Information Management and Business Review
Vol. 3, No. 2, pp. 48-56, Aug 2011 (ISSN 2220-3796)

Accuracy of Web Survey Data: The State Of Research on Factual Questions in Surveys

Lenandlar Singh
Department of Computer Science, University of Guyana
lenandlar.singh@uog.edu.gy

Abstract: Surveys have long become a standard data collection tool. In more recent times web surveys have become particularly popular given their cost-effectiveness on one hand, and the ubiquitous nature of the internet and the World Wide Web on the other. Regardless of the type of survey, a number of issues still challenge survey researchers and practitioners. The accuracy of data collected for socio-demographic and other factual –type research questions is of utmost importance if the researcher is to make any claim about the data collected. Accuracy as characteristic of data quality is perhaps the most important issue of all. This paper specifically reviews the body of literature on work completed on data quality and identifies and analyzes studies on accuracy and reliability of data. This paper critically examines the most significant published research in the literature that addresses the issue of accuracy and reliability of survey data. Specifically, this review addresses and critique existing research methods, identifies and discusses important limitations, and concludes with a discourse on key questions to be answered and suggestions for future research.

Keywords: *Surveys, Web Surveys, Data Accuracy, Factual Questions, Data Quality*

1. Introduction

The objective of this article is to review the state of research on the quality of data supplied by respondents. This is approached through a careful examination of completed and published empirical research on the reliability and accuracy of data collected for socio-demographics variables and knowledge-based questions in surveys. The question of accuracy and reliability of survey data is vitally important for a number of reasons. Survey instruments have long become the most popular means of collecting large volumes of data. The ability of surveys to collect large sets of data is further enhanced with advent of the World Wide Web which is increasingly used for administering surveys and collecting data for research and other purposes (Elo, 2009). Surveys are a cost effective approach to large scale data collection. Specifically, web surveys offers a comparatively more cost effective method for reaching large groups of potential respondents in a short turnaround time(Sanders et al., 2007; Cobanoglu, Warde, & Moreo, 2001; Couper, 2000; Dillman, 2000).

Surveys, whether web or other forms, if not carefully designed and administered, may lead to a number of methodological and quality issues. Data accuracy and reliability, identification of representative samples and survey design issues are important quality characteristics that have engaged methodology researchers in the last 60 years. However, to date, a significantly small number of all the studies focus on the issue of accuracy and reliability. This paper identifies and addresses the issue of data accuracy and reliability through a systematic review of the literature on data accuracy. This review adds to the existing literature on survey data quality by analyzing respondents' patterns to socio-demographic and other categories of factual data questions. Further, particular issues to be addressed by survey researchers as they solicit responses to factual questions, are discussed.

Survey researches centered on the accuracy issue (Sanders et al., 2007; Gibson & McAllister, 2002) have explored a number of items and categories of questions that fit the 'accuracy' classification. The following categories of questions emerged from the literature; factual-type knowledge questions survey on a particular topic – knowledge-based question items (Elo, 2009); spending patterns on utility bills (Warriner, 1991); socio-demographic questions such as 'the age question' (Gendall & Healey (2008). Irrespective of the type of factual questions, the goal remains the same and the challenges appear to increase with the advent of new means of data collection. As Knapp and Kirk (2003) intimates, obtaining accurate data from respondents has been a key challenge facing researchers for many decades and have lead to efforts to address this need.

One of the earliest published studies on the accuracy of responses highlights the importance of the truthfulness of responses by respondents as they engage with survey items (Hyman, 1944). He suggested that it is necessary to know whether respondents are telling the truth as they respond to survey questions. He further noted that researchers ought to be able to ascertain that data supplied is a true picture of the behavior of respondents (Hyman, 1944). Indeed, researchers and practitioners would expect that when a survey is administered, the best possible quality of data is supplied by respondents. However, this has not often been the case. Researchers and practitioners have added responsibility to ensure that the best quality of data is reported. Managing the process of data collection to ensure the most accurate responses are supplied by respondents is therefore a critical consideration. Minimizing survey errors and reducing inaccuracies of responses should be the objective of every data collection process.

