates of Internet access indicate that roughly one third of the U.S. adult population does not use the Internet on a regular basis. A few probability-based panels try to minimize undercoverage by recruiting via traditional methods (high-quality sample frames and telephone or face-to-face contact) and supplying Internet access to panel members who do not already have it. However, the majority of online panels rely almost completely on those who already are online. Thus, all

nonprobability online panels have inherent and significant coverage error, primarily in the form of undercoverage.

Although there is little hard data to go by, what little we do know suggests that there also is an extremely high level of nonresponse at the various stages of building a nonprobability panel and delivering respondents to individual studies. Panel companies continually display a good deal of creativity in placing offers to join panels across the Internet suggesting that a significant portion of Internet users are exposed to them. Yet, for example, even a relatively large U.S. panel of three million members has only about two percent of adult Internet users enrolled at any given time. Further, the response rates for surveys from nonprobability panels have fallen markedly over the last several years to a point where in many cases they are 10 percent or less. This combination of major undercoverage and high nonresponse presumably results in substantial bias in surveys using nonprobability panels, bias that thus far is not well understood in the literature.

A large number of studies have compared results from surveys using nonprobability panels with those using more traditional methods, most often telephone. These studies almost always find major differences. Those differences are sometimes attributed to the change in mode from interviewer administration to self-administration by computer. In many instances these explanations are conceptually grounded in the survey literature (e.g., social desirability or satisficing) and empirical testing often confirms that survey administration by computer elicits higher reports of socially undesirable behavior and less satisficing than interviewer administration. Unfortunately, the designs of most of these studies make it difficult to determine whether mode of administration or sample bias is greater cause of the differences. In those instances where comparisons to external benchmarks such as Census or administrative records are possible, the results suggest that studies using probability sampling methods continue to be more accurate than those using nonprobability methods.

One special case is electoral polling where studies using nonprobability panels sometimes have yielded results that are as accurate as or more accurate than some surveys using probability samples. However, these studies are especially difficult to evaluate because of the myriad of design choices (likely voter models, handling of item nonresponse, weighting, etc.) pollsters face, the proprietary character of some of those choices, and the idiosyncratic nature of the resulting surveys.

ADJUSTMENTS TO REDUCE BIAS

Researchers working with nonprobability panels generally agree that there are significant biases. Some attempt to correct bias through standard demographic weighting. Others use more sophisticated techniques either at the sample design stage or at the post-survey weighting stage. Simple purposive sampling that uses known information about panel members to generate demographically balanced samples is widely practiced, as is standard quota sampling. More sophisticated model-based and sample matching methods are sometimes used. These methods have been successfully used in other disciplines but have yet to be widely adopted by survey researchers.

Arguably the greatest amount of attention has focused on the use of propensity models in post-stratification adjustments. These models augment standard demographic weighting with attitudinal or behavioral measures thought to be predictors of bias. A probability-based reference survey typically is used to determine the magnitude of the adjustments. There is a growing literature aimed at evaluating and

refining these measures. That literature suggests that effective use of these techniques continues to face a number of unresolved challenges.

CONCERNS ABOUT PANEL DATA QUALITY

Over about the last five years, market researchers working extensively with nonprobability online panel sample sources have voiced a number of concerns about panel data quality. These concerns arose out of increasing evidence that some panelists were completing large numbers of surveys, that some were answering screener questions in ways to maximize their chances of qualifying, and that there were sometimes alarming levels of satisficing.

The industry's response has come at three levels. First, industry and professional associations worldwide have stepped up their efforts to promote online data quality through a still evolving set of guidelines and standards. Second, panel companies are actively designing new programs and procedures aimed at validating panelists more carefully and eliminating duplicate members or false identities from their databases. Finally, researchers are doing more research to understand the drivers of panelist behaviors and to design techniques to reduce the impact of those behaviors on survey results.

CONCLUSIONS

The Task Force's review has led us to a number of conclusions and recommendations:

- Researchers should avoid nonprobability online panels when one of the research objectives is to accurately estimate population values. There currently is no generally accepted theoretical basis from which to claim that survey results using samples from nonprobability online panels are projectable to the general population. Thus, claims of "representativeness" should be avoided when using these sample sources.
- The majority of studies comparing results from surveys using nonprobability online panels with those using probability-based methods (most often RDD telephone) often report significantly different results on a wide array of behaviors and attitudes. The degree to which those differences might be due to mode effects versus the nonprobability character of online panels is a matter of ongoing debate. The few studies that have disentangled mode of administration from sample source indicate that nonprobability samples are generally less accurate than probability samples.
- There are times when a nonprobability online panel is an appropriate choice. Not all research is intended to produce precise estimates of population values and so there may be survey purposes and topics where the generally lower cost and unique properties of Web data collection is an acceptable alternative to traditional probability-based methods.
- Research aimed at evaluating and testing techniques used in other disciplines to make population inferences from nonprobability samples is interesting and valuable. It should continue.

- Users of online panels should understand that there are significant differences in the composition and practices of individual panels that can affect survey results. Researchers should choose the panels they use carefully.
- Panel companies can inform the public debate considerably by sharing more about their methods and data describing outcomes at the recruitment, enrollment, and survey-specific stages.
- Full and complete disclosure of how results were obtained is essential. It is the only means by which the quality of research can be judged and results replicated.
- AAPOR should consider producing its own “Guidelines for Internet Research” or incorporate more specific references to online research in its code. Its members and the industry at large also would benefit from a single set of guidelines that describe what AAPOR believes to be appropriate practices when conducting research online across the variety of sample sources now available.
- There are no widely-accepted definitions of outcomes and methods for calculation of rates similar to AAPOR’s *Standard Definitions* (2009) that allow us to judge the quality of results from surveys using online panels. AAPOR should consider revising *Standard Definitions* accordingly.
- Research should continue. AAPOR, by virtue of its scientific orientation and the methodological focus of its members is uniquely positioned to encourage research and disseminate its findings. It should do so deliberately.

2. BACKGROUND AND PURPOSE OF THIS REPORT

The dramatic growth of online survey research is one of the most compelling stories of the last decade. Virtually nonexistent just 10 years ago, *Inside Research* (2009) estimates the total spend on online research in 2009 at about \$2 billion, the vast majority of which is supported by online panels. About 85 percent of that research replaces research that previously would have been done with traditional methods, principally by telephone or face-to-face. The vast majority of this research is in market research applications such as product testing, sales tracking, advertising and brand tracking, and customer satisfaction. Political polling has also become a very visible online application, although its overall share of online research is quite small (2 percent or less according to *Inside Research*). The rapid rise of online has been partly due to its generally lower cost and faster survey turnaround time, but also to rapidly escalating costs, increasing nonresponse, and, more recently, concerns about coverage in other modes.

Later in this report we describe the types of online panels being used. We distinguish between two types: (1) those recruited by probability based methods and (2) those taking a nonprobability approach. The former use random sampling methods such as RDD or area probability. They also use traditional methods such as telephone or face-to-face to recruit people to join panels and agree to do future studies. In this report we generally refer to these as *probability-based panels*. A second type of panel mostly relies on a nonprobability approach and uses a wide variety of methods (e.g., Web site banner ads, email, and direct mail) to make people aware of the opportunity in the hope that they elect to join the panel and participate in surveys. We generally refer to these as *nonprobability* or *volunteer online panels*. These sometimes are called *opt-in panels* in the literature, although that is potentially confusing since all panels, regardless of how they are recruited, require that a respondent opt in, that is, agree to participate in future surveys. The term *access panel* is also sometimes used as a way to describe *nonprobability volunteer online panels* although it is not used in this report.

Although both probability and nonprobability panels are discussed in this report, the overwhelming emphasis is on the latter. The approaches that have developed over the last decade to build, use, and maintain these panels are distinctly different from the probability-based methods traditionally used by survey researchers and therefore are most in need of detailed evaluation of those factors that may affect the reliability and validity of their results.

This report also has a U.S. focus. While online panels are now a global phenomenon, U.S. companies have been especially aggressive in developing the techniques for building them and using them for all kinds of research. Over about the last five years the amount of online research with panels has increased markedly in Europe and from time to time in this report we reference some especially useful European studies. Nonetheless, this report is primarily concerned with the pros and cons of online panels in the U.S. setting.

From the beginning, much of the scientific survey community has viewed online research with nonprobability online panels with skepticism. This has been especially true in academia and in organizations doing a substantial amount of government-funded work. The use of representative random samples of a larger population has been an established practice for valid survey research for over 50

years. The nonprobability character of volunteer online panels runs counter to this practice and violates the underlying principles of probability theory. Given this history, the reluctance of many practitioners in academia, government, and even parts of commercial research to embrace online is understandable.

But time marches on and the forces that created the opportunity for online research to gain traction so quickly—increasing nonresponse in traditional methods, rising costs and shrinking budgets, dramatic increases in Internet penetration, the opportunities in questionnaire design on the Web, and the lower cost and shorter cycle times of online surveys—continue to increase the pressure on all segments of the survey industry to adopt online research methods.

This report is a response to that pressure. It has a number of objectives:

1. To educate the AAPOR membership about how online panels of all kinds are constructed and managed.
2. To evaluate online panels from the traditional Total Survey Error perspective.
3. To describe the application of some newer techniques for working with nonprobability samples.
4. To review the empirical literature comparing online research using nonprobability volunteer online panels to traditional methods.
5. To provide guidance to researchers wishing to understand the tradeoffs involved when choosing between a nonprobability online panel and a traditional probability-based sample.

Finally, even though online research with panels has been adopted on a broad scale, it is by no means a mature methodology. The methods and techniques for creating survey samples online continue to evolve. Panels may prove to be only the first stage of this development. Researchers increasingly look to deeper and more sustainable sources such as expanded river sampling, social networks, and even “offline” sources such as mobile phones. Blending multiple panel sources into a single sample is being practiced on a wider scale.

At the same time, there is arguably more methodological research about online surveys being executed and published today than at any time since its introduction in the mid-1990s. Although there was a good deal of such research done in the commercial sector that work has generally not found its way into peer-reviewed journals. More recently, academic researchers have begun to focus on online surveys and that research is being published.

Despite this activity, a great deal still needs to be done and learned. We hope that this report will introduce the key issues more broadly across the industry and in doing so stimulate additional research.

3. AN OVERVIEW OF ONLINE PANELS

One of the first challenges the researcher encounters in all survey modes is the development of a sample frame for the population of interest. In the case of online, survey researchers face the additional challenge of ensuring that sample members have access to the mode of questionnaire administration, that is, the Internet. For some potential populations of interest, almost all members are online (e.g., software developers, college students, and business executives). But if the research requires a general population sample of the US, not everyone is online. Estimates of Internet use and penetration in the U.S. household population can vary widely. Arguably the most accurate are those collected face-to-face by the Current Population Survey (CPS). The CPS reports that as of October, 2009, 69 percent of U.S. households had an Internet connection while 77 percent had household members who reported that they connected to the Internet from home or some other location such as their workplace (Current Population Survey, 2009). This comports with the most recent data from the Pew Research Center showing that, as of December of 2009, 74 percent of U.S. adults use the Internet either at home or some other location (Rainie, 2010). However, the Pew data also report that only 72 percent of Internet users actually go online at least once a week. Further, Internet access tends to be positively associated with income and education and negatively associated with age (younger people are more likely to be online than older people). Some demographic groups are also less likely to be online (e.g., blacks, Hispanics, and undocumented immigrants).

There also is no comprehensive list of those who are online. While virtually all Internet users have an email address, no complete list of these addresses exists. Individuals may have several email addresses or they may share a single address with other family members. Addresses may fall into disuse without being deactivated. In addition, the non-standardized format of e-mail addresses precludes the generation of RDD-like methods for Web surveys.

Even if such a comprehensive list existed there are both legal prohibitions and industry practices that discourage the kind of mass emailing that might be akin to the calling we do for an RDD telephone survey. The CAN-SPAM Act of 2003 established clear guidelines in the U. S. around the use of email addresses in terms of format, content, and process. Internet Service Providers (ISPs) generally take a dim view of mass emails and will sometimes block suspected spammers from sending email to their subscribers. The Council of American Survey Research Organizations (CASRO) has established the standard for its members that all email communications must be “permission-based.” Specifically, the CASRO Code of Standards and Ethics for Survey Research (2009) requires that its members only mail to potential respondents with whom the research organization or its client has a pre-existing relationship. Examples include individuals who have previously agreed to receive email communications either from the client or the research organization or customers of the client.

There are many specialized populations for which a full list of email addresses might be available and usable. Some examples include members of an organization (e.g., employees of a company or students at a university), users of a particular Web site, or customers of an online merchant. These circumstances are now relatively common. However, gaining access to representative samples of the general population for online research continues to be problematic.

Online panels have become a popular solution to the sample frame problem for those instances in which there is no usable and complete list of email addresses for the target population. For purposes of

this report we use the definition of online panel from ISO 26362: Access Panels in Market, Opinion, and Social Research. It reads: “A sample database of potential respondents who declare that they will cooperate with future [online] data collection if selected” (International Organization for Standardization 2009). As noted in Section 2 of this report, we further distinguish between *nonprobability-based panels* and *probability-based panels*. The former are panels constructed through the widespread placement of offers to join for the purpose of participating in future surveys. Anyone seeing the panel offering may choose to join and become a panel member provided they meet requirements specified by the panel builder. Probability-based panels select potential members in advance from a sampling frame of the target population and then attempt to recruit only those sampled individuals to join the panel and participate in future surveys, sometimes providing Internet access to those sampled members without access at the time they are contacted.

Probably the most familiar type of online panel is a *general population panel*. These panels typically include hundreds of thousands to several million members and are used for both general population studies as well as for reaching respondents with low incidence events or characteristics (e.g., owners of luxury vehicles or people suffering from Stage II pancreatic cancer). The panel serves as a frame from which samples are drawn to meet the specific needs of particular studies. The design of these study-specific samples may vary depending on the survey topic and population of interest.

Census-balanced samples are designed to reflect the basic demographics of the U.S. population (and the target proportions could be based on distributions as they occur in the larger population for some combination of sex, age, region of the country, income, education, ethnicity, or other relevant demographic characteristics). Another common example of an online panel is a *specialty panel*. A specialty panel could be a group of people who are selected because they own certain products (e.g., big screen TVs), are a specific demographic group (e.g., Spanish-speaking Hispanics in the U.S.), are in a specific profession (e.g., physicians), engage in certain behaviors (e.g., watch sports), hold certain attitudes or beliefs (e.g., identify as Republicans), or are customers of a particular company (e.g., buy Dr. Pepper).

A *proprietary panel* is a type of specialty panel in which the members of the panel participate in research for a particular company (e.g., a vehicle manufacturer gathers email addresses of high-end vehicle owners who have volunteered to take surveys about vehicles).

Targeted samples may select panel members who have characteristics of specific interest to a researcher such as auto owners, specific occupational groups, persons suffering specific diseases, or households with children (often this information has been gathered in prior surveys). In addition, regardless of how the sample is selected, individual surveys may include specific screening criteria to further select for low incidence populations.

In the following sections we describe the most commonly-used approaches to building, managing, and maintaining online panels. While Couper (2001) has offered a comprehensive typology of Internet samples, this report focuses primarily on online panels. Other online sample sources such as customer lists or Web options in mixed-mode surveys are not discussed. Because the vast majority of panels do not rely on probability-based methods for recruitment we discuss those first. We then describe the probability-based model. We conclude with a discussion of *river sampling*. Although this method does not involve panel development per se, it has become an increasingly popular technique for developing online samples.

NONPROBABILITY/VOLUNTEER ONLINE PANELS

The nonprobability volunteer online panel concept has its origins in the earlier mail panels developed by a number of market research companies, including Market Facts (now Synovate), NOP, and NFO. These panels generally had the same nonprobability design as contemporary online panels and were recruited in much the same way only relying on offline sources. While many of these panels originally were built to support syndicated research¹, they came to be widely used for custom research² as well. Their advantages for researchers were much the same as those touted for online panels: (1) lower cost; (2) faster response; and (3) the ability to build targeted samples of people who would be low incidence in a general population sample (Blankenship, Breen, and Dutka, 1998).

Today, companies build and manage their online panels in a number of different ways and draw on a wide variety of sources. There is no generally-accepted best method for building a panel and many companies protect the proprietary specifics of their methods with the belief that this gives them a competitive advantage. There are few published sources (see, for example, Miller, 2006; or Comley, 2007) to turn to and so in this section we also rely on a variety of informal sources, including information obtained in RFPs, technical appendices in research reports, and informal conversations with panel company personnel.

Overall we can generalize to five major areas of activity: (1) recruitment of members; (2) joining procedures and profiling; (3) specific study sampling; (4) incentive programs; and (5) panel maintenance.

RECRUITMENT OF MEMBERS. Nonprobability-based panels recruit their members in a variety of ways, but ultimately all involve a voluntary self-selection process on the part of the person wanting to become a member. Panel companies all try to put the invitation to join in front of as many people as possible. The choice of where and how to do this is guided by a combination of cost effectiveness and the desired demographic and behavioral characteristics of recruits. Both online and offline methods may be used. For specialty panel recruitment, this may mean a very selective recruitment campaign (e.g., involving contact with specific organizations or ad buying for specific Web sites, magazines, or TV programs that are targeted to draw people with the characteristics of interest). Regardless of the medium, the recruitment campaign typically appeals to some combination of the following motivations to complete surveys:

1. A contingent incentive, either fixed (money or points received for each completed survey that can be redeemed after a certain number of completed surveys) or variable (sweepstakes with the potential to win prizes for each completed survey);
2. Self-expression (the importance of expressing/registering one's opinions);
3. Fun (the entertainment value of taking surveys);
4. Social comparison (the opportunity to find out what other people think); and
5. Convenience (the ease of joining and participating).

¹ Also sometimes referred to as "multi-client research," this is a market research product that focuses on a specific topic or population but is conducted without a specific client. Instead, the company designs and conducts the research on its own and then sells it to a broad set of clients.

² Research done on a specific population and topic and sponsored by a client.

Although it is widely assumed that earning incentives is the primary motive for joining a panel, there are a few studies to help us understand which motivations are most prominent. Poynter and Comley (2003) report a mix of respondent motives in their study, with incentives topping the list (59 percent) but significant numbers reporting other factors such as curiosity (42 percent), enjoying doing surveys (40 percent), and wanting to have their views heard (28 percent). In a follow-on study, Comely (2005) used results from an online study of panelist motivation to assign respondents to one of four segments: the “Opinionated” (35 percent) who want to have their views heard and enjoy doing surveys; “Professionals” (30 percent) who do lots of surveys and generally will not respond unless there is an incentive; the “Incentivized” (20 percent) who are attracted by incentives but will sometimes respond when there isn’t one; and “Helpers” (15 percent) who enjoy doing surveys and like being part of the online community.

One very popular method for online panel development is through co-registration agreements. Many sites compile email databases of their visitors through a voluntary sign-up process. Portals, e-commerce sites, news sites, special interest sites, and social networks are all examples of sites with a large volume of traffic that the site owner might choose to “monetize” by presenting offers that include those to join a research panel (meaning that the site receives some financial compensation for each person recruited from its site). As a visitor registers with the site, he or she may be offered the opportunity to join other “partner” company databases. The invitation to join additional email lists often asks the user if he or she is interested in receiving special offers or information from partner companies. Panel companies purchase these email address lists from the site owner and contact individuals with offers to join their panel, or users may be taken immediately to the panel company’s registration page. A classic example of this approach is the original panel developed by Harris Black International. This panel was initially recruited through a co-registration agreement with the Excite online portal (Black and Terhanian, 1998).

Another commonly used approach is the use of “affiliate hubs.” These are sites that offer access to a number of different online merchants. A visitor who buys merchandise or services from a listed merchant through the hub receives points that can then be redeemed for merchandise at the hub. (One example typical of such hubs is www.mypoints.com.) Panel companies will sometimes post their offers on these hubs alongside those of various online merchants. Interested visitors can click through from the hub to panel’s registration page. In addition, hubs often conduct email campaigns with registered visitors advertising opportunities to earn points. These might include survey initiations in which potential participants are offered incentives in the site’s currency.

Panel companies may also recruit online via display ads or banners placed across a variety of sites. The panel company generally will not place these ads directly. Rather, the company buys placements from one of the major online advertising companies who in turn place ads where they expect to get the best return, that is, click-throughs to the panel’s registration page.

Still another method relies on search engines. A panel company may buy text ads to appear alongside search engine results with the expectation that some visitors will see the ad and click through to join the panel. These frequently are tied to the use of specific search terms such as “survey” or “market research.” Search for “survey research” on search engines like Yahoo, Google, or Bing and you likely will see at least one offer on to join a panel on the advertisement section of the search results page.

Though not used by any reputable company in the U.S., other email recruitment tactics include blasting or spamming methods – sending unsolicited commercial email in mass quantities. Sending these

requests en masse creates the potential for violation of the CAN-SPAM Act of 2003 and may result in ISP blacklisting, which blocks the sender's email systems from sending out any further emails.

Finally, as has been often done in research in other modes under the names of snowball recruiting or viral recruiting, some panel companies encourage their members to recruit friends and relatives. These programs often offer the member a per-recruit reward for each new member recruited.

There is an equally wide variety of methods for offline recruitment. Many full-service research companies conduct surveys in a variety of modes (including paper-pencil and RDD telephone research as well as online surveys). Sometimes they will field a survey that has as its primary purpose the recruitment of panel members (known as *purposeful recruitment*). Other times, they may field a study that will recruit panel members as a byproduct of their traditional research activities (known as *incidental recruitment*). For example, they might routinely ask a respondent at the end of a RDD telephone survey whether he or she is interested in joining a panel to take future surveys online. Or, a company might borrow a common technique from telemarketers and use an IVR system to autodial numbers and invite people to join a panel. In the event the contacted person is interested, the system asks for a name and email address for future contact that can then be used to send an email invitation to join the panel.

Companies also may include invitations to join a panel in offline marketing or advertising campaigns, such as messaging on cash register receipts, event tickets, monthly account statements, and various other traditional media. Sometimes surveys are conducted for a specific company (with invitations to the survey on cash register receipts), and upon completion of the survey, respondents are offered the opportunity to join a panel (whether general population or specialty panel). Merchants who allow panel companies to do this often get a discount on their research costs for each recruited respondent. Yet another source for panel members occurs with some companies that are so large and feature such diverse products that they convert their databases into panels and develop a supplemental line of business.

No two panels recruit their members with the same mix of techniques. Individual panels set targets for the specific populations and target respondents they feel will represent the most compelling value proposition to research clients and design their recruitment strategies appropriately. However, the variation in recruitment sources and methods by the various companies can lead to differences in research findings, something we address in Section 4 on errors of nonobservation.

Panel companies rarely disclose the success rates from their recruitment strategies. One exception is a study by Alvarez, , and Van Beselaere (2003) based on a project conducted by a team of academics who built an online panel for the Internet Surveys of American Opinion project. Through a co-registration agreement with ValueClick, Inc., an online marketing company with access to a wide variety of Web sites, Internet users who were registering for various services were provided a check box on their registration form to indicate their interest in participating in Web surveys. Some 21,378 Web users (out of an unknown total number of people passing through the targeted registration sites) checked the box. Among this group, 6,789 completed the follow-up profile and were enrolled in the panel (a 32 percent yield).

Alvarez et al. also collected data on the effectiveness of banner ads. Their banner ad was displayed over 17 million times, resulting in 53,285 clicks directing respondents to the panel Web site, and ultimately 3,431 panel members. The percentage yield of panel members per click was 6.4 percent, and the percentage member yield per banner display (a.k.a. impression) was 0.02 percent.

