Title
Measurement Bias with Mixed-mode Patient-reported Outcome (PRO) Survey Administration: Measurement Equivalence, Cost, and Data Quality

Author
Broering, Jeanette Marie

Publication Date
2012

Peer reviewed|Thesis/dissertation

Measurement Bias with Mixed-mode Patient-reported Outcome (PRO) Survey Administration: Measurement Equivalence, Cost, and Data Quality

by

Jeanette Marie Broering, RN, PhD, MPH

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Nursing

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

[redacted]

Acknowledgements

Completion of a course of doctoral studies and dissertation would not be possible without numerous sources of encouragement and support. I would like to acknowledge the longstanding contributions of the men who have so diligently and unselfishly given of their time to complete their patient-reported outcomes to the Cancer of the Prostate Strategic Urologic Research Endeavor (CaPSURE**)** study**,** a disease-specific prostate cancer registry that was established in May of 1995. To the 40 community-based, academic, and Veteran's Administration urology practices for their longstanding dedication to participate in the CaPSURE study. The onsite physician clinical investigators, and the respective research staff, have demonstrated enormous dedication to the project by submission of clinician reported outcome metrics about the diagnosis and treatment of their patient population.

For the measurement equivalence study I would like to acknowledge the interdisciplinary team who were involved with that study. Specifically, the research staff at the University of California San Francisco (UCSF) Genitourinary Cancer Epidemiology and Population Science (G-CEPS) data coordinating center: Lydia K. Moody, BA; Bryan Lanzman, BS; Paige L. Marr, MSc.; Suzanne C. Lessard, BA; and Eric P. Elkin, MPH for their participation in the project design and study operations. To acknowledge Kenneth Lewis, Jr. and Gordon Olesek, at Net Outcomes Corporation in Santa Clara, California, for their contributions with the design of the web application. To acknowledge Michael Cole at METROFORM in Gold River, California for his technical assistance with the TeleForm® optical character recognition software program and

printing services of the TeleForm® surveys. Finally to acknowledge Janeen DuChane, PhD, formerly of TAP Pharmaceutical Products in Lake Forest, Illinois, for funding the software development of the web-mode survey and funding the measurement equivalence study.

For the time and motion study that supported the cost analyses, I would like to acknowledge the G-CEPS data operations team. Specifically, Daniel Donohue, BA; Gary Havlicek, BS; Suzanne C. Lessard, BA; Paige L. Marr, MSc.; Lydia K. Moody, BA; Tatyana Noller, Ada Sanchez, BA; Paula Solomon, BS; and Ali Zargham, BA. This team was dedicated to exploring our business processes and always willing to innovate with the improvement of the data quality assurance operations of CaPSURE.

I would like to acknowledge the wonderful assistance with statistical analyses and SAS output interpretation of my analytic finding by the G-CEPS biostatistics group specifically, Erick P. Elkin, MPH and Alan Paciorek, BS. Alan, your understanding of complex mathematical constructs and willingness to discuss output, were just the best!

I would like to acknowledge the funding support for CaPSURE. From 1995 to 2006 CaPSURE was generously supported by TAP Pharmaceutical Products Inc. of Lake Forrest, Illinois. In 2007, the CaPSURE registry was donated to the Regents of the University of California; where it currently is supported by the Department of Urology at the University of California San Francisco under the leadership of the principal investigator, Peter R. Carroll MD, MPH. Currently, CaPSURE is supported by an unrestricted educational gift from Abbott Laboratories (Chicago, IL), by the National Institutes of Health/National Cancer Institute (5RC1CA146596), and by the Agency for Healthcare Research and Quality (1U01CA88160).

I would like to acknowledge the support and encouragement provided to me by my family, specifically, my husband, Jacob K. Moody, and our children, Paul and Lydia Moody. Additionally, to my father, Leander A. Broering (1918-2011), who always wanted his daughter to be a doctor. They ALL were supportive throughout this five year journey.

I would like to acknowledge the faculty who provided support and mentorship throughout the past five years. I would like to thank my academic advisor, Robert Newcomer, PhD, for being the chair of my qualifying examination committee. Bob's guidance on survey research methods and encouraged me to explore the nuances of mixed-mode survey administration. I would like to acknowledge the other members of my qualifying and dissertation committees. Leslie S. Wilson, PhD, Professor from the UCSF School of Pharmacy, whose input on the development of the time and motion study and the conduct of the cost analyses, were essential to the content in Chapter 3 of this dissertation. To Mark S. Litwin, MD, MPH, Professor of Urology at the David Geffen School of Medicine at the University of California Los Angeles and the developer of the UCLA-Prostate Cancer Index, for his willingness to participant as a committee member on my qualifying examination papers and my dissertation, despite all of his numerous responsibilities at UCLA. His careful reading of drafts, comments, and constructive critiques were always welcomed. To Christine Miaskowski, RN, PhD, FAAN, Professor and Associate Dean at the UCSF School of Nursing, who so generously acted as the chair of my dissertation committee. Her knowledge of oncology nursing and guidance on seeing the "big picture" helped to keep me on track and moving forward. She provided the best mentorship a doctoral student could ever want. Her tireless review

of drafts and revisions was most appreciated. Her insights and professional guidance were exceptional. I can only hope to aspire to be a nurse scientist of equal stature to what she has become. Dr. Miaskowski's personal passion and commitment to improving the quality of life and delivery of the highest quality healthcare to patients with cancer is truly admirable.

Abstract

Background: Measurement with mixed-mode administration (i.e., paper-mode versus web-mode) of patient-reported outcome (PRO) health surveys may vary by psychometric measurement equivalence (ME), differential response rates, costs, and data quality.

Purpose: Three data-based papers evaluated ME, cost, and data quality when a mixed-mode approach for self-administration was used.

Methods: Data were derived from the Cancer of the Prostate Strategic Urologic Research Endeavor (CaPSURE) study, a multi-institutional United States based longitudinal prostate cancer registry. ME study used a randomized cross-over design of 209 participants. Cost and data quality analyses used a cross-sectional time frame of 5,008 participants. The Short Form-36 (SF-36) and the UCLA-Prostate Cancer Index (UCLA-PCI) were examined.

Results: ME study, participants were White (97%), college educated (66%), reported an annual income > $75,000 (46%), and a median age of 69 years. Intraclass correlation coefficients were high (*ICC* = 0.66–0.97). Exact percent agreement was high (≥ 0.89). For the cost and data quality analyses, 90% opted for paper-mode and 10% for web-mode. Total costs to process 5,008 surveys was $75,216 or on average $15.02 per survey with web-mode costs significantly higher ($18.47/survey) than paper-mode ($14.66/survey). Web-mode surveys had lower error rate (9% versus 14%, NS). Overall response rate was 77%, paper-mode 76% versus 88% web-mode (p <0.01). Predictors of response were older age, being Caucasian, having attained college education, and living in a significant relationship. Paper-mode participants were less likely to respond (OR 0.69, 95%CI .68–.70) and had significantly more missing data on the SF-36, the UCLA-

PCI, and fewer computable scale scores (all significant $p < 0.001$). All mean scale scores were lower among paper-mode participants but effect sizes for clinically meaningful differences between modes were small. CaPSURE participants had higher physical and mental function when compared to US based age-stratified norms.

Discussion: The use of a mixed-mode approach found support for ME; costs to administer were lower for paper-mode but web-mode had higher accuracy; and significant but small differences in data quality. Findings suggest that mixed-mode administration did not introduce significant measurement bias but allowed for participation by more diverse participants (e.g., older, poorer health, non-Caucasian).

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction and Dissertation Aims

This dissertation includes three data-based papers focused on the contemporary issues of measurement bias in the conduct of longitudinal assessment of patient-reported outcomes (PRO) among men diagnosed and treated for prostate cancer (PCa). Data for all analyses were derived from the Cancer of the Prostate Strategic Urologic Research Endeavor (CaPSURE) study, a multi-institutional United States (US) based longitudinal registry database of men with PCa. CaPSURE has collected physician-reported clinical outcomes, PRO data, and economic data, on over 14,450 men predominately from community-based practice settings (Lubeck, Litwin, Henning, et al., 1996). Administrative data about data quality assurance and operations have been collected by the University of California San Francisco (UCSF) data coordinating center since 1999. The aims of this dissertation research were to:

1. Examine measurement equivalence (ME) between PRO instruments that were self-administered using paper-mode versus web-mode for systematic differences in correlation coefficients and mean scale scores by mode of administration, order of administration, and the interaction of mode and order of administration (Arpinelli & Bamfi, 2006; Coons, Gwaltney, Hays, et al., 2009; Gwaltney, Shields, & Shiffman, 2008).
    a. To describe sociodemographic and clinical characteristics of web-mode respondents.
    b. To describe participant preferences for mode of administration.

2. Examine the cost of employing mixed-mode questionnaire administration within a single cross-sectional administration of the CaPSURE questionnaire.
    a. Fixed and variable costs for each mode were estimated that accounts for differences in volume, mode mix, and response rates.
    b. A cost-comparison analysis was conducted that examined the cost per survey, the cost per response, the cost per error free response, and total costs framed from the perspective of the CaPSURE data coordinating center. All costs were based in 2006 U.S. dollars.
    c. Additional cost analyses were performed from the four perspectives of the cost per survey, the cost per response, the cost per error free response, and total costs to administer both modes. The benefits and cost-effectiveness of mixed-mode was examined by varying the volume and the proportion of users by mode for a hypothetical sample of 5,000 survey units (Torrance, Siegel, & Luce, 1996).
3. Examine issues of data quality for administration of a generic patient-reported health related quality of life (HRQOL) instrument (i.e., the Short Form 36 (SF-36) version 1), and a disease-specific instrument (i.e., the UCLA-Prostate Cancer Index (UCLA-PCI)), (Hays, Sherbourne, & Mazel, 1993; Litwin, Hays, Fink, et al., 1998; Litwin, Hays, Fink et al., 1995; Lubeck, Litwin, Henning, & Carroll, 1997;).
    a. Important parameters of data quality in mixed-mode survey administration that can affect generalizability of results were examined, including:

i. Response rate by mode;
ii. Characteristics of non-respondents versus respondents by mode (coverage bias);
iii. Comparison by mode for data completeness (item non-response error);
iv. Examination of response effects (ceiling and floor effects) and mean scale scores by mode compared to national norm-based data for men by relevant age groups.
v. Anchor-based comparison of the CaPSURE sample against United States age-stratified norms of men using the physical and mental component scale scores

Background

Prevalence: Approximately 11.4 million persons in the United States (US) are living with a diagnosis of cancer. The majority of cancer survivors (61%) are age 65 or older (Ries, Melbert, Krapcho, et al., 2008). As the population of older persons in the US increases, it is expected that elderly cancer survivors may experience a reduction in their HRQOL due to treatment associated toxicities (Beck, Towsley, Caserta, Lindau, & Dudley, 2009).

PCa Clinical Characteristics: Approximately 91% of newly diagnosed PCa cases will be local or regionally confined within the pelvis with an estimated 5-year survival rate of nearly 100%. Increases in survival are attributable to improvements in early detection and treatment (Siegel, Ward, Brawley, & Jemal, 2011).

Multiple treatment modalities exist for PCa, each with known toxicities or trade-offs and with no single preferred therapy recommended (Bannuru, Dvorak, Obadan, et al., 2011; Wilt, MacDonald, Rutks, et al., 2008). Active surveillance as a treatment option is defined as a period of medically managed observation to monitor disease progression. Improvements in transrectal ultrasound imaging and biomarker monitoring through the prostate specific antigen test have contributed to the viability of this surveillance strategy. Surgical options include radical prostatectomy performed by one of several different approaches (i.e., open via perineal, retropubic approaches, laparoscopic with or without a robot) or freezing of the prostate by cryosurgery. Radiation options include external beam (3-D conformal, intensity modulated radiotherapy, cyberknife) or implantation of permanent or temporary radioactive seeds. Hormonal therapies include variations in type of compound and interval of administration. Treatment of late stage metastatic disease includes chemotherapy, palliative care, and emerging therapies such as vaccine trials (Penson & Chan, 2007).