Managing survey errors have received significant attention over the past decades. Measurement error is one of three types of errors that can affect that quality of survey data (Couper, 2000); the other two being coverage and sampling error. Self- reporting errors in the form of respondents refusal to provide a response (non response bias) or respondents reporting inaccurate data (response bias) (Warriner, 1991), affect the overall quality of data. Elo (2009) suggested that measurement error has received little attention compared to the other types of errors. Minimizing incorrect data is will reduce measurement errors and therefore improve the overall quality of data. Many studies have explored an array of different methods to this end. However, on the subject of improving accuracy to factual and knowledge-based questions, significantly less progress is reported.

Research with objectives to improve accuracy of responses to factual questions has yielded positive results in some respects. However, a number of important questions are yet to be definitively answered and centers around the essential concept of verification of data. What level of guarantee can be placed on a data set supplied by respondents on any given survey? How can researchers verify the accuracy and validity of these data sets? Are there available benchmarks and if so, how accurate are they and how accessible is it to researchers? How useful and reliable are benchmark samples? Are there legal and ethical considerations that serve as barriers or threats to verification and validation of data sets? Do researchers have the time and resources available to conduct validations and verifications? These are important questions that researchers have begun to explore and questions to which there are no definitive answers at the moment. These questions together form the basis of the discourse in this review. In order for large volumes of data to be useful, and indeed any quantity of data, quality is critically important factor. When researchers and practitioners conduct important analyses and make compelling conclusions which may eventually lead to decision making, they want to be very certain that the input to their analysis and decision making process are reliable datasets. Poor quality data as input, will undermine any attempt at sophisticated and comprehensive analysis, and may indeed result in faulty conclusions and eventually poor decision making. Conversely, high quality datasets may serve as the solid base upon which careful analysis, conclusions and decision making are build.

In this paper the core argument is rooted in the premise that the result of empirical research and the subsequent conclusions drawn thereafter; regardless of the quality of analysis and interpretation, is only as good as the quality of the data that entered into the process. This quality is largely defined by the accuracy and reliability of the data collected. Specifically, paper focuses on the importance of examining the methodological issues that affect quality of data sets collected from Web Surveys. The capacity to collect large data sets using the World Wide Web brings with it a number of methodological issues. Elo (2009) summarizes the importance of these issues by suggesting that if per chance the web replaces all other forms of data collection, we ought to be sure that it does allow us to collect reliable data. Therefore exploring web surveys' potential for collecting accurate data is vitally important and all methodological approaches must be explored, refined and deployed to ensure data quality improves at every step of the data collection process.

2. Literature Review

Quality Indicators in Web Surveys: Quality of data collected is an issue that is foremost on the agenda of survey researchers. Collecting quality data from respondents is high on the agenda of researchers and has motivated data collection teams to deploy a variety of strategies to improve data quality (Ganassali, 2008). Data of the highest quality is the ultimate goal for researchers. However, on one hand, the question as to what constitutes quality data is in some ways, still very much a work-in-progress even though some

clearly defined metrics are established. On the other hand, even though researchers might be keen to ensure the best quality data and reject data that is of poor quality, indicators and benchmarks to verify the collected data sets prove time and again to be an equally challenge task for researchers.

Many quality characteristics have been studied over the last fifty years or so and they serve as the main indicators of the quality of surveys in general, and specifically, the quality of data collected from survey instruments. The number of completed and partially completed questionnaires, number of blocks of questions completed, dropout profile, time required to complete the entire survey and number of missing values were variables investigated by Heerwegh and Loosveldt (2002). Don't know responses (DKs), item missing data, compliance with survey instruction; survey completion time was investigated by Heerwegh et al., (2005). Kwak and Radler (2002) explored response rate and response speed, respondent profile, item non-response and length of open-ended questions. Don't knows, response rate, response differentiation score, acquiescence score, item non response rate, was furthered studied by Heerwegh (2009); accuracy of answers on factual knowledge questions by Elo (2009); response rate, drop-out rate, filling-up rate, abundance of responses (response length and number of responses, variety of differentiation of responses, satisfaction of the respondent by Ganassali (2008) and accuracy of data against benchmarks by Yeager et al., (2009).