JOINING PROCEDURES AND PROFILING. Joining a panel is typically a two-step process. At a minimum, most reputable research companies in the U.S. follow what is called a *double opt-in* process whereby a person first indicates his/her interest in joining the panel (either signing up or checking a box on a co-registration site). The panel company then sends an email to the listed address and the person must take a positive action indicating the intent to join the panel. At this second stage of confirmation some panel companies will ask the new member to complete a profiling survey that collects a wide variety of background, demographic, psychographic, attitudinal, experiential, and behavioral data that can be used later to select panelist for specific studies. This double opt-in process is required by ISO 26362 (International Organization for Standardization, 2009) and the industry has come to accept it as defining the difference between a panel and simply a database of email addresses. This international standard was released in 2009 as a supplement to the previously-released ISO 20252 – Market, Opinion, and Social Research (International Organization for Standardization, 2006). ISO 26362 specifies a vocabulary and set of service quality standards for online panels.

Upon agreeing to join, panelists typically are assigned a unique identification number used to track the panelist throughout their lifetime on the panel. Panel companies then assign respondents to their correct geography and demographic groups so that they can provide samples by DMA, MSA, ZIP code, and other geographic identifiers.

As part of this initial recruitment most panel companies now have validation procedures to ensure that individuals are who they say they are and are allowed to join the panel only once. Checks at the joining stage may include: verification against third party databases, e-mail address validity (via format checks and checks against known ISPs), postal address validity (via checks against postal records), “reasonableness” tests done via data mining (appropriate age compared to age of children, income compared to profession, etc.), duplication checks, or digital fingerprint checks to prevent duplication by IP address and ensure correct geographic identification.

Panel companies vary in the type of information they collect in their respondent profiles. A company may try to create a panel database with hundreds of variables so that they can select efficient, targeted samples. Profiles might include shopping habits, product brand ownership, health ailments, occupation, and household roles, to name a few. Companies then price their services based on the incidence of the specific groups of interest to sample buyers. Panel companies typically refresh these profiles on a regular basis, sometimes simply to update and other times to add still more information about panelists. They also may capture screening information from individual surveys and incorporate those data into member profiles.

SPECIFIC STUDY SAMPLING. Simple random samples from panels are rare because of the tendency for them to be highly skewed toward certain demographic characteristics (e.g., just as older people are more likely to answer landline phones, younger people are more likely to be online and respond to an email survey invitation). *Purposive sampling* (discussed in more detail in Section 6) is the norm and the sampling specifications used to develop these samples can be very detailed and complex, including not just demographic characteristics but also specific behaviors or even previously-expressed positions for specific attitudes.

Although the empirical evidence is mixed, some researchers believe that having respondents take too many surveys in too short of a time interval can yield results that are less than optimal (Garland, Santus, and Uppal, 2009). To prevent some respondents from “over-participating” and dominating results, some

panel suppliers include so-called “lockout” criteria that prevent some panel members from being sampled. These lockouts might specify that no one who has completed a survey on the same topic during a specific time frame can be sampled for the current study (e.g. cannot participate in another political survey for three months). Or the panel company may limit the frequency with which a member can be solicited for a survey (e.g., they will not invite any member to a new survey for at least 10 days).

With the sample drawn, the panel company sends an email invitation to the sampled member. The content of these emails varies widely and generally reflects the information the panel company’s client wishes to put in front of the respondent to encourage participation. At a minimum the email will include the link to the survey and a description of the incentive. It might also specify a closing date for the survey, the name of the organization conducting or sponsoring the survey, an estimate of the survey length, and even a description of the survey topic.

INCENTIVE PROGRAMS. To combat panel attrition, and to increase the likelihood that panel members will complete a survey, panelists are typically offered compensation of some form. These include cash, points redeemed for various goods (e.g., music downloads, airline miles, etc.), sweepstakes drawings, or instant win games. Incentive programs vary from company to company but the amount of incentive is typically tied to the length or topic of the survey. Large incentives may be used when the expected incidence of qualifiers is low. These incentives are paid contingent on completion, although some panels also pay out partial incentives when a member starts a survey but fails to qualify.

PANEL MAINTENANCE. Once people have been recruited to join a panel, the challenge is to keep them active. People who join expect to take surveys. Each panel company has its own definition as to what it deems as an active panelist. However, nearly all of the definitions are based on a calculation that balances the date a person joined and the number of surveys taken in a specified time period. ISO 26362 defines an active member as a panelist who either has participated in at least one survey or updated his/her profile in the last year. As such, the claims of panel size depend greatly on value of the variables in the algorithm used to calculate the status of a member.

Most panel companies have a multi-faceted approach to maintain a clean panel comprised of members whose information is current and who can be contacted successfully to take a survey. These “hygiene” procedures include treatment plans for undeliverable e-mail addresses, “mailbox full” status, syntactically undeliverable email addresses, non-responding panelists, panelists with missing data, panelists who repeatedly provide bad data, and duplicate panelists. In addition, panel companies may expend considerable effort to maintain deliverability of panelist email addresses via white-listing³ with major ISPs.

Attrition of panel members is a natural process with the levels varying considerably from one panel to the next. As occurs with most joining activities, the greatest attrition most likely occurs among the newest members. Attrition can come from multiple sources – people change email addresses but do not update with the panel company, they can drop out due to long or boring surveys, they switch panels because they may have better rewards or more interesting or shorter surveys, etc. Once a panel has been established, panel recruitment becomes an essential ongoing activity to ensure a sufficient number of respondents are available for ongoing projects.

³ An email white list is a list of email addresses from which an individual or an ISP will accept email messages.

In light of attrition and underutilized panelists, firms attempt to reengage people who appear to be reluctant to participate in surveys and those who have left the panel altogether. But there appears to be nothing in the research literature reporting the effectiveness of such approaches.

PROBABILITY-BASED RECRUITMENT

Despite the early emergence of the Dutch Telepanel in 1986 (Saris 1998), online panels that recruit using traditional probability-based methods have been slow to appear and are fewer in number than volunteer panels, although they are now gaining in prevalence. These panels follow roughly the same process as that described for volunteer panels with the exception that the initial contact with a potential member is based on a probability design such as RDD or area probability. To account for the fact that not everyone in the sample may have Internet access, some of these panels provide the necessary computer hardware and Internet access or may conduct surveys with the panels using a mix of modes (Web, telephone, mail, IVR, etc.). In one study of the four stages of the recruitment process (recruitment, joining, study selection, participation), Hoogendoorn and Daalmans (2009) report that there is differential non-response and engagement at each stage for a probability panel associated with a number of demographic variables, including age and income. The effects of this self-selection on population estimates for survey measures in subsequent surveys remain unexplored.

Aside from the key differences in sampling and provision of Internet access to those who are not already online, probability-based panels are built and maintained in much the same way as nonprobability panels (Callegaro and DiSogra, 2008). Once contacted, a respondent can indicate his/her willingness to join the panel in a number of ways – by agreeing in the telephone interview, calling a toll-free number, completing and returning the mail solicitation, or going to the registration page of the panel's Web site. Once registered, the panelist is asked to complete an online profile questionnaire much like those used by panel companies generally. Sampling methods and incentive structure vary depending on individual study requirements. As with any panel, attrition creates the need for ongoing recruitment and rebalancing. In addition, because the cost of acquisition of panel members is higher than it is for nonprobability panels, probability panels may require completion of a minimum number of surveys on a regular basis to remain in the panel.

The combination of the cost of recruitment and the requirement of some panels to provide Internet access for those who are not already online generally translates to these panels being more expensive to build and maintain than those that use nonprobability methods. As a consequence, their members may number in the tens of thousands rather than the millions often claimed by volunteer panels. Studies executed with these panels tend toward smaller sample sizes, although, depending on the target respondent, studies with large sample sizes are possible. It also can be difficult to get large sample sizes for low-incidence populations or smaller geographic areas unless the panel was designed with these criteria in mind. Nonetheless, probability-based panels are attractive to researchers who require general population samples and a basis in probability theory to ensure their representativeness.

RIVER SAMPLING

River sampling is an online sampling method that recruits respondents when they are online and may or may not involve panel construction. Sometimes referred to as *intercept interviewing* or *real-time sampling*, river sampling most often will present a survey invitation to a site visitor while he/she is engaged in some other online activity.⁴ In the vast majority of river sampling applications, the target population is much broader than visitors to a single site and determining how many and on which Web sites to place the invitation is a complex task. Knowledge about each site's audience and the response patterns of their visitors is a key piece of information needed for effective recruiting. Companies that do river sampling seldom have access to the full range of sites they need or the detailed demographic information on those sites' visitors, and so they work through intermediaries. Companies such as Digitas and DoubleClick serve up advertising across the Internet and also will serve up survey invitations. ISPs such as AOL or Comcast might also be used.

Once the sites are selected, promotions and messages of various types are randomly placed within those sites. The invitation may appear in many forms, including within a page via a randomized banner, an *n*-th user pop-up, or pop-under page. In many cases, these messages are designed to recruit respondents for a number of active surveys rather than a single survey. Visitors who click through are asked to complete a short profile survey that collects basic demographics and any behavioral data that may be needed to qualify the respondent for one or more of the active surveys. Once qualified, the respondent is assigned to one of the open surveys via a process sometimes referred to as *routing*.

Upon completion of the survey respondents are generally, but not always, rewarded for their participation. River sampling incentives are the same as those offered to online panel members. They include cash, PayPal reimbursements, online merchant gift codes and redeemable points, frequent flyer miles, and deposits to credit card accounts.

There are some indications that river sampling may be on the rise as researchers seek larger and more diverse sample pools and less-frequently surveyed respondents than those provided by online panels.

⁴ Many Web sites use pop-up questionnaires to survey visitors to ask them about their own Web site (e.g., are the features easy to access, did they obtain the information they were looking for, etc.) and these surveys are known as *Web site evaluations*. Distinct from Web site evaluations, the invitations to a river sample direct the visitor away from the originating Web site to a survey that is not about the Web site or Web site company.

4. ERRORS OF NONOBSERVATION IN ONLINE PANEL SURVEYS

Any estimate of interest, whether a mean, a proportion, or a regression coefficient, is affected by sample-to-sample variations. These variations lead to some amount of imprecision (variance) in estimating the true parameters and this type of error is known as sampling error. In addition to sampling error, error of estimates (bias and variance) can result from the exclusion relevant types of respondents, a type of error known as nonobservation. Two types of nonobservation error, coverage and nonresponse, affect all modes of surveys, regardless of sample source, but have the potential to be more severe with online panels than with other types of surveys.

BASIC CONCEPTS REGARDING COVERAGE ISSUES IN ONLINE PANELS

The *target population* is the group of elements which the survey investigator wants to describe using the sample statistics. There are three important features of most target populations for sample surveys:

- Target populations are finite in size (i.e., at least theoretically, they can be counted).
- They have some time restrictions (i.e., they exist within a specified time frame).
- They are observable (i.e., they can be accessed).

For any survey, clearly specifying these aspects of target populations is desirable for clarity of the survey's purpose and for "replicability" of the survey. With regard to surveys using online panels, one potential target population is the full set of persons living in housing units in the United States during a specific time period. This target population would resemble that often chosen when other survey methods are used, such as RDD. In RDD, the target population might be altered to a smaller subset – those having access to a telephone (or more restrictively, access to a landline telephone). In Web surveys, the target population might be restricted to those having access to the Internet in their homes, at work, or at school.

A *housing unit* is a house, an apartment, a mobile home, a group of rooms, or a single room that is occupied (or if vacant, is intended for occupancy) as separate living quarters. Separate living quarters are those in which the occupants live and eat separately from any other people in the building and have direct access from the outside of the building or through a common hall. The occupants may be a single family, one person living alone, two or more families living together, or any other group of related or unrelated persons who share living arrangements. Not all people in the United States at any moment are adults; not all adults reside in housing units (e.g., some live in dormitories, prisons, long-term care medical facilities, or military barracks; still others are homeless). Online panels might choose a target population that includes those in such institutions.

Since the population changes over time, the time of the survey also defines the target population. Since many household surveys are conducted over a period of several days, weeks, or even months, and since the population is changing daily as persons move in and out of U.S. households, the target population of many household surveys is the set of persons in the household population during the survey

period. In practice, the members of households typically are “fixed” at the time of first contact in many surveys.

A *sampling frame* is a set of materials used to identify the elements of the target population. Sampling frames are lists or procedures intended to identify all elements of a target population. The frames may be maps of areas in which elements can be found, time periods during which target events would occur, or records in filing cabinets, among others. Sampling frames, at their simplest, consist of a list of population elements. As previously noted in Section 3, a sampling frame consisting of all Internet users or containing all possible U.S. email addresses does not exist.

All sampling frames have various features, each of which can affect the quality of sample survey estimates based on them: (1) undercoverage; (2) multiple mappings; and (3) duplication. *Coverage error* arises when such features produce systematic under- or over-estimation of target population parameters, or act to inflate the standard errors of estimates.

Undercoverage is the weakness of sampling frames prompting the greatest possibility of coverage error. It threatens to produce errors of nonobservation in survey statistics from failure to include parts of the target population in any survey using the frame. By definition, online panels seeking to represent the total U.S. population suffer from undercoverage because non-Internet users are not members. This is akin to the problem posed by cell-phone-only households for landline-only telephone surveys: Persons with only cell phones are not members of the landline frame (AAPOR, 2008).

Multiple mappings of frame to population (clustering) or population to frame (duplication) are problems in sample selection. For example, a telephone directory lists telephone households in order by surname, given name, and address. When sampling adults from this frame, an immediately obvious problem is the clustering of eligible persons that occurs. Clustering means that multiple elements (e.g., people) of the target population are represented by the same frame element (e.g., telephone number). A residential telephone listing in the directory may have a single adult or two or more adults living there. The same is true of email addresses which can be shared by persons (e.g., smithfamily257@comcast.net). And so there can be ambiguity about which person connected to that address is being asked to participate. However, many online panels, as a matter of policy, insist that each member have his or her own email address, unshared with another person.

Duplication means that a single target population element (e.g., a person) is associated with multiple frame elements (e.g., multiple telephone numbers – business, cell, home, etc.). The problem that arises with this kind of frame problem is similar to that encountered with clustering. Target population elements with multiple frame units have higher chances of selection and will be overrepresented in the sample, relative to the population. If there is a correlation between duplication and the survey variables of interest, then the estimates of the survey measures will be biased. In survey measure estimation, the problem is that both the presence of duplication and the correlation between duplication and survey variables are often unknown. In the case of online panels, different email addresses may be held by the same person. While panel companies attempt to enforce a one email/one member rule at the joining stage, the effectiveness of those measures is unclear. In addition, it is not unusual for a member of one panel to also be a member of other panels. Surveys that draw on multiple panels therefore run the risk of sampling the same person multiple times. A recent study by The Advertising Research Foundation that conducted the same survey across 17 different panel companies found a duplication rate of 40 percent or 16 percent, depending on how it is measured (Walker, Pettit, and Rubinson, 2009). In general, the more

that large samples from different panels are combined for a given study, the greater the risk of respondent duplication.

ONLINE PANEL SURVEYS, FRAMES, AND COVERAGE ISSUES

Now we can apply more specifically the general notions of target populations, sampling frames, and coverage issues to online panel surveys. As previously noted, there is no full and complete list of email addresses that can be used as a sampling frame for general population Web surveys. Even if such a frame existed, it would fail to cover a significant portion of the U.S. adult population, a common target population of interest to commercial and social researchers. It would have duplication problems because a person can have more than one email address. It would have clustering problems because more than one person can share an email address.

As described in Section 3, volunteer panels do not attempt to build a complete sampling frame of email addresses connected to persons. Their approach differs from classic sampling in a number of ways.

First, often the entire notion of a sample frame is skipped. Instead, the panel company focuses on the recruitment and sampling steps. Persons with Internet access are solicited in a wide variety of ways to acquire as diverse and large a group as possible.

Second, a common evaluative criterion of a volunteer panel is not full coverage of the household population but the collection of a set of persons with sufficient diversity on attributes related to the type of surveys the panel supports. For a panel built to support research on household products, this might include diversity on income, household size, ethnic sub-cultural identity, age, and various other lifestyle attributes. For a panel focused on public opinion, the diversity of interest might be income, age, gender, ethnicity, party identification, and social engagement. The only way to evaluate whether the panel has the desired diversity is to compare the assembly of volunteers to the full target population. To do this most thoroughly would require census data on all variables, and an assessment of means, variances, and covariances for all combinations of variables. Even then, there is no statistical theory that would offer assurance that some other variable not assessed would have the desired diversity.

Third, within the constraints of lockout periods, online panels can repeatedly sample from the same set of assembled willing survey participants and will send survey solicitations to a member as long as the member responds. There generally is little attention to systematically reflecting dynamic change in the full population of person-level email addresses. To date, the practice of systematically rotating the sample so that new entrants to the target population are properly represented is seldom used.

In short, without a universal frame of email addresses with known links to individual population elements, some panel practices will ignore the frame development step. Without a well-defined sampling frame, the coverage error of resulting estimates is unknowable.

One might argue that lists built using volunteers could be said to be frames and are sometimes used as such. However, the nature of these frames versus probability sample frames is quite different. The goal is usually to make inferences, not to the artificial population of panel members, but to the broader population of U.S. households or adults.

Further, defining coverage for panels is not quite as straightforward as defining coverage for other modes such as telephone or face-to-face frames. Every adult in a household with a telephone is

considered to be covered by the frame, as is every adult in an occupied dwelling unit selected for an area probability sample. Although people may have access to the Internet, either at home or at some other location, not everyone in a household may actually use the Internet.

In a household with Internet access and at least one user, non-users in the household could be considered as covered since the user can either answer for the other adults or bring another adult to the computer in the same way as the person who answers the telephone or is interviewed in person has the opportunity to engage other members of the household in the interview process. Individuals who only have Internet access at a location other than at home may or may not be considered as being covered. For example, an individual's access at work may be limited by corporate policy to business-related activities.

As we noted in the previous section, estimates of access to the Internet can vary widely. Mediamark Research and Intelligence (MRI) in their in-person *Survey of the American Consumer*TM (Piekariski et al, 2008) estimated that in 2008 approximately 85 percent of the adult population of the continental U.S. has *some* Internet access. This is considerably higher than the previously-cited 77 percent reported by the 2009 CPS and the 74 percent reported by Pew. The demographics of those without Internet access in some cases differ significantly from those with access and these differences may persist even if statistical adjustments are made. Adults without Internet access are more than twice as likely to be over the age of 65 as the general adult population. Those without Internet access are also more likely to be a member of a minority group, have incomes less than \$25,000, have a high school education or less, be unemployed or retired, not own their home, live in rural counties, or in the South Census Region.

Since participating in an online survey or belonging to an online panel is probably limited to those individuals who actually use the Internet, it is reasonable to assume that *coverage* really means using the Internet. The MRI, CPS, and Pew estimates are all lower than the household (88 percent) and population (89 percent) telephone coverage estimates from 1970 that led to the acceptability of using telephone surveys in place of in-person surveys.⁵

The foregoing discussion has focused on the interplay of sample frames and coverage issues within the context of the recruiting practices commonly employed for volunteer panels. Of course, email addresses are not the only form of sampling frame for Web surveys. As described in Section 3, people are sometimes recruited by telephone or mail and asked to join an online panel. In such a design, the coverage issues are those that pertain to the sampling frame used (e.g., RDD). Panels recruited in this manner often try to ameliorate the coverage problems deriving from not everyone contacted having Internet access either by providing the needed equipment and access or by conducting surveys with this subgroup using offline methods (telephone or mail). The extent to which providing Internet access might change the behaviors and attitudes of panel members remains unknown.

UNIT NONRESPONSE AND NONRESPONSE ERROR

⁵ We note that due to the emergence of cell-only households landline coverage has dropped significantly in recent years and is now less than 80 percent and may soon approach or even be less than household Internet penetration. However, it also is becoming common practice to include cell phone samples in telephone studies that aim to represent the full population.

Unit nonresponse, in contrast to coverage of the target population, concerns the failure to measure a unit in a sample. Unit nonresponse occurs when a person selected for a sample does not respond to the survey. This is distinguished from *item nonresponse*, which occurs when a respondent skips a question within a survey – either intentionally or unintentionally. In our treatment of nonresponse in this section we are referring to unit nonresponse. In traditional survey designs, this is a nonobservation error that arises after the sampling step from a sampling frame that covers a given target population. Unlike online panels that are established using probability sampling methods, volunteer panels are not established using probabilistic sampling techniques. This in turn affects various aspects of how nonresponse and nonresponse bias in such panels are conceptualized and measured, and what strategies are effective in trying to reduce these problems.

There are four stages in the development, use, and management of volunteer panels where nonresponse can become an issue: (1) recruitment; (2) joining and profiling; (3) specific study sampling; and (4) panel maintenance.

RECRUITMENT STAGE. As described in Section 3, none of the means by which nonprobability online panel members are recruited allow those who are establishing and managing the panel to know with any certainty from what base (i.e., target population) their volunteer members come. Because of this, there is no way of knowing anything precise about the size or nature of the nonresponse that occurs at the recruitment stage. However, the size of the nonresponse is very likely to be considerable. Similarly, the nature of the nonresponse may not be random with respect to variables of interest to health, social science, consumer and general marketing researchers. The latter follows from the known demographics of those who enroll as members, and their differences from the general population. The motives that lead some who are exposed to invitations to join online panels often reflect attitudinal and personality differences between them and others in the general population (e.g., Bosnjak, Tuten, and Wittman 2005). Psychographic variables (attitudes and personality measures of individuals) are starting to be used by online panel companies in their efforts to create panels and samples that are potentially more representative of the larger population.

Empirical evaluations of online panels abroad and in the U.S. leave no doubt that those who choose to join online panels differ in important and nonignorable ways from those who do not. For example, researchers directing the Dutch online panel comparison study (Vonk, van Ossenbruggen, and Willems, 2006) report that ethnic minorities and immigrant groups are systematically underrepresented in Dutch panels. They also found that, relative to the general population, online panels contained disproportionately more voters, more Socialist Party supporters, more heavy Internet users, and fewer church-goers.

Similarly, researchers in the U.S. have documented that online panels are disproportionately comprised of whites, more active Internet users, and those with higher levels of educational attainment (Couper, 2000; Dever et al., 2008; Chang and Krosnick, 2004 and 2008; and Malhotra and Krosnick, 2007). In other words, the membership of a panel generally reflects the demographic bias in Internet use. Attitudinal and behavioral differences similar to those reported by Vonk et al. also exist for Internet users in the U.S. based on an analysis of the online population conducted by Piekarski, et al. (2008) using data from the MRI in-person *Survey of the American Consumer*TM. After standard demographic weighting, U.S. Internet users (any use) who reported being online five or more times per day were found to be considerably more involved in civic and political activities than the general U.S. population. The researchers also found that frequent users in the U.S. placed less importance on religion, traditional

gender roles and more importance on environmental issues. In this study, panel members reported even greater differences from the general population for these activities when the unweighted data were examined.

If online panel members belonging to under-represented groups are similar to group members who are not in the panel, then the risk of bias is diminished under an appropriate adjustment procedure. However, there is evidence to suggest that such within-group homogeneity may be a poor assumption. In the Dutch online panel comparison study, some 62 percent of respondents were members of multiple panels. The mean number of memberships for all respondents was 2.7 panels. Frequent participation in online surveys does not necessarily mean that an individual will be less representative of a certain group than he or she would otherwise be, but panel members clearly differed on this activity. Vonk et al. (2006) concluded that “panels comprise a specific group of respondents that differ on relevant criteria from the national population. The representative power of online panels is more limited than assumed so far.” Based on their analyses, the Dutch panels also had a number of differences between them, much like the house effects we have seen in RDD telephone samples (Converse and Traugott, 1986).