Survey Mode Approaches: Mode of survey administration has evolved from face-to-face interviewer administered, to self-administered questionnaires (SAQ) sent by postal mail, and finally to telephone administration of surveys with and without automation. Studies that provide empiric evidence of the effect of alternate mode designs on data quality are often limited (Groves, Fowler, Couper, et al., 2009). Two design options for mode of survey administration are important in the planning phase. First, what is the most appropriate mode of data collection method for a particular question? Second, what is the impact of a particular data collection method on survey errors and administration costs? The advantages of a web administered survey can be reduced cost,

increased timeliness, and improvements in measurement. The disadvantages are coverage of the target sample, non-responses, and challenges associated with construction of a suitable sampling frame (deLeeuw, 2005; Gwaltney, et al., 2008). For mailed SAQ, the respondent needs basic literacy skills and minimal equipment (pencil or pen). In contrast, computer surveys require specialized skills such as access to a computer, Internet service provider (ISP) access, and computer literacy.

deLeeuw and colleagues define mixed-mode as the use of multiple modes to communicate with survey respondents either in the contact or data collection phase of survey administration (deLeeuw, Hox, & Dillman, 2008; deLeeuw, 2005). Using a blend of approaches (i.e., mixed-mode) may have advantages as a tactic to reduce costs and maximize response rates which is important in longitudinal studies. Groves and colleagues argue that the logic of the mixed-mode approach is to exploit the advantages of one mode (to reduce costs) while neutralizing the disadvantage of the other (reduced coverage) (Groves, et al., 2009). Empiric evidence of measurement equivalence to assure survey validity as well as an examination of data quality issues such as coverage, response rates, and item non-response are critical factors to evaluate in longitudinal studies that may start with a single SAQ mode but may modify their method of data collection to provide surveys by a mixed-mode approach (Arpinelli & Bamfi, 2006; Coons, et al., 2009; deLeeuw, et al., 2008; Gwaltney, et al., 2008).

Measurement Equivalence (ME): The American Psychological Association has defined ME as demonstrating the ability of alternative modes of survey administration to closely approximate each other such that "the means, dispersion, and shapes of the score distributions are approximately the same" (American Psychological Association, 1986).

Coons and colleagues similarly defined ME as "the function of comparability of the psychometric properties of the data obtained via the original and the adapted administration mode." (Coons, et al., 2009). Differences in derived scores from instruments can occur due to differences in how the items are presented to respondents and due to potential difficulties that the respondent may have in interacting with the computer or electronic device (i.e., computer anxiety or lack of computer literacy) (Gwaltney, et al., 2008). In a meta-analysis of 43 studies that assessed ME between paper self administered survey and electronic devices only, the mean age of the participants was 48.0 (± 13.9) years old with few studies thoroughly examining ME in older participants (≥ 65 years old) (Gwaltney, et al., 2008). The International Society for Pharmacoeconomic Outcomes Research (ISOPR) published best practice recommendations and guidance about the assessment of ME when there has been minor, moderate, or extensive changes are made to an instrument that was designed for paper administration but modified for electronic administration. The US Food & Drug Administration requires validation of an instrument when migrating from a paper mode to an electronic mode to insure ME (Arpinelli & Bamfi, 2006).

Mixed-mode PRO in Oncology: The use a of mixed-mode approach in oncology has been restricted to use in clinical care settings within academic medical centers to monitor adverse events associated with active treatment (Abernethy, Herndon, Wheeler, et al., 2008; Basch, Artz, Iasonos, et al., 2007; Dupont, Wheeler, Herndon, et al., 2009; Farnell, Routledge, Hannon, et al., 2010; Velikova, Wright, Smith, et al., 1999). None of these studies evaluated the use of a complex SAQ administered outside a clinical

environment where hands-on technical assistance is absent to support use and web navigation.

Cost of Mixed-mode: Web-administered surveys can result in substantial reduction of costs since it reduces the costs of printing, postage, data processing, and may result in higher data quality (Ekman & Litton, 2007; Kypri, Gallagher, & Cashell-Smith, 2004; McAlindon, Formica, Kabbara, LaValley, & Lehmer, 2003). Other studies suggest that the costs of web administered surveys are unknown (Ekman & Litton, 2007; Rodriguez, von Glahn, Rogers, et al., 2006). Cost evaluation of mixed-mode administration is limited and suggests that mixed-mode may be more expensive but may result in higher response rates (Greenlaw & Brown-Welty, 2009; McHorney, Kosinski, & Ware, 1994). While the use of a web-based approach may make this process more efficient; no study has compared the costs and benefits of mixed-mode approach for complex longitudinal studies of health or illness.

Assessment of Data Quality: A single study by McHorney and colleagues compared data quality when the Short Form-36 (SF-36) was administered for a national survey using a mixed-mode approach (mailed survey versus telephone) to examine the trade-offs experienced with each mode (McHorney, et al., 1994). The costs of telephone administration were 77% higher for telephone mode ($47.87 vs. $27.07, respectively). Response rates were higher for the mailed mode (79.2% vs. 68.9%). Nonresponse bias was evident for both modes. The rate of missing data was higher for paper compared to telephone (1.59 vs. 0.49 missing items) (McHorney, et al., 1994). No studies have compared mixed-mode using paper versus web-mode of administration for the SF-36.

Theoretical Framework

The total survey error framework or paradigm served as the theoretical approach to guide the examination of data quality components of survey administration (Figure 1) (Groves, et al., 2009). Mixed-mode administration can be subject to a number of measurement biases which can introduce error which affects the quality of survey statistics specifically through coverage error, non-response error, and data quality errors (i.e., missing data and systematic differences between modes that cannot be adjusted for by statistical weighting strategies). While cost is not explicitly included in the total survey error paradigm, it is relevant to the survey quality paradigm. Researchers must carefully select data collection methods (modes) which attempt to minimize both costs and errors (Groves, et al., 2009). Therefore, this dissertation research was designed to focus on the constructs of ME, costs, and data quality metrics when administering complex PRO instruments using a mixed-mode approach of paper-mode versus web-mode within a large PCa registry.

Dissertation Chapters

Chapter 1 provides an introduction the dissertation, presents the research aims, background and significance and the theoretical framework used to guide this dissertation. Three data-based papers, prepared for publication, report the findings of measurement bias when a mixed-mode approach to survey administration was used within a context of a complex longitudinal cohort cancer registry from the perspectives of ME, costs, and data quality.

Chapter 2 presents a paper that was prepared for publication which contains the findings of the measurement equivalence study for mixed-mode administration of patient-

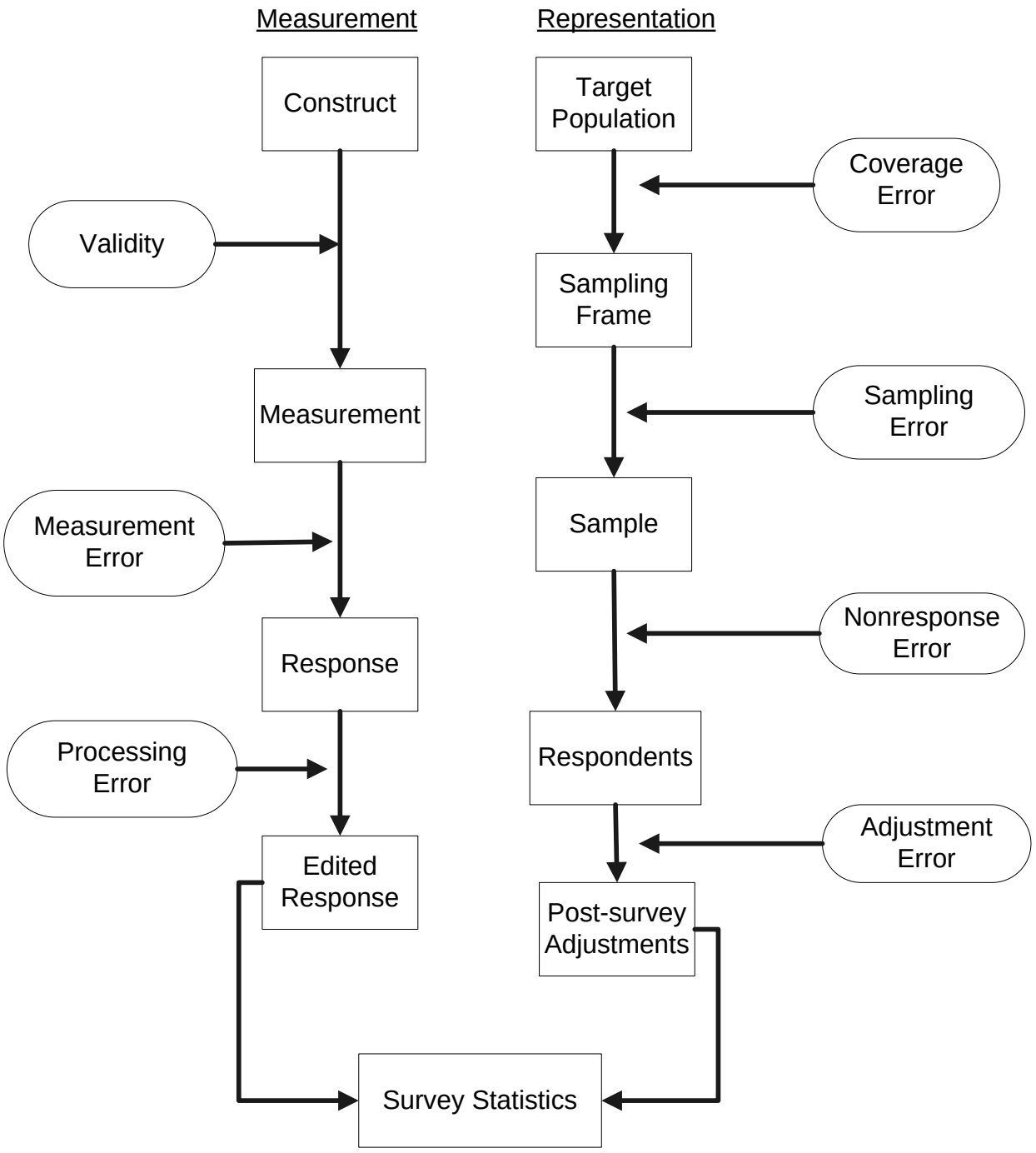
reported outcomes (PRO) instruments using classic test-retest cross over design method (Coons, et al., 2009; Gwaltney, et al., 2008). A subset of 209 CaPSURE participants volunteered to participate in the ME. Participants were randomized to mode of administration (paper then web versus web then paper) to cancel out any ordering effects associated with mode of administration.

Chapter 3 presents a paper that was prepared for publication that reports the results of a cross-sectional cohort study which compared the costs and accuracy of collecting PRO data by mixed-mode during a single administration period for CaPSURE. The mailed approach was defined as a combination of pencil-and-paper surveys mailed to participants, returned by mail, and then digitized into an electronic database using optical character recognition technology. The web approach was defined as a digital version of the same survey to be self-completed on a personal computer and stored on a secure server via the World Wide Web.

Chapter 4 presents a paper prepared for publication that the results from a cross-sectional study that evaluated differences in the data quality parameters using the same single timeframe of administration as used in the cost analyses by the two modes of administration (January to July, 2006). Comparison of data quality by administration mode used four criteria: response rate (RR), nonresponse bias, data quality, and response effects.

Chapter 5 summarizes the findings from these three papers on mixed-mode administration when employed within the CaPSURE registry. This chapter concludes with the implications for observational research methods as well as recommendations for future research.

Figure 1. Total Survey Error Paradigm



Adapted from (Groves, et al., 2009), Reprinted with permission.

References

Abernethy, A. P., Herndon, J. E., 2nd, Wheeler, J. L., Patwardhan, M., Shaw, H., Lyerly, H. K., et al. (2008). Improving health care efficiency and quality using tablet personal computers to collect research-quality, patient-reported data. *Health Services Research, 43*, 1975-1991.

American Psychological Association. (1986). *Guidelines for computer-based tests and interpretations*. Washington, D.C.: American Psychological Association.

Arpinelli, F., & Bamfi, F. (2006). The FDA guidance for industry on PROs: the point of view of a pharmaceutical company. *Health and Quality of Life Outcomes, 4*, 85-90.

Bannuru, R. R., Dvorak, T., Obadan, N., Yu, W. W., Patel, K., Chung, M., et al. (2011). Comparative evaluation of radiation treatments for clinically localized prostate cancer: an updated systematic review. *Annals of Internal Medicine, 155*, 171-178.

Basch, E., Artz, D., Iasonos, A., Speakman, J., Shannon, K., Lin, K., et al. (2007). Evaluation of an online platform for cancer patient self-reporting of chemotherapy toxicities. *Journal of the American Medical Informatics Association, 14*, 264-268.