While it is relatively straightforward to measure the quality of data collected from some questions, it is a much more complex task for others. Factual questions can be validated through comparisons with correct answers (Elo, 2009). This process presupposes the availability and accessibility of correct answers, and importantly, that the validity of these answers are known. Other socio-demographic variables such as age, gender, education level and so on, can be similarly validated against existing benchmarks (Yeager et al., 2009). This is also dependent on the availability of benchmarks and that these benchmarks have been validated and are proven to be reliable. In some cases however, is more problematic to validate other variables when no benchmarks or correct answers to validate responses, are available. This often leads to the deployment of sometimes 'questionable' quality assumptions. Quality characteristics that appear to be best-fit for the moment are often deployed. A Word count for examples is a commonly used quality measure. But this raises an important question. Does a higher 'word count' from a response necessarily indicate a better quality of response compared to a lower work count from another response? A similar question could be asked of 'higher response rates'. Even though higher response rates provides insurance against non response bias (Heerwegh, 2009), can we say with a high degree of certainty that a higher response rate lends itself to higher quality of data? Further, higher completion rates in non-probability samples do not necessarily correlate with a more accurate data set. This was found to be consistent with previous studies (Yeager et al., 2009). The opposite of these scenarios lead to further questions. Lower response rates do not necessarily mean that respondents are not providing accurate responses. However, tangible evidence to support either of the two set of scenarios are yet to emerge in overwhelming proportions from empirical studies.

Strategies to Improve Data Quality in Surveys: Given the abundance of character definitions of data and survey quality explored, researchers have experimented with various strategies for improving quality using different number of techniques with various combinations and permutations. Most common among these are incentive use of in various forms, personalization of different degrees and questionnaire design from a technical perspective. Researchers and Practitioners have worked tirelessly to refine, improve and invent new and innovative designs for Web Surveys to aid and simplify the process of data collection from survey respondents. This is with the hope that better designs will simplify the data entry process for respondents and therefore increase the quality of response. Indeed a huge body of research exists that speaks to the issue of design of the survey instruments from a technical perspective and the quality of data collected (Couper, 2000; Manfreda et al., 2002, & Dillman et al., 2005). However, the best designed survey, with the most articulated questions is not a guarantee that valid or accurate data will be supplied by the respondents (Knapp & Kirk, 2003).

Optimized design parameters and conditions, and the right environment for collecting data does not in itself guarantee accurate submissions where requested. In a study that explored responses to the 'age question', Gendall and Healey (2008) perhaps provide the most visible experimental study using web surveys. Using a survey design approach, the study focuses exclusively on the accuracy of data collected from respondents on the 'age' variable. This study suggested that web survey designers solicit respondents' age by asking for date of birth or the year in which the respondent was born. The study further warned against the use of drop down menu designs because they can lead of unintended incorrect

responses. However, a number of issues might arise as a result of asking respondents for their age or date of birth. This is a more direct form of arriving at respondents' age and may result in deliberate misreporting by respondents.

Other measures that are thought to be representative of quality are documented. Heerwegh (2002) examined response time for two different formats of web survey elements (radio buttons vs. drop-down boxes). In this study, accuracy of data was not the core measure of quality. Response time was a used as a function of data quality. This is an interesting metric and lends itself to questioning. Faster response times may or may not be a measure of the quality of all factual questions, or any question type at all. In some cases, respondents may need to recall responses from memory and that may account for some of the response time.

Personalization and incentives have been used in many studies as a means of improving the overall quality of data (see Heerwegh & Loosveldt, 2006; Heerwegh, 2006; Bosnjak & Tuten 2003). The proxy use of these interventions may have improved data quality but they were not specifically used to test accuracy; that is to say that these interventions were not isolated and studied systematically for their effects on accuracy. No known study have attempted to explicitly experiment with any of these interventions to investigate their effects on validity and reliability of web survey data and specifically socio-demographic data. And even though (Yeager et al., 2009) used incentives in a number of experiments in studies of accuracy of data, they did not explicitly test for accuracy outcomes as a result of incentives. So an important question emerges - do incentives affect respondents' responses to factual questions? The following sections address these issues and questions in the preceding section.