JOINING AND PROFILING STAGES. As described in Section 3, many panels require individuals wishing to enroll first to indicate their willingness join by clicking through to the panel company’s registration page and entering some personal information, typically their email address and key demographics (minimally age to ensure “age of consent”). Then he or she typically is sent an email to which the volunteer must respond to indicate that he/she in fact did sign up for the panel. This two-step process constitutes the “double opt-in” the vast majority of online panels require before someone is officially recognized as a member and available for specific studies. Potential members who initially enroll may choose not to complete the profiling questionnaire. Alvarez et al. (2003) report that just over 6 percent of those who clicked through a banner ad to the panel registration page eventually completed all the steps required to become a panel member. Those building and managing online panels can learn something about this nonresponse by comparing the limited data about the prospective panel member gathered at the recruitment stage between those who complete the profiling stage versus those who do not. However, very little has been reported.

SPECIFIC STUDY STAGE. Once a person has joined the panel, he or she likely will be selected as one of the sampled panel members invited to participate in specific surveys. There are several reasons why a sampled member may not end up participating fully or at all in a specific survey. These include:

- Refusal due to any number of reasons such as lack of interest, survey length, or a heavy volume of survey invitations;
- Failure to qualify due either to not meeting the study’s eligibility criteria or not completing within the survey’s defined field period;
- Technical problems that prevent either delivery of the survey invitation or access to and completion of the online questionnaire.

Those building and managing online panels can learn a great deal about nonresponse at this stage by using the extensive data about their panel members gathered at the initial recruitment and profiling stages, or from information gleaned from any previous surveys the members may have completed as panel members. Analyses can compare those sampled members who complete the specific survey questionnaire with those panel members who were sampled for the specific survey but did not complete the

questionnaire. It is not unusual to find that response rates vary across subgroups that are of analytic interest, particularly demographic groups. For telephone, mail, and face-to-face surveys, nonresponse has often been reported to be higher among those less educated, older, less affluent, or male (Dillman, 1978; McCabe et al. 2002; Suchman and McCandless, 1940; Wardle, Robb, and Johnson, 2002). The pattern for nonprobability panels may be somewhat different – in one study, nonresponse was higher among panel members who were elderly, racial or ethnic minorities, unmarried, less educated, or highly affluent (Knapton and Myers, 2005). Gender was found to be unrelated to nonresponse in the online panel they studied. Different panels can vary substantially in their composition and in their recruitment and maintenance strategies, so the results from this one study may not generalize to other panels. Despite a great deal of data being available to investigate this issue, little has yet been publically reported.

Some panel companies attempt to address differential nonresponse at the sampling stage, i.e., before data collection even begins. In theory, one can achieve a final responding sample that is balanced on the characteristics of interest by disproportionately sampling panel members belonging to historically low response rate groups at higher rates (cf. Rivers, 2009). For example, Hispanic panel members might be sampled for a specific survey at a higher rate than other members in anticipation of disproportionately more Hispanics not responding. Granted, this sampling approach is possible outside of the online panel context, but it requires certain information (e.g., demographics) about units on the frame – information that is also often available in other survey designs (certain telephone exchanges and blocks of numbers may have a high density of Hispanics or blacks, or be more likely to be higher or lower income households; demographics have often been linked with zip codes so they can be used extensively in sampling). Bethlehem and Stoop (2007) note that this practice of *preemptive differential nonresponse adjustment* is becoming more challenging as Web survey response rates decline. Control over the final composition can only be achieved by taking into account differential response propensities of many different groups, using information about response behavior from previous, similar surveys. *However, even when balance is achieved on the desired dimensions, there is no guarantee that nonresponse error has been eliminated or even reduced.* The success of the technique relies on the assumption that nonresponding panel members within specific groups are similar to respondents within the same specific groups on the measures of interest.

PANEL MAINTENANCE AND PANEL ATTRITION. As mentioned earlier, excessive attrition can be a problem for replicating results within a panel. There are two types of panel attrition: *forced* and *normal*. Some panels may set a fixed length of time (e.g., 24 months, 100 surveys) during which a person can remain a panel member, but many do not. Forced attrition occurs within panels that have a maximum duration of elapsed time or a threshold number of surveys for which any one member can remain in the panel. After that criterion has been reached the member is removed (dropped) from the panel. Whether that person is eligible to rejoin the panel after some passage of time differs from panel to panel. Forced turnover is not a form of nonresponse. Rather, it is one of the criteria that determines who remains eligible for continued panel membership.

In contrast, so-called “normal” or unforced turnover is a form of nonresponse, in that people who are panel members and who have not reached the end of their eligibility leave the panel even though panel management might ideally like them to remain active members. The panel company may drop a member because he or she is not participating in enough surveys, may be continually providing data of questionable quality or engaging in some other forms of response that the panel’s users find objectionable. Panel members also may choose to opt out on their own for any number of reasons.

Many approaches could be used by those who strive to maintain panel membership among those members eligible to continue such membership. But there appears to be nothing reported in the research literature reporting the effectiveness of such possible approaches to reduce panel attrition or whether reducing attrition is desirable (obviously low attrition reduces new panel member acquisition costs which can make panel costs lower, but attrition that is too low may create its own difficulties, e.g., panel conditioning).

Those managing nonprobability panels can learn a great deal about attrition-related nonresponse by using the extensive data about their members gathered at the recruitment and profiling stages along with a host of other information that can be gleaned from past surveys the member may have completed. Analyses comparing those sampled members who do not drop out during their eligibility with those who do drop out due to reasons of normal turnover might lead to a clearer understanding of the reasons for the high rates of attrition. We are aware of no published accounts to verify that this is being done or what might be being learned.

RESPONSE METRICS

Callegaro and DiSogra (2008) point out that there currently are no widely-accepted metrics that can be used to accurately quantify or otherwise characterize the nonresponse that occurs at the recruitment stage for nonprobability online panels. This is because the base (denominator) against which the number of people who joined the panel (numerator) can be compared is often unknown. Furthermore, recruitment for many nonprobability online panels is a constant, ongoing endeavor. Thus, the concept of a recruitment response rate has a “moving target” aspect to it that precludes any standardization of its calculation. Although only sparsely documented in the literature, we believe that nonresponse at this stage is very high. At the profiling stage, a “profile rate” can be calculated which is similar to AAPOR’s Response Rate 6 (AAPOR RR6). The numerator is comprised of the number of those who completed profile questionnaires and possibly those who partially completed profile questionnaires. The denominator is comprised of the number of all people who initially enrolled in the panel, regardless of whether they started or completed the profiling questionnaire. The profiling rate is also an ever-changing number, since persons often are constantly coming into the panel over time. Thus, one could envision a profile rate being computed for set periods of time; e.g., a seven-day period, for many consecutive weeks. These rates could then be plotted to inform the panel managers how the profile rate is changing over time.

Thinking about the study-specific stage (where panel members are selected for participation in a specific study), Callegaro and DiSogra (2008) recommend several different rates that can be calculated:

- *Absorption rate* (Lozar, Manfreda, and Vehovar; 2002), which is the rate at which emails invitations sent by the panel managers to the panel membership actually reach the panel members. This is a function of the number of network-error undeliverable emails that are returned and the number of bounce-back undeliverable emails;
- *Completion rate*, which essentially is the AAPOR RR6 formula mentioned above, but limited to those panel members who are sampled for the specific survey;
- *Break-off rate*, which is the portion of specific survey questionnaires that were begun but never completed during the field period;

- *Screening completion rate* (Ezzati-Rice, Frankel, Hoaglin, Loft, Coronado, et al., 2000), which is the proportion of those panel members invited to participate in a specific survey, but who are deemed eligible or ineligible for the full questionnaire through a series of study-specific screening questions that they answer at the beginning of the questionnaire;
- *Eligibility rate*, which is the number of sampled panel members for a specific study that completed the screening and were found qualified compared to that same number plus the number who completed the screening and were found to be ineligible.

Finally, in terms of metrics that address panel maintenance, Callegaro and DiSogra (2008) suggest that the computation of the *attrition rate* be defined as “the percentage of [panel] members who drop out of the panel in a defined time period” (also see Clinton, 2001; and Sayles and Arens, 2007).

Of note, Bethlehem and Stoop (2007) point out that use of the term “response rate” in the context of a nonprobability panel survey can be misleading. Response rates can be boosted by pre-selecting the most cooperative panel members. This can invalidate comparisons with response rates from samples used with a different survey design. ISO 26362 (2009) recommends the use of the term *participation rate* rather than response rate because of the historical association of response rate with probability samples. The participation rate is defined as “the number of respondents who have provided a usable response divided by the total number of initial personal invitations requesting participation.”

COVERAGE ERRORS VERSUS NONRESPONSE BIAS

Errors of nonobservation are generally classified into either coverage or unit nonresponse errors (distinguished from item nonresponse errors). The section above on coverage errors notes that for most online panels there is no knowable target population or sampling frame that is well-defined. Without a sampling frame there can be no assessment of whether the frame covers the target population well—that is, whether those eligible for the sampling step are in some sense a microcosm of the full target population. Given the absence of a sampling frame in online panels, the conceptual difference between coverage error and nonresponse error also gets blurred, making it extremely difficult to ascribe errors in estimates to either source.

As discussed above, there are several mechanisms that can lead to unit nonresponse in online panel surveys. In this section we comment on the implications of nonresponse for survey quality, specifically the accuracy of survey estimates. Nonresponse may or may not result in biased estimates depending on (1) how likely a typical sample member is to participate (i.e., the average response likelihood, also known as response propensity) and (2) the relationship between the survey measure and response behavior (Bethlehem, 2002; Lessler and Kalsbeek, 1992).

The first factor, average response propensity, is a continuing and increasingly serious problem for all survey modes and samples, and online panels appear to be no exception. As occurs even with RDD samples, response rates in online panel surveys can vary dramatically, depending on the nature of the sample, the topic, the incentives, and other factors. Although response rate alone may be a poor indicator of error (Curtin, Presser, and Singer, 2000; Groves, 2006; Keeter, Miller, Kohut, Groves, and Presser, 2000; Merkle and Edelman, 2002), low response rates typically signal an increased concern for the possibility of nonresponse bias.

Research in the U.S. demonstrated only a weak association between response rate and nonresponse bias (Groves, 2006; Keeter, Kennedy, Dimock, Best, and Craighill, 2006). Additional research in the Netherlands has replicated these findings in online panels. One Dutch online panel comparison study (Vonk, van Ossenbruggen, and Willems, 2006) featured a common survey fielded independently by 19 different panels in the Netherlands. The study-specific response rates ranged from 18 percent to 77 percent, with an overall completion rate of 50 percent. The investigators found no meaningful differences between point estimates from the surveys with low response rates and those with high response rates. Similar results are reported by Yeager, Krosnick, Chang, Javitz, Levindusky, Simpser, and Wang (2009), although there was considerable variation between panels in terms of which estimates were closer to the “true” value. In other words, some panels were not reliably more accurate than others.

The second factor that is a concern with any survey, no matter what the mode or sample source, is the relationship between the survey measure (responses on the questionnaire, e.g., attitude ratings) and response behavior. This relationship has two critical properties from the researcher’s perspective. First, it is specific to each survey measure. That is, the likelihood of participating may be strongly associated with one factor (e.g., level of interest in the survey topic) and only weakly related or unrelated with a different factor (e.g., product satisfaction, purchase likelihood, presidential approval). The other important property is that it is generally very difficult, if not impossible, to measure the extent of this relationship since it requires knowledge of each sample member’s value on the survey measure, but typically these values would be known only for those who responded to the specific survey. Because unit nonresponse may be very high with a nonprobability panel, the relationship between response behavior and survey measures may be incalculable. However, since a good deal is known about panel members there is at least the opportunity to characterize the differences between responders and nonresponders to a given survey. As far as we can tell, this type of analysis is seldom done.

One potential reason for nonresponse bias is differential interest in the topic influencing participation decisions in online surveys. Early on, some believed that online panel surveys would be immune from this effect because the decision to participate in panels is general and response rates on individual surveys were high (Bethlehem and Stoop, 2007). However, as response rates declined, concern increased that advertising the survey topic in invitations may induce bias (and ethical guidelines concerning respondent treatment indicate that generally respondents should be informed about the topic of the survey, the length of the survey, and the incentives for the survey).

Unexpectedly, a recent experiment on this topic suggests that topic interest may have little if any effect on participation decisions. Tourangeau and his colleagues (2009) conducted a repeated measures experiment on topic-induced nonresponse bias using a pooled sample from two different online panels. They found that membership in multiple panels and a high individual participation rate were strong predictors of response in a follow-up survey but interest in the topic was not a significant predictor. Additional empirically rigorous studies of this sort are needed to confirm this null effect pertaining to topic interest, but the only available evidence suggests that general online survey taking behavior may be more influential in participation decisions than are attitudes about the survey topic.

NONRESPONSE BIAS AT THE RECRUITMENT STAGE. Given that we currently have little empirical evidence about nonresponders at the panel recruitment stage, the only chance to estimate potential effects from non-cooperation on bias at this stage is from what can be surmised about how those who joined the panel are likely to differ along key variables of interest compared to those who did not choose to join. Whether selective sampling and/or weighting can “correct” for such bias is addressed in Section 6.

DIFFERENCES BETWEEN ONLINE PANEL RESPONDERS AND NONRESPONDERS IN SPECIFIC STUDIES.

In contrast to the lack of information we have about nonresponders at the panel recruitment stage, those managing online panels often have large amounts data they can use to study possible nonresponse bias that may exist on key variables in specific surveys of their members. We remind the reader of the framework used in this report: The sampling frame for a specific study is the database of online panel members, not a theoretical list of all persons who use the Internet. This is important for understanding our definition of nonresponse error and how it differs from coverage error. *Coverage error* refers to error resulting from the failure of the sampling frame (the panel) to adequately include all members of the target population. *Nonresponse error* arises because not all persons sampled for a study actually complete the questionnaire.

The low response rates sometimes observed in specific studies using nonprobability panels may signal a general risk of nonresponse bias but, ultimately, this error operates at the level of the individual survey question. We believe it is the researcher's responsibility to assess the threat to key survey measures from nonresponse using the extensive data they have available about nonresponders, knowledge of the survey subject matter and what has been documented about nonresponse bias in the literature.

5. MEASUREMENT ERROR IN ONLINE PANEL SURVEYS

A primary interest in social science research is to understand how and why people think, feel, and act in the ways they do. Much of the information that we use to help us describe and explain people's behavior comes from surveys. One way we gather this information is to ask people about the occurrence of events or experiences using nominal classification (Stevens, 1946; 1951), often in the form of "did the event occur?" or "which event(s) occurred?" We also ask respondents to evaluate their experiences along some underlying quality or dimension of judgment using ordinal, interval, or ratio scales. As an example, when we ask a person to indicate his/her attitude toward a governmental policy, we assume that the attitude can be represented along some dimension of judgment (e.g., 'Very bad' to 'Very good'). In the process, there are a number of sources for potential errors that can influence the accuracy of these kinds of measurements.

Measurement error is commonly defined as the difference between an observed response and the underlying true response. There are two major types of measurement error: random and systematic. *Random error* occurs when an individual selects responses other than his/her true response without any systematic direction in the choices made. With random error, a person is just as likely to select a response higher on the continuum as select a lower response than his/her true position. *Systematic measurement error* (also known as "bias") occurs when respondents select responses that are more often in one direction than another and these responses are not their true responses. Random error tends to increase the dispersion of observed values around the average (most often the mean) but does not generally affect the average value when there is a sufficiently large sample. Systematic error may or may not increase the dispersion of values around the average, but will generally shift the measure of central tendency in one direction or another. In general, systematic error is of greater concern to the researcher than is random error.

There are a number of potential causes of measurement error. They include: how the concepts are measured (the questions and responses used – typically referred to as questionnaire design effects), the mode of interview, the respondents, and the interviewers.

QUESTIONNAIRE DESIGN EFFECTS

The influence of questionnaire design on measurement error has received attention in a number of publications (e.g., Dillman, Smyth, and Christian, 2009; Galesic and Bosnjak, 2009; Lugtigheid and Rathod, 2005; Krosnick, 1999; Groves, 1989; Tourangeau, 1984) and the design of Web questionnaires has introduced a new set of challenges and potential problems. Much of the literature specific to Web along with its implications for survey design has been conveniently summarized in Couper (2008). That literature has demonstrated a wide range of response effects due to questionnaire and presentation effects in Web surveys. However, there is no empirical evidence to tie those effects to sample source (RDD-recruited, nonprobability recruited, river, etc.). Although researchers doing research by Web should familiarize themselves with the research on questionnaire design effects those findings are beyond the scope of this report. The primary concern in this section is the possibility of measurement error arising either out of mode of administration or the respondents themselves.

MODE EFFECTS

The methodologies employed by online panels involve two shifts away from the most popular methodologies preceding them: (1) the move from interviewer-administered questionnaires to self-completion questionnaires on computers and (2) in the case of online volunteer panels, the move from probability samples to non-probability samples. A substantial body of research has explored the impact of the first shift, assessing whether computer self-completion yields different results than face-to-face or telephone interviewing. Other studies have considered whether computer self-completion yields different results with nonprobability samples than with probability samples. Some studies combined the two shifts, examining whether computer self-completion by non-probability samples yields different results than face-to-face or telephone interviewing of probability samples.

This section reviews this research. In doing so it considers whether computer self-completion might increase or decrease the accuracy of reports that respondents provide when answering survey questions and how results from non-probability samples compare to those from probability samples in terms of their accuracy in measuring population values.⁶ We note that a number of these studies have focused on pre-election polls and forecasting. We view these as a special case and discuss them last.

THE SHIFT FROM INTERVIEWER ADMINISTRATION TO SELF-ADMINISTRATION BY COMPUTER. In a study by Burn and Thomas (2008) the same respondents answered a set of attitudinal questions both online and by telephone, counter-balancing the order of the modes. The researchers observed notable differences in the distributions of responses to the questions, suggesting that mode alone can affect answers (and perhaps answer accuracy). However, in a similar study by Hasley (1995), equivalent answers were obtained in both modes. So differences between modes may occur sometimes and not others, depending on the nature of the questions and response formats.

Researchers have explored two specific hypotheses about the possible impact of shifting from one mode (interviewer administration) to another (computer self-administration). They are *social desirability* response bias and *satisficing*.

The *social desirability* hypothesis proposes that in the presence of an interviewer, some respondents may be reluctant to admit embarrassing attributes about themselves and/or may be motivated to exaggerate the extent to which they possess admirable attributes. The risk of having an interviewer frown or sigh when a respondent says he/she cheated an income tax return, or inadvertently convey a sign of approval when hearing that the respondent gave money to charity may be the source of such intentional misreporting. An even more subtle influence could be characteristics of the interviewer. For example, consider a situation in which an interviewer asks a respondent whether he/she thinks that the federal government should do more to ensure that women are paid as much as men are for doing the same work. If a female interviewer asks the question, respondents might feel some pressure to answer affirmatively, because saying so would indicate support for government effort to help a social group to which the interviewer belongs. But if asked the same question by a male interviewer the respondent might feel no pressure to answer affirmatively and perhaps even the reverse. Thus, the social desirability hypothesis states that respondents may be more honest and accurate when reporting confidentially on a computer than when providing reports orally to an interviewer.

⁶ We do not review an additional large literature that has compared paper-and-pencil self-completion to other modes of data collection (e.g., interviewer administration, computer self-completion).

A number of studies have explored the idea that computer self-completion yields more honest reporting of embarrassing attributes or behaviors and less exaggeration of admirable ones. For the most part, this research finds considerable evidence in support of the social desirability hypothesis. However, many of these studies simply demonstrate differences in rates of reporting socially desirable or undesirable attributes, without providing any direct tests of the notion that the differences were due to intentional misreporting inspired by social desirability pressures.

For example, Link and Mokdad (2004, 2005) conducted an experiment in which participants were randomly assigned to complete a questionnaire by telephone or via the Internet. After weighting to yield demographic equivalence of the two samples, the Internet respondents reported higher rates of diabetes, high blood pressure, obesity, and binge drinking, and lower rates of efforts to prevent contracting sexually transmitted diseases when compared to those interviewed by telephone. This is consistent with the social desirability hypothesis, assuming that all of these conditions are subject to social desirability pressures. The telephone respondents also reported more smoking than did the Internet respondents, which might seem to be an indication of more honesty on the telephone. However, other studies suggest that adults' reports of smoking are not necessarily subject to social desirability pressures (see Aguinis, Pierce, and Quigley, 1993; Fendrich, Mackesy-Amity, Johnson, Hubbell, and Wislar, 2005; Patrick, Cheadle, Thompson, Diehr, Koepsell, and Kinne, 1994)

Mode comparison studies generally have used one of three different designs. A first set of studies (Chang and Krosnick, 2010; Rogers, Willis, Al-Tayyib, Villarroel, Turner, Ganapathi, et al., 2005) have been designed as true experiments. Their designs called for respondents to be recruited and then immediately assigned to a mode, either self-completion by computer or oral interview, making the two groups equivalent in every way, as in all true experiments. A second set of studies (Newman, Des Jarlais, Turner, Gribble, Cooley, and Paone, 2002; Des Jarlais, Paone, Milliken, Turner, Miller, Gribble, Shi, Hagan, and Friedman, 1999; Riley, Chaisson, Robnett, Vertefeuille, Strathdee, and Vlahov, 2001) randomly assigned mode at the sampling stage, that is, prior to recruitment. Because assignment to mode was done before respondent contact was initiated, the response rates in the two modes differed, introducing the potential for confounds in the mode comparisons. In a final series of studies, (Cooley, Rogers, Turner, Al-Tayyib, Willis, and Ganapathi, 2001; Metzger, Koblin, Turner, Navaline, Valenti, Holte, Gross, Sheon, Miller, Cooley, Seage, and HIVNET Vaccine Preparedness Study Protocol Team, 2000; Waruru, Nduati, and Tylleskar, 2005; Ghanem, Hutton, Zenilman, Zimba, and Erbeding, 2005) respondents answered questions both in face-to-face interviews and on computers. All of these studies, regardless of design, found higher reports of socially stigmatized attitudes and behaviors in self-administered computer-based interviews than in face-to-face interviews.

This body of research is consistent with the notion that self-administration by computer elicits more honesty, although there is no direct evidence of the accuracy of those reports (one notable exception being Kreuter, Presser, and Tourangeau, 2008). They are assumed to be accurate because the attitudes and behaviors are assumed to be stigmatized.

The *satisficing* hypothesis focuses on the cognitive effort that respondents devote to generating their answers to survey questions. The foundational notion here is that providing accurate answers to such questions usually requires that respondents carefully interpret the intended meaning of a question, thoroughly search their memories for all relevant information with which to generate an answer, integrate that information into a summary judgment in a balanced way, and report that judgment accurately. But some respondents may choose to shortcut this process, generating answers more superficially and less

accurately than they might otherwise (Krosnick, 1991; 1999). Some specific respondent behaviors generally associated with satisficing include response non-differentiation (“straightlining”), random responding, responding more quickly than would be expected given the nature of the questions and responses (“speeding”), response order effects, or item non-response (elevated use of non-substantive response options such as “don’t know” or simply skipping items).

Some have argued that replacing an interviewer with a computer for self-administration has the potential to increase the likelihood of satisficing due to the ease of responding (simply clicking responses without supervision). If interviewers are professional and diligent and model their engagement in the process effectively for respondents, this may be contagious and may inspire respondents to be more effortful than they would be without such modeling. Likewise, the presence of an interviewer may create a sense of accountability in respondents, who may feel that they could be asked at any time to justify their answers to questions. Elimination of accountability may allow respondents to rush through a self-administered questionnaire without reading the questions carefully or thinking thoroughly when generating answers. Such accountability is believed to inspire more diligent cognitive effort and more accurate answering of questions.

Although much of the literature on satisficing has often focused on the characteristics of respondents (e.g., male, lower cognitive skills, younger), the demands of the survey task also can induce higher levels of satisficing. Computer-based questionnaires often feature extensive grid response formats (items in rows, responses in columns) and may ask more responses than what might occur in other modes. In addition, some researchers leverage the interactive nature of online to design response tasks and formats (such as slider bars and complex conjoint designs) that may be unfamiliar to respondents or increase respondent burden.