Beck, S. L., Towsley, G. L., Caserta, M. S., Lindau, K., & Dudley, W. N. (2009). Symptom experiences and quality of life of rural and urban older adult cancer survivors. *Cancer Nursing, 32*, 359-369.

Coons, S. J., Gwaltney, C. J., Hays, R. D., Lundy, J. J., Sloan, J. A., Revicki, D. A., et al. (2009). Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO)

measures: ISPOR ePRO Good Research Practices Task Force report. *Value in Health, 12*, 419-429.

deLeeuw, E., Hox, J., & Dillman, D. (2008). Mixed-mode surveys: When and why. In E. deLeeuw, J. Hox & D. Dillman (Eds.), *International handbook of survey methodology* (pp. 299-316). New York: Psychology Press Taylor and Francis Group.

deLeeuw, E. D. (2005). To mix or not to mix data collection modes in surveys. [Review]. *Journal of Official Statistics, 21*, 233-255.

Dupont, A., Wheeler, J., Herndon, J. E., 2nd, Coan, A., Zafar, S. Y., Hood, L., et al. (2009). Use of tablet personal computers for sensitive patient-reported information. *Journal of Supportive Oncology, 7*, 91-97.

Ekman, A., & Litton, J. E. (2007). New times, new needs; e-epidemiology. *European Journal of Epidemiology, 22*, 285-292.

Farnell, D. J., Routledge, J., Hannon, R., Logue, J. P., Cowan, R. A., Wylie, J. P., et al. (2010). Efficacy of data capture for patient-reported toxicity following radiotherapy for prostate or cervical cancer. *European Journal of Cancer, 46*, 534-540.

Greenlaw, C., & Brown-Welty, S. (2009). A comparison of web-based and paper-based survey methods: testing assumptions of survey mode and response cost. *Evaluation Review, 33*, 464-480.

Groves, R. M., Fowler, J., F.J, Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). Hoboken, New Jersey: A. John Wiley & Sons, Inc.

Gwaltney, C. J., Shields, A. L., & Shiffman, S. (2008). Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: a meta-analytic review. *Value in Health, 11*, 322-333.

Hays, R. D., Sherbourne, C. D., & Mazel, R. M. (1993). The RAND 36-Item Health Survey 1.0. *Health Economics, 2*, 217-227.

Kypri, K., Gallagher, S. J., & Cashell-Smith, M. L. (2004). An internet-based survey method for college student drinking research. *Drug and Alcohol Dependence, 76*, 45-53.

Litwin, M. S., Hays, R. D., Fink, A., Ganz, P. A., Leake, B., & Brook, R. (1998). The UCLA Prostate Cancer Index: development, reliability, and validity of a health-related quality of life measure. *Medical Care, 36*, 1002-1012.

Litwin, M. S., Hays, R. D., Fink, A., Ganz, P. A., Leake, B., Leach, G. E., et al. (1995). Quality-of-life outcomes in men treated for localized prostate cancer. *Journal of the American Medical Association, 273*, 129-135.

Lubeck, D. P., Litwin, M. S., Henning, J. M., & Carroll, P. R. (1997). Measurement of health-related quality of life in men with prostate cancer: the CaPSURE database. *Quality of Life Research, 6*, 385-392.

Lubeck, D. P., Litwin, M. S., Henning, J. M., Stier, D. M., Mazonson, P., Fisk, R., et al. (1996). The CaPSURE database: a methodology for clinical practice and research in prostate cancer. CaPSURE Research Panel. Cancer of the Prostate Strategic Urologic Research Endeavor. *Urology, 48*, 773-777.

McAlindon, T., Formica, M., Kabbara, K., LaValley, M., & Lehmer, M. (2003). Conducting clinical trials over the internet: feasibility study. *BMJ, 327*, 484-487.

McHorney, C. A., Kosinski, M., & Ware, J. E., Jr. (1994). Comparisons of the costs and quality of norms for the SF-36 health survey collected by mail versus telephone interview: results from a national survey. *Medical Care, 32*, 551-567.

Penson, D. F., & Chan, J. M. (2007). Prostate Cancer. In M. S. Litwin & C. S. Saigal (Eds.), *Urologic Diseases in America* (pp. 73-120). Washington, D.C.: U.S. Government Printing Office. U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases.

Ries, L., Melbert, D., Krapcho, M., Stinchcomb, D., Howladern, N., Horner, M., et al. (2008). SEER cancer statistics review, 1975-2005. Retrieved March 24, 2009, from http://seer.cancer.gov/csr/1975_2005 .

Rodriguez, H. P., von Glahn, T., Rogers, W. H., Chang, H., Fanjiang, G., & Safran, D. G. (2006). Evaluating patients' experiences with individual physicians: a randomized trial of mail, internet, and interactive voice response telephone administration of surveys. *Medical Care, 44*, 167-174.

Siegel, R., Ward, E., Brawley, O., & Jemal, A. (2011). Cancer statistics, 2011: the impact of eliminating socioeconomic and racial disparities on premature cancer deaths. *CA: A Cancer Journal for Clinicians, 61*, 212-236.

Torrance, G. W., Siegel, J. E., & Luce, B. R. (1996). Framing and designing the cost-effectiveness analysis. In M. R. Gold, J. E. Siegel, L. B. Russell & W. C. Weinstein (Eds.), *Cost-effectiveness in health and medicine* (pp. 54-80). New York: Oxford University Press.

Velikova, G., Wright, E. P., Smith, A. B., Cull, A., Gould, A., Forman, D., et al. (1999). Automated collection of quality-of-life data: a comparison of paper and computer touch-screen questionnaires. *Journal of Clinical Oncology, 17*, 998-1007.

Wilt, T. J., MacDonald, R., Rutks, I., Shamliyan, T. A., Taylor, B. C., & Kane, R. L. (2008). Systematic review: comparative effectiveness and harms of treatments for clinically localized prostate cancer. *Annals of Internal Medicine, 148*, 435-448.

Chapter 2

Measurement Equivalence Using a Mixed-Mode Approach To Administer Patient-reported Outcome Surveys

Jeanette M. Broering[1†], Alan Paciorek[1], Peter R. Carroll[1], Leslie S. Wilson[2], Mark S. Litwin[3], and Christine Miaskowski[4]

[1] Jeanette M. Broering, RN, PhD, MPH
†Corresponding author
UCSF Helen Diller Family Comprehensive Cancer Center
Department of Urology, School of Medicine
University of California San Francisco (UCSF)
San Francisco, CA 94143-1355
Tel (415) 514-0203
Fax (415) 514-0210
jbroering@urology.ucsf.edu

[1]Alan Paciorek, BS
Department of Urology, School of Medicine
University of California San Francisco (UCSF)
apaciorek@urology.ucsf.edu

[1]Peter R. Carroll, MD, MPH
Department of Urology, School of Medicine
University of California San Francisco (UCSF)
pcarroll@urology.ucsf.edu

[2] Leslie S. Wilson, PhD
Departments of Clinical Pharmacy and Medicine, Schools of Pharmacy and Medicine
University of California San Francisco (UCSF)
wilsonl@pharmacy.ucsf.edu

[3] Mark S. Litwin, MD, MPH
Department of Urology, David Geffen School of Medicine
Department of Health Services, Fielding School of Public Health
University of California Los Angeles (UCLA)
mlitwin@mednet.ucla.edu

[4] Christine Miaskowski, RN, PhD, FAAN
Department of Physiological Nursing, School of Nursing
University of California San Francisco (UCSF)
Chris.miaskowski@nursing.ucsf.edu

Abstract

Purpose: To determine whether questionnaires that measure health-related quality of life (i.e., Medical Outcomes Study Short Form-36 and the UCLA Prostate Cancer Index), when self-administered by paper-mode and web–mode, exhibited measurement equivalence.

Methods: A sample of 245 men with prostate cancer was recruited from CaPSURE, a longitudinal health outcomes study. Participants were randomly assigned to two groups (paper-mode-then-web-mode or web-mode-then-paper-mode) with a 3–day washout period between modes. Cognitive debriefing was performed to evaluate participant's mode preference.

Results: Eighty-five percent of participants completed both modes. The majority were White (97%), college educated (66%), reported an annual income > $75,000 (46%), and a median age of 69 years. Intraclass correlation coefficients for the multi-level items on both instruments were high (0.66–0.97). Exact percent agreement for yes/no items was very high ($\geq$ 0.89). Participants rated both methods of administration favorably. However, the web-mode was rated as somewhat more convenient and faster to complete. Seventy percent of participants preferred the web-mode, 21% had no preference and 9% preferred the paper-mode.

Conclusions: Results suggest that paper-mode and web-mode questionnaires were equivalent. Both modes were well accepted among this highly select group of older men with prostate cancer.

Purpose

Concerns have been raised about psychometric measurement equivalence (ME) between health surveys that are developed and tested in one mode (e.g., paper-and-pencil administration that is migrated to web-mode) [1-4]. Two primary ME concerns, that may introduce measurement bias, are differences in how the items and responses are presented to a participant and potential difficulties an individual has in interacting with a computer (i.e., computer anxiety) [5]. Best practice guidelines were published in *Value in Health* on methods to assess ME between paper-mode health surveys that are redesigned for electronic administration using a personal computer [1, 6-7]. These best practices include recommendations for both study design and statistical methods to evaluate ME [1]. A randomized parallel group design or a randomized crossover design with adequate time between first and second administration to minimize the effects of memory or carry over testing effects should be used. Statistical methods for evaluating ME include the use of the intraclass correlation coefficient (ICC), a weighted kappa, a comparison of mean scores, or differential item functioning [1]. The ICC can only be used with a randomized crossover design. The ICC has the advantage of being able to assess both the covariance and degree of agreement between score distributions [1].

The advantages of adopting web-based methods include better data quality, less missing data, easier implementation of skip patterns for data quality assurance monitoring, the immediate creation of electronic data sets, the ability to see real-time data reporting, and the avoidance of secondary data entry errors [5]. Potential disadvantages include selection bias since the method requires computer literacy, patient burden, the need for training, expense, validation requirements for instrument comparability, and

technology or communication requirements specifically in developing countries [5, 8]. In a meta-analytic review that assessed ME for patient-reported outcomes (PRO) between paper and computerized modes, the average age of participants in the meta-analysis was 48 ±13.9 [5]. For the eight studies in the meta-analysis that assessed ME among person's over 65 years of age, a slight decrease in correlation coefficients between modes was noted as age increased (r = 0.02 decrease for each year increase). This finding suggests that the correlation concordance between paper and web-mode of administration may vary more for older persons [5].

Visual design effects, or changes in the appearance of questionnaire items and responses, can alter the respondent's answers thereby introducing error [9-12]. Difficulties interacting with the computer may impact ME (i.e., computer anxiety may negatively influence mood state) [13-14]. If these factors affect ME, then threats to internal validity occur that can affect one's ability to draw statistical inference from the health-related quality of life (HRQOL) questionnaire, a type of PRO, on external validity.

Prostate cancer is the most common solid tumor diagnosed in men. Over 90% of newly diagnosed cases will be local or regionally confined within the pelvis with an estimated 5–year survival rate of nearly 100% [15-16]. Given the high number of prostate cancer survivors, the potential for treatment associated morbidity, and the extended duration of cancer survivorship, the measurement of HRQOL is an integral part of any evaluation of the potential deleterious effects of prostate cancer treatment [17-18]. However, no studies have evaluated ME of HRQOL instruments used in older men with prostate cancer.

The Cancer of the Prostate Strategic Urologic Research Endeavor (CaPSURE) is a multi-institutional longitudinal registry of men with prostate cancer [19]. Since its inception, CaPSURE participants complete a self-administered questionnaire (SAQ) using a single paper-mode approach to measure HRQOL outcomes. The introduction of computer-based methods to administer HRQOL questionnaires promises to reduce the burden on study participants as well as the CaPSURE data coordination center. Increased access to high-speed broad band Internet has made offering of mixed-mode administration of research questionnaires not only a possibility but a necessity for many survey situations [3-4].

However, contemporary surveys of Internet usage indicate that persons over 65 years of age, those with lower incomes, and those with less education were less likely to have access to the Internet [20]. Less well known is how older persons perform when completing a SAQ in paper-mode versus web-mode. Only four studies were found that evaluated the use of mixed-mode approaches (paper-mode or web-mode) in oncology practices within academic medical centers to monitor adverse events associated with active treatment [21-24]. None of these studies evaluated the ability of individuals to complete a web-based SAQ in the home where hands-on technical assistance is absent to support web navigation. Therefore, we undertook this research to assess ME when mixed-mode approaches are used within longitudinal research to insure that no systematic measurement error is inadvertently introduced into the HRQOL outcomes data.