Threats to Accuracy, Validity and Reliability: Given that all conditions under the control of the survey administrator are perfect and that all is in place for respondents to participate in a survey, why would respondents choose to respond with incorrect data or not to supply any data at all? Various theories have been proposed to explain respondents' participation and question answering behaviors on surveys. Social Exchange Theory, (Blau, 1964; Dillman, 2000) provide insights into the participation or non-participation of respondents in self-administered surveys. The core of this theory describes response in terms of the benefits it is expected to bring to the respondent. Further, accurately responding to surveys is a measure of trust that expected rewards outweigh anticipated costs (Dillman, 2000). The Theory of Planned Behavior (Ajzen, 1985; 1991) provides another core explanation for respondents behavior towards surveys and explains the respondents intention to perform the behavior in question based on plans and actions, beliefs about likely consequences which leads to a decision that is very goal-oriented (Bosnjak et al., 2005).

Krosnick (1991) theory on satisficing provides a strong basis for explaining response errors in surveys as noted by Heerwegh and Loosveldt (2008). The very core of this theory suggests that it takes a great deal of cognitive effort in order to optimally engage in the question-answering process (Krosnick, 1991). Tourangeau, Rips, and Rasinski (2000) describe four processes that they consider necessary in the process; interpretation of meaning and intent of each question, retrieval of relevant information from memory, use the retrieved information for creating a summary judgment, report the judgment considering available alternatives. Further, Heerwegh and Loosveldt (2008) stated that while some respondents may take all the steps suggested, others may opt for a short-cut cognitive approach to reduce required effort and in the process of doing so exhibit satisfying behavior. In testing the theory of satisfying across survey modes, Heerwegh and Loosveldt (2008) found lower quality of data supplied by respondents in web surveys compared to face-to-face interviews. Measures of quality used were 'don't know', response rates, rating scale values and item non-response. These quality measures are more applicable for opinion type questions than factual questions. Nevertheless, the implications of 'item non-response' may also affect factual questions, with response patterns correlating to those of opinion-based questions.

Chang and Krosnick (2009) found web survey to produce less satisfying behavior and higher concurrent validity when compared to telephone-administered survey. However, the study reveals discrepancies between probability and non-probability samples. They found that practice at completing surveys online and therefore experience increases accuracy of responses. This could perhaps be attributed to increasing level of trust place in the survey process. They also found that topic of interest lead to higher levels of reported accuracy. Respondents are more likely to be truthful when they become engaged or generally interested in the topic. The need to report accurately may be as a result of intrinsic and extrinsic benefits

that might result from the survey process. However, not all surveys will be of interest to respondents and this might conversely affect the accuracy of responses. The non-probability samples reported the highest level of accuracy. This is likely due to self-interest in the topic of the survey and therefore a desire to self-report as accurately as possible.

Social Desirability Response Bias is a bias in respondents' responses caused by the individual respondent's desire or tendencies to favorably confirm to the actual or perceived values and norms of the society or culture which they belong(Leggett et al., 2003; Nunnally, 1978). Respondents desire for approval or acceptance is achieved through culturally acceptable and appropriate behaviors; a need for social approval. This theory provides another plausible explanation for the way respondents may behave towards surveys especially when confronted with sensitive questions and when there is apparent lack of privacy and personalization of salutations (Heerwegh, 2005). This has implications for factual-type questions, where respondents could be tempted to respond in a manner that appears to satisfy the minimum needs of the survey administrators, and in the process report what they think is required and not necessarily what is actually correct responses. Nevertheless, it is suggested that ensuring maximal privacy is one way of addressing this form of measurement errors. However, maximizing privacy rules out any opportunity for verification of collected data and this creates a further dilemma for researchers; a recursive dilemma. Notwithstanding this problem, researchers have continued to explore ways to solve this problem. Newman et al., (2002) found a positive correlation between the degrees of granted anonymity (privacy) and with the amount of personal information respondents are likely to reveal. But this does not quite answer the question about the actual accuracy of what is provided. Galesic et al., (2006) suggested that the absence of interviewers in web surveys may reduce social desirability bias but again there is no guarantee that respondents will report accurately and there is even less of a change for survey administrators to perform validation and verification.

3. Survey of Previous Work on Accuracy of Survey Data

For many decades researchers have investigated accuracy and reliability of data collected from various modes of surveys. However, analyses have revealed mixed results. Hyman (1944) provides an account of one of the first studies on accuracy and reliability on a number of behavioral variables considered to be of high prestige. In three separate studies, varying levels of distortion between the data supplied by respondents when compared to recorded data were observed. He suggested that for the behaviors considered to have a high prestige character; results from opinion polls should be used with greatest caution.