It is also possible that removing interviewers may improve the quality of the reports that respondents provide. As we noted at the outset of this section, interviewers themselves can sometimes be a source of measurement error. For example, if interviewers model only a superficial engagement in the interviewing process and suggest by their non-verbal (and even verbal) behavior that they want to get the interview over with as quickly as possible, this approach may also be contagious and may inspire more satisficing by respondents. When allowed to read and think about questions at their own pace during computer self-completion, respondents may generate more accurate answers. Further, while some have proposed that selection of neutral responses or the use of non-substantive response options reflect lower task involvement, it may be that such choices are more accurate reflections of people’s opinions. People may feel compelled to form an attitude in the presence of an interviewer but not so when taking a self-administered questionnaire (Fazio, Lenn, and Effrein, 1984). Selection of these non-substantive responses might also be more detectable in an online survey when it is made an explicit response rather than an implicit response as it often occurs in interviewer administered surveys.

Chang and Krosnick (2010) conducted a true experiment, randomly assigning respondents to complete a questionnaire either on a computer or to be interviewed orally by an interviewer. They found that respondents assigned to the computer condition manifested less non-differentiation and were less susceptible to a response order effects.

Other studies not using true random assignment yielded more mixed evidence. Consistent with the satisficing hypothesis, Chatt and Dennis (2003) observed more non-differentiation in telephone interviews than in questionnaires completed online. Fricker, Galesic, Tourangeau, and Yan (2005) found less item

non-response among people who completed a survey via computer than among people interviewed by telephone.

On the other hand, Heerwegh and Loosveldt (2008) found more non-differentiation and more "don't know" responses in computer-mediated interviews than in face-to-face interviews. Fricker, Galesic, Tourangeau, and Yan (2005) found more non-differentiation in data collected by computers than in data collected by telephone and no difference in rates of acquiescence. Miller (2000; see also Burke, 2000) found equivalent non-differentiation in computer-mediated interviews and telephone interviews. And Lindhjem and Navrud (2008) found equal rates of "don't know" responses in computer and face-to-face interviewing. Because the response rates in these studies differed considerably by mode (e.g., in Miller's, 2000, study, the response rate for the Internet completion was one-quarter the response rate for the telephone surveys), it is difficult to know what to make of differences or lack of differences between the modes.

Speed of survey completion is another potential indicator of satisficing. If we assume that rapid completion reflects less cognitive effort then most research reinforces the argument that administration by computer is more prone to satisficing. In a true experiment done in a lab, Chang and Krosnick (2010) found that computer administration was completed more quickly than oral interviewing. In a field study that was not a true experiment, Miller (2000; see also Burke, 2000) described a similar finding. A telephone survey lasted 19 minutes on average, as compared to 13 minutes on average for a comparable computer-mediated survey. In a similar comparison, Heerwegh and Loosveldt (2008) reported that a computer-mediated survey lasted 32 minutes on average, compared to 48 minutes for a comparable face-to-face survey. Only one study, by Christian, Dillman, and Smyth (2008) found the opposite: Their telephone interviews lasted 12 minutes, whereas their computer self-completion questionnaire took 21 minutes on average.

Alternatively, one could argue that speed of completion, in and of itself, compared to completion in other modes is not necessarily an indication that quality suffers in self-administration modes. Perhaps respondents answer a set of questions in a visual self-administered mode more quickly than in an aural format primarily because people can read and process visual information more quickly than they can hear and process spoken language.

Primacy and *recency* effects are also linked to satisficing. *Primacy* is the tendency for respondents to select answers offered at the beginning of a list. *Recency* is the tendency for respondents to select answers from among the last options offered. Nearly all published primacy effects have involved visual presentation, whereas nearly all published recency effects have involved oral presentation (see, e.g., Krosnick and Alwin, 1987). Therefore, we would expect computer administration and oral administration to yield opposite response order effects, producing different distributions of responses. Chang and Krosnick (2010) reported just such a finding, although the computer mode was less susceptible to this effect than was oral administration, consistent with the idea that the latter is more susceptible to satisficing.

As the foregoing discussion shows, the research record relative to the propensity for respondents to satisfice across survey modes is conflicted. True experiments show reduced satisficing in computer responses than in telephone or face-to-face interviews. Other studies have not always found this pattern, but those studies were not true experiments and involved considerable confounds with mode. Therefore, it seems reasonable to conclude that the limited available body of evidence supports the notion that there

tends to be less satisficing in self-administration by computer than in interviewer administration.

Another way to explore whether interviewer-administered and computer-administered questionnaires differ in their accuracy is to examine *concurrent and predictive validity*, that is, the ability of measures to predict other measures to which they should be related on theoretical grounds. In their experiment, Chang and Krosnick (2010) found higher concurrent or predictive validity for computer-administered questionnaires than for interviewer-administered questionnaires. However, among non-experimental studies, some found the same pattern (Miller, 2000; see also Burke, 2000), whereas others found equivalent predictive validity for the two modes (Lindhjem and Navrud, 2008).

Finally, some studies have assessed validity by comparing results with nonsurvey measurements of the same phenomena. In one such study Bender, Bartlett, Rand, Turner, Wamboldt, and Zhang (2007) randomly assigned respondents to report on their use of medications either via computer or in a face-to-face interview. The accuracy of their answers was assessed by comparing them to data in electronic records of their medication consumption. The data from the face-to-face interviews proved more accurate than the data from the self-administered by computer method.

Overall, the research reported here generally suggests higher data quality for computer administration than for oral administration. Computer administration yields more reports of socially undesirable attitudes and behaviors than does oral interviewing, but no evidence that directly demonstrates that the computer reports are more accurate. Indeed, in one study, computer administration compromised accuracy. Research focused on the prevalence of satisficing across modes also suggests that satisficing is less common on computers than in oral interviewing, but more true experiments are needed to confirm this finding. Thus, it seems too early to reach any firm conclusions about the inherent superiority or equivalence of one mode vis-a-vis the other in terms of data accuracy.

THE SHIFT FROM INTERVIEWER ADMINISTRATION WITH PROBABILITY SAMPLES TO COMPUTER SELF-COMPLETION WITH NON-PROBABILITY SAMPLES. A large number of studies have examined survey results when the same questionnaire was administered by interviewers to probability samples and online to nonprobability samples (Taylor, Krane, and Thomas, 2005; Crete and Stephenson, 2008; Braunsberger, Wybenga, and Gates, 2007; Klein, Thomas, and Sutter, 2007; Thomas, Krane, Taylor, & Terhanian, 2008; Baker, Zahs, and Popa, 2004; Schillewaert and Meulemeester, 2005; Roster, Rogers, Albaum, and Klein, 2004; Loosveldt and Sonck, 2008; Miller, 2000; Burke, 2000; Niemi, Portney, and King, 2008; Schonlau, Zapert, Simon, Sanstad, Marvus, Adams, Spranca, Kan, Turner, and Berry, 2004; Malhotra and Krosnick, 2007; Sanders, Clarke, Stewart, and Whiteley, 2007; Berrens, Bohara, Jenkins-Smith, Silva, and Weimer, 2003; Sparrow, 2006, Cooke, Watkins, and Moy, 2007; Elmore-Yalch, Busby, and Britton, 2008). Only one of these studies yielded consistently equivalent findings across methods, and many found differences in the distributions of answers to both demographic and substantive questions. Further, these differences generally were not substantially reduced by weighting.

Once again, *social desirability* is sometimes cited as a potential cause for some of the differences. A series of studies comparing side-by-side probability sample interviewer-administered surveys with nonprobability online panel surveys found that the latter yielded higher reports of:

- Opposition to government help for blacks among white respondents (Chang and Krosnick, 2009);
- Chronic medical problems (Baker, Zahs, and Popa, 2004);
- Motivation to lose weight to improve one's appearance (Baker, Zahs, and Popa, 2004);

- Feeling sexually attracted to someone of the same sex (Taylor, Krane, and Thomas, 2005);
- Driving over the speed limit (Taylor, Krane, and Thomas, 2005);
- Gambling (Taylor, Krane, and Thomas, 2005);
- Cigarette smoking (Baker, Zahs, and Popa, 2004; Klein, Thomas, and Sutter, 2007);
- Being diagnosed with depression (Taylor, Krane, and Thomas, 2005);
- Consuming beer, wine, or spirits (Taylor, Krane, and Thomas, 2005).

Conversely, compared to interviewer-administered surveys using probability-based samples, online surveys using nonprobability panels have documented fewer reports of:

- Excellent health (Baker, Zahs, and Popa, 2004; Schonlau, Zapert, Simon, Sanstad, Marcus, Adams, Spranca, Kan, Turner, and Berry, 2004; Yeager, Krosnick, Chang, Javitz, Levendusky, Simpser, and Wang, 2009);
- Having medical insurance coverage (Baker, Zahs, and Popa, 2004);
- Being motivated to lose weight for health reasons (Baker, Zahs, and Popa, 2004);
- Expending effort to lose weight (Baker, Zahs, and Popa, 2004);
- Giving money to charity regularly (Taylor, Krane, and Thomas, 2005);
- Doing volunteer work (Taylor, Krane, and Thomas, 2005)
- Exercising regularly (Taylor, Krane, and Thomas, 2005);
- Going to a church, mosque, or synagogue most weeks (Taylor, Krane, and Thomas, 2005);
- Believing in God (Taylor, Krane, and Thomas, 2005);
- Cleaning one's teeth more than twice a day (Taylor, Krane, and Thomas, 2005).

It is easy to imagine how all the above attributes might be tinged with social desirability implications and that self-administered computer reporting might have been more honest than reports made to interviewers. An alternative explanation may be that the people who join online panels are more likely to truly have socially undesirable attributes and to report that accurately. And computer self-completion of questionnaires could lead to more accidental misreading and mistyping, yielding inaccurate reports of socially undesirable attributes. More direct testing is required to demonstrate whether higher rates of reporting socially undesirable attributes in Internet surveys is due to increased accuracy and not due to alternative explanations.

Thus, the bulk of this evidence can again be viewed as consistent with the notion that online surveys with nonprobability panels elicit more honest reports, but no solid body of evidence documents whether this is so because the respondents genuinely possess these attributes at higher rates or because the data collection mode elicits more honesty than interviewer-based methods.

As with computer administration generally, some researchers have pointed to *satisficing* as a potential cause of the differences observed in comparisons of results from Web surveys using nonprobability online panels with those from probability samples by interviewers. To test this proposition Chang and Krosnick (2009) administered the same questionnaire via RDD telephone and a Web survey using a nonprobability online panel. They found that the online survey yielded less non-differentiation, which is consistent with the claim that Web surveys elicit less satisficing.

Market research practitioners often use the term “inattentives” to describe respondents suspected of satisficing (Baker and Downes-LeGuin, 2007). In his study of 20 nonprobability panels, Miller (2008) found an average incidence of nine percent inattentives (or, as he refers to them, “mental cheaters”) in a

20-minute customer experience survey. The maximum incidence for a panel was 16 percent and the minimum 4 percent. He also fielded the same survey online to a sample of actual customers provided by his client and the incidence of inattentives in that sample was essentially zero. These results suggest that volunteer panelists may be more likely to satisfice than online respondents in general.

Thus far in this section we have considered research that might help us understand more clearly why results from nonprobability online panels might differ from those obtained by interviewers from probability samples. Much of this research has compared results from the two methods and simply noted differences but without looking specifically at the issue of accuracy. Another common technique for evaluating the accuracy of results from these different methods has been to compare results with external benchmarks established through non-survey means such as Census data, election outcomes, or industry sales data. In comparisons of nonprobability online panel surveys with RDD telephone and face-to-face probability sample studies, a number of researchers have found the latter two modes to yield more accurate measurements when compared to external benchmarks in terms of voter registration (Niemi, Portney, and King, 2008; though see Berrens, Bohara, Jenkins-Smith, Silva, and Weimer, 2003), turnout (Malhotra and Krosnick, 2007; Sanders, Clarke, Stewart, and Whiteley, 2007), vote choice (Malhotra and Krosnick, 2007; though see Sanders, Clarke, Stewart, and Whiteley, 2007), and demographics (Crete and Stephenson, 2008; Malhotra and Krosnick, 2007; Yeager et al., 2009). Braunsberger, Wybenga, and Gates (2007) reported the opposite finding: greater accuracy online than in a telephone survey.⁷ Krosnick, Nie, and Rivers (2005) found that while a single telephone RDD sample was off an average of 4.5 percent from benchmarks, six different nonprobability online panels were off an average of five percent to 12 percent, depending on the nonprobability sample supplier. In an extension of this same research, Yeager et al. (2009) found that the probability sample surveys (whether by telephone or Web) were consistently more accurate than the nonprobability sample surveys even after post-stratification by demographics. Results from a much larger study by the Advertising Research Foundation (ARF) using 17 panels have shown even greater divergence, although release of those results is only in the preliminary phase (Walker and Pettit, 2009).

Findings such as those showing substantial differences among nonprobability online panel suppliers inevitably lead to more questions about the overall accuracy of the methodology. If different firms independently conduct the same survey with nonprobability online samples simultaneously and the various sets of results closely resemble one another then researchers might take some comfort in the accuracy of the results. But disagreement would signal the likelihood of inaccuracy in some if not most such surveys. A number of studies in addition to those cited in the previous paragraph have arranged for the same survey to be conducted at the same time with multiple nonprobability panel firms (e.g., Almore-Yalch, Busby, and Britton, 2008; Baim, Galin, Frankel, Becker, and Agresti, 2009; Ossenbruggen, Vonk, and Williams, 2006). All of these studies found considerable variation from firm-to-firm in the results obtained with the same questionnaire, raising questions about the accuracy of the method.⁸

⁷ Braunsberger et al. (2007) did not state whether their telephone survey involved pure random digit dialing – they said it involved “a random sampling procedure” from a list “purchased from a major provider of such lists” (p. 761). And Braunsberger et al. (2007) did not describe the source of their validation data.

⁸ A series of studies at first appeared to be relevant to the issues addressed in this literature review, but closer inspection revealed that their data collections were designed in ways that prevented them from clearly addressing the issues of interest here (Boyle, Freeland, & Mulvany, 2005; Schillewaert & Meulemeester, 2005; Gibson & McAllister, 2008; Jackman, 2005; Stirton & Robertson, 2005; Kuran & McCaffery, 2004, 2008; Elo, 2010; Potoglou

Finally, a handful of studies have looked at concurrent validity across method. These studies administered the same questionnaire via RDD telephone interviews and via Web and nonprobability online panels and found evidence of greater concurrent validity and less measurement error in the Internet data (Berrens, Bohara, Jenkins-Smith, Silva, and Weimer, 2003; Chang and Krosnick, 2009; Malhotra and Krosnick, 2007; Thomas, Krane, Taylor, & Terhanian, 2008). Others found no differences in predictive validity (Sanders, Clarke, Stewart, and Whiteley, 2007; Crete and Stephenson, 2008).

In sum, the existing body of evidence shows that online surveys with nonprobability panels elicit systematically different results than probability sample surveys on a wide variety of attitudes and behaviors. Mode effects are one frequently-cited cause for those differences, premised on research showing that self-administration by computer is often more accurate than interviewer administration. But while computer administration offers some clear advantages, the literature to date also seems to show that the widespread use of nonprobability sampling in Web surveys is the more significant factor in the overall accuracy of surveys using this method. The limited available evidence on validity suggests that while volunteer panelists may describe themselves more accurately than do probability sample respondents, the aggregated results from online surveys with nonprobability panels are generally less accurate than those using probability samples.

Although the majority of Web surveys being done worldwide are with nonprobability samples, a small number are being done with probability samples. Studies that have compared these results from these latter surveys to RDD telephone surveys have sometimes found equivalent predictive validity (Berrens, Bohara, Jenkins-Smith, Silva, and Weimer, 2003) and rates of satisficing (Smith and Dennis, 2005) and sometimes found higher concurrent and predictive validity and less measurement error, satisficing, and social desirability bias in the Internet surveys, as well as greater demographic representativeness (Chang and Krosnick, 2009; Yeager et al., 2009) and greater accuracy in aggregate measurements of behaviors and attitudes (Yeager et al., 2009).

THE SPECIAL CASE OF PRE-ELECTION POLLS. Pre-election polls are perhaps the most visible context in which probability sample and non-probability sample surveys compete and can be evaluated against an objective benchmark – specifically, an election outcome. However, as tempting as it is to compare the accuracy of final polls across modes of data collection, one special aspect of this context limits the usefulness of such comparisons. Analysts working this area must make numerous decisions about how to identify likely voters, how to handle respondents who decline to answer vote choice questions, how to weight data, how to order candidate names on questionnaires, and more, so that differences between polls in their accuracy may reflect differences in these decisions rather than differences in the inherent accuracy of the data collection method. The leading pollsters rarely reveal the details of how they make these decisions for each poll, so it is impossible to take them fully into account.

A number of publications have compared the accuracy of final pre-election polls forecasting election outcomes (Harris Interactive, 1998, 2000, 2004, 2008; Stirton and Robertson, 2005; Taylor, Bremer, Overmeyer, Sigeel, and Terhanian, 2001; Twyman, 2008; Vavreck and Rivers, 2008). In general, these publications document excellent accuracy of online nonprobability sample polls (with some notable exceptions), some instances of better accuracy in probability sample polls, and some instances of lower accuracy than probability sample polls.

& Kanaroglou, 2008; Duffy, Terhanian, Bremer, & Smith, 2005; Sparrow & Curtice, 2004; Marta-Pedroso, Freitas, & Domingos, 2007)

RESPONDENT EFFECTS

No matter how we recruit respondents for our surveys, respondents will vary from each other in terms of their cognitive capabilities, motivations to participate, panel-specific experiences, topic interest and experience, and survey satisficing behaviors. These respondent-level factors can influence the extent of measurement error on an item-by-item basis and over the entire survey as well. While demographic variables may influence respondent effects, other factors likely have greater influence.

COGNITIVE CAPABILITIES. People who enjoy participating in surveys may have higher cognitive capabilities or a higher need for cognition (Cacioppo and Petty, 1982). If respondents join a panel or participate in a survey based on their cognitive capabilities or needs then it can lead to differences in results compared to samples selected independent of cognitive capabilities or needs. For example, in a self-administered survey, people are required to read and understand the questions and responses. The attrition rate of those who have lower education or lower cognitive capabilities is often higher in a paper-and-pencil or Web survey. Further, if the content of the survey is related to cognitive capabilities of the respondents (e.g., attitudes toward reading newspapers or books), then there may also be significant measurement error. A number of studies have indicated that people who belong to volunteer online panels are more likely to have higher levels of education than those in the general population (Malhotra and Krosnick, 2007). To the extent that this is related to their responses on surveys, such differences may either reduce or increase measurement error depending on the survey topic or target population.

MOTIVATION TO PARTICIPATE. Respondents also can vary in both the types of motivation to participate and in the strength of that motivation. For example, offering five dollars to a respondent to participate in a survey might be more enticing to those who make less money. Others might be more altruistic or curious and like participating in research more than others. Still others may want to participate in order to find out how their opinions compare with those of other people. There may be some people who are more than willing to participate in surveys about political issues but not consumer issues, so topical motivation can vary between respondents. Participation in a survey or in a panel may not be motivated by a single motive but by multiple motives that vary in strength and across time and topics. To the extent that motivation affects who participates and who does not, results may be biased and less likely to reflect the population of interest. While this potentially biasing effect of motivation occurs with other survey methods,⁹ it may apply even more to online surveys with nonprobability panels where people have self-selected into the panel to begin with and then can pick and choose the surveys to which they wish to respond. This is especially true when people are made aware of the nature and extent of incentives and even the survey topic prior to their participation by way of the survey invitation.

The use of incentives, in particular, whether to induce respondents to join a panel, to maintain their membership in a panel, or take a particular survey may lead to measurement error (Jäckle and Lynn, 2008). Respondents may over-report behaviors or ownership of products in order to obtain more rewards for participation in more surveys. Conversely, if they have experienced exceptionally long and boring

⁹ For example, people who answer the phone and are willing to complete an interview may be substantially different (e.g., older, more likely to be at home, poorer, more altruistic, more likely to be female, etc.) than those who do not.

surveys resulting from their reports of behaviors or ownership of products, they may under-report these things in subsequent surveys.

One type of respondent behavior observed with nonprobability online panels is false qualifying. In the language of online research these respondents are often referred to as “fraudulents” or “gamers” (Baker and Downes-LeGuin, 2007). These are individuals who assume false identities or simply misrepresent their qualifications either at the time of panel registration or in the qualifying questions of individual surveys. Their primary motive is assumed to be incentive maximization. They tend to be seasoned survey takers who can recognize filter questions and answer them in ways that they believe will increase their likelihood of qualifying for the survey. One classic behavior is selection of all options in multiple response qualifying questions; another is overstating purchase authority or span of control in a business-to-business (B2B) survey. Downes-Le Guin, Mechling, and Baker (2006) describe a number of first hand experiences with fraudulent panelists. For example, they describe a study in which the target respondents were both home and business decision makers to represent potential purchasers of a new model of printer. The study was multinational with a mix of sample sources including a U.S. customer list provided by the client, a U.S. commercial panel, a European commercial panel, and an Asian phone sample that was recruited to do the survey online. One multiple response qualifying question asked about the ownership of ten home technology products. About 14 percent of the U.S. panelists reported owning all 10 products, including the Segway Human Transporter, an expensive device known to have a very low incidence (less than 0.1 percent) of ownership among consumers. This response pattern was virtually nonexistent in the other sample sources.

The above examples are within the range of fraudulent reporting reported by Miller (2008). He found an average of about five percent fraudulent respondents across the 20 panels he studied with a maximum of 11 percent on one panel and a minimum of just 2 percent on four others. Miller also points out that while about five percent of panelists are likely to be fraudulent on a high incidence study that number can grow significantly—to as much as 40 percent—on a low incidence study where a very large number of panelist respondents are screened.

PANEL CONDITIONING. The experience of repeatedly taking surveys may lead to some respondents experiencing changes in attitudes or even behaviors as a consequence of survey participation. For example, completing a series of surveys about electoral politics might cause a respondent to pay closer attention to news stories on the topic, to become better informed and even to express different attitudes on subsequent surveys. Respondents who frequently do surveys about various kinds of products may become aware of new brands and report that awareness in future surveys. This type of change in respondent behavior or attitudes due to repeated survey completion is known as *panel conditioning*.

Concerns about panel conditioning arise because of the widespread belief that members of online panels complete substantially more surveys than, say, RDD telephone respondents. For example, Miller (2008), in his comparison study of 20 U.S. online panels, found an average of 33 percent of respondents reported taking 10 or more online surveys in the previous 30 days. Over half of the respondents on three of the panels he studied fell into this hyperactive group. One way for panelists to maximize their survey opportunities is by joining multiple panels. A recent study of 17 panels involving almost 700,000 panelists by ARF analyzed multi-panel membership and found either a 40 percent or 16 percent overlap in respondents, depending on how one measures it (Walker and Pettit, 2009). Baker and Downes-LeGuin (2007) report that in general population surveys rates of multi-panel membership (based on self reports)

of around 30 percent are not unusual. By way of comparison, they report that on surveys of physicians rates of multipanel membership may be 50 percent or higher, depending on specialty. General population surveys of respondents with few qualifying questions often show the lowest levels of hyperactivity, while surveys targeting lower incidence or frequently surveyed respondents can be substantially higher.

Whether this translates into measurable conditioning effects is still unclear. Coen, Lorch, and Piekarski (2005) found important differences in measures such as likelihood to purchase based on previous survey taking history with more experienced respondents generally being less positive than new panel members. Nancarrow and Cartwright (2007) found a similar pattern, although they also found that purchase intention or brand awareness was less affected when time between surveys was sufficiently long. Other research has found that differences in responses due to panel conditioning can be controlled when survey topics are varied from study to study within a panel (Dennis, 2001; Nukulkij, Hadfield, Subias, and Lewis, 2007).