Given the paucity of research on ME in HRQOL studies of older adults with cancer, this instrument validation study examined ME between paper-mode and web-

mode administration of two HRQOL instruments (i.e., the Medical Outcomes Study Short Form-36 (SF-36) version 1 [25-28], the UCLA-Prostate Cancer Index (PCI) [29-30] in a sample of men with prostate cancer who participated in the CaPSURE registry. Evidence of differences in HRQOL scores by mode, order of administration, and the interaction of mode and order of administration were evaluated. Examination of missing data by mode was performed to assess data quality. Finally participant's preference for mode was assessed.

Methods

Participants

For this administration period, there were 4,948 CaPSURE active participants. Five participants were excluded who received a Spanish language HRQOL questionnaire. A letter of invitation was sent to all active English language study participants (N = 4,943) which was inserted into their mailed semi-annual CaPSURE questionnaire packet of which 3,888 (79%) returned their survey and 245 (6%) respondents volunteered to participate in the ME study (Figure 1). These 245 volunteers were randomized into group A (n = 123) or group B (n = 122) for order of administration (Figure 2). Twenty persons who volunteered to participate did not respond to any follow-up communications. An additional sixteen persons were excluded from this analysis because of incomplete data (nine did online only and seven completed paper only). A final sample of 209 (85%) participants completed all four surveys and had evaluable data (Figures 1 & 2). The final sample contained 105 participants from group A and 104 participants from group B (Figure 2). Participants were compensated $25.00 for their participation. The Committee on Human Research at UCSF approved this ME study.

Data Collection

Participants were randomized into one of two groups using a simple randomization algorithm. The rationale for randomizing the order of administration between the two modes was to eliminate any order effects which might arise (i.e., a participant may prefer the first mode, regardless of which mode is presented first). Participants randomized to group A completed the paper-mode first whereas participants in group B completed the web-mode using a secure Internet website. For this analysis, each participant completed four instruments (i.e., the SF-36 and the UCLA-PCI by each mode; a 5–item health stability questionnaire; and a 15–item survey designed to measure preference for mode of administration). All paper-mode questionnaires were sent via express mail with a return air bill. Participants completed the 5–item health stability questionnaire prior to the administration of the second set of instruments to assess the occurrence of any significant illness that could have affected responses on the second mode. After completion of both modes, participants were asked to complete a paper survey that measured their mode of administration preference (Figure 2).

Instruments

The SF-36 version 1 consists of 36 items that evaluate overall physical and mental health status. Items are scored into eight health status scales that measure physical functioning (PF), role limitations due to physical health (RP), bodily pain (BP), social functioning (SF), general mental health (MH), role limitations due to emotional problems (RE), vitality, energy or fatigue (VT), general health perceptions (GH) and a single item that compares transition in health status over the past year (HT) [31]. Additionally, the eight scales, excluding the HT item, can be aggregated to achieve summary measures of

physical and mental health status. The PF, RP, BP, and GH scales correlate most highly with the physical component summary (PCS) score. The VT, SF, RE, and MH scales correlate most highly with the mental component summary (MCS) score [25].

The UCLA-PCI consists of 20 items with six dimensions that assess urinary, bowel, and sexual function and bother [29-30]. Both the SF-36 and the UCLA-PCI instruments were scored from zero to 100, with a higher score indicating better HRQOL. Both the SF-36 and the UCLA-PCI use a standard 4–week recall period.

Statistical analysis

A power calculation at a beta of 0.80 was performed to estimate the minimum sample size needed to be able to detect an effect between modes. The power calculation was performed using historical data from the role physical (RP) domain of the SF-36 which had the largest standard deviation observed in CaPSURE (± 29.4). This study was powered conservatively to detect a 5 point difference in HRQOL scores between paper-mode and web-mode versions of administration. For the RP scale, a minimum of 133 participants were needed to detect a 5 point difference in scores.

To assess differences by mode, two statistical approaches were used, namely the ICC and the paired *t*-test. An ICC which assesses the degree of equivalence between the score distributions, were used to evaluate the strength of the association between the two modes [5, 32]. The ICC statistic, with 95% confidence intervals (CI), was estimated for each of the 56 items contained within the SF-36 and UCLA-PCI. A test-retest ICC correlation of ≥0.75 is considered to demonstrates excellent agreement and was used as the standard for comparison [33-34]. For the paired *t*-test, change scores were calculated by subtracting scores on the paper-mode from scores on the web-mode versions of the

questionnaires for each participant [35]. To examine the effects of mode, order of administration, and any two-way interactions between mode and order of administration, a linear mixed methods approach was used [36]. Post-hoc analyses comparing differences between mean scale scores were performed. No formal adjustment was made for multiple comparisons and a *p*-value of < .05 was considered statistically significant. To examine differences in data quality, missing data were examined using a two tailed *t*-test. Participant mode preference was assessed using the paired *t*-test for differences. All analyses were generated using SAS® software Version 9.2 of the SAS System for Microsoft Windows (SAS Institute, Cary, NC, USA).

Results

Table 1 presents the participants' demographic and clinical characteristics. Participant median age was 69 years (range 50 to 94 years**).** Participants were predominately white, well educated, and indicated a relatively high annual income. The majority had opted for definitive treatment with radical prostatectomy and had low or intermediate-risk disease [37]. The median duration of participation in CaPSURE was 54 months (range 5 to 118 months). All participants had considerable prior experience completing the CaPSURE questionnaire with the median number of ten prior semi-annual questionnaires completed (range 1 to 27 prior surveys completed). Randomization to group assignment allowed for evaluation of the order of administration effects. No differences were observed between mode groups for demographic or clinical characteristics, or in prior experience with HRQOL questionnaire completion (Table 1).

Participants observed a mean washout period of 4.4 days (SD ± 4.6 days) and a median washout period of three days (range 0 to 35 days) between the first and second

mode of administration. The CaPSURE questionnaire contained a proxy administration question. Five men reported receiving assistance from a proxy. Three men reported getting assistance with completing the paper-mode and three men reported receiving assistance with the web-mode, of which one person reported assistance with both modes. These participants were retained in the analyses.

Twelve participants reported minor illnesses during the washout period on the health stability questionnaire (e.g., recovering from an upper respiratory infection or flu). No participants reported any major injuries or serious illnesses during the washout period.

ME by Mode of Administration

With several exceptions for the SF-36, ICC coefficients indicated an excellent level of agreement (*ICC* = 0.70–0.95), indicative of high reproducibility (Table 2). Three items within the RP and RE domains had lower correlation coefficients, specifically, reduction in time worked (*ICC* = 0.69, 95%CI 0.61–0.75), limitations in activity (*ICC* = 0.68, 95%CI 0.60–0.74), and inability to do activities as carefully as usual (*ICC* = 0.54, 95%CI 0.43–0.63). Items in these two domains have a dichotomous response option (Yes/No) which reduces variability in the response. The proportion of responses with exact matches for each item response was very high at 74% to 98% (Table 2).

For the individual items on the UCLA-PCI, the ICC coefficients were very high (*ICC* = 0.66 to 0.97), indicative of high reproducibility. A single item, rectal urgency, had a lower correlation value between modes (*ICC* = 0.66, 95%CI 0.58–0.73). The proportion of agreement for exact match for each response item was very high at 70% to 96% (Table 2).

Comparison of mean SF-36 scale scores demonstrated equivalence between the paper and web-modes with the exception of the MCS score and the RE domain (*p*-values 0.02 and 0.005, respectively) (Figure 3). However, the effect size (ES) for these two scales were small (0.2) suggesting a difference which was statistically significant but not clinically meaningful [38]. Comparison of mean UCLA-PCI scale scores between modes for the six domains of urinary, bowel, and sexual function and bother revealed no differences (Figure 4).

ME and Ordering Effects

For the SF-36, when compared to participants who did the web-mode first, participants who completed the paper-mode first reported lower scores for both the PF and the PCS scores ($p = 0.02$ and 0.05, respectively) (Table 3). For the UCLA-PCI, a significant ordering effect was found for the sexual bother domain. Participants who completed the paper-mode first reported less bother at both time points than those who did the web-mode first ($p = 0.03$). For participants who completed the web-mode first, the mean difference between modes for sexual bother was almost five points. However, the effect size was small (ES = 0.10) (Table 4).

Interaction of Mode and Ordering

For the SF-36, significant two-way interactions for mode by order of administration were found for the MCS, RE, SF, VT, and MH domains (Table 3). Post-hoc comparisons for order of administration found that the MCS score was lower when web-mode of administration was done first (54.9 vs. 56.3, $p < 0.001$). A similar pattern was observed for the various subscale scores of the MCS. In terms of the RE scores, for those who completed the web-mode first, RE score was lower than the subsequent paper-

mode (88.8 vs. 93.9, $p = 0.0004$). For the UCLA-PCI scales, no significant interactions were found between mode and order of administration (Table 4).

Data Quality and Completeness

Missing data were examined by mode at the questionnaire, item, and computable scale levels. For the 56 items examined among the 209 web-mode questionnaires, a higher proportion of these questionnaires had missing data. Specifically, 53 of the 209 (25%) web-mode questionnaires had 136 missing items. Whereas for the paper-mode, 24 (11%) of the questionnaires had 114 missing items. For all 56 items examined, the mean number of missing responses was 0.65 (SD 3.9) for the web-mode compared to 0.55 (SD 2.3) for the paper-mode ($p = 0.75$). For the SF-36, 18 web-mode questionnaires had missing data compared to 13 for the paper-mode. The mean number of missing SF-36 items was 0.28 (SD 2.5) for the web-mode versus 0.36 (SD 2.1) for the paper-mode ($p = 0.72$).

For the UCLA-PCI, 38 web-mode questionnaires had missing data compared to 12 paper-mode questionnaires. The mean number of missing items was 0.37 (SD 1.5) for the web-mode versus 0.19 (SD 1.0) for the paper-mode ($p = 0.15$). Missing data for the web-mode was assumed to be missing at random. However, manual review of the paper-mode surveys revealed two systematic reasons for missing data. The paper-mode survey was printed using an 11 by 17 inch booklet format. For the 13 paper-mode questionnaires with missing SF-36 data, it appeared that a page stick within the booklet with subsequent omission of the first three items which contained the physical function, role physical, and role emotional scales. For the UCLA-PCI, four (33%) of the 12 paper-

mode questionnaires had "not applicable" written next to the selected items in the urinary and sexual function domains.

A decision rule used for computation of a scale score required that a participant answer ≥ 50% of the items contained within the scale in order to generate a computable score. Missing data were more problematic for the calculation of the PCS and MCS summary scores of the SF-36. For the paper-mode, computable scale scores were missing for eleven (5%) of the participants compared to six participants (2.9%) who used the web-mode. For the three UCLA-PCI function scales, inability to calculate a scale score was equivalent between modes (i.e., 1 participant with missing scale scores for the three function domains). For the single-item bother scales, missing data for the web-mode were observed for three participants for the urinary bother and two participants for bowel bother compared to one participant for the paper-mode on these same two scales. For the sexual bother scale, three participants had missing computable scale scores for the paper-mode compared to one for the web-mode.

Participant Mode Preference

Post completion of both modes, participants completed a fifteen item mode preference questionnaire as part of their cognitive debriefing. Respondents rated both methods favorably. No differences were found between the two modes for ease of reading or ease of answer completion. They evaluated both modes as easy to navigate and neither mode was confusing or stressful to complete. Participants rated the web-mode as somewhat more convenient and faster to complete (p = 0.005 and 0.02, respectively) (Table 5). While participants rated both modes favorably, 70% preferred the web-mode, 21% had no preference, and 9% preferred the paper-mode.

Fifty (24%) participants provided qualitative comments about their experience. Over half of these comments (n=26) were negative in terms of their experience with use of the web-mode application. Negative comments pertained to three areas of application usability. Specifically, problems were related to connectivity issues; the speed of page loading and page saves as they advanced forward online; and problems with scrolling up the page to see choice options. Positive comments addressed the ease of use, convenience, and the efficiency of the mode (e.g., "let's stop cutting down trees"). The application was designed to save data from each page as the participant advanced through the application and it allowed the participant to move both backwards and forwards similar to the paper-mode. One participant found this forward and backwards navigation option to be a desirable feature (e.g., "The Internet questionnaire allowed me to back up and change/clarify my answer").