Cahalan(1968), conducted a probability mail survey called the Denver Validity Survey to measure against official records, the relative accuracy of various respondents factual characteristics; age, sex, and socio-economic status. Varying levels of accuracy were reported with a tendency for younger respondents to be less accurate. A number of potential sources of errors which could have influenced the accuracy of data were noted. In a later study, Peterson (1984) empirically studied 'how to best ask the age question' for most accurate data by experimenting with four (4) different versions of the age question and compared results for accuracy from each version with data from available records. He found high levels of reported accurate data for all questions. Miller and Groves (1985) conducted a telephone survey of on a sample of victims who had reported to the police. The aim of this study is to scrutinize the process of 'record checks' in an effort to investigate 'record checks' and 'matching issues' with the self- reporting process. Some discrepancies were found in 'matching' records.

Warriner (1991) investigated the quality of self reports of mail surveys by examining the quality of self-reported data of financial expenditure on utility services. Benchmark data for this study was obtained from supplier records. Overall accuracy, direction of bias and influence on relationships were the three quality assessments conducted. Analysis revealed accurate self-reports at the overall and aggregate level but inaccurate at the individual and disaggregate level. The distinction between aggregate level benchmarks and individual level benchmarks is important and the issues that are likely to result must be understood. Aggregate level benchmarks are more likely to be obtained from companies and institutions since no individual piece of identity data is shared. Privacy is paramount and companies will not let up in this respect. Summary data is more likely to be sold or shared with researchers than detailed data.

However, aggregates can be misleading. Take for example, 10 respondents on a survey resulted in an average age of 35 years. However, 5 respondents over-reported their age by5 years and 5 respondents

under-reported their true age by 5 years. The aggregate, in this case, an average, is accurate, however all 10 respondents individually reported their true age incorrectly. The same could hold true for non-numeric data. How are we sure that a male respondent did not select female instead? Here lies the difficulty of privacy and data validation. Misleading benchmarks are problematic for many reasons other than those listed above. Analysis and conclusions that uses key variables such as age and gender can be compromised significantly. After all, if a researcher uses incorrect data in an analysis, the analysis technique will not help to correct for this inaccurate input and may render an entire study highly questionable. Of course this is often unknown since researchers operate from the premise that the data they have is the best possible data-set from respondents and therefore results are accepted as is.

In a review on the validity of self reports for mentally ill patients(Crisanti et al., 2003) accepted from previous studies, the suggestion that sex and race seems to have little effect on the validity of self reports and as such did not consider those variables in the review (Crisanti et al., 2003). There are no similar studies reported on this particular behavior pattern arising from Web Surveys. Therefore this remains a potential area to investigate as sex and race may be variables that affect how respondents behave online. Gendall and Healey (2008) experimented with various design elements using a mail and an online survey to study the use of the four different versions of the 'age question' for accuracy of age responses. They found that the option 'what age range are you in?' together with the use of drop-down menus, in an online survey produced significantly higher proportions of incorrectly reported ages by respondents. This was partially inconsistent with a previous finding by Peterson (1984) and Smith et al., (1997). There's no clear-cut indicator of what might be causing the fluctuation in accuracy of response to this question; whether it's the design, question wording or respondent issues. As Presser (1984) pointed that it is not clear whether invalidity is respondent-specific or item-specific.

Elo (2009) evaluated the reliability of answers of factual knowledge questions using a probability-based offline sample survey and an offline non-probability based web survey. Results suggest that online surveys are capable of providing reliable information for different groups of respondents but should not be used for drawing conclusions on the subject studied in a population. Non-probability samples are likely to attract respondents with interest in the topic of the survey. This may in turn affect the levels of reported accuracy as suggested by Chang and Krosnick (2009). In cases where the factual questions are knowledge-based, particular group of respondents may be drawn to the survey simply because they know the answers to the questions. Other respondents may not, and is likely to not respond if there is little or no incentive. However, we still have little evidence on factual questions that are socio-demographic in nature and how probability and non-probability samples respond to them.