On the other hand, a number of studies of consumer spending (Bailar, 1989; Silberstein and Jacobs, 1989), medical care expenditures (Corder and Horvitz, 1989), and news media consumption (Clinton, 2001) have found few differences attributable to panel conditioning. Studies focused on attitudes (rather than behaviors) across a wide variety subjects (Bartels, 2006; Sturgis, Allum, and Brunton-Smith, 2008; Veroff, Hatchett, and Douvan, 1992; Waterton and Lievesley, 1989; Wilson, Kraft, and Dunn, 1989) also have reported few differences.

Completing a large number of surveys might also cause respondents to approach survey tasks differently than those with no previous survey taking experience. It might lead to "bad" respondent behavior, including both weak and strong satisficing (Krosnick, 1999). Or the experience of completing many surveys might also lead to more efficient survey and accurate survey completion (Chang and Krosnick, 2009; Waterton and Lievesley, 1989; Schlackman, 1984). Conversely, Toepfel, Das, and van Soest (2008) compared the answering behavior of more experienced panel members with those less experienced and found few differences.

TOPIC INTEREST AND EXPERIENCE. Respondent experience with a topic can influence reactions to the questions about the topic. For example, if a company wants to measure people's feelings and thoughts about the company, they often invite people from all backgrounds to the survey. If, however, invitations or the survey itself screens respondents on the basis of their experience with the company, the resulting responses will generally tend to be more positive than if all respondents familiar with the company were asked to respond. Among the people who have not purchased in the past 30 days we are more likely to find people who have had negative experiences or who feel less positively toward the company. Therefore, the results from the sample will not give as accurate a picture of the company's reputation as it exists in the general population.

People who join online panels that complete consumer-based surveys may have greater consumer orientation, either through self-selection at the outset or through attrition. If this tendency exists and remains uncorrected, surveys using that panel might yield substantially different results concerning consumer attitudes and product demand among the general population.

While self-selection occurs in both RDD-recruited and nonprobability panels (as it does even for single-occasion randomly-selected samples), self-selection is more likely to be a stronger factor for respondents in nonprobability panels since there is strong self-selection at the first stage of an invitation to join the panel and at the single study stage where the survey topic is sometimes revealed. Stronger self-

selection factors can also yield respondents with a higher likelihood to be significantly different from all possible respondents in the larger population of interest. People who join panels voluntarily can differ from a target population in a number of ways (e.g., they may have less concern about their privacy, be more interested in expressing their opinions, be more technologically interested or experienced, be more involved in the community or political issues). For a specific study sample, this may be especially true when the topic of the survey is related to how the sample differs from the target population. For example, results from a survey assessing people's concerns about privacy may be significantly different in a volunteer panel than in the target population (Couper, Singer, Conrad, and Groves, 2008). For nonprobability online panels, attitudes toward technology may be more positive than the general population since respondents who are typically recruited from those who already have a computer and spend a good deal of time online. As a consequence, a survey concerning government policies toward improving computing infrastructure in the country may yield more positive responses in a Web nonprobability panel than in a sample drawn at random from the general population (Duffy et al., 2005). Chang and Krosnick (2009) and Malhotra and Krosnick (2007) found that in surveys using nonprobability panels, respondents were more interested in the topic of the survey (politics) than were respondents in face-to-face and telephone probability sample surveys.

6. SAMPLE ADJUSTMENTS TO REDUCE ERROR AND BIAS

While there may be considerable controversy surrounding the merits and proper use of nonprobability online panels, one thing virtually everyone agrees to is that the panels themselves are not representative of the general population. This section describes three techniques sometimes used to attempt to correct for this deficiency with the goal of making these results projectable to the general population.

PURPOSIVE SAMPLING

Purposive sampling is a non-random selection technique that has as its goal a sample that is representative of a defined target population. Anders Kiaer generally is credited for first advancing this sampling technique at the end of the 19th century with what he called “the representative method.” Kiaer argued that if a sample is representative of a population for which some characteristics are known then that sample also will be representative of other survey variables (Bethlehem and Stoop, 2007). Kish (1965) used the term *judgment sampling* to convey the notion that the technique relies on the judgments of experts about the specific characteristics needed in a sample for it to represent the population of interest. It presumes that an expert can make choices about the relationship between the topic of interest and the key characteristics that influence responses and their desired distributions based on knowledge gained in previous studies.

The most common form of purposive sampling is *quota sampling*. This technique has been widely used for many years in market and opinion research as a means to protect against nonresponse in key population subgroups and to reduce costs. Quotas typically are defined by a small set of demographic variables (age, gender, and region are common) and other variables thought to influence the measures of interest.

Purposive sampling is widely used by online panel companies to offer samples that correct for known biases in the panel itself. In the most common form of purposive sampling, the panel company provides a “census-balanced sample” meaning samples that are drawn to conform to the overall population demographics (typically age, gender, and perhaps, race) as measured by the U.S. Census. Individual researchers may request that the sample be stratified by other characteristics or they may implement quotas at the data collection stage to ensure the achieved sample meets their requirements.

More aggressive forms of purposive sampling use a wider set of both attitudinal and behavioral measures in sample selection. One advantage of panels is that the panel company often knows a good deal about individual members via profiling and past survey completions, and this information can be used in purposive selection. For example, Kellner (2008) describes the construction of samples for political polls in the UK that are drawn to ensure not just a demographic balance but also “the right proportions of past Labour, Conservative, and Liberal Democrat voters and also the right number of readers of each national newspaper.”

The use of purposive sampling and quotas, especially when demographic controls are used to set the quotas, is the basis on which results from online panel surveys are sometimes characterized as being “representative.”

The merits of purposive or quota sampling versus random probability sampling have been debated for decades and will not be reprised here. However, worthy of note is the criticism that purposive sampling relies on the judgment of an expert and so to a large degree the quality of the sample in the end depends on the soundness of that judgment. Where nonprobability online panels are concerned, there appears to be no research that focuses specifically on the reliability and validity of the purposive sampling aspects of online panels when comparing results with those from other methods.

MODEL-BASED METHODS

Probability sampling has a rich tradition, a strong empirical basis, and a well-established theoretical foundation, but it is by no means the only statistical approach to making inferences. Many sciences, especially the physical sciences, have rarely used probability sampling methods and yet they have made countless important discoveries using statistical data. These studies typically have relied on statistical models and assumptions and might be called model-based.

In the survey realm, small area estimation methods (Rao, 2003) have been developed to produce estimates for areas for which there are few or no observations. Prediction-based methods and Bayesian methods that either do not require probability sampling or ignore the sampling weights at the analysis stage have also been proposed (Valliant et al., 2001).

Epidemiological studies (Woodward, 2004) may be closely related to the types of studies that employ probability sampling methods. These studies often use some form of matching and adjustment methods to support inferences rather than relying on probability samples of the full target population (Rubin, 2006). An example is a case-control study in which controls (people without a specific disease) are matched to cases (people with the disease) to make inferences about the factors that might cause the disease.

Some online panels use approaches that are related to these methods. The most common approach of online panels has been to use propensity or other models to make inferences to the target population. At least one online panel (Rivers, 2007) has adopted a sample matching method for sampling and propensity modeling to make inferences in a manner closely related to the methods used in epidemiological studies.

Online panels are relatively new, and these ideas are still developing. Clearly, more theory and empirical evidence is needed to determine whether these approaches may provide valid inferences that meet the goals of the users of the data. Major hurdles that face nonprobability online panels are related to the validity and reproducibility of the inferences from these sample sources. To continue the epidemiological analogy, (external) validity refers to the ability to generalize the results from the study beyond the study subjects to the population while reproducibility (internal validity) refers the ability to derive consistent findings within the observation mechanism. Since many users of nonprobability online panels expect the results to be reproducible and to generalize to the general population, the ability of these panels to meet these requirements is a key to their utility.

In many respects the challenges for nonprobability panels are more difficult than those faced in epidemiological studies. All panels, even those that are based on probability samples, are limited in their ability to make inferences to dynamic populations. Changes in the population and attrition in the panel

may affect the estimates. In addition, online panels are required to produce a wide variety of estimates, as opposed to modeling a very specific outcome such as the incidence of a particular disease in most epidemiological studies. The multi-purpose nature of the requirement significantly complicates the modeling and the matching for the panel.

POST-SURVEY ADJUSTMENT

Without a traditional frame, the burden of post-survey adjustment for online nonprobability panels is much greater than in surveys with random samples from fully-defined frames. The gap between the respondents and the sample (that arises from nonresponse) is addressed through weighting procedures that construct estimates that give less weight to those respondents from groups with high response rates and more weight to those respondents from groups with low response rates. The gap between the sample and the sampling frame with probability samples is handled through very strong probability theory principles. The gap between the sampling frame and the target population is handled by using full target population counts from censuses or other sources, in an attempt to repair omissions from the frame.

Although a researcher working with a sample from an online volunteer panel may have counts or estimates for the full target population, there is no well-defined frame from which the respondents emerge. For that reason, post-survey adjustments to results from online panels take on the burden of moving directly from the respondents to the full target population.

WEIGHTING TECHNIQUES. For all surveys, regardless of mode or sample source, there is a chance that the achieved sample or set of respondents may differ from the target population of interest in nonignorable ways. This may be due to study design choices (e.g., the choice of the frame; analysis goals that require over-sampling) or to factors not easily controlled (e.g., the coverage of any given frame, nonresponse to the survey). Weights are often used to adjust survey data to help mitigate these compositional differences.

Compositional differences are an indication of possible bias. If people differ in their responses based upon some set of underlying characteristics and the people are not represented in their true proportions based upon these characteristics, estimates obtained will be biased or not representative of the population as a whole.

There are three main reasons why people might not be represented in their proper proportions in the survey and weighting adjustments may be needed to compensate for this over- or underrepresentation. First, weights may be needed to compensate for differences in selection probabilities of individual cases. Second, weights can help compensate for subgroup differences in response rates. Even if the sample selected is representative of the target population differences in response rate can compromise representation without adequate adjustments.

For either of the above situations, weighting adjustments can be made by using information from the sample frame itself. However, even if these types of weights are used, the sample of respondents may still fluctuate from known population characteristics, which leads to another type of weighting adjustment.

The third type of weight involves comparing the sample characteristics to an external source of data that is deemed to have high degree of accuracy. For surveys of individuals or households this information

often comes from sources such as the U.S. Census Bureau. This type of weighting adjustment is commonly referred to as a *post-stratification adjustment* and it differs from the first two types of weighting procedure in that it utilizes information external to the sample frame.

Online panels can have an underlying frame that is a probability based or nonprobability based frame. If the frame is probability based, all of the weighting methods mentioned above might apply and weights could be constructed accordingly.

Things are a bit different for a frame that is not probability based. Although cases may be selected at different rates from within the panel, knowing these probabilities tell us nothing about the true probabilities that would occur in the target population. The same basic problem holds true for sub-group response rates. Although sub-group response rates can often be measured for online panels, as with selection probabilities, it is difficult to tie them to the target population. For these reasons, weights for nonprobability panels typically rely solely upon post-stratification adjustments to external population targets.

The most common techniques to make an online panel more closely mirror the population at large occur either at the sample selection stage or after all data has been collected. At the selection stage, panel administrators may use purposive sampling techniques to draw samples that match the target population on key demographic measures. Panel administrators also may account and adjust for variation in response rates (based upon previous studies) related to these characteristics. The researchers may place further controls on the make-up of the sample through the use of quotas. Thus, a sample selected from this panel and fielded will yield a set of respondents that more closely matches the target population than a purely “random” sample from the online panel.

After data collection, post-stratification can be done by a weighting adjustment. Post-stratification can take different forms, the two most common of which are: (1) *cell-based weighting* where one variable or a set of variables is used to divide the sample into mutually exclusive categories or cells, with adjustments being made so that the sample proportions in each cell match the population proportions; or (2) *marginal-based weighting* whereby the sample is matched to the marginal distribution of each variable in a manner such that all the marginal distributions for the different categories will match the targets. For example, assume a survey uses three variables in the weighting adjustment: age (18-40 years old, 41-64 years old, and 65 years old or older), sex (male and female) and race/ethnicity (Hispanic, non-Hispanic white, non-Hispanic black, and non-Hispanic other race). Cell-based weighting will use 24 ($3 \times 2 \times 4$) cross-classified categories, where the weighted sample total of each category (e.g., the total number of Hispanic 41-64 year old males) will be projected the known target population total. By contrast, marginal-based weighting, which is known by several names including iterative proportional fitting, raking, and rim weighing, will make adjustments to match the respective marginal proportions for each category of each variable (e.g., Hispanic). Post-stratification relies on the assumption that people with similar characteristics on the variables used for weighing will have similar response characteristics for other items of interest. Thus, if samples can be put into their proper proportions, the estimates obtained from them will be more accurate (Berinsky, 2006). Work done by Dever, Rafferty, and Valliant (2008) however suggests that post-stratification based on standard demographic variables alone will likely fail to adequately adjust for all the differences between those with and without Internet access at home, but with the inclusion of sufficient variables they found that statistical adjustments alone could eliminate any coverage bias. However, their study did not address the additional differences associated with belonging

to a nonprobability panel. A study by Schonlau and colleagues (2009) casts doubt on using only a small set of variables in the adjustment.

PROPENSITY WEIGHTING. Weighting based on *propensity score adjustment* is another technique that is used in an attempt to make online panels selected as nonprobability samples more representative of the population. Propensity score adjustment was first introduced as a post-hoc approach to alleviate the confounding effects of the selection mechanism in observational studies by achieving a balance of covariates between comparison groups (Rosenbaum and Rubin, 1983). It is widely used in biostatistical applications involving quasi-experimental designs in an attempt to equate non-equivalent groups. It has its origin as a statistical solution to the selection problem (Caliendo and Kopeinig, 2008). It has been adopted in survey statistics mainly for weighting adjustment of telephone, mail, and face-to-face surveys (Lepkowski et al., 1989, Czajka et al, 1992, Iannacchione et al, 1991, Smith et al., 2001, Göksel, et al, 1991, Garren and Chang, 2002, Duncan and Stasny, 2001, Lee and Valliant, 2008) but not necessarily for sample selection bias issues.

Propensity score weighting was first introduced for use in online panels by Harris Interactive (Taylor, 2000, Terhanian and Bremer, 2000) and further examined by Lee and her colleagues (Lee, 2004, 2006; Lee and Valliant, 2008), Schonlau and his colleagues (Schonlau et al. 2004), Loosveldt and Sonck (2008), and others. Its purpose is to use propensity score models to reduce or eliminate the selection biases in samples from nonprobability panels by aligning the distributions of certain characteristics (covariates) within the panel to those of the target population.

Propensity score weighting differs from the traditional weighting techniques in two respects. First, it is based on explicitly specified models. Second, it requires the use of a supplemental or reference survey that is probability-based. The reference survey is assumed to be conducted parallel to the online survey with the same target population and survey period. Better coverage and sampling properties and higher response rates than the online survey are also expected for a reference survey. Furthermore, it is assumed that there are measurement error differences between the reference survey and the online survey. For instance, the reference survey may be conducted using traditional survey modes, such as RDD telephone in Harris Interactive's case (Terhanian and Bremer, 2000). The reference survey must include a set of variables that are also collected in the online surveys. These variables are used as covariates in propensity models. The hope is to use the strength of the reference survey to reduce selection biases in the online panel survey estimates. Schonlau, van Soest, and Kapteyn (2007) give an example of this.

By using data combining both reference and online panel surveys, a model (typically logistic regression) is built to predict whether a sample case is from the reference or the online survey. The covariates in the model can include items similar to the ones used in post-stratification, but other items are usually included that more closely relate to the likelihood of being on an online panel. Furthermore, propensity weighting can utilize not only demographic characteristics, but attitudinal characteristics as well. For example, people's opinions about current events can be used as these might relate a person's likelihood of choosing to be on an online panel. Because the technique requires a reference survey, items can be used that often don't exist from traditional population summaries, like the decennial census.

Once the model is developed, each case can then be assigned a predicted propensity score of being from the reference sample (a predicted propensity of being from the online sample could also be used). First, the combined sample cases are divided into equal sized groups based upon their propensity scores. (One might also consider using only reference sample cases for this division.) Ideally, all units in a given

subclass will have about the same propensity score or, at least, the range of scores in each class is fairly narrow. Based on the distribution of the proportions of reference sample cases across divided groups, the online sample is then assigned adjustment factors that can be applied to weights reflecting their selection probabilities.

Propensity score methods can be used alone or along with other methods, such as post-stratification. Lee and Valliant (2008) showed that weights that combine propensity score and calibration adjustments with general regression estimators were more effective than weighting by propensity score adjustments alone for online panel survey estimates that have a sample selection bias.

Propensity weighting still suffers from some of the same problems as more traditional weighting approaches and adds a few as well. The reference survey needs to be high quality. To reduce cost one reference study is often used to adjust a whole set of surveys. The selection of items to be used for the model is critical and can be depend on the topic for the survey. The reference study is often done with a different mode of administration, such as a telephone survey. This can complicate the modeling process if there are mode effects on responses, though items can be selected or designed to function equivalently in different modes. Moreover, the bias reduction from propensity score adjustment comes at the cost of increased variance in the estimates, therefore, decreasing the effective sample sizes and estimate precision (Lee, 2006). When the propensity score model is not effective, it can increase variance without decreasing bias, increasing the overall error in survey estimates. Additionally, the current practice for propensity score adjustment for nonprobability online panels is to treat the reference survey as though it were not subject to sampling errors, although typically reference surveys have small sample sizes. If the sampling error of the reference survey estimates is not taken into account, the precision of the online panel survey estimates using propensity score adjustment will be overstated (Bethlehem, 2009).

While propensity score adjustment can be applied to reduce biases, there is no simple approach for deriving variance estimates. As discussed previously, because online panels samples do not follow the randomization theory, the variance estimates cannot be interpreted as repeated sampling variances. Rather, they should be considered as reflecting the variance with respect to an underlying structural model that describes the volunteering mechanism and the dependence of a survey variable on the covariates used in adjustment. Lee and Valliant (2008) showed that naïvely using variance estimators derived from probability sampling may lead to a serious underestimation of the variance, erroneously causing Type 1 error. Also, when propensity score weighting is not effective in reducing bias, estimated variances are likely to have poor properties, regardless of variance estimators.

7. THE INDUSTRY-WIDE FOCUS ON PANEL DATA QUALITY

Over the last four or five years there has been a growing emphasis in the market research sector on online panel data quality (Baker, 2008). A handful of high profile cases in which online survey results did not replicate despite use of the same questionnaire and the same panel caused deep concern among some major client companies in the market research industry. One of the most compelling examples came from Kim Dedeker, Vice President for Global Consumer and Market Knowledge at Procter and Gamble when she announced at the 2006 Research Industry Summit on Respondent Cooperation, “Two surveys a week apart by the same online supplier yielded different recommendations ... I never thought I was trading data quality for cost savings.” At the same time, researchers working with panels on an ongoing basis began uncovering some of the troubling behaviors among panel respondents described in Section 5 (Downes-LeGuin, Mechling, and Baker, 2006). As a consequence, industry trade associations, professional organizations, panel companies, and individual researchers have all focused on the data quality issue and created differing responses to deal with it.

INITIATIVES BY PROFESSIONAL AND INDUSTRY TRADE ASSOCIATIONS

All associations in the survey research industry share a common goal of encouraging practices that promote quality research and the credibility of results in the eye of consumers of that research, whether clients or the public at large. Virtually every association both nationally and worldwide has incorporated some principles for conducting online research into their codes and guidelines. Space limitations make it impossible to describe them all here and so we note just four that seem representative.

The Council of American Survey Research Organization (CASRO) was the first US-based association to modify its “Code of Standards and Ethics for Survey Research” to include provisions specific to online research. A section on Internet research generally was added in 2002, and was revised in 2007 to include specific clauses relative to online panels. The portion of the CASRO code related to Internet research and panels is reproduced in Appendix A.

One of the most comprehensive code revisions has come from ESOMAR. Originally the European Association for Opinion and Market Research, the organization has taken on a global mission and now views itself as “the world association for enabling better research into markets, consumers, and societies.” In 2005 ESOMAR developed a comprehensive guideline titled, “Conducting Market and Opinion Research Using the Internet.” and incorporated it into their “International Code on Market and Social Research.” As part of that effort ESOMAR developed their “25 Questions to Help Research Buyers.” This document was subsequently revised and published in 2008 as “26 Questions to Help Research Buyers of Online Samples.” Questions are grouped into seven categories:

- Company profile;
- Sources used to construct the panel;
- Recruitment methods;

- Panel and sample management practices;
- Legal compliance;
- Partnership and multiple panel partnership;
- Data quality and validation.

The document specifies the questions a researcher should ask of a potential online panel sample provider along with a brief description of why the question is important. It is reproduced in Appendix B.

The ISO Technical Committee that developed ISO 20252 – Market, Opinion and Social Research also developed and subsequently deployed in 2009 an international standard for online panels, ISO 26362 – Access Panels in Market, Opinion, and Social Research (International Organization for Standardization, 2009). Like the main 20252 standard, ISO 26362 requires that panel companies develop, document, and maintain standard procedures in all phases of their operations and that they willingly share those procedures with clients upon request. The standard also defines key terms and concepts in an attempt to create a common vocabulary for online panels. It further details the specific kinds of information that a researcher is expected to disclose or otherwise make available to a client at the conclusion of every research project.

Finally, in 2008 the Advertising Research Foundation (ARF) established the Online Research Quality Council which in turn designed and executed The Foundations of Quality research project. The goal of the project has been to provide a factual basis for a new set of normative behaviors governing the use of online panels in market research. With data collection complete and analysis ongoing the ARF has turned to implementation via a number of test initiatives under the auspices of their Quality Enhancement Process (QeP). It is still too early to tell what impact the ARF initiative will have.

AAPOR has yet to incorporate specific elements related to Internet or online research into its code. However, it has posted statements on representativeness and margin of error calculation on its Web site. These are reproduced in Appendix C and D.

PANEL DATA CLEANING

Both panel companies and the researchers who conduct online research with nonprobability panels have developed a variety of elaborate and technically sophisticated procedures to remove “bad respondents.” The goal of these procedures, in the words of one major panel company (MarketTools, 2009), is to deliver respondents who are “real, unique, and engaged.” To be more specific, this means taking whatever steps are necessary to ensure that all panelists are who they say they are, that the same panelist participates in the same survey only once, and that the panelist puts forth a reasonable effort in survey completion.

ELIMINATING FRAUDULENTS. Assuming a false identity or multiple identities on the same panel is one form of fraud or misrepresentation. Validating the identities of all panelists is a responsibility that typically resides with the panel company. Most companies do this at the enrollment stage and a prospective member is not available for surveys until their identity has been verified. The specific checks vary from panel to panel but generally involve verifying information provided at the enrollment stage (e.g., name, address, telephone number, email address) against third-party databases. When identity

cannot be verified, the panelist is rejected. The specific checks done vary from one panel company to the next, but all reputable companies will supply details on their validation procedures if asked.

A second form of fraudulent behavior consists of lying in the survey's qualifying questions as a way to ensure participation. Experienced panel respondents understand that the first questions in a survey typically are used to screen respondents and so they may engage in behaviors that maximize their chances of qualifying. Market research surveys, for example, may be targeted at people who own or use certain products and those surveys often ask about product usage in a multiple response format. As described earlier, fraudulent respondents sometimes will select all products in the list in an attempt to qualify for the survey. Examining the full set of responses for respondents who choose all options in a multiple response qualifier is one technique for identifying fraudulent respondents. Surveys may also qualify people based on their having engaged in certain types of activities, the frequency with which they engage, or the number of certain products owned. When qualifying criteria such as these are collected over a series of questions the researcher can perform consistency checks among items or simply perform reasonableness tests to identify potential fraudulent respondents.