Discussion

Participants in this ME study were a highly select group of older men who were motivated to self-administer both a paper-mode and a web-mode survey outside of a clinical environment where technical assistance was not readily available. The correlation data provide strong support for ME of the SF-36 and the UCLA-PCI questionnaires administered by web and paper-modes. For the SF-36, significant interactions were found for mode by order of questionnaire administration for the physical and mental health domains. While these interactions were statistically significant, the differences were not clinically meaningful [39-42].

Two prior studies evaluated the validity of and patient preference for electronic administration of the SF-36 using a personal computer among older participants who

completed the survey within a supervised environment [43-44]. In a randomized cross-over study, that evaluated mean scale scores for the eight domains of the SF-36 completed electronically, no mode or ordering effects were found [42]. However, this sample was limited to eighteen volunteers who were members of a community-based senior citizens club. In a pilot study that evaluated SF-36 scale scores reported by 55 patients recruited consecutively in a cardiology clinic using a touch screen device, ICC coefficients were lower (0.54 to 0.75) than in the current study [43]. Reasons for the differences in results may be related to participant characteristics including: younger age (median age 51–years); more ethnically diverse (44% minority participants); lower literacy levels (i.e., 7^{th}-grade reading level); and poorer health (mean SF-36 scale scores were 20–48 points lower than those of the CaPSURE sample). Both studies conducted their test-retest within the same visit (test-retest time within 5 to 55 minutes) which raises concerns about carry over effects.

No studies have evaluated ME for the UCLA-PCI. A significant ordering effect was observed for sexual bother (i.e., those who completed the web-mode first followed by the paper-mode reported more sexual bother using the web-mode). This finding is consistent with previous studies that found that individuals tend to report sensitive personal information (e.g., sexual function) more accurately using a web-based approach [21, 23].

While the proportion of participants who completed the surveys and the number of items with missing data were higher for the web-mode participants, no statistically significant differences between modes were found when comparisons were made at the item level. Web-mode participants had more missing data on three items specifically,

urinary leakage interfering with sexual activity and difficulties with erection and orgasm in the past 4 weeks.

This study has several strengths. The study incorporated self-selected participants who were highly motivated older men with prostate cancer who were able to successfully use computer based technology to report their HRQOL using a web-based questionnaire. While no formal assessment of reading or computer literacy was performed on this sample, reading literacy was considered to be high given the high proportion with some college education or greater. The researchers invoked visual design principles to enhance comparability in the visual appearance of the formatting of items and screen layout for a personal computer. The study employed a rigorous cross-over design methodology to examine mode-by-order findings, had adequate sample size for statistical power, and used the ICC and linear mixed methods approach to assess ME.

This study has several limitations. Because this convenience sample of participants who had high physical and mental health status, findings may not generalize to patients with poorer health status. Only 6% of eligible active CaPSURE participants volunteered for this study. The participants were predominately white, college educated, and more affluent men. This study did not assess participants' experience with using a personal computer. All participants had extensive experience with self-administration of the CaPSURE semi-annual questionnaire. These characteristics may influence the sample's preference for the web-mode.

Future research needs to replicate web-mode offering with a sample of participants who have more diverse characteristics in terms of education, ethnicity, and

income to establish ME. In addition, future studies could evaluate ME using other electronic devices such a smart phones or tablets.

In conclusion, findings from this study provide evidence for ME between the paper and web-modes in a select sample of men with prostate cancer. This evidence supports the introduction of a mixed-mode approach to administer HRQOL questionnaires within the CaPSURE registry without concerns about the introduction of measurement bias that could affect the internal and external validity.

Figure 1. CaPSURE Measurement Equivalence Study Participants

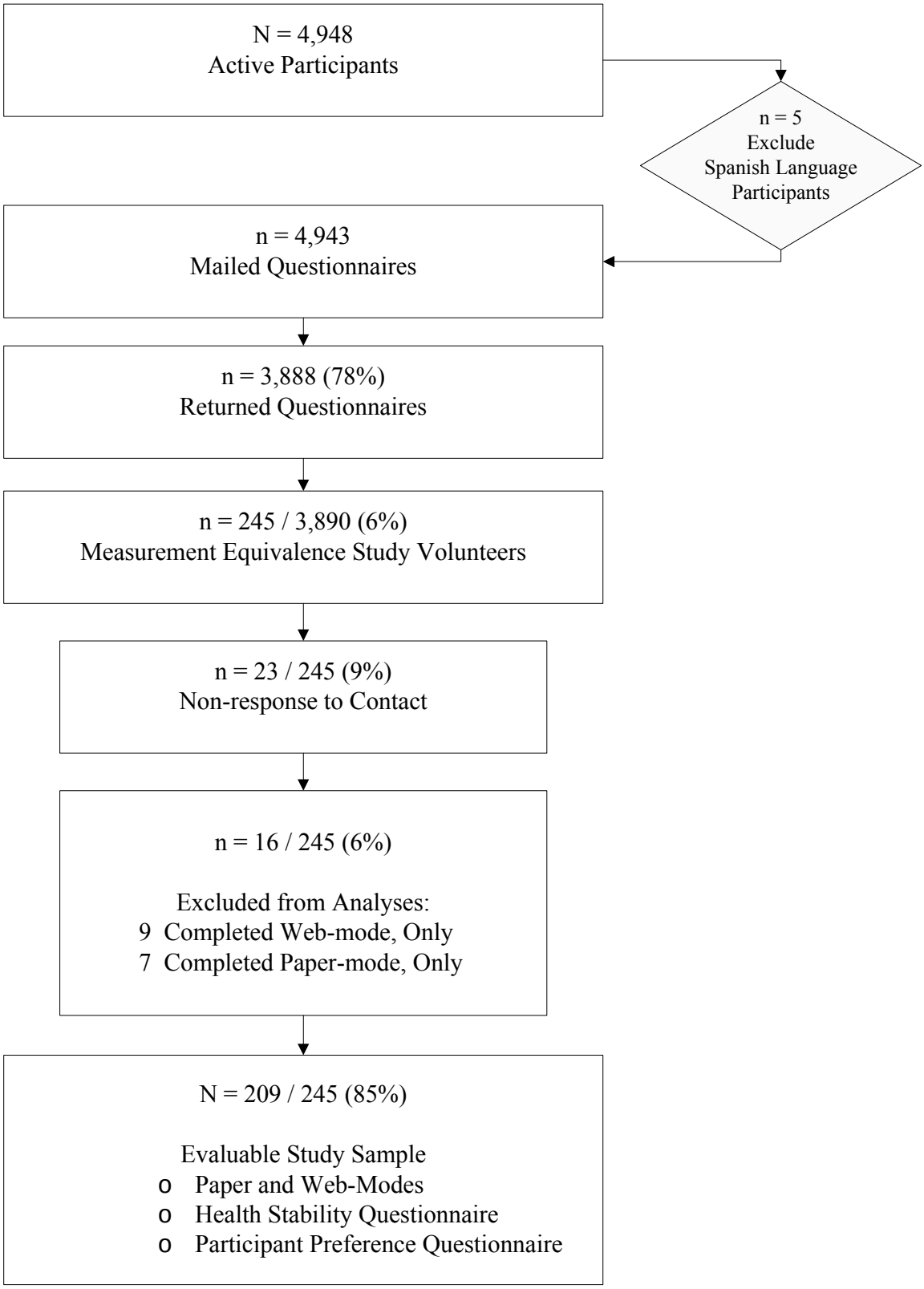


Figure 2. Schema for Randomized Cross-Over Study Design

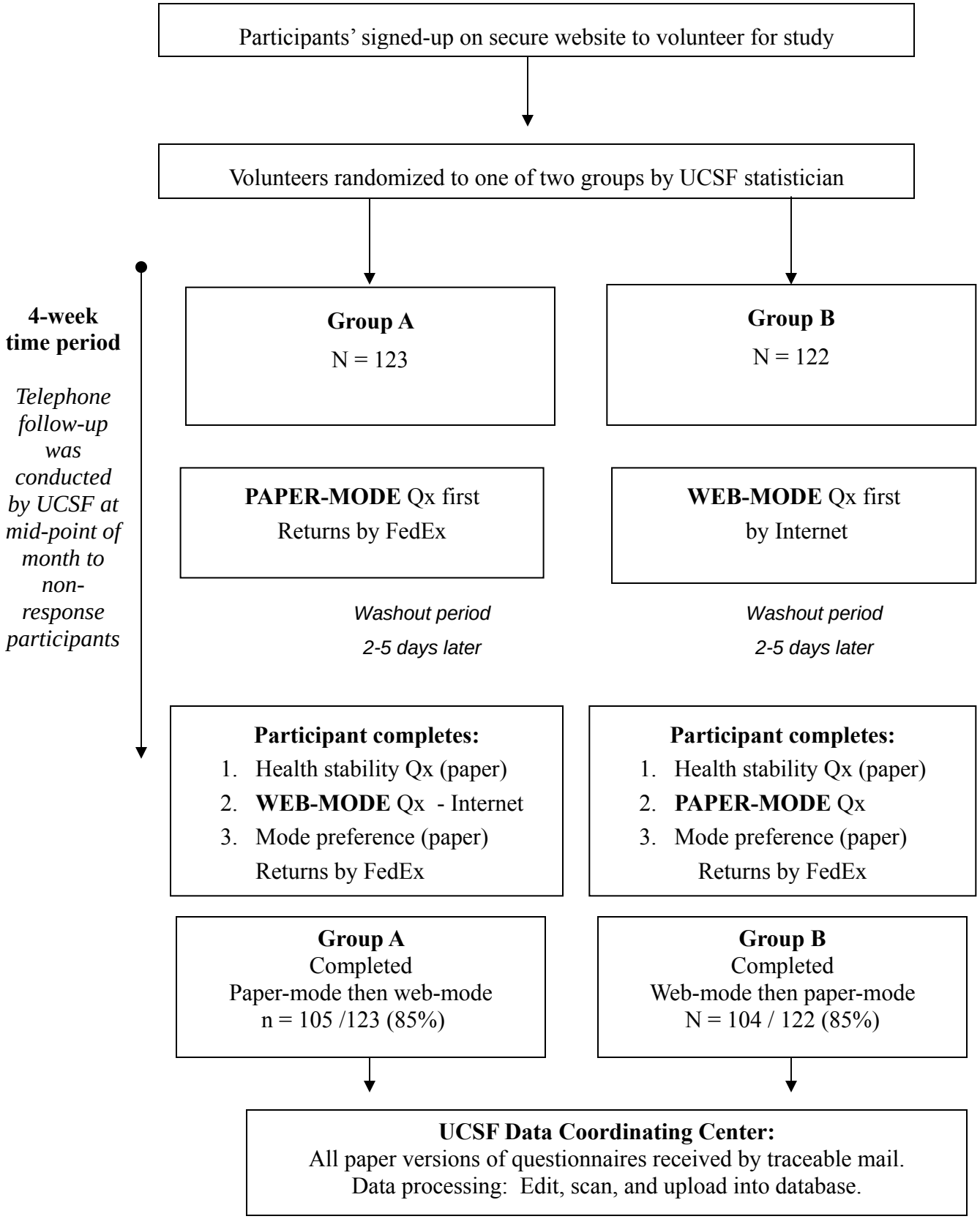


Table 1. Demographic and Clinical Characteristics

Characteristic	Full Sample N = 209	Paper-mode First n = 105	Web-mode First n = 104	Pearson Chi-square *p*-value
Age at Time of Study (*years*)	N (%)	N (%)	N (%)	
<60	32 (15)	15 (14)	17 (16)	0.92
60-69	79 (38)	40 (38)	39 (38)	
70+	98 (47)	50 (48)	48 (46)	
Race/Ethnicity				
Non-White	6 (3)	4 (4)	2 (2)	0.41
White	203 (97)	101 (96)	102 (98)	
Education				
No College	27 (13)	16 (16)	11 (11)	0.11
Some College	43 (21)	26 (25)	17 (17)	
College Graduate	135 (66)	60 (59)	75 (73)	
Unknown	4	3	1	
Annual Household Income				
< $50,000	58 (31)	35 (36)	23 (25)	0.18
$50-75,000	45 (24)	24 (24)	21 (23)	
> $75,000	87 (46)	39 (40)	48 (52)	
Unknown	19	7	12	
Marital Status				
With a Partner	193 (96)	95 (96)	98 (95)	0.78
Single	9 (4)	4 (4)	5 (5)	
Unknown	7	6	1	
Risk Group†				
Low	83 (42)	43 (43)	40 (41)	0.28
Intermediate	82 (42)	44 (44)	38 (39)	
High	32 (16)	12 (12)	20 (20)	
Unknown	12	6	6	
Type of Initial PCa Treatment				
Radical Prostatectomy	151 (72)	76 (74)	75 (73)	0.71
Radiation	35 (17)	18 (17)	17 (17)	
Hormone Therapy	7 (3)	2 (2)	5 (5)	
Cryotherapy	7 (3)	3 (3)	4 (4)	
Active Surveillance	6 (3)	4 (4)	2 (2)	
Unknown	3	2	1	
Number of Prior Questionnaires				
1-2	7 (3)	2 (2)	5 (5)	0.33
3-4	13 (6)	5 (5)	8 (8)	
5-27	189 (90)	98 (93)	91 (88)	