Yeager et al., (2009) investigated the accuracy of data from probability and non-probability web surveys and Random Digital Dialing (RDD) survey on American Adults. Respondents' data were compared with benchmark data from previously conducted surveys. They found that probability sample surveys were consistently more accurate than non-probability sample surveys. However points of interest lie in the quality of benchmark data. Benchmarks for probability samples are likely to be more robust than benchmarks for non-probability samples. Since probability samples are drawn from known populations, greater opportunities exist to develop and verify more accurate benchmarks. It therefore means that the reported levels of accuracy across the probability and non-probability respondents may partially be a function of the accuracy of the benchmarks and not necessarily on the reported accuracy of responses. However, this is largely speculative. Chang and Krosnick (2009) suggested that self-interest may encourage respondents to complete surveys with accurate responses. But with limited opportunities for verification and validation on non-probability samples, the actual effect of self-interest also remains highly speculative a rationale for respondents providing accurate responses. Additionally, it could be argued that benchmarks from previous studies are only as accurate as the responses provided in those studies and are therefore limited. Comparisons using those benchmarks will be limited to the scope of those benchmarks and the studies they are derived from, unless these studies were also benchmarked.

A small number of studies have emerged on how respondents report on factual questions. Inconclusive results dominate the literature and are evident across the studies published. In most cases, no particular 'theory' provides explanations as to why respondents might choose to report inaccurately to factual questions in a survey. Web Surveys, being fairly new to the survey landscape, and conducted with mainly virtually unknown populations, complicates the issue of verification and validation of data further. In particular, generating representative samples and having available benchmarks to conduct these checks

have proved fairly elusive for researchers mainly because benchmarks are either unavailable or the benchmarking process itself it too costly or impractical to do.

4. Discussion and Conclusion

Dating back to the early 1940s, survey researchers have attempted to investigate the accuracy and reliability of responses to survey questions. The basic premise upon which such investigations are rooted in the principle that quality data is a pre-requisite for quality analyses, careful conclusions and quality decision making. In more recent times, Web Surveys have become a common mode to collect large amounts of data from online samples. However, this mode has brought with it added challenges. Difficulties with sampling together with verification of actual respondents are two of the most important issues being addressed. The literature on responses to factual and knowledge-based questions from Web Surveys is sparse and inconclusive at best. However, a number of recent studies highlight the magnitude of the issues faced by researchers. Yeager et al., (2009) conducted probability-based web and telephone surveys and non-probability web surveys. They found that probability web surveys were more highly accurate across demographics and non-demographics that non-probability surveys. Probability surveys imply known benchmarks, as it is reasonable to suggest that data is available for known populations from which samples are drawn. Potential respondents signing-on to organizations for the purpose of completing surveys may required to provide accurate data, given that this is usually tied to rewards for completing surveys. However, this is not altogether an established fact from the literature and deserves further investigation. Similarly, benchmarks derived from organizations' employee databases are likely to be highly accurate, with some form of prior verification expected. Accessing these databases is problematic as they raise the issues of privacy and security and usually result in restricted-use policies by organizations.

On the other hand non-probability samples would suggest that respondents' socio-demographic data are relatively unknown. This is likely to affect the quality of benchmark data available and subsequently the quality of benchmarking. The challenge therefore is for researchers to develop strategies that encourage respondents to report accurately as it might become extremely impractical and even impossibly to verify the data collect. One method that might be useful to benchmark non-probability samples is through a test-retest approach, using previous benchmark datasets as a means of verifying new datasets. There is no guarantee that previous non-probability samples will be similar to the next non-probability samples. However, in cases where the overall population is know, using benchmarks from previous samples might be a consideration but no major conclusions can be drawn.