Increasingly researchers are designing questionnaires in ways to make it easier to find respondents who may be lying to qualify. For example, they may include false brands, nonexistent products, or low incidence behaviors in multiple choice questions. They might construct qualifying questions in ways that make it easier to spot inconsistencies in answers.

IDENTIFYING DUPLICATE RESPONDENTS. While validation checks performed at the join stage can prevent the same individual from joining the same panel multiple times, much of the responsibility for ensuring that the same individual does not participate in the same survey more than once rests with the researcher. It is reasonable to expect that a panel company has taken the necessary steps to eliminate multiple identities for the same individual and a researcher should confirm that prior to engaging the panel company. However, no panel company can be expected to know with certainty whether a member of their panel is also a member of another panel. In those instances where the researcher feels it is necessary or wise to use sample from multiple panels on the same study the researcher must also have a strategy for identifying and removing potential duplicate respondents.

The most common technique for identifying duplicate respondents is *digital fingerprinting*. Specific applications of this technique vary, but they all involve the capture of technical information about a respondent's IP address, browser, software settings and hardware configuration to construct a unique ID for that computer. (See Morgan (2008) for an example of a digital fingerprinting implementation.) Duplicate IDs in the same sample signal that the same computer was used to complete more than one survey and so a possible duplicate exists. False positives are possible (e.g., two persons in the same household), and so it is wise to review the entire survey for expected duplicates prior to deleting any data.

To be effective, digital fingerprinting must be implemented on a survey-by-survey basis. Many survey organizations have their own strategies and there are several companies who specialize in these services.

MEASURING ENGAGEMENT. Perhaps the most controversial set of techniques are those used to identify satisficing respondents. Four are commonly used:

1. Researchers often look at the full distribution of survey completion times to identify respondents with especially short times;

2. Grid or matrix style questions are a common feature of online questionnaires and respondent behavior in those questionnaires is another oft-used signal of potential satisficing. “Straightlining” answers in a grid by selecting the same response for all items in a grid or otherwise showing low differentiation in the response pattern can be an indicator of satisficing (though it could also indicate a poorly designed questionnaire). Similarly, random selection of response options can be a signal although this is somewhat more difficult to detect (high standard deviation around the average value selected by a respondent may or may not signal random responding). Trap questions in grids that reverse polarity are another technique (though this may reflect questions that are more difficult to read and respond to).;
3. Excessive selection of non-substantive responses such as “don’t know” or “decline to answer” are still another potential indicator of inattentiveness (though they could also reflect greater honesty).;
4. Finally, examination of responses to open-ended questions can sometimes identify problematic respondents. Key things to look for are gibberish or answers that appear to be copied and then repeatedly pasted in question after question.

PUTTING IT ALL TOGETHER. There is no widely accepted industry standard for editing and cleaning panel data. Which, if any, of these techniques is used for a given study is left to the judgment of individual researchers. Similarly, how the resulting data are interpreted and the action taken against specific cases varies widely. Failure on a specific item such as a duplication check or fraudulent detection sequence may be enough to cause a researcher to delete a completed survey from the dataset. Others may use a scoring system where a case must fail on multiple tests before it is eliminated. This is especially true with attempts to identify inattentive responses. Unfortunately, there is nothing in the research literature to help us understand how significantly any of these respondent behaviors affect estimates.

This editing process may strike researchers accustomed to working with probability samples as a strange way to ensure data quality. Eliminating respondents because of their response patterns is not typically done with these kinds of samples. On the other hand, interviewers are trained to recognize some of the behaviors and take steps to correct them during the course of the interview.

We know of no research that shows the effect these kinds of edits either on the representativeness of these online surveys or their estimates. Nonetheless, these negative respondent behaviors are widely believed to be detrimental to data quality.

8. CONCLUSIONS/RECOMMENDATIONS

We believe that the foregoing review, while not exhaustive of the literature, is at least comprehensive in terms of the major issues researchers face with online panels. But research is ongoing and both the panel paradigm itself and the methods for developing online samples more generally continue to evolve. On the one hand, the conclusions that follow flow naturally from the state of the science as we understand it today. Yet they also are necessarily tentative as that science continues to evolve.

Researchers should avoid nonprobability online panels when one of the research objectives is to accurately estimate population values. There currently is no generally accepted theoretical basis from which to claim that survey results using samples from nonprobability online panels are projectable to the general population. Thus, claims of “representativeness” should be avoided when using these sample sources. Further, empirical research to date comparing the accuracy of surveys using nonprobability online panels with that of probability-based methods finds that the former are generally less accurate when compared to benchmark data from the Census or administrative records. From a total survey error perspective, the principal source of error in estimates from these types of sample sources is a combination of the lack of Internet access in roughly one in three U.S. households and the self-selection bias inherent in the panel recruitment processes.

Although mode effects may account for some of the differences observed in comparative studies, the use of nonprobability sampling in surveys with online panels is likely the more significant factor in the overall accuracy of surveys using this method. The majority of studies comparing results from surveys using nonprobability online panels with those using probability-based methods (most often RDD telephone) report significantly different results on a wide array of behaviors and attitudes. Explanations for those differences sometimes point to classic measurement error phenomena such as social desirability response bias and satisficing. And indeed, the literature confirms that in many cases self administration by computer results in higher reports of socially undesirable behavior and less satisficing than in interviewer administered modes. Unfortunately, many of these studies confound mode with sample source, making it difficult to separate the impact of mode of administration from sample source. A few studies have attempted to disentangle these influences by comparing survey results from different modes to external benchmarks such as the Census or administrative data. These studies generally find that surveys using nonprobability online panels are less accurate than those using probability methods. Thus, we conclude that while measurement error may explain some of the divergence in results across methods the greater source of error is likely to be the undercoverage and self selection bias inherent in nonprobability online panels.

There are times when a nonprobability online panel is an appropriate choice. To quote Mitofsky (1989), “...different surveys have different purposes. Defining standard methodological practices when the purpose of the survey is unknown does not seem practical. Some surveys are conducted under circumstances that make probability methods infeasible if not impossible. These special circumstances require caution against unjustified or unwarranted conclusions, but frequently legitimate conclusions are possible and sometimes those conclusions are important.” The quality expert J. M. Juran (1992) expressed this concept more generally when he coined the term “fitness for use” and argued that any

definition of quality must include discussion of how a product will be used, who will use it, how much it will cost to produce it, and how much it will cost to use it. Not all survey research is intended to produce precise estimates of population values. For example, a good deal of research is focused on improving our understanding of how personal characteristics interact with other survey variables such as attitudes, behaviors and intentions. Nonprobability online panels also have proven to be a valuable resource for methodological research of all kinds. Market researchers have found these sample sources to be very useful in testing the receptivity of different types of consumers to product concepts and features. Under these and similar circumstances, especially when budget is limited and/or time is short, a nonprobability online panel can be an appropriate choice. However, researchers also should carefully consider any biases that might result due to the possible correlation of survey topic with the likelihood of Internet access, the propensity to join an online panel, or to respond to and complete the survey and qualify their conclusions appropriately.

Research aimed at evaluating and testing techniques used in other disciplines to make population inferences from nonprobability samples is interesting but inconclusive. Model-based sampling and sample management have been shown to work in other disciplines but have yet to be tested and applied more broadly. While some have advocated the use of propensity weighting in post-survey adjustment to represent the intended population, the effectiveness of these different approaches has yet to be demonstrated consistently and on a broad scale. Nonetheless, this research is important and should continue.

Users of online panels should understand that there are significant differences in the composition and practices of individual panels that can affect survey results. It is important to choose a panel sample supplier carefully. One obvious difference among panels that is likely to have a major impact on the accuracy of survey results is method of recruitment. Panels using probability-based methods such as RDD telephone or addressed-based mail sampling are likely to be more accurate than those using nonprobability-based methods, assuming all other aspects of survey design are held constant. Other panel management practices such as recruitment source, incentive programs, and maintenance practices also can have major impacts on survey results. Arguably the best guidance available on this topic is the ESOMAR publication, *26 Questions to Help Research Buyers of Online Samples*, included as Appendix B to this report. Many panel companies have already answered these questions on their Web sites, although words and practices sometimes do not agree. Seeking references from other researchers may also be helpful.

Panel companies can inform the public debate considerably by sharing more about their methods and data describing outcomes at the recruitment, join, and survey-specific stages. Despite the large volume of research that relies on these sample sources we know relatively little about the specifics of undercoverage or nonresponse bias. Such information is critical to fit-for-purpose design decisions and attempts to correct bias in survey results.

Disclosure is critical. O’Muircheartaigh (1997) proposed that error be defined as “work purporting to do what it does not do.” Much of the controversy surrounding use of online panels is rooted in claims that may or may not be justified given the methods used. Full disclosure of the research methods used is a bedrock scientific principle and a requirement for survey research long-championed by AAPOR. Disclosure is the only means by which the quality of research can be judged and results replicated. Full and complete disclosure of how results were obtained is a requirement for all survey research regardless of methodology. The disclosure standards included in the *AAPOR Code of Professional Ethics and*

Practice are an excellent starting. Researchers also may wish to review the disclosure standards required in ISO 20252 and, especially, ISO 26362. Of particular interest is the calculation of a within-panel “participation rate” in place of a response rate, the latter being discouraged by the ISO standards except when probability samples are used. The participation rate is defined as “the number of respondents who have provided a usable response divided by the total number of initial personal invitations requesting participation.”¹⁰

AAPOR should consider producing its own “Guidelines for Internet Research” or incorporate more specific references to online research in its code. AAPOR has issued a number of statements on topics such as representativeness of Web surveys and appropriateness of margin of error calculation with nonprobability samples. These documents are included as Appendix C and D respectively. AAPOR should consider whether these statements represent its current views and revise as appropriate. Its members and the industry at large also would benefit from a single set of guidelines that describe what AAPOR believes to be appropriate practices when conducting research online across the variety of sample sources now available.

Better metrics are needed. There are no widely-accepted definitions of outcomes and methods for calculation of rates similar to AAPOR’s *Standard Definitions* (2009) that allow us to judge the quality of results from surveys using online panels. For example, while the term “response rate” is often used with nonprobability panels the method of calculation varies and it is not at all clear how analogous those methods are to those described in *Standard Definitions*. Although various industry bodies are active in this area we are still short of consensus. AAPOR may wish to take a leadership position here much as it has with metrics for traditional survey methods. One obvious action would be to expand *Standard Definitions* to include both probability and nonprobability panels.

Research should continue. Events of the last few years have shown that despite the widespread use of online panels there still is a great deal about them that is not known with confidence. There continues to be considerable controversy surrounding their use. The forces that have driven the industry to use online panels will only intensify going forward, especially as the role of the Internet in people’s lives continues to expand. AAPOR, by virtue of its scientific orientation and the methodological focus of its members is uniquely positioned to encourage research and disseminate its findings. It should do so deliberately.

¹⁰ We should note that while response rate is a measure of survey quality participation rate is not. It is a measure of panel efficiency.

9. REFERENCES AND ADDITIONAL READINGS

- AAPOR (2009). *Standard Definitions: Final Dispositions of Case Codes and Outcomes for Surveys*.
- AAPOR (2008). *Guidelines and Considerations for Survey Researchers When Planning and Conducting RDD and Other Telephone Surveys in the U.S. With Respondents Reached via Cell Phone Numbers*.
- Abate, T. (1998). "Accuracy of On-line Surveys May Make Phone Polls Obsolete." *The San Francisco Chronicle*, D1.
- Aguinis, H., Pierce, C.A., & Quigley, B.M. (1993). "Conditions under which a bogus pipeline procedure enhances the validity of self-reported cigarette smoking: A meta-analytic review." *Journal of Applied Social Psychology* 23: 352-373.
- Alvarez, R.M., Sherman, R., & VanBeselaere, C. (2003). "Subject acquisition for Web-based surveys," *Political Analysis*, 11, 1, 23-43.
- Baim, J., Galin, M., Frankel, M. R., Becker, R., & Agresti, J. (2009). *Sample surveys based on Internet panels: 8 years of learning*. New York, NY: Mediamark.
- Baker, R. (2008). "A Web of Worries," *Research World*, 8-11.
- Baker, R. & Downes-LeGuin, T. (2007). "Separating the Wheat from the Chaff: Ensuring Data Quality in Internet Panel Samples." in *The Challenges of a Changing World; Proceedings of the Fifth International Conference of the Association of Survey Computing*. Berkeley, UK: ASC.
- Baker, R., Zahs, D., & Popa, G. (2004). "Health surveys in the 21st century: Telephone vs. web" in *Cohen SB, Lepkowski JM, eds., Eighth Conference on Health Survey Research Methods*, 143-148. Hyattsville, MD: National Center for Health Statistics.
- Bandilla, W., Bosnjak, M. & Altdorfer, P. (2003). "Survey Administration Effects?: A Comparison of Web and Traditional Written Self-Administered Surveys Using the ISSP Environment Module." *Social Science Computer Review* 21: 235-243
- Bartels, L. M. (2006). "Three Virtues of Panel Data for the Analysis of Campaign Effects." *Capturing Campaign Effects*, ed. Henry E. Brady and Richard Johnston. Ann Arbor, MI: University of Michigan Press.
- Bender B, Bartlett SJ, Rand CS, Turner CF, Wamboldt FS, & Zhang L. (2007) "Impact of Reporting Mode on Accuracy of Child and Parent Report of Adherence with Asthma Controller Medication." *Pediatrics*, 120: 471-477.
- Berinsky, A.J. (2006). "American Public Opinion in the 1930s and 1940s: The Analysis of Quota-controlled Sample Survey Data." *Public Opinion Quarterly* 70:499-529.
- Berrens, R. P., Bohara, A. K., Jenkins-Smith, H., Silva, C., & Weimer, David L. (2003). "The Advent of Internet Surveys for Political Research: A Comparison of Telephone and Internet Samples." *Political Analysis* 11:1-22.

- Bethlehem, J. (2008a). "Can we make official statistics with self-selection web surveys?" in *Statistics Canada's International Symposium Series*. Catalogue Number 11-522-X.
- Bethlehem, J. (2008b). "How accurate are self-selection web surveys?" Discussion paper 08014, Statistics Netherlands. The Hague: Statistics Netherlands.
- Bethlehem, J. & Stoop, I. (2007), "Online Panels – A Theft of a Paradigm?" in *The Challenges of a Changing World; Proceedings of the Fifth International Conference of the Association of Survey Computing*. Berkeley, UK: ASC.
- Biemer, P.P. (2001). Nonresponse bias and measurement bias in a comparison of face to face and telephone interviewing. *Journal of Official Statistics*, 17, 295-320.
- Biemer, P. P., Groves, R. M., Lyberg, L. E., Mathiowetz, N. A., and Sudman, S. (Eds.) (1991). *Measurement errors in surveys*. New York: John Wiley and Sons.
- Biemer, P., & Lyberg, L.E. (2003). *Introduction to Survey Quality*. Wiley Series in Survey Methodology. Hoboken, NJ: John Wiley and Sons, Inc.
- Black, G. S., & Terhanian, G. (1998). "Using the Internet for Election Forecasting." *The Polling Report* October 26.
- Blankenship, A.B., Breen, G., & Dutka, A. (1998). *State of the Art Marketing Research*. Second edition, Chicago, IL: American Marketing Association.
- Blumberg, S.J. & Luke, J.V. (2008). "Wireless substitution: Early release of estimates from the National Health Interview Survey, July to December 2007." National Center for Health Statistics. Available from: <http://www.cdc.gov/nchs/nhis.htm>. May 13, 2008.
- Birnbaum, M. H. (2004). "Human Research and Data Collection via the Internet." *Annual Review of Psychology* 55:803-822.
- Bowling, A. (2005). Mode of questionnaire administration can have serious effects on data quality. *Journal of Public Health*, 27, 281-291.
- Boyle, J.M., Freeman, G. & Mulvany, L. "Internet Panel Samples: A Weighted Comparison of Two National Taxpayer Surveys." Paper presented at the 2005 Federal Committee on Statistical Methodology Research Conference.
- Braunsberger, K., Wybenga, H., & Gates, R. (2007). "A comparison of reliability between telephone and web-based surveys." *Journal of Business Research* 60, 758-764.
- Burke. (2000). "Internet vs. telephone data collection: Does method matter?" Burke White Paper 2(4).
- Burn, M., & Thomas, J. (2008). "Do we really need proper research any more? The importance and impact of quality standards for online access panels." ICM White Paper. London, UK: ICM Research.
- Cacioppo, J, T. & Petty, R. (1982). "The Need for Cognition." *Journal of Personality and Social Psychology* 42:116-131.
- Caliendo, M. & Kopeinig, S. (2008), "Some Practical Guidance for the Implementation of Propensity Score Matching." *Journal of Economic Surveys* 22: 31-72

- Callegaro, M., & Disogra, C., (2008). "Computing Response Metrics for Online Panels." *Public Opinion Quarterly*, 72(5): 1008-32.
- CAN-SPAM. <http://www.ftc.gov/bcp/edu/pubs/business/e-commerce/bus61.shtm>
- Cartwright, T., & Nancarrow, E. (2006). "The effect of conditioning when re-interviewing: A pan-European study," Panel Research 2006: ESOMAR World Research Conference. Amsterdam: ESOMAR.
- CASRO. (2009). <http://www.casro.org/codeofstandards.cfm>.
- Chang, L.C. & Krosnick, J.A. (2001). "National Surveys via RDD Telephone Interviewing vs. the Internet: Comparing Sample Representativeness and Response Quality." Paper presented at the 56th Annual Conference of the American Association for Public Opinion Research, Montreal.
- Chang, L., & Krosnick, J.A. (2010). "Comparing oral interviewing with self-administered computerized questionnaires: An experiment." *Public Opinion Quarterly*.
- Chang, L., & Krosnick, J.A. (2009). "National surveys via RDD telephone interviewing versus the internet: Comparing sample representativeness and response quality." *Public Opinion Quarterly*.
- Chatt, C., & Dennis, J. M. (2003). "Data collection mode effects controlling for sample origins in a panel survey: Telephone versus internet." Paper presented at the 2003 Annual Meeting of the Midwest Chapter of the American Association for Public Opinion Research, Chicago, IL.
- Christian, L. M., Dillman, D. & Smyth, J.D. (2006). "The Effects of Mode and Format on Answers to Scalar Questions in Telephone and Web Surveys." Paper presented at the 2nd International Conference on Telephone Survey Methodology, Miami, Florida.
- Christian, L. M., Dillman, D. A., and Smyth, J. D. (2008). "The effects of mode and format on answers to scalar questions in telephone and web surveys." In *J. M. Lepkowski et al. (Eds.). Advances in telephone survey methodology*. New York: John Wiley and Sons, 250-275.
- Clinton, J.D. (2001). "Panel Bias from Attrition and Conditioning: A Case Study of the Knowledge Networks Panel." Stanford.
- Coen, T., Lorch, J. & Piekarski, L. (2005). "The Effects of Survey Frequency on Panelists' Responses." *Worldwide Panel Research: Developments and Progress*. Amsterdam: ESOMAR.
- Comly, P. (2007). "Online Market Research." In *Market Research Handbook*, ed. ESOMAR, pp. 401-20, Hoboken, NJ: Wiley.
- Comly, P. (2005). "Understanding the Online Panelist." *Worldwide Panel Research: Developments and Progress*. Amsterdam: ESOMAR.
- Converse, P. E., & Traugott, M.W. (1986). "Assessing the Accuracy of Polls and Surveys." *Science* 234: 1094-1098.
- Cooke, M., Watkins, N., & Moy, C. (2007). "A hybrid online and offline approach to market measurement studies." *International Journal of Market Research*, 52, 29-48.
- Cooley, P.C., Rogers, S.M., Turner, C.F., Al-Tayyib, A.A., Willis, G., & Ganapathi, L. (2001). "Using touch screen audio-CASI to obtain data on sensitive topics." *Computers in Human Behavior*, 17: 285-293.

- Couper, M. P. (2008). *Designing Effective Web Surveys*. New York: Cambridge University Press.
- Couper, M. P. (2000). "Web Surveys: A Review of Issues and Approaches." *Public Opinion Quarterly* 64:464-494.
- Couper, M. P., Traugott, M.W., & Lamias, M.J. (2001). "Web Survey Design and Administration." *Public Opinion Quarterly* 65:230-253.
- Couper, M.P., Singer, E., Conrad, F., & Groves, R. (2008). "Risk of Disclosure, Perceptions of Risk, and Concerns about Privacy and Confidentiality as Factors in Survey Participation." *Journal of Official Statistics* 24: 255-275.
- Crete, J., & Stephenson, L. B. (2008). "Internet and telephone survey methodology: An evaluation of mode effects." Paper presented at the annual meeting of the MPSA, Chicago, IL.
- Curtin, R., Presser, S., & Singer, E. (2005). "Changes in Telephone Survey Nonresponse over the Past Quarter Century." *Public Opinion Quarterly* 69:87-98.
- Dever, Jill A., Rafferty, Ann, & Valliant, Richard. (2008). "Internet Surveys: Can Statistical Adjustments Eliminate Coverage Bias?" *Survey Research Methods* 2: 47-62.
- De Leeuw, Edith D. (2005). "To Mix or Not to Mix Data Collection Modes in Surveys." *Journal of Official Statistics* 21:233-255.
- Denscombe, Martyn. (2006). "Web Questionnaires and the Mode Effect." *Social Science Computer Review* 24: 246-254.
- Des Jarlais DC, Paone D, Milliken J, Turner CF, Miller H, Gribble J, Shi Q, & Hagan H, Friedman. (1999). "Audio-computer interviewing to measure risk behaviour for HIV among injecting drug users: a quasi-randomised trial." *Lancet*, 353(9165): 1657-61.
- Dillman, Don A. & Leah Melani Christian. (2005). "Survey Mode as a Source of Instability in Responses Across Surveys." *Field Methods* 17:30-52.
- Dillman, Don A., Smyth, Jolene D., & Christian, Leah Melani. (2009). *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method (3rd Ed.)*, Hoboken, NJ: Wiley.
- DMS Research. (2009). "The Devil Is in the Data". *Quirks Marketing Research Review April 2009*.<http://www.quirks.com/search/articles.aspx?search=DMS+Research&searched=39711685>.
- Downes-Leguain, T., Meckling, J., & Baker, R. (2006). "Great results from ambiguous sources: Cleaning Internet panel data." *Panel Research 2006: ESOMAR World Research Conference*. Amsterdam: ESOMAR.
- Duffy, Bobby, Smith, Kate, Terhanian, George, & Bremer, John. (2005). "Comparing Data from Online and Face-to-Face Surveys." *International Journal of Market Research* 47:615-639.
- Elmore-Yalch, R., Busby, J., & Britton, C. (2008). "Know thy customer? Know thy research!: A comparison of web-based & telephone responses to a public service customer satisfaction survey." Paper presented at the TRB 2008 Annual Meeting.
- Elo, Kimmo. (2010). "Asking Factual Knowledge Questions: Reliability in Web-Based, Passive Sampling Surveys." *Social Science Computer Review* 28.