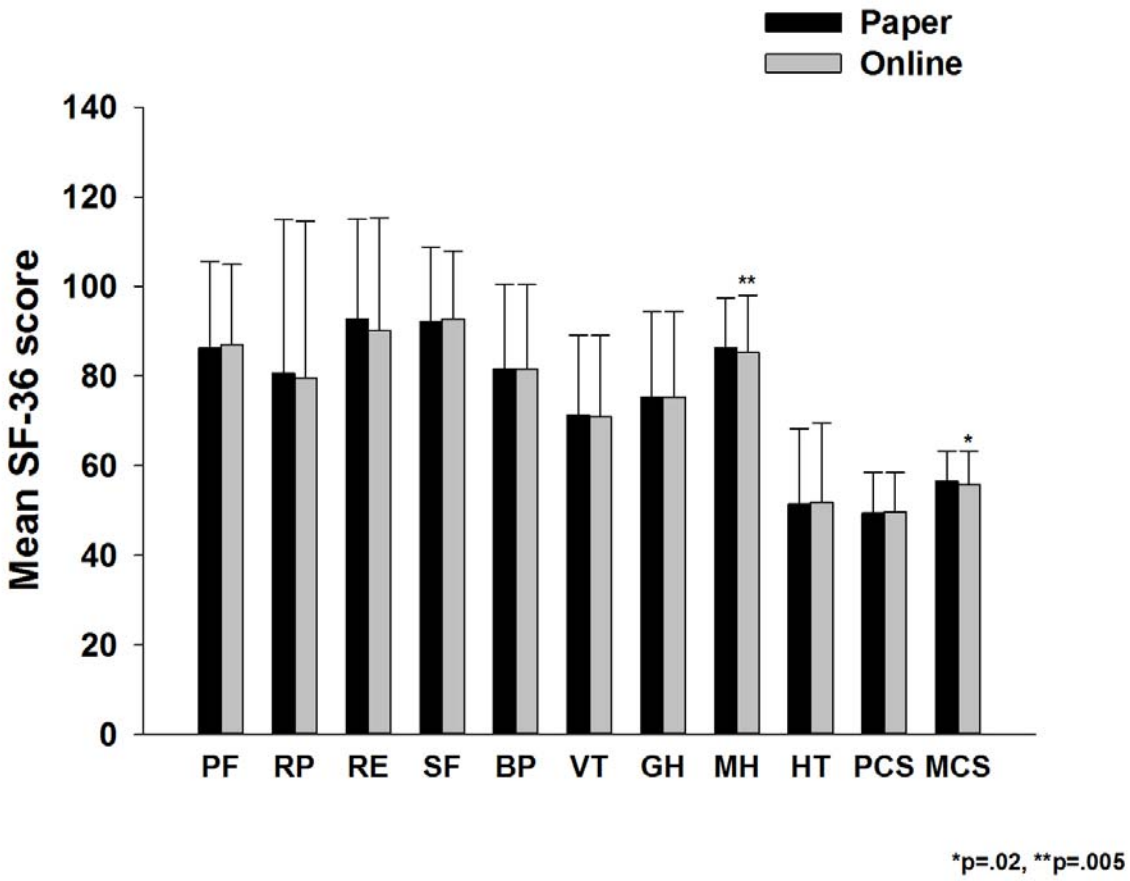
† Using UCSF modified D'Amico [37] criteria risk, defined as: **Low risk** – PSA ≤10 ng/ml; Gleason Total 2-6 with no 4/5 pattern; 2002 Tumor staging cT1, cT2a; **Intermediate risk** - PSA 10.1-20 ng/ml; Gleason Total 7 OR secondary 4/5 pattern; 2002 Tumor staging cT2b, cT2c; **High risk** – PSA >20 ng/ml; Gleason Total 8-10 OR primary 4/5 pattern; 2002 Tumor staging cT3a+.

Table 2. Percent Agreement and Intraclass Correlation Coefficient (*ICC*) Between Paper-mode and Web-mode

Scale Description	# Items	% Agree Range †	ICC Range
SF-36			
Physical Functioning (PF)	10	87 - 98	.78 - .93
Role Physical (RP) ‡	4	88 - 92	.68 - .79
Role Emotional (RE) ‡	3	94 - 95	.54 - .78
Vitality (VT)	4	74 – 79	.78 - .88
Mental Health (MH)	5	77 – 89	.70 - .83
Bodily Pain (BP)	2	77 – 85	.83 - .86
Social Functioning (SF)	2	88 – 92	.78 - .89
General Health (GH)	5	81 - 88	.79 - .93
Health Transition (HT)	1	90	.87
UCLA-PCI			
Urinary Function (UF)	5	78 - 92	.81 - .95
Urinary Bother (UB)	1	84	.84
Bowel Function (BF)	4	87 - 90	.66 - .89
Bowel Bother (BB)	1	86	.88
Sexual Function (SF)	8	83 – 96	.94 - .97
Sexual Bother (SB)	1	70	.84
Total items compared by *ICC*	56		

† Percent agree is the proportion of items with an exact match on the response value for each item
‡Contains dichotomous Yes/No response items

Figure 3. Short Form 36 Mean Scale Scores by Mode of Administration



PF = Physical Function, RP = Role Physical, RE = Role Emotional, SF = Social Function, BP = Bodily Pain, VT = Vitality, GH = General Health, MH = Mental Health, HT = Health Transition, PCS is Physical Component Score, MCS = Mental Component Score.

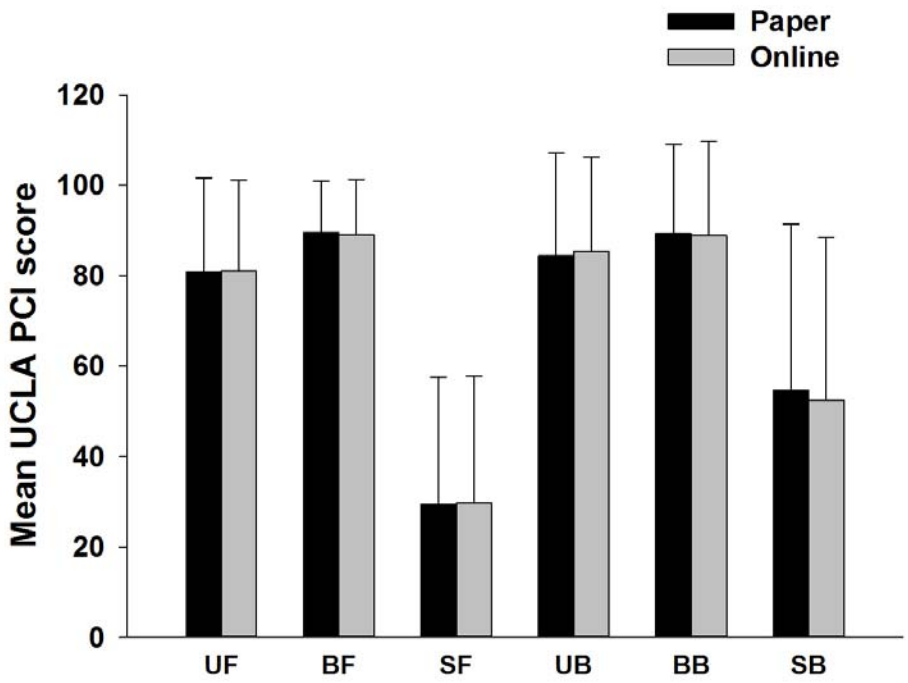
Domains scored 0-100 with higher equals optimal functioning.

P-value for difference used paired Student's t-test significant at ≤ 0.05.

* Mental Component Score (MCS) significant at $p = 0.02$.

** Role Emotional (RE) significant at $p = 0.005$

Figure 4. UCLA- Prostate Cancer Index Mean Scale Scores by Mode of Administration



Score 0-100 with higher equals optimal functioning

P-value for difference used paired Student's t-test significant at ≤ 0.05.

Table 3. SF-36 Measurement Equivalence by Mode, Ordering, and Interaction Effects

Mode Effect		**Group Order**		***Pr > F****
SF-36 Domain	Paper-mode First[†] N = 105 Mean (SD)[‡]		Web-mode First [††] N = 104 Mean (SD)[‡]	
Physical Composite Score				
Paper	48.3 (12.7)	Online	50.8 (12.5)	
Online	48.5 (12.5)	Paper	50.8 (12.7)	
			Mode effect	0.65
			Ordering effect	**0.05**
			Interaction (mode * order)	0.79
Mental Composite Score				
Paper	56.2 (10.0)	Online	54.9 (10.4)	
Online	56.5 (10.4)	Paper	56.3 (10.0)	
			Mode effect	**0.01**
			Ordering effect	0.41
			Interaction (mode * order)	**0.0002**
Physical Function				
Paper	83.2 (26.8)	Online	89.6 (26.3)	
Online	83.5 (26.2)	Paper	89.5 (26.9)	
			Mode effect	0.64
			Ordering effect	**0.02**
			Interaction (mode * order)	0.78
Role Physical				
Paper	76.9 (47.7)	Online	81.3 (49.6)	
Online	77.5 (49.5)	Paper	84.9 (48.1)	
			Mode effect	0.27
			Ordering effect	0.20
			Interaction (mode * order)	0.14
Role Emotional				
Paper	91.7 (31.3)	Online	88.8 (35.9)	
Online	91.1 (35.8)	Paper	93.9 (31.5)	
			Mode effect	**0.005**
			Ordering effect	0.98
			Interaction (mode * order)	**0.03**
Social Function				
Paper	90.5 (23.3)	Online	93.0 (22.0)	
Online	92.0 (21.9)	Paper	93.9 (23.5)	
			Mode effect	0.53
			Ordering effect	0.30
			Interaction (mode * order)	**0.02**
Bodily Pain				
Paper	79.6 (26.8)	Online	81.9 (27.0)	
Online	80.8 (27.0)	Paper	83.3 (26.9)	
			Mode effect	0.88

Mode Effect		**Group Order**		***Pr > F****
SF-36 Domain	Paper-mode First† N = 105 Mean (SD)‡		Web-mode First †† N = 104 Mean (SD)‡	
			Ordering effect	0.35
			Interaction (mode * order)	0.06
Vitality				
Paper	69.0 (25.7)	Online	70.9 (26.3)	
Online	70.2 (26.1)	Paper	72.9 (25.8)	
			Mode effect	0.42
			Ordering effect	0.35
			Interaction (mode * order)	**0.0009**
General Health				
Paper	75.2 (26.9)	Online	76.1 (27.3)	
Online	74.4 (27.3)	Paper	75.3 (26.9)	
			Mode effect	0.90
			Ordering effect	0.74
			Interaction (mode * order)	0.08
Mental Health				
Paper	86.0 (16.7)	Online	83.9 (18.0)	
Online	86.6 (17.9)	Paper	85.7 (16.7)	
			Mode effect	0.11
			Ordering effect	0.38
			Interaction (mode * order)	**0.002**
Health Transition				
Paper	50.9 (23.7)	Online	51.7 (25.0)	
Online	51.7 (24.9)	Paper	51.9 (23.8)	
			Mode effect	0.68
			Ordering effect	.084
			Interaction (mode * order)	.042

† Paper-mode first was Group A, who then completed web-mode.
†† Web-mode first was Group B, who then completed paper-mode.
* F-test statistic derived from random effects model, significant at *Pr* of F ≤0.05
‡Score 0-100 with higher equals optimal functioning

Table 4. UCLA-Prostate Cancer Index Measurement Equivalence by Mode, Ordering, and Interaction Effect