Elo (2009) found that non-probability, self-selected respondents provided more accurate answers to knowledge-based questions. These types of questions are likely to attract respondents that are learned in the topic of the survey. As a result, a higher overall level of accuracy to questions is expected. Probability samples may not be as accurate since sampled respondents may have little or no interest in the subject matter. However, knowledge-based questions administered via the World Wide Web are susceptible to 'cheating'. Respondents, whether probability-based or non-probability could check correct answers by conducting basic online search. Reported correct answers may have little to do with the type of sampling.
Future work should explore other populations where response patterns might be very different. Probability Web Surveys are desirable; however this is not always possible. When non-probability samples are used, methods to ensure best quality, most-accurate responses, must be deployed. Given that techniques to verify collected data is limited and given that benchmarks are likely to be unavailable and inadequate, approaches to encourage accurate responses from non-probability samples in web surveys and strategies to benchmark and verify collected data present opportunities for web survey researchers. Online populations are difficult to track and any form of verification is difficult to conduct. However, if researchers wish to rely on the studies conducted with web surveys, strategies to validate data from web surveys must be carefully developed and executed.

Factual socio-demographic and knowledge-based questions can be checked against correct answers to verify their accuracy. Variables such as socio-demographic can be checked for accuracy if existing benchmark data exists for these variables. One major hindrance to this process is the availability of benchmarking data. This does not prevent researchers from asking these types of questions when and where necessary. However, if benchmarks are not readily available, data sets are accepted as is, and may in some instances lead to faulty analysis and conclusions. The greatest threat to the successful deployment of these strategies is the development and availability of quality benchmarks and trust in the

validity of these benchmarks. The development and availability of benchmarks are affected by issues such as privacy and security of data. Verification of data from web surveys is complicated by anonymity, making it difficult to develop benchmarks that are reliable.

References

Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In *J. Kuhl & J. Beckman (Eds.), Action-control: From cognition to behavior* (11–39). Heidelberg: Springer.

Ajzen, I. (1991). The Theory of Planned Behavior. *Organizational Behavior and Human Decision Processes*, *50*, 179-211.

Blau, P. (1964). Exchange and Power. New York: John Wiley and Sons.

Bosnjak, M., & Tuten, T. L. (2003). Prepaid and Promised Incentives in Web Surveys: An Experiment. Social Science Computer Review, 21(2), 208-217.

Bosnjak, M., Tuten, T. L., & Wittmann, W. W. (2005). Unit (Non) Response in Web-Based Access Panel Surveys: An Extended Planned-Behavior Approach. *Psychology & Marketing*, 22(6), 489–505.

Cahalan, D. (1968). Correlates of respondent inaccuracy in the Denver validity survey. *Public Opinion Quarterly*, 32, 607-21.

Chang, L. & Krosnick, J. A. (2009). National Surveys via RDD Telephone Interviewing Versus the Internet: Comparing Sample Representativeness and Response Quality. *Public Opinion Quarterly*, 73(4), 641–678.

Cobanoglu, C., Warde, B., & Moreo, P. J. (2001). A comparison of mail, fax and web survey methods. *International Journal of Market Research*, 43(4), 441–452.

Couper, M. P. (2000). Web Surveys. A Review of Issues and Approaches. *Public Opinion Quarterly* 64(4), 464–94.

Crisanti, A., Laygo, R., & Junginger, J. (2003). A review of the validity of self-reported arrests among persons with mental illness. *Current Opinion in Psychiatry*, 16, 565–569.

Dillman, D. A. (2000). *Mail and Internet Surveys: The Tailored Design Method*. New York: Wiley.

Dillman, D. A., Gertseva, A., & Mahon-Haft, T. (2005). Achieving Usability in Establishment Surveys through the Application of Visual Design Principles. *Journal of Official Statistics*, 21(2), 183–214.

Elo, K. (2009). Asking Factual Knowledge Questions: Reliability in Web-Based, Passive Sampling Surveys. *Social Science Computer Review*, 0(0), 1-16.

Galesic, M., Tourangeau, R. & Couper, M. P. (2006). Complementing Random-Digit-Dial Telephone Surveys with other Approaches to Collecting Sensitive Data. *American Journal of Preventive Medicine*, 31(5), 437–443.

Ganassali, S. (2008). The Influence of the Design of Web Survey Questionnaires on the Quality of Responses. *Survey Research Methods*, 2(1), 21-32.

Gendall, P., & Healey, B. (2008). Asking the age questions in mail and online surveys. *International Journal of Market Research*, 50(3), 309-317.

Gibson, R., & McAllister, I. (2002). The future of national election surveys? Paper presented at the 98th Annual Meeting of the American Political Science Association (APSA). Evaluating online election surveys in Australia, Boston, M A.

Heerwegh, D. (2002.). Describing response behavior in web surveys using client-side paradata. Paper presented at the International Workshop on Web Surveys, Mannheim, Germany.

Heerwegh, D., & Loosveldt, G. (2002). Web Surveys: The effects of Controlling Survey Access Using PIN Numbers. *Social Science Computer Review*, 20(1), 10-21.

Heerwegh, D. (2005). Effects of Personal Salutations in Email Invitations to Participate in a Web Survey. *Public Opinion Quarterly*, 73(4), 641–678.

Heerwegh, D., Vanhove, T., Matthijs, K. & Loosveldt, G. (2005). The Effect of Personalization on Response Rates and Data Quality in Web Surveys. *International Journal of Social Research Methodology*, 8(2), 85-99.

Heerwegh, D. (2006). An Investigation of the Effect of Lotteries on Web Survey Response Rates. *Field Methods*, 18(2), 205–220.

Heerwegh, D., & Loosveldt, G. (2006). An Experimental Study on the Effects of Personalization, Survey Length Statements, Progress Indicators, and Survey Sponsor Logos in Web Surveys. *Journal of Official Statistics*, 22(2), 191–210.

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(5), 836-846.

Heerwegh, D. (2009). Mode Differences Between Face-to-Face and Web Surveys: An Experimental Investigation of Data Quality and Social Desirability Effects. *International Journal of Public Opinion Research,* 21(1), 111-121.

Hyman, H. (1944). Do They Tell The Truth? *Public Opinion Quarterly,* 8(4), 557-559.

Knapp, H., & Kirk, S. A. (2003). Using pencil and paper, Internet and touch-tone phones for self-administered surveys: does methodology matter? *Computers in Human Behavior*, 19(1), 117–134.

Krosnick, J. A. (1991). Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology,* 5, 213–36.

Kwak, N. & Radler, B. (2002). A Comparison between Mail and Web Surveys: Response Pattern, Respondent Profile, and Data Quality. *Journal of Official Statistics*, 18(2), 257-273.

Leggett, C. G., Kleckner, N. S., Boyle, K J., Duffield, J .W. & Mitchell, R. C.(2003). Social desirability bias in contingent valuation surveys administration through in-person interviews. *Land Economics*, 79 (4), 561–575.

Manfreda, K. L., Batagelj, Z., & Vehovar, V. (2002). Design of Web survey questionnaires: Three basic experiments. *Journal of Computer-Mediated Communication,* 7(3). Retrieved January 5, 2010 from http://jcmc.indiana.edu/vol7/issue3/vehovar.html.

Miller, P. V., & Groves, R. M. (1985). Matching Survey Responses to Official Records: An Exploration of Validity in Victimization Reporting. *Public Opinion Quarterly*, 49(3), 366-380.

Newman, J. C., Des Jarlais, D. C., Paone, D., Turner, C. F., Gribble, J., & Cooley, P. (2002). The differential effects of face-to-face and computer interview modes. *American Journal of Public Health, 92,* 294–297.

Nunnally, J. (1978). Psychometric Theory. New York: McGraw-Hill.

Peterson, R. A. (1984) Asking the age question: a research note. *Public Opinion Quarterly*, 48(1), 379–383.

Presser, S. (1984). Is Inaccuracy on Factual Survey Items Item-Specific or Respondent-Specific? *Public Opinion Quarterly*, 48(1), 344-355.

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, 257-285.

Smith, J. H., Deeg, D. J. H. & Schmand, B. A. (1997) Asking the age question in elderly populations: a reverse record check study. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 52(4), 175–177.

Tourangeau, R., Rips, L. J., & Rasinski, K. A. (2000). *The Psychology of Survey Response*. Cambridge: Cambridge University Press.

Warriner, G. K. (1991). Accuracy of self-reports to the burdensome question: survey response and non response error trade-offs. *Quality & Quantity*, 25, 253-269.

Yeager, D. S., Krosnick, J. A., Chang, L., Javitz, H. S., Levindusky, M. A., Simpser,A., & Wang, R. (2009). Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-Probability Samples. Knowledge Networks. Retrieved January 2, 2010, from http://www.knowledgenetworks.com/insights/docs/Mode-04_2.pdf