- Emerson, Ralph Waldo. 1841. *Essays: First Series*.
- Fazio, Russell H., Lenn, T. M., & Effrein, E. A. (1984). "Spontaneous Attitude Formation." *Social Cognition* 2:217-234.
- Fendrich, M., Mackesy-Amity, M. E., Johnson, T. P., Hubbell, A., & Wislar, J. S. (2005). Tobacco-reporting validity in an epidemiological drug-use survey. *Addictive Behaviors*, 30, 175–181.
- Fine, Brian, Menictas, Con, & Casdas, Dimitrio. (2006). "Comparing people who belong to multiple versus single panels," Panel Research 2006: ESOMAR World Research Conference. Amsterdam: ESOMAR.
- Fricker, Scott, Galesic, Mirta, Tourangeau, Roger, & Yan, Ting. (2005). "An Experimental Comparison of Web and Telephone Surveys." *Public Opinion Quarterly* 69:370-392.
- Frisina, Laurin T., Krane, David, Thomas, Randall K., & Taylor, Humphrey. (2007). "Scaling social desirability: Establishing its influence across modes." Paper presented at the 62nd Annual Conference of the American Association for Public Opinion Research in Anaheim, California.
- Galesic, M. & M. Bosnjak,. (2009). "Effects of Questionnaire Length on Participation and Indicators of Response Quality in Online Surveys." *Public Opinion Quarterly* 73: 349-360.
- Garland, P., Santus, D., & Uppal, R.. (2009). "Survey Lockouts: Are we too cautious." Survey Sampling Intl. white paper. http://www.surveysampling.com/sites/all/files/SSI_SurveyLockouts_0.pdf
- Ghanem, K.G., Hutton, H. E., Zenilman, J. M., Zimba, R., & Erbeling, E. J. (2005). "Audio computer assisted self interview and face to face interview modes in assessing response bias among STD clinic patients." *Sex Transm Infect*, 81: 421–425.
- Gibson, R., & McAllister, I. (2008). "Designing online election surveys: Lessons from the 2004 Australian Election." *Journal of Elections, Public Opinion, and Parties* 18: 387-400.
- Gosling, S. D., Vazire, S., Srivastava, S. & John, O. P. (2004). "Should we trust Web studies?: A comparative analysis of six preconceptions about Internet questionnaires." *American Psychologist* 59:93-104.
- Groves, R. M. (1989). *Survey Errors and Survey Costs*. New York: John Wiley and Sons.
- Groves, R. M. (2006). "Nonresponse Rates and Nonresponse Bias in Household Surveys." *Public Opinion Quarterly* 70: 646–675.
- Groves, R., Brick, M., Smith, R., & Wagner, J. (2008). "Alternative Practical Measures of Representativeness of Survey Respondent Pools," presentation at the 2008 AAPOR meetings.
- Gundersen, D. A. (2007). "Mode effects on cigarette smoking estimates: Comparing CAPIR and CATI responders in the 2001/02 Current Population Survey." Paper presented at the APHA Annual Meeting and Expo, Washington, DC.
- Harris Interactive (2008). "Election results further validate efficacy of Harris Interactive/s Online Methodology." Press Release from Harris Interactive, November 6, 2008.
- Harris Interactive. (2004). "Final pre-election Harris Polls: Still too close to call but Kerry makes modest gains." The Harris Poll #87, November 2, 2004. http://www.harrisinteractive.com/harris_poll/index.asp?pid=515.

- Hasley, S. (1995). "A comparison of computer-based and personal interviews for the gynecologic history update." *Obstetrics and Gynecology*, 85, 494-498.
- Hennigan, K. M., Maxson, C. L., Sloane, D., & Ranney, M. (2002). "Community views on crime and policing: Survey mode effects on bias in community surveys." *Justice Quarterly* 19, 565-587.
- Heerwegh, D. & Loosveldt, G. (2008). "Face-to-face versus web surveying in a high Internet coverage population: differences in response quality," *Public Opinion Quarterly* 72, 836-846.
- Holbrook, Allyson L., Green, Melanie C. & Krosnick, Jon A. (2003). "Telephone versus Face-to-Face Interviewing of National Probability Samples with Long Questionnaires: Comparisons of Respondent Satisficing and Social Desirability Response Bias." *Public Opinion Quarterly* 67:79-125.
- Hoogendoorn, Adriaan W. & Daalmans, Jacco (2009). "Nonresponse in the Recruitment of an Internet Panel Based on Probability Sampling." *Survey Research Methods* 3: 59-72.
- Jäckle, Annette & Lynn, Peter. (2008). "Respondent incentives in a multi-mode panel survey: cumulative effects on non-response and bias." *Survey Methodology*, 34.
- Jackman, S. (2005). "Pooling the polls over an election campaign." *Australian Journal of Political Science* 40: 499-517.
- Inside Research (2009). "U.S. Online MR Gains Drop." 20(1), 11-134.
- International Organization for Standardization, (2009). *ISO 26362:2009 Access panels in market, opinion, and social research- Vocabulary and service requirements*. Geneva.
- International Organization for Standardization. (2006). *ISO 20252: 2006 Market, opinion, and social research- Vocabulary and service requirements*. Geneva.
- Juran, J.M. (1992). *Juran on Quality by Design: New Steps for Planning Quality into Goods and Services*. New York: The Free Press.
- Keeter, S, Kennedy, C., Dimock, M., Best, J., & Craighill. P. (2006). "Gauging the Impact of Growing Nonresponse on Estimates from a National RDD Telephone Survey." *Public Opinion Quarterly* 70:759-779.
- Keeter, S., Miller, C., Kohut, A., Groves, R.M. & Presser, S.. (2000). "Consequences of Reducing Nonresponse in a National Telephone Survey." *Public Opinion Quarterly* 64:125-148
- Kellner, P. (2008). "Down with random samples." *Research World*, June, 31.
- Kellner, P. (2004). "Can Online Polls Produce Accurate Findings?" *International Journal of Market Research* 46: 3-21.
- Kish, L. (1965). *Survey Sampling*. New York: John Wiley and Sons.
- Klein, J. D., Thomas, R. K., & Sutter, E. J. (2007). "Self-reported smoking in online surveys: Prevalence estimate validity and item format effects." *Medical Care* 45: 691-695.
- Kohut, A., Keeter, S., Doherty, C. & Dimock, M.. (2008). "The Impact of 'Cell-onlys' on Public Opinion Polling: Ways of Coping with a Growing Population Segment. " The Pew Research Center. Available from <http://people-press.org>, January 31, 2008.

- Knapton, K. & Myers, S. (2005). "Demographics and Online Survey Response Rates." *Quirk's Marketing Research Review*: 58-64.
- Krane, D., Thomas, R.K., & Taylor, H. (2007). "Presidential approval measures: Tracking change, predicting behavior, and cross-mode comparisons." Paper presented at the 62nd Annual Conference of the American Association for Public Opinion Research in Anaheim, California.
- Kreuter, F., Presser, S., & Tourangeau, R. (2008). "Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity." *Public Opinion Quarterly* 72:847-865.
- Krosnick, J. A. (1999). "Survey Research." *Annual Review of Psychology* 50:537-567.
- Krosnick, J. A. (1991). "Response Strategies for Coping with Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology* 5: 213-236.
- Krosnick, J. A., Nie, N., & Rivers, D. (2005). "Web Survey Methodologies: A Comparison of Survey." Paper presented at the 60th Annual Conference of the American Association for Public Opinion Research in Miami Beach, Florida.
- Krosnick, J. A., & Alwin, D.F. (1987). "An Evaluation of a Cognitive Theory of Response Order Effects in Survey Measurement." *Public Opinion Quarterly*, 51, 201-219.
- Kuran, Timur & McCaffery, Edward J. (2004). "Expanding Discrimination Research: Beyond Ethnicity and to the Web." *Social Science Quarterly* 85.
- Lee, S. & Valliant, R. (2009). "Estimation for Volunteer Panel Web Surveys Using Propensity Score Adjustment and Calibration Adjustment." *Sociological Methods and Research* 37: 319-343
- Lindhjem, H., & Stale, N. (2008). "Internet CV surveys – a cheap, fast way to get large samples of biased values?" *Munich Personal RePEc Archive Paper* 11471. <http://mpra.ub.uni-muenchen.de/11471>.
- Link, M. W., & Mokdad, A. H. (2005). "Alternative modes for health surveillance surveys: An experiment with web, mail, and telephone." *Epidemiology* 16, 701-704.
- Link, M., & A. Mokdad (2004). "Are Web and Mail Feasible Options for the Behavioral Risk Factor Surveillance System?" in *Cohen SB, Lepkowski JM, eds., Eighth Conference on Health Survey Research Methods*, 149-158. Hyattsville, MD: National Center for Health Statistics.
- Link, M. W., Battaglia, M.P., Frankel, M. R., Osborn, Larry, & Mokdad, Ali H.. (2008). "A Comparison of Address-based Sampling (ABS) versus Random-Digit Dialing (RDD) for General Population Surveys." *Public Opinion Quarterly* 72:6-27.
- Loosveldt, G. & Sonck, N. (2008). "An Evaluation of the Weighting Procedures for an Online Access Panel Survey." *Survey Research Methods* 2: 93-105.
- Lozar Manfreda, K. & Vehovar, V. (2002). "Do Mail and Web Surveys Provide Same Results?" *Metodološki zvezki* 18:149-169.
- Lugtigheid, A., & Rathod, S. (2005). *Questionnaire Length and Response Quality: Myth or Reality*, Survey Sampling International.

- Malhotra, N. & Krosnick, J. A. (2007). "The Effect of Survey Mode and Sampling on Inferences about Political Attitudes and Behavior: Comparing the 2000 and 2004 ANES to Internet Surveys with Nonprobability Samples." *Political Analysis* 286-323.
- MarketTools. (2009). "MarketTools TrueSample."
http://www.markettools.com/pdfs/resources/DS_TrueSample.pdf.
- Marta-Pedroso, Cristina, Freitas, Helena & Domingos, Tiago. (2007). "Testing for the survey mode effect on contingent valuation data quality: A case study of web based *versus* in-person interviews." *Ecological Economics* 62 (3-4), 388-398.
- Merkle, D. & Langer, G. (2008). "How Too Little Can Give You a Little Too Much." *Public Opinion Quarterly* 72:114-124.
- Metzger, D.S., Koblin, B., Turner, C., Navaline, H., Valenti, F., Holte, S., Gross, M., Sheon, A., Miller, H., Cooley, P., Seage, G.R., & HIVNET Vaccine Preparedness Study Protocol Team (2000). "Randomized controlled trial of audio computer-assisted self-interviewing: utility and acceptability in longitudinal studies." *American Journal of Epidemiology*, 152(2): 99-106.
- Miller, Jeff. (2008). "Burke Panel Quality R and D." Cincinnati: Burke, Inc.
- Miller, Jeff. (2006). "Online Marketing Research." In *The Handbook of Marketing Research. Uses, Abuses and Future Advances*, eds. Rjaiv Grover and Marco Vriens, pp.110-31. Thousand Oaks, CA: Sage.
- Miller, J. (2000). "Net v phone: The great debate." *Research*, 26-27.
- Mitofsky, Warren J. (1989). "Presidential Address: Methods and Standards: A Challenge for Change." *Public Opinion Quarterly* 53:446-453.
- Morgan, Alison. (2008). "Optimus ID: Digital Fingerprinting for Market Research." San Francisco: PeanutLabs.
- Nancarrow, C. & Cartwright, Trixie. (2007). "Online access panels and tracking research: the conditioning issue." *International Journal of Market Research* 49 (5).
- Newman JC, Des Jarlais DC, Turner CF, Gribble J, Cooley P, & Paone D. (2002). "The differential effects of face-to-face and computer interview modes." *Am J Public Health* 92(2): 294-7.
- Niemi, R. G., Portney, K., & King, D. (2008). "Sampling young adults: The effects of survey mode and sampling method on inferences about the political behavior of college students." Paper presented at the annual meeting of the American Political Science Association, Boston, MA.
- Nukulki, P., Hadfield, J. & Subias, S. (2007). "An Investigation of Panel Conditioning with Attitudes Toward U.S Foreign Policy." Presented at the AAPOR 62nd Annual Conference.
- Olson, Kristen. (2006). "Survey participation, nonresponse bias, measurement error bias, and total bias." *Public Opinion Quarterly* 70, 737-758.
- O'Muircheartaigh, C. (1997). "Measurement Error in Surveys: A Historical Perspective," in L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, and D. Trewin, eds., *Survey Measurement and Process Quality*. New York: Wiley.

- Ossenbruggen, R. van, Vonk, T., & Willems, P. (2006). "Results Dutch Online Panel Comparison Study (NOPVO)." Paper presented at the open meeting "Online Panels, Goed Bekeken", Utrecht, the Netherlands. www.nopvo.nl.
- Partick, P.L., Cheadle, A., Thompson, D.C., Diehr, P., Koepsell, T. & Kinne, S. (1994). "The Validity of Self-Reported Smoking: A Review and Meta-Analysis," *American Journal of Public Health* 84,7,1086-1093.
- Pew Research Center for the People and the Press. (2009). <http://www.pewInternet.org/static-pages/trend-data/whos-online.aspx>.
- Pew Research Center for the People and the Press. (2008). <http://www.pewInternet.org/Data-Tools/Download-Data/~media/Infographics/Trend%20Data/January%202009%20updates/Demographics%20of%20Internet%20Users%201%206%2009.jpg>
- Peirce, Charles Sanders. (1877). "The Fixation of Belief." *Popular Science Monthly* 12: 1–15.
- Piekarski, L., Galin, M., Baim, J., Frankel, M., Augemberg, K. & Prince, S. (2008). "Internet Access Panels and Public Opinion and Attitude Estimates." Poster Session presented at 63rd Annual AAPOR Conference, New Orleans LA.
- Pineau, Vicki & Slotwiner, Daniel. (2003). "Probability Samples vs. Volunteer Respondents in Internet Research: Defining Potential Effects on Data and Decision-making" in Marketing Applications. Knowledge Networks White Paper.
- Potoglou, D. & Kanaroglou, P. S. (2008). "Comparison of phone and web-based surveys for collecting household background information." Paper presented at the 8th International Conference on Survey Methods in Transport, France.
- Poynter, R. & Comley, P. (2003). "Beyond Online Panels." *Proceedings of the ESOMAR Technovate Conference*. Amsterdam: ESOMAR.
- Rainie, L. (2010). "Internet, Broadband, and Cell Phone Statistics," Pew Internet and American Life Project. Pew Research Center.
- Rao, J.N.K. (2003). *Small Area Estimation*. Hoboken, New Jersey, John Wiley & Sons.
- Riley, Elise D., Chaisson, Richard E., Robnett, Theodore J., Vertefeuille, John, Strathdee, Steffanie A. & Vlahov, David. (2001). "Use of Audio Computer-assisted Self-Interviews to Assess Tuberculosis-related Risk Behaviors." *Am. J. Respir. Crit. Care Med.* 164(1): 82-85
- Rivers, Douglas. (2007). "Sample matching for Web surveys; Theory and application." Paper presented at the 2007 Joint Statistical Meetings.
- Rogers, S.M., Willis, G., Al-Tayyib, A., Villarroel, M.A., Turner, C.F., Ganapathi, L. et al. (2005). "Audio computer assisted interviewing to measure HIV risk behaviours in a clinic population." *Sexually Transmitted Infections* 81(6): 501-507.
- Rosenbaum, Paul R. & Rubin, Donald B. (1983). "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70:41-55.

- Rosenbaum, Paul R. & Rubin, Donald B. (1984). "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score." *Journal of the American Statistical Association* 79:516-524.
- Roster, Catherine A., Rogers, Robert D., Albaum, Gerald & Klein, Darin. (2004). "A Comparison of Response Characteristics from Web and Telephone Surveys." *International Journal of Market Research* 46:359-373.
- Rubin, D.R. (2006). *Matched Sampling for Causal Effects*. New York, Cambridge University Press
- Ryzin, Gregg G. (2008). "Validity of an On-Line Panel Approach to Citizen Surveys." *Public Performance and Management Review* 32:236-262.
- Sanders, D., Clarke, H. D., Stewart, M. C., & Whiteley, P. (2007). "Does Mode Matter for Modeling Political Choice? Evidence from the 2005 British Election Study." *Political Analysis* 15 (3): 257-285.
- Saris, Wilem E. (1998). "Ten Years of Interviewing without Interviewers: the Telepanel." In *Computer Assisted Survey Information Collection*, eds. Mick Couper, Reginald P. Baker, Jelke Bethlehem, Cynthia Z. F. Clark, Jean Martin, William L. Nicholls, and James O'Reilly, pp. 409-29. New York: Wiley.
- Sayles, H. & Arens, Z. (2007). "A Study of Panel member Attrition in the Gallup Panel." Paper presented at 62nd AAPOR Annual Conference, Anaheim, CA.
- Schillewaert, Niels & Meulemeester, Pascale. (2005). "Comparing Response Distributions of Offline and Online Data Collection Methods." *International Journal of Market Research* 47:163-178.
- Schlackman, W. (1984) "A Discussion of the use of Sensitivity Panels in Market Research." *Journal of the Market Research Society* 26: 191 - 208.
- Schonlau, Matthias, van Soest, Arthur & Kapteyn, Arie. (2007). "Are 'Webographic' or Attitudinal Questions Useful for Adjusting Estimates from Web Surveys Using Propensity Scoring?" *Survey Research Methods* 1: 155-163.
- Schonlau, Matthias, van Soest, Arthur, Kapteyn, Arie & Couper, Mick. (2009). "Selection Bias in Web Surveys and the Use of Propensity Scores." *Sociological Methods and Research* 37: 291-318.
- Schonlau, Matthias, Zapert, Kinga, Simon, Lisa P., Sanstad, Katherine H., Marcus, Sue M., Adams, John, Spranca, Mark, Kan, Hongjun, Turner, Rachel, & Berry, Sandra H. (2004). "A Comparison Between Responses From a Propensity-Weighted Web Survey and an Identical RDD Survey." *Social Science Computer Review* 22:128-138.
- Smith, Renee & Brown, Hofman. (2006). "Panel and data quality: Comparing metrics and assessing claims." Panel Research 2006: ESOMAR World Research Conference. Amsterdam: ESOMAR.
- Smith, Tom W. (2001). "Are Representative Internet Surveys Possible?" Proceedings of Statistics Canada Symposium 2001, Achieving Data Quality in a Statistical Agency: A Methodological Perspective.

- Smith, T. W., & Dennis, J. M. (2005). "Online vs. In-person: Experiments with mode, format, and question wordings." *Public Opinion Pros*.
- Smyth, Jolene D., Christian, Leah Melani, & Dillman, Don A. (2008). "Does "Yes or No" on the Telephone Mean the Same as "Check-All-That-Apply" on the Web?" *Public Opinion Quarterly* 72: 103-113.
- Snell, Laurie, J., Peterson, Bill & Grinstead, Charles. (1998). "Chance News 7.11." Accessed August 31, 2009: http://www.dartmouth.edu/~chance/chance_news/recent_news/chance_news_7.11.html.
- Squire, Peverill. (1988). "Why the 1936 Literary Digest Poll Failed." *Public Opinion Quarterly* 52, 125-133.
- Sparrow, Nick. (2006). "Developing reliable online polls." *International Journal of market Research* 48, 659-680.
- Sparrow, Nick & Curtice, John. (2004). "Measuring the Attitudes of the General Public via Internet Polls: An Evaluation." *International Journal of Market Research* 46:23-44.
- Speizer, H., Baker, R., & Schneider, K. (2005). "Survey Mode Effects; Comparison between Telephone and Web." Paper presented at the annual meeting of the American Association For Public Opinion Association, Fontainebleau Resort, Miami Beach, FL.
- Stevens, S.S. (1946). "On the theory of scales of measurement." *Science* 103: 677-680.
- Stevens, S.S. (1951). "Mathematics, measurement and psychophysics." In S.S. Stevens (Ed.), *Handbook of experimental psychology*, pp. 1-49. New York: Wiley.
- Stirton, J. & Robertson, E. (2005). "Assessing the viability of online opinion polling during the 2004 federal election." *Australian Market and Social Research Society*.
<http://www.enrollingthepeople.com/mumblestuff/ACNielsen%20AMSRS%20paper%202005.pdf>
- Sturgis, P., Allum, N. & Brunton-Smith, I. (2008) "Attitudes Over Time: The Psychology of Panel Conditioning." In P. Lynn (ed) *Methodology of Longitudinal Surveys*, Wiley.
- Sudman, S., Bradburn, N., & Schwarz, N. (1996) *Thinking About Answers*. San Francisco: Jossey-Bass.
- Taylor, Humphrey. (2000). "Does Internet Research Work?: Comparing Online Survey Results with Telephone Surveys." *International Journal of Market Research* 42:51-63.
- Taylor, Humphrey. (2007). "The Case For Publishing (Some) Online Polls." *Polling Report*. Accessed August 31, 2009 from http://www.pollingreport.com/ht_online.htm.
- Taylor, Humphrey, Bremer, John, Overmeyer, Cary, Siegel, Jonathan W., & Terhanian, George. (2001). "The Record of Internet-based Opinion Polls in Predicting the Results of 72 Races in the November 2000 U.S. Elections." *International Journal of Market Research* 43:127-136.
- Taylor, Humphrey, Krane, David, & Thomas, Randall K. (2005) "Best Foot Forward: Social Desirability in Telephone vs. Online Surveys." *Public Opinion Pros*, Feb. Available from http://www.publicopinionpros.com/from_field/2005/feb/taylor.asp.
- Terhanian, George, Smith, Renee, Bremer, John, & Thomas, Randall K. (2001). "Exploiting analytical advances: Minimizing the biases associated with internet-based surveys of non-random samples." *ARF/ESOMAR: Worldwide Online Measurement* 248: 247-272.

- Thomas, R. K., Krane, D., Taylor, H., and Terhanian, G. (2006). "Attitude measurement in phone and online surveys: Can different modes and samples yield similar results?" Paper presented at the Joint Conference of the Society for Multivariate Analysis in the Behavioural Sciences and European Association of Methodology, Budapest, Hungary.
- Thomas, R. K., Krane, D., Taylor, H., & Terhanian, G. (2008). "Phone and Web interviews: Effects of sample and weighting on comparability and validity." Paper presented at ISA-RC33 7th International Conference, Naples, Italy.
- Toepoel, Vera, Das, Marcel, & van Soest, Arthur. (2008). "Effects of design in web surveys: Comparing trained and fresh respondents." *Public Opinion Quarterly* 72:985-1007.
- Tourangeau, Roger. (2004). "Cognitive science and survey methods." In T. Jabine et al. (Eds.), *Cognitive Aspects of Survey Design: Building a Bridge Between Disciplines*. Washington: National Academy Press, pp.73-100.
- Tourangeau, Roger. (2004). "Survey Research and Societal Change." *Annual Review of Psychology* 55:775-801.
- Tourangeau, R., R.M. Groves, R. M., Kennedy, C., & Yan, T. (2009). "The Presentation of a Web Survey, Nonresponse and Measurement Error among Members of Web Panel." *Journal of Official Statistics* 25: 299-321.
- Twyman, J. (2008). "Getting it right: YouGov and online survey research in Britain." *Journal of Elections, Public Opinion, and Parties* 18: 343-354.
- Valliant, R., Royall, R., & Dorfman, A. (2001). *Finite Population Sampling and Inference: A Prediction Approach*. New York, Wiley.
- Vavreck, L., & Rivers, D. (2008) "The 2006 Cooperative Congressional Election Study." *Journal of Elections, Public Opinion, and Parties* 18: 355-366.
- Vonk, T.W.E., Ossenbruggen, R & Willems, P. (2006). "The effects of panel recruitment and management on research results." ESOMAR Panel Research 2006.
- Walker, R.& Pettit, R. (2009). "ARF Foundations of Quality: Results Preview." New York: The Advertising Research Foundation.
- Walker, R., Pettit, R., & Rubinson, J. (2009). "The Foundations of Quality Study Executive Summary 1: Overlap, Duplication, and Multi Panel Membership." New York: The Advertising Research Foundation.
- Waksberg, J. (1978). "Sampling Methods for Random Digit Dialing," *Journal of the American Statistical Association* 73:40-46.
- Waruru, A. K., Nduati, R., & Tylleskar, T. (2005). "Audio computer-assisted self-interviewing (ACASI) may avert socially desirable responses about infant feeding in the context of HIV." *Medical Informatics and Decision Making* 5: 24-30.
- Waterton, J. & Lievesley, D. (1989). "Evidence of conditioning effects in the British Social Attitudes Panel." in Kasprzyk, D., Duncan, G., Kalton, G., and Singh, M.P. *Panel Surveys*. New York, John Wiley and Sons.