Domain		**Group Order**		***Pr > F****
UCLA-PCI	Paper-mode First† N = 105 Mean (SD)‡		Web-mode First†† N = 104 Mean (SD)‡	
Urinary Function				
Paper	80.5 (29.2)	Online	80.9 (28.6)	
Online	81.4 (28.5)	Paper	81.4 (29.3)	
			Mode effect	0.66
			Ordering effect	0.93
			Interaction (mode * order)	0.06
Bowel Function				
Paper	89.6 (16.1)	Online	88.4 (17.5)	
Online	89.5 (17.4)	Paper	89.5 (16.2)	
			Mode effect	0.12
			Ordering effect	0.70
			Interaction (mode * order)	0.21
Sexual Function				
Paper	30.1 (39.7)	Online	29.1 (39.6)	
Online	29.9 (39.4)	Paper	28.6 (39.9)	
			Mode effect	0.52
			Ordering effect	0.76
			Interaction (mode * order)	0.26
Urinary Bother				
Paper	83.8 (32.1)	Online	84.8 (29.7)	
Online	86.0 (29.6)	Paper	85.1 (32.2)	
			Mode effect	0.26
			Ordering effect	0.98
			Interaction (mode * order)	0.15
Bowel Bother				
Paper	88.3 (27.9)	Online	88.9 (29.6)	
Online	89.0 (29.4)	Paper	90.4 (28.0)	
			Mode effect	0.60
			Ordering effect	0.72
			Interaction (mode * order)	0.12
Sexual Bother				
Paper	58.7 (51.9)	Online	45.9 (50.4)	
Online	59.2 (50.2)	Paper	50.8 (52.1)	
			Mode effect	0.13
			Ordering effect	**0.03**
			Interaction (mode * order)	0.06

† Paper-mode first was Group A, who then completed web-mode.
†† Web-mode first was Group B, who then completed paper-mode.
* F-test statistic derived from random effects model, significant at *Pr* of F ≤0.05
‡Score 0-100 with higher equals optimal functioning

Table 5. Mean Differences in Participant Mode Preference[†]

Dimension	Mean Difference (SD)[‡]	*P*-value
Readability	-0.06 (0.99)	0.37
Difficulty with item completion	-0.01 (1.44)	0.89
Confusion with navigation	0.06 (0.99)	0.40
Convenient to complete	0.50 (2.52)	**0.005**
Stressful to complete	0.02 (1.77)	0.88
Speed of completion	0.45 (2.72)	**0.02**
Comparable – paper-mode vs. web-mode	1.32 (2.81)	**<0.001**
Mode Preference	0.90 (1.02)	**<0.001**

†15–item instrument with Likert scale response items scaled from 1 = easy to 10 = very difficult.

‡Items 1-4, 7-8 and 11-14 were reverse scored so that 10 = best and 1 = worse. The mean, SD, median, and range was calculated for respondents by each item. Then a difference score was calculated by subtracting the paper-mode score from the web-mode score for each of the dimensions contained within the instrument. Difference scores for mode preference ranged from -2 (strong paper-mode), neutral = 0 (no preference) and +2 (strong web-mode).

References

1. Coons, S.J., et al., *Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report.* Value in Health, 2009. **12**(4): p. 419-29.
2. Bowling, A., *Mode of questionnaire administration can have serious effects on data quality.* Journal of Public Health 2005. **27**(3): p. 281-91.
3. deLeeuw, E., J. Hox, and D. Dillman, *Mixed-mode surveys: When and why*, in *International handbook of survey methodology*, E. deLeeuw, J. Hox, and D. Dillman, Editors. 2008, Psychology Press Taylor and Francis Group: New York. p. 299-316.
4. deLeeuw, E.D., *To mix or not to mix data collection modes in surveys.* Journal of Official Statistics, 2005. **21**(2): p. 233-255.
5. Gwaltney, C.J., A.L. Shields, and S. Shiffman, *Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: a meta-analytic review.* Value in Health, 2008. **11**(2): p. 322-33.
6. Arpinelli, F. and F. Bamfi, *The FDA guidance for industry on PROs: the point of view of a pharmaceutical company.* Health and Quality of Life Outcomes, 2006. **4**: p. 85.
7. FDA, *Guidance for industry patient-reported outcome measures: Use in medical product development to support labeling claims*. 2006, U.S. Department of Health and Human Services, Food and Drug Administration: Rockville, MD. p. i-32.

8. Couper, M.P., et al., *Noncoverage and nonresponse in an Internet survey.* Social Science Research, 2007. **36**: p. 131-148.

9. Dillman, D. and J. Smyth, *Design effects in the transition to web-based surveys.* American Journal of Preventive Medicine, 2007. **32**(5 Suppl): p. S90-6.

310. Dillman, D., J. Smyth, and L. Christian, *Internet, mail and mixed-mode surveys: The tailored design method*. 3rd ed. 2009, Hoboken, NJ: John Wiley & Sons, Inc.

11. Dillman, A., A. Gertseva, and T. Mahon-Haft, *Achieving usability in establishment surveys through the application of visual design principles.* Journal of Official Statistics, 2005. **21**(2): p. 183-214.

12. Kwak, N. and B. Radler, *A comparison between mail and web surveys: Response pattern, respondent profile and data quality.* Journal of Official Statistics, 2002. **18**(2): p. 257-273.

13. Schulenberg, S.E. and B.A. Yutrzenka, *The equivalence of computerized and paper-and-pencil psychological instruments: implications for measures of negative affect.* Behav Res Methods Instrum Comput, 1999. **31**(2): p. 315-21.

14. Tseng, H.M., et al., *Computer anxiety: a comparison of pen-based personal digital assistants, conventional computer and paper assessment of mood and performance.* British Journal of Psychology, 1998. **89 (Pt 4)**: p. 599-610.

15. Siegel, R., et al., *Cancer statistics, 2011: the impact of eliminating socioeconomic and racial disparities on premature cancer deaths.* CA: A Cancer Journal for Clinicians, 2011. **61**(4): p. 212-36.

16. *Urologic Diseases in America*, in *Chapter 3 Prostate Cancer, Tables 3-10 & 3-11*, Litwin MS, Saigal CS, and et.al., Editors. 2012, US Department of Health and

Human Services, Public Health Service, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases: Washington, DC. p. 73-96.

17. Lipscomb, J., C. Gotay, and C. Snyder, *Introduction to outcomes assessment in cancer*, in *Outcome assessment in cancer: Measures, methods and applications*, J. Lipscomb, C. Gotay, and C. Snyder, Editors. 2005, Cambridge University Press: New York. p. 1-13.

18. Lipscomb, J., C. Snyder, and C. Gotay, *Cancer outcomes measurement: Through the lens of the Medical Outcomes Trust framework.* Quality of Life Research, 2007. **16**(1): p. 143-64.

19. Lubeck, D.P., et al., *The CaPSURE database: a methodology for clinical practice and research in prostate cancer. CaPSURE Research Panel. Cancer of the Prostate Strategic Urologic Research Endeavor.* Urology, 1996. **48**(5): p. 773-7.

20. Jones, S. and S. Fox. *Generations online in 2009*. 2009 November 22, 2009]; Available from: http://www.pewinternet.org/Reports/2009/Generations-Online-in-2009.aspx.

21. Abernethy, A.P., et al., *Improving health care efficiency and quality using tablet personal computers to collect research-quality, patient-reported data.* Health Services Research, 2008. **43**(6): p. 1975-91.

22. Basch, E., et al., *Evaluation of an online platform for cancer patient self-reporting of chemotherapy toxicities.* Journal of the American Medical Informatics Association, 2007. **14**(3): p. 264-8.

23. Dupont, A., et al., *Use of tablet personal computers for sensitive patient-reported information.* Journal of Supportive Oncology, 2009. **7**(3): p. 91-7.

24. Farnell, D.J., et al., *Efficacy of data capture for patient-reported toxicity following radiotherapy for prostate or cervical cancer.* European Journal of Cancer, 2010. **46**(3): p. 534-40.

25. Ware, J.E., Jr. and M. Kosinski, *SF-36® physical and mental health summary scales: A manual for users of version 1.0*. 2nd ed. 2001, Lincoln, RI. : QualityMetric Incorporated.

26. McHorney, C.A., J.E. Ware, Jr., and A.E. Raczek, *The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs.* Medical Care, 1993. **31**(3): p. 247-63.

27. Stewart, A.L., R.D. Hays, and J.E. Ware, Jr, *The MOS short-form general health survey. Reliability and validity in a patient population.* Medical Care, 1988. **26**(7): p. 724-35.

28. Hays, R.D., C.D. Sherbourne, and R.M. Mazel, *The RAND 36-Item Health Survey 1.0.* Health Economics, 1993. **2**(3): p. 217-27.

29. Litwin, M.S., et al., *The UCLA Prostate Cancer Index: development, reliability, and validity of a health-related quality of life measure.* Medical Care, 1998. **36**(7): p. 1002-12.

30. Litwin, M.S., et al., *Quality-of-life outcomes in men treated for localized prostate cancer.* Journal of the American Medical Association, 1995. **273**(2): p. 129-35.

31. Ware, J.E., Jr., M. Kosinski, and B. Gandek, *SF-36® health survey: Manual & Interpretation guide*. 2004, Lincoln, RI: QualityMetric Incorporated.

32. Rosner, B., *The Intraclass correlation coefficient*, in *Fundamentals of Biostatistics, 6th Ed.* 2006, Thomson Higher Education Belmont, CA. p. 613-617.

33. Fleiss, J.L., *Statistical methods for rates and proportions*. 2nd ed. 1981, New York: Wiley.

34. Cicchetti, D.V. and S.A. Sparrow, *Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior.* American Journal of Mental Deficiency, 1981. **86**(2): p. 127-37.

35. Rosner, B., *The paired t test*, in *Fundamentals of Biostatistics, 6th Ed.* 2006, Thomson Higher Education Belmont, CA. p. 298-302.

36. Snijders, T. and R. Bosker, *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. 1999, Thousand Oaks, CA: SAGE Publications Inc.

37. D'Amico, A.V., et al., *Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer.* Journal of the American Medical Association, 1998. **280**(11): p. 969-74.

38. Cohen, J., *Calculation of effect size*, in *Statistical power analysis for the behavioral sciences*, 2nd, Editor. 1988, Lawrence Erlbaum Associates: Hillsdale, New Jersey. p. 24-27.

39. Farivar, S.S., H. Liu, and R.D. Hays, *Half standard deviation estimate of the minimally important difference in HRQOL scores?* Expert Review of Pharmacoeconomics & Outcomes Research, 2004. **4**(5): p. 515-23.

40. Norman, G.R., J.A. Sloan, and K.W. Wyrwich, *Interpretation of changes in health-related quality of life: The remarkable universality of half a standard deviation.* Medical Care, 2003. **41**(5): p. 582-592.

41. Revicki, D.A., et al., *Responsiveness and minimal important differences for patient reported outcomes.* Health and Quality of Life Outcomes, 2006. **4**(70): p. 1-5.

42. Sloan, J.A., et al., *The clinical significance of quality of life assessments in oncology: A summary for clinicians.* Supportive Care in Cancer, 2006. **14**(10): p. 998-98.

43. Ryan, J.M., et al., *A comparison of an electronic version of the SF-36 General Health Questionnaire to the standard paper version.* Quality of Life Research, 2002. **11**(1): p. 19-26.

44. Bliven, B.D., S.E. Kaufman, and J.A. Spertus, *Electronic collection of health-related quality of life data: validity, time benefits, and patient preference.* Quality of Life Research, 2001. **10**(1): p. 15-22.

Chapter 3

Cost Comparison of Mixed-Mode Administration of Patient-reported Outcome Surveys

Jeanette M. Broering[1†], Luke Tso[2], Suzanne C. Lessard[1], Ali Zargham[1], Alan Paciorek[1], Peter R. Carroll[1], Leslie S. Wilson[3]

Institutional Affiliations:

[1]Department of Urology, School of Medicine at the University of California San Francisco (UCSF) Helen Diller Family Comprehensive Cancer Center; San Francisco, CA

[2]School of Pharmacy, at the University of California San Francisco, San Francisco, CA

[3] Schools of Pharmacy and Medicine, at the University of California San Francisco, San Francisco, CA

Financial support: Funded by the Department of Urology, UCSF

[†]Correspondence:
University of California, San Francisco (UCSF)
Department of Urology
Box 1355
3333 California Street; Ste 282
San Francisco, CA 94143-1355
Tel (415) 514-0203
Fax (415) 514-0210
Email jbroering@urology.ucsf.edu

Abstract

Objective: To compare costs and outcomes between mailed paper-mode and web-mode survey methods, accounting for differences in volume, method mix, and response rates for the CaPSURE longitudinal prostate cancer registry.

Design: A cross-sectional convenience sample included participants' who self-selected to complete a patient-reported outcome (PRO) survey by paper (4,541 (90.6%)) or web (467 (9.4%)) modes.

Measurement and analysis: Semi-annual mailed paper-mode and web-mode workflow processes were defined for the time period January 1 – June 30, 2006. Fixed costs for optical character recognition (OCR) scanner ownership and web hosting costs were determined. A time trial was conducted to assess processing time. Variable costs were compared for staff wages associated with questionnaire throughput for pre-processing (distribution); processing (receiving, triage, editing, follow-up, OCR processing); and post-processing (double data entry (DDE), filing, storage). Outcomes were total costs, cost per survey, cost per response, and cost per error-free response.

Results: For web-mode surveys, the response rate was higher (88% vs. 76%, $P < 0.01$) and DDE error rate was lower (9% vs. 14%, NS). Total costs to process this 91:9 mix was $75,216 or on average $15.02 per survey with web-mode costs significantly higher ($18.47/survey) than paper-mode ($14.66/survey). Web-mode costs per response and per error-free response were higher.

Conclusion: The efficiency of web-mode surveys depends on improved response rates and error reductions. Web-mode becomes cheaper as volume increases and becomes a higher proportion of the survey mix. Although web-mode is more cost-efficient,

participants' often require both methods to mitigate coverage bias and to accommodate preference.

Background and Significance

Researchers and healthcare organizations can conduct comparative effectiveness research (CER) using observational data from registries, electronic health records, or claims data [1-5]. Inclusion of patient-reported outcomes (PRO) adds an important dimension to CER analyses. While collection of detailed PRO data for the creation of a disease-focused registry for research purposes is both costly and time intensive, it provides value when outcomes from multiple treatment paradigms are evaluated simultaneously [6-8].

Collection of PRO data is a reliable and efficient tool to evaluate people's behaviors, functional status after treatment, and preferences for health care [9]. Modes of survey administration have evolved from highly personal modes of face-to-face interviews, to highly tailored personalized modes through mailed paper-and-pencil or telephone administration, to a highly impersonal experience mediated through a computer interface [10]. The use of more than one mode, or mixed-mode, to communicate with study participants provides alternatives to collect data that generally yields similar results [11]. Few studies have evaluated its use in older cohorts [9-13]. With the rapid expansion of broad band Internet access, the use of mixed-mode to collect PRO data is possible. However, mixed-mode administration may result in different effects on sample coverage, response rates, costs, and adds other complexities to the workflow process. Additionally, despite advances in Internet accessibility, the exclusive use of web-based surveys in large epidemiologic research occurred in less than one percent of published studies [14].

Web-administered surveys can result in substantial reductions in costs associated with printing, postage, data processing, and may result in higher data quality [15-16].

Other studies suggest that the costs of web administered surveys are unknown [17-18]. Cost evaluation of mixed-mode administration is limited and suggests that mixed-mode may be more expensive but may result in higher response rates by participants [19-20]. While using a web-based approach may be more efficient, no study has compared the costs and benefits of mixed-mode administration of PRO questionnaires delivered by mail or through a web-based interface.

Objective

The purpose of this study was to compare fixed and variable costs and data accuracy between paper-mode and web-mode administered PRO surveys. The costs between modes were compared while accounting for differences in volume, method mix, and response rates during a single semi-annual mailing for the Cancer of the Prostate Strategic Urologic Research Endeavor (CaPSURE), a multi-institutional United States (US) based longitudinal registry database of men with prostate cancer [21-22].

Materials and Methods

Study Design. A cross-sectional cohort from CaPSURE was used. The mailed approach was defined as pencil-and-paper surveys mailed to participants, returned by mail, and then uploaded into the secure web application, using optical character recognition (OCR) technology. The web approach was defined as a digital version of the same survey that was administered on a personal computer and stored on a secure server via the World Wide Web.

Primary Data Source. Participating CaPSURE sites were predominantly community-based (n=34); with several Veterans Administration (n=3) and academic urology practices (n=3) [22]. Participants were recruited by their treating urologist and

given a paper baseline survey by the onsite study coordinator. Follow-up surveys, using mixed-mode, were administered by the data coordinating center at the University of California, San Francisco (UCSF).

Study Cohort. A web-based questionnaire for CaPSURE was initiated in July, 2005 after a randomized cross-over design study was performed to validate measurement equivalence between mixed-mode administration [11, 23]. A letter of invitation was inserted into the survey packet, which offered all participants the opportunity to complete their semi-annual questionnaire using the Internet. Participants were instructed to navigate to a secure web site where they registered for this option. For the subsequent administration period of January 1, 2006, a similar recruitment strategy was used. Inclusion criteria for the web-mode were: English language literacy; access to broadband Internet connectivity using Internet Explorer® or Mozilla Firefox browser; and willingness to self-administer via the secure web site. For this study, we used the six month phase from 1/1/2006 to 6/30/2006 where 91% of participants (n=4,541) opted to complete their survey by paper-mode and 9% (n=467) by web-mode (Figure 1). Approval for the mixed-mode questionnaire administration was obtained by the Committee on Human Research at UCSF.

Research Instruments. The PRO questionnaire was 32-pages in length and was a composite of several validated instruments that contained a total of 411 response fields of varying complexity for data processing (Table 1). The PRO included general and prostate-specific quality of life domains, satisfaction with treatment, bounded recall for health care utilization (past six-months), worker productivity, and demographic and insurance status questions [24-30]. The mailed survey was designed using Autonomy

Cardiff TeleForm® version 9.2 OCR software [31]. OCR was proven to be more efficient and accurate than manual data entry for the CaPSURE questionnaire [32].

The web version of the questionnaire was built using Microsoft .NET™ framework version 2.0 software and was hosted on a secure 128-bit encryption https secure server supported by two third-party commercial vendors, a web developer, and a secure hosting facility [33]. The web survey was designed to be visually comparable to the paper version [9, 12].

Prior to data collection, assumptions about framing the cost methods for the analytic plan were outlined [34]. The cost model assumed that all labor costs were for time spent for continuous work at average efficiency. The UCSF data coordination center research staff was experienced in processing PRO data for the CaPSURE study. Therefore, no costs to train staff were included in the model. The study assumed that there were no set up costs to develop the web-mode nor any costs associated with creation of the paper mode using the OCR software. All costs were framed from CaPSURE's actual staff model hourly wage costs in 2006 dollars. Web hosting contract fees can be quite variable both across geographic regions, across time, and hosting vendors but the contract fee for the CaPSURE web-mode hosting cost was averaged across the year based on cost estimates provided by the third-party web developer (K. Lewis, personal communication, January 16, 2008).

Cost Data. Because the business processes for each mode differs, processing steps for each mode were outlined and associated costs were determined from a time and motion study or obtained from assumptions derived from the prior mailing. Microsoft Excel®, 2003 was used to create cost equations for the decision model. Cost variables were both fixed, (i.e., those for the implementation of both methods which did not vary

by volume) and variable, (i.e., those associated with producing one survey unit and which varied by survey volume).

Questionnaire Processing Variables. For paper-mode questionnaires, workflow was divided into six quantifiable steps: distribution, receiving, editing, OCR processing, double data entry (DDE) on 20% of responses and filing; and storage. Each step contained multiple subtasks of varying complexity in order for a survey to be completed. Web-mode surveys needed fewer steps by eliminating OCR processing and required fewer subtasks (i.e., reduced filing). For web-mode, DDE was performed on 10% of responses (see Supplemental Tables 1 and 2).

Fixed Costs. Fixed costs for paper were defined as the OCR software costs; and for the web-based, as the web-hosting costs to post the survey. Costs to develop each survey mode were assumed to be equally cost intensive and were excluded from the analyses.

Variable Costs. Variable costs (VC) were based on the number of surveys sent and returned. Variable costs were calculated as: VC = (Probability * (Unit Wage Cost + Unit Production). Variable costs included probability estimates for the frequency of how often an activity occurred, assumed that not every step occurred for each survey, and were either estimated from historical project records or determined directly. Probabilities may not sum to 1.0 since many activities can happen to the same survey. Unit wage costs are the costs associated with the actual paid wages of the staff for producing one survey unit for each activity. This estimate was used to calculate the Minute Wage (MW) (i.e., MW = (Wage / Minute * Time (in minutes) / activity)). Unit production cost is the non-wage cost associated with producing the task (i.e., the cost to print a single questionnaire).

Total Costs. Total costs were the sum of the fixed and variable costs for all tasks.

Cost Comparison Analysis. A cost comparison analysis by mode was performed [34]. The primary outcome measure was financial costs associated with implementing and processing a single 6-month phase that occurred from January 1 to June 30, 2006. Cost estimates of data collection were calculated as: cost per survey; cost per response; cost per error-free response; and total costs. The analytic model was framed from the perspective of the CaPSURE data coordinating center at UCSF. All costs were converted to 2006 U.S. dollars.

Statistical Analysis. To describe participants and test for differences in the characteristics of participants who chose paper versus web-modes Pearson chi-square statistics were used. A *P* value of < .05 was considered statistically significant. The analysis for this study was generated using SAS® software, Version 9.2 of the SAS system for Microsoft Windows®.

Results

A total of 5,008 participants were sent a PRO (Figure 1). As shown in Table 2, web-mode respondents were significantly more likely to be younger; to report a college education and higher income; to experience lower risk disease, and to have radical prostatectomy as their primary treatment (all, $p < 0.01$).

Questionnaire complexity for processing could vary depending on the need to look-up prescription medication/s, to petition hospital audits, and to code outpatient procedures with ICD-9 procedure codes. With the exception of the number of self-reported prescription medications, no differences were found in complexity between modes (Table 2).

Fixed Costs. Tables 3 and 4 report all fixed and variable costs. While the fixed cost for equipment was the highest single cost for both methods of data collection, it was twice as high for the web-mode ($6,000). The web-mode survey was designed and built by a third-party software vendor who in turn contracted with another commercial web hosting vendor. The web-mode fixed processing costs consisted of the software web hosting contract which included both processing and storage. For the mailed approach, a scanner was used to process the paper surveys which cost only $3,000; half of the web-mode development and hosting contract fees. However, the web-mode fixed cost was twice as high as the OCR costs and was used to process only nine percent of the total eligible surveys.

Variable Costs of Distribution. Costs of distribution, which included printing and mailing, were the most different between modes of all variable costs ($6.85/paper-mode compared with $2.38/web-mode). The cost of production (paper, envelopes, printing, collating, and sales tax) was the highest contributor ($4.60 vs. $1.55, respectively). While the web-based approach had one mailing, the paper approach required at least two or more reminder mailings for non-responses. The ability to send e-mail reminders to web non-responders provided additional savings (Tables 3 and 4).

Variable Costs of Troubleshooting, Receiving, and Triage. The cost of receiving and processing was twice as expensive for paper-mode versus web-mode questionnaire ($0.60/paper vs. $0.28/web). The paper-mode required processing time that included envelope opening, logging-in using a two-dimensional barcode, and triaging for complexity. Sixteen percent (n = 549) of paper-mode respondents required phone calls to troubleshoot incomplete or illegible responses. These steps were eliminated with the web-based approach, except for the 2% (n = 8) of web-mode respondents who

requested a mode switch back to paper. The web-based processing was largely completed on-line, which required less personnel time (Tables 3 and 4).

Variable Costs of Editing. Different time costs were found for editing returned surveys by complexity which modified time to edit. Uncomplicated surveys required no coding or look-up for hospitalization, ICD-9 procedure codes, or prescription medication. Complicated surveys varied by the need for additional follow-up such as coding hospitalization, look-up of new prescription medications, coding of ICD-9 procedures and/or outlier events (i.e., phone call to participant or data meeting discussion)). Survey editing costs were higher for the web-based approach ($2.69 for web-mode vs. $2.12/paper-mode) because web-mode was found to be more visually difficult for staff to edit online. The web-based mode had fewer complicated responses (47% vs. 49%, respectively). However, both modes were equal in the proportion of surveys that required substantial editing (13%) (Tables 3 and 4).

Variable Costs of OCR. Our mailed surveys were processed electronically with OCR software in batches of 15 surveys. Batches required time for preparation