Weijters, B., Schillewaert, N., & Geuens, M. (2008). „Assessing response styles across modes of data collection.” *Academy of Marketing Science* 36: 409-422.

Woodward, M. (2004). *Study Design and Data Analysis*, 2nd Edition. Boca Rotan, Florida Chapman & Hall.

Yeager, D. S., Krosnick, J.A., Chang, L., Javitz, H.S., Levindusky, M.S., Simpser, A., & Wang, R. (2009). “Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Nonprobability Samples.” Stanford University.

APPENDIX A: PORTION OF THE CASRO CODE OF STANDARDS AND ETHICS DEALING WITH INTERNET RESEARCH

3. Internet Research

The unique characteristics of Internet research require specific notice that the principle of respondent privacy applies to this new technology and data collection methodology. The general principle of this section of the Code is that survey Research Organizations will not use unsolicited emails to recruit survey respondents or engage in surreptitious data collection methods. This section is organized into three parts: (A) email solicitations, (B) active agent technologies, and (C) panel/sample source considerations.

a. Email Solicitation

(1) Research Organizations are required to verify that individuals contacted for research by email have a reasonable expectation that they will receive email contact for research. Such agreement can be assumed when ALL of the following conditions exist:

a. A substantive pre-existing relationship exists between the individuals contacted and the Research Organization, the Client supplying email addresses, or the Internet Sample Providers supplying the email addresses (the latter being so identified in the email invitation);

b. Survey email invitees have a reasonable expectation, based on the pre-existing relationship where survey email invitees have specifically opted in for Internet research with the research company or Sample Provider, or in the case of Client-supplied lists that they may be contacted for research and invitees have not opted out of email communications;

c. Survey email invitations clearly communicate the name of the sample provider, the relationship of the individual to that provider, and clearly offer the choice to be removed from future email contact.

d. The email sample list excludes all individuals who have previously requested removal from future email contact in an appropriate and timely manner.

e. Participants in the email sample were not recruited via unsolicited email invitations.

(2) Research Organizations are prohibited from using any subterfuge in obtaining email addresses of potential respondents, such as collecting email addresses from public domains, using technologies or techniques to collect email addresses without individuals' awareness, and collecting email addresses under the guise of some other activity.

(3) Research Organizations are prohibited from using false or misleading return email addresses or any other false and misleading information when recruiting respondents. As stated later in this Code, Research Organizations must comply with all federal regulations that govern survey research activities. In addition, Research Organizations should use their best efforts to comply with other federal regulations that govern unsolicited email contacts, even though they do not apply to survey research.

(4) When receiving email lists from Clients or Sample Providers, Research Organizations are required to have the Client or Sample Provider verify that individuals listed have a reasonable expectation that they will receive email contact, as defined, in (1) above.

(5) The practice of “blind studies” (for sample sources where the sponsor of the study is not cited in the email solicitation) is permitted if disclosure is offered to the respondent during or after the interview. The respondent must also be offered the opportunity to “opt-out” for future research use of the sample source that was used for the email solicitation.

(6) Information about the CASRO Code of Standards and Ethics for Survey Research should be made available to respondents.

b. Active Agent Technology

(1) Active agent technology is defined as any software or hardware device that captures the behavioral data about data subjects in a background mode, typically running concurrently with other activities. This category includes tracking software that allows Research Organizations to capture a wide array of information about data subjects as they browse the Internet. Such technology needs to be carefully managed by the research industry via the application of research best practices.

Active agent technology also includes direct to desktop software downloaded to a user's computer that is used solely for the purpose of alerting potential survey respondents, downloading survey content or asking survey questions. A direct to desktop tool does not track data subjects as they browse the Internet and all data collected is provided directly from user input.

Data collection typically requires an application to download onto the subjects' desktop, laptop or PDA (including personal wireless devices). Once downloaded, tracking software has the capability of capturing the data subject's actual experiences when using the Internet such as Web page hits, web pages visited, online transactions completed, online forms completed, advertising click-through rates or impressions, and online purchases.

Beyond the collection of information about a user's Internet experience, the software has the ability to capture information from the data subject's email and other documents stored on a computer device such as a hard disk. Some of this technology has been labeled “spyware,” especially because the download or installation occurs without the data subject's full knowledge and specific consent. The use of spyware by a member of CASRO is strictly prohibited.

A cookie (defined as a small amount of data that is sent to a computer's browser from a web server and stored on the computer's hard drive) is not an active agent. The use of cookies is permitted if a description of the data collected and its use is fully disclosed in a Research Organizations' privacy policy.

(2) Following is a list of unacceptable practices that Research Organizations should strictly forbid or prevent. A Research Organization is considered to be using spyware when it fails to adopt all of the practices in set forth in Section 3 below or engages in any in the following practices:

a. Downloading software without obtaining the data subject's informed consent.

b. Downloading software without providing full notice and disclosure about the types of information that will be collected about the data subject, and how this information may be used. This notice needs to be conspicuous and clearly written.

- c. Collecting information that identifies the data subject without obtaining affirmed consent.
- d. Using keystroke loggers without obtaining the data subject's affirmed consent.
- e. Installing software that modifies the data subject's computer settings beyond that which is necessary to conduct research providing that the software doesn't make other installed software behave erratically or in unexpected ways.
- f. Installing software that turns off anti-spyware, anti-virus, or anti-spam software.
- g. Installing software that seizes control or hijacks the data subject's computer.
- h. Failing to make commercially reasonable efforts to ensure that the software does not cause any conflicts with major operating systems and does not cause other installed software to behave erratically or in unexpected ways.
- i. Installing software that is hidden within other software that may be downloaded.
- j. Installing software that is difficult to uninstall.
- k. Installing software that delivers advertising content, with the exception of software for the purpose of ad testing.
- l. Installing upgrades to software without notifying users
- m. Changing the nature of the active agent program without notifying user
- n. Failing to notify the user of privacy practice changes relating to upgrades to the software

(3) Following are practices Research Organizations that deploy active agent technologies should adopt. Research Organizations that adopt these practices and do not engage in any of the practices set forth in Section 2 above will not be considered users of spyware.

a. Transparency to the data subject is critical. Research companies must disclose information about active agents and other software in a timely and open manner with each data subject. This communication must provide details on how the Research Organization uses and shares the data subject's information.

i. Only after receiving an affirmed consent or permission from the data subject or parent's permission for children under the age of 18, should any research software be downloaded onto the individual's computer or PDA.

ii. Clearly communicate to the data subject the types of data if any, that is being collected and stored by an active agent technology.

iii. Disclosure is also needed to allow the data subject to easily uninstall research software without prejudice or harm to them or their computer systems.

iv. Personal information about the subject should not be used for secondary purposes or shared with third parties without the data subject's consent.

v. Research Organizations are obligated to ensure that participation is a conscious and voluntary activity. Accordingly, incentives must never be used to hide or obfuscate the acceptance of active agent technologies.

vi. Research Organizations that deploy active agent technologies should have a method to receive queries from end-users who have questions or concerns. A redress process is essential for companies if they want to gauge audience reaction to participation on the network.

vii. On a routine and ongoing basis, consistent with the stated policies of the Research Organization, data subjects who participate in the research network should receive clear periodic notification that they are actively recorded as participants, so as to insure that their participation is voluntary. This notice should provide a clearly defined method to uninstall the Research Organization's tracking software without causing harm to the data subject.

b. Stewardship of the data subject is critical. Research companies must take steps to protect information collected from data subjects.

i. Personal or sensitive data (as described in the Personal Data Classification Appendix) should not be collected. If collection is unavoidable, the data should be destroyed immediately. If destruction is not immediately possible, it: (a) should receive the highest level of data security and (b) should not be accessed or used for any purpose.

ii. Research Organizations have an obligation to establish safeguards that minimize the risk of data security and privacy threats to the data subject.

iii. It is important for Research Organizations to understand the impact of their technology on end-users, especially when their software downloads in a bundle with other comparable software products.

iv. Stewardship also requires the Research Organization to make commercially reasonable efforts to ensure that these “free” products are also safe, secure and do not cause undue privacy or data security risks.

v. Stewardship also requires a Research Organization that deploys active agent technologies to be proactive in managing its distribution of the software. Accordingly, companies must vigorously monitor their distribution channel and look for signs that suggest unusual events such as high churn rates.

vi. If unethical practices are revealed, responsible research companies should strictly terminate all future dealings with this distribution partner.

c. Panel/Sample Source Considerations

The following applies to all Research Organizations that utilize the Internet and related technologies to conduct research.

(1) The Research Organization must:

- a. Disclose to panel members that they are part of panel.
- b. Obtain panelist's permission to collect and store information about the panelist.
- c. Collect and keep appropriate records of panel member recruitment, including the source through which the panel member was recruited.
- d. Collect and maintain records of panel member activity.

(2) Upon Client request, the Research Organization must disclose:

- a. Panel composition information (including panel size, populations covered, and the definition of an active panelist).
- b. Panel recruitment practice information.
- c. Panel member activity.
- d. Panel incentive plans.
- e. Panel validation practices.
- f. Panel quality practices.
- g. Aggregate panel and study sample information (this information could include response rate information, panelist participation in other research by type and timeframe, see Responsibilities in Reporting to Clients and the Public).
- h. Study related information such as email invitation(s), screener wording, dates of email invitations and reminders, and dates of fieldwork.

(3) Stewardship of the data collected from panelists is critical:

- a. Panels must be managed in accordance with applicable data protection laws and regulations.
- b. Personal or sensitive data should be collected and treated as specified in the Personal Data Classification Appendix.
- c. Upon panelist request, the panelist must be informed about all personal data (relating to the panelist that is provided by the panelist, collected by an active agent, or otherwise obtained by an acceptable method specified in a Research Organization's privacy policy) maintained by the Research Organization. Any personal data that is indicated by panel member as not correct or obsolete must be corrected or deleted as soon as practicable.

(4) Panel members must be given a straightforward method for being removed from the panel if they choose. A request for removal must be completed as soon as practicable and the panelist must not be selected for future research studies.

(5) A privacy policy relating to use of data collected from or relating to the panel member must be in place and posted online. The privacy policy must be easy to find and use and must be regularly communicated to panelists. Any changes to the privacy policy must be communicated to panelists as soon as possible.

(6) Research Organizations should take steps to limit the number of survey invitations sent to targeted respondents by email solicitations or other methods over the Internet so as to avoid harassment and response bias caused by the repeated recruitment and participation by a given pool (or panel) of data subjects.

(7) Research Organizations should carefully select sample sources that appropriately fit research objectives and Client requirements. All sample sources must satisfy the requirement that survey participants have either opted-in for research or have a reasonable expectation that they will be contacted for research.

(8) Research Organizations should manage panels to achieve the highest possible research quality. This includes managing panel churn and promptly removing inactive panelists.

(9) Research Organizations must maintain survey identities and email domains that are used exclusively for research activities.

(10) If a Research Organization uses a sample source (including a panel owned by the Research Organization or a subcontractor) that is used for both survey research and direct marketing activities, the Research Organization has an obligation to disclose the nature of the marketing campaigns conducted with that sample source to Clients so that they can assess the potential for bias.

(11) All data collected on behalf of a Client must be kept confidential and not shared or used on behalf of another Client (see also Responsibilities to Clients).

APPENDIX B: ESOMAR 26 QUESTIONS TO HELP RESEARCH BUYERS OF ONLINE SAMPLES

These questions, in combination with additional information, will help researchers consider issues which influence whether an online sampling approach is fit for purpose in relation to a particular set of objectives; for example whether an online sample will be sufficiently representative and unbiased. They will help the researcher ensure that they receive what they expect from an online sample provider.

These are the areas covered:

- Company profile
- Sample source
- Panel recruitment
- Panel and sample management
- Policies and compliance
- Partnerships and multiple panel partnership
- Data quality and validation

Company profile

1. What experience does your company have with providing online samples for market research? 

This answer might help you to form an opinion about the relevant experience of the sample provider. How long has the sample provider been providing this service and do they have for example a market research, direct marketing or more technological background? Are the samples solely provided for third party research, or does the company also conduct proprietary work using their panels?

Sample Source

2. Please describe and explain the types of source(s) for the online sample that you provide (are these databases, actively managed panels, direct marketing lists, web intercept sampling, river sampling or other)? 

The description of the type of source a provider uses for delivering an online sample might provide insight into the quality of the sample. An actively managed panel is one which contains only active panel

members - see question 11. Note that not all online samples are based on online access panels.

3. What do you consider to be the primary advantage of your sample over other sample sources in the marketplace? 

The answer to this question may simplify the comparison of online sample providers in the market.

4. If the sample source is a panel or database, is the panel or database used solely for market research? If not, please explain. 

Combining panellists for different types of usage (like direct marketing) might cause survey effects.

5. How do you source groups that may be hard-to-reach on the internet? 

The inclusion of hard-to-reach groups on the internet (like ethnic minority groups, young people, seniors etc.) might improve the quality of the sample provided.

6. What are people told when they are recruited? 

The type of rewards and proposition could influence the type of people who agree to answer a questionnaire or join a specific panel and can therefore influence sample quality.

Panel Recruitment

7. If the sample comes from a panel, what is your annual panel turnover/attrition/retention rate and how is it calculated? 

The panel attrition rate may be an indicator of panellists' satisfaction and (therefore) panel management, but a high turnover could also be a result of placing surveys which are too long with poor question design. The method of calculation is important because it can have a significant impact on the rate quoted.

8. Please describe the opt-in process. 

The opt-in process might indicate the respondents' relationship with the panel provider. The market generally makes a distinction between single and double opt-in. Double opt-in describes the process by which a check is made to confirm that the person joining the panel wishes to be a member and understands what to expect.

9. Do you have a confirmation of identity procedure? Do you have procedures to detect fraudulent respondents at the time of registration with the panel? If so, please describe. 

Confirmation of identity might increase quality by decreasing multiple entries, fraudulent panellists, etc.

10. What profile data is kept on panel members? For how many members is this data collected and

how often is this data updated? 

Extended and up-to-date profile data increases the effectiveness of low incidence sampling and reduces pre-screening of panellists.¹¹

11. What is the size and/or the capacity of the panel, based on active panel members on a given date? Can you provide an overview of active panellists by type of source? 

The size of the panel might give an indication of the capacity of a panel. In general terms, a panel's capacity is a function of the availability of specific target groups and the actual completion rate. There is no agreed definition of an active panel member, so it is important to establish how this is defined. It is likely that the new ISO for access panels which is being discussed will propose that an active panel member is defined as a member that has participated in at least one survey, or updated his/her profile data, or registered to join the panel, within the last 12 months. The type and number of sources might be an indicator of source effects and source effects might influence the data quality. For example, if the sample is sourced from a loyalty programme (travel, shopping, etc.) respondents may be unrepresentatively high users of certain services or products.

Panel and Sample Management

12. Please describe your sampling process including your exclusion procedures if applicable. Can samples be deployed as batches/replicates, by time zones, geography, etc? If so, how is this controlled? 

The sampling processes for the sample sources used are a main factor in sample provision. A systematic approach based on market research fundamentals may increase sample quality.

13. Explain how people are invited to take part in a survey. What does a typical invitation look like? 

Survey results can sometimes be influenced by the wording used in subject lines or in the body of an invitation.

14. Please describe the nature of your incentive system(s). How does this vary by length of interview, respondent characteristics, or other factors you may consider? 

The reward or incentive system might impact on the reasons why people participate in a specific panel and these effects can cause bias to the sample.

15. How often are individual members contacted for online surveys within a given time period? Do you keep data on panellist participation history and are limits placed on the frequency that members are contacted and asked to participate in a survey? 

Frequency of survey participation might increase conditioning effects whereas a controlled survey load environment can lead to higher data quality.

Policies and Compliance

16. Is there a privacy policy in place? If so, what does it state? Is the panel compliant with all regional, national and local laws with respect to privacy, data protection and children e.g. EU Safe Harbour, and COPPA in the US? What other research industry standards do you comply with e.g. ICC/ESOMAR International Code on Market and Social Research, CASRO guidelines etc.? 

Not complying with local and international privacy laws might mean the sample provider is operating illegally.

17. What data protection/security measures do you have in place? 

The sample provider usually stores sensitive and confidential information on panellists and clients in databases. These need to be properly secured and backed-up, as does any confidential information provided by the client.

18. Do you apply a quality management system? Please describe it. 

A quality management system is a system by which processes in a company are described and employees are accountable. The system should be based on continuous improvement. Certification of these processes can be independently done by auditing organizations, based for instance, on ISO norms.

19. Do you conduct online surveys with children and young people? If so, please describe the process for obtaining permission. 

The ICC/ESOMAR International Code requires special permissions for interviewing children.

20. Do you supplement your samples with samples from other providers? How do you select these partners? Is it your policy to notify a client in advance when using a third party provider? Do you de-duplicate the sample when using multiple sample providers? 

Many providers work with third parties. This means that the quality of the sample is also dependent on the quality of sample providers that the buyer did not select. Transparency is a key issue in this situation. Overlap between different panel providers can be significant in some cases and de-duplication removes this source of error, and frustration for respondents.

Partnerships and Multiple Panel Membership

21. Do you have a policy regarding multi-panel membership? What efforts do you undertake to ensure that survey results are unbiased given that some individuals belong to multiple panels? 

It is not that uncommon for a panellist to be a member of more than one panel nowadays. The effects of multi-panel membership by country, survey topic, etc., are not yet fully known. Proactive and clear policies on how any potential negative effects are minimized by recruitment, sampling, and weighting

practices is important.

Data Quality and Validation

22. What are likely survey start rates, drop-out and participation rates in connection with a provided sample? How are these computed? 

Panel response might be a function of factors like invitation frequency, panel management (cleaning) policies, incentive systems and so on. Although not a quality measure by itself these rates can provide an indication of the way a panel is managed. A high start rate might indicate a strong relationship between the panel member and the panel. A high drop-out rate might be a result of poor questionnaire design, questionnaire length, survey topic or incentive scheme as well as an effect of panel management. The new ISO for access panels will likely propose that participation rate is defined as the number of panel members who have provided a usable response divided by the total number of initial personal invitations requesting members to participate.

23. Do you maintain individual level data such as recent participation history, date of entry, source, etc., on your panelists? Are you able to supply your client with a per job analysis of such individual level data? 

This type of data per respondent increases the possibility of analysis for data quality, as described in ESOMAR's Guideline on Access Panels.

24. Do you use data quality analysis and validation techniques to identify inattentive and fraudulent respondents? If yes, what techniques are used and at what point in the process are they applied? 

When the sample provider is also hosting the online survey, preliminary data quality analysis and validation is usually preferable.

25. Do you measure respondent satisfaction? 

Respondent satisfaction may be an indicator of willingness to take future surveys. Respondent reactions to your survey from self-reported feedback or from an analysis of suspend points might be very valuable to help understand survey results.

26. What information do you provide to debrief your client after the project has finished? 

One might expect a full sample provider debrief report, including gross sample, start rate, participation rate, drop-out rate, the invitation text, a description of the field work process, and so on.

APPENDIX C: AAPOR STATEMENT - WEB SURVEYS UNLIKELY TO REPRESENT ALL VIEWS

Non-scientific polling technique proliferating during Campaign 2000

September 28, 2000 -- Ann Arbor ---Many Web-based surveys fail to represent the views of all Americans and thus give a misleading picture of public opinion, say officials of the American Association for Public Opinion Research (AAPOR), the leading professional association for public opinion researchers.

One of the biggest problems with doing online surveys is that half the country does not have access to the Internet," said AAPOR president Murray Edelman. "For a public opinion survey to be representative of the American public, all Americans must have a chance to be selected to participate in the survey."

Edelman released a new statement by the AAPOR Council, the executive group of the professional organization, giving its stance on online surveys.

Examples of recent Web-based polls that produced misleading findings include:

* Various online polls during the presidential primaries showed Alan Keyes, Orrin Hatch, or Steve Forbes as the favored Republican candidate. No scientifically conducted public opinion polls ever corroborated any of these findings.

* At the same time that a Web-based poll reported that a majority of Americans disapproved of the government action to remove Elian Gonzalez, a scientific poll of a random national sample of Americans showed that 57% approved of that action.

Edelman said that AAPOR is seeking to alert journalists and the public in advance of the upcoming presidential debates that many post-debate polls taken online may be just as flawed and misleading as these examples.

Lack of universal access to the Internet is just one problem that invalidates many Web-based surveys. In some applications of the technology, individuals may choose for themselves whether or not to participate in a survey, and in some instances, respondents can participate in the same survey more than once. Both practices violate scientific polling principles and invalidate the results of such surveys.

"Many online polls are compromised because they are based on the responses of only those people who happened to volunteer their opinions on the survey," said Michael Traugott, past president of AAPOR. "For a survey to be scientific, the respondents must be chosen by a carefully designed sampling process that is completely controlled by the researcher."

Because of problems such as these, AAPOR urges journalists and others who evaluate polls for public dissemination to ask the following questions:

1. Does the online poll claim that the results are representative of a specific population, such as the American public?
2. If so, are the results based upon a scientific sampling procedure that gives every member of the population a chance to be selected?

3. Did each respondent have only one opportunity to answer the questions?

4. Are the results of the online survey similar to the results of scientific polls conducted at the same time?

5. What was the response rate for the study?

Only if the answer to the first four questions is "yes" and the response rate is reported, should the online poll results be considered for inclusion in a news story.

Only when a Web-based survey adheres to established principles of scientific data collection can it be characterized as representing the population from which the sample was drawn. But if it uses volunteer respondents, allows respondents to participate in the survey more than once, or excludes portions of the population from participation, it must be characterized as unscientific and is unrepresentative of any population.

APPENDIX D: AAPOR STATEMENT - OPT-IN SURVEYS AND MARGIN OF ERROR

The reporting of a margin of sampling error associated with an opt-in or self-identified sample (that is, in a survey or poll where respondents are self-selecting) is misleading.

When we draw a sample at random—that is, when every member of the target population has a known probability of being selected—we can use the sample to make projective, quantitative estimates about the population. A sample selected at random has known mathematical properties that allow for the computation of sampling error.

Surveys based on self-selected volunteers do not have that sort of known relationship to the target population and are subject to unknown, non-measurable biases. Even if opt-in surveys are based on probability samples drawn from very large pools of volunteers, their results still suffer from unknown biases stemming from the fact that the pool has no knowable relationships with the full target population.

AAPOR considers it harmful to include statements about the theoretical calculation of sampling error in descriptions of such studies, especially when those statements mislead the reader into thinking that the survey is based on a probability sample of the full target population. The harm comes from the inferences that the margin of sampling error estimates can be interpreted like those of probability sample surveys.

All sample surveys and polls are subject to multiple sources of error. These include, but are not limited to, sampling error, coverage error, nonresponse error, measurement error, and post-survey processing error. AAPOR suggests that descriptions of published surveys and polls include notation of all possible sources of error.

For opt-in surveys and polls, therefore, responsible researchers and authors of research reports are obligated to disclose that respondents were not randomly selected from among the total population, but rather from among those who took the initiative or agreed to volunteer to be a respondent.

AAPOR recommends the following wording for use in online and other surveys conducted among self-selected individuals: *Respondents for this survey were selected from among those who have [volunteered to participate/registered to participate in (company name) online surveys and polls]. The data (have been/have not been) weighted to reflect the demographic composition of (target population). Because the sample is based on those who initially self-selected for participation [in the panel] rather than a probability sample, no estimates of sampling error can be calculated. All sample surveys and polls may be subject to multiple sources of error, including, but not limited to sampling error, coverage error, and measurement error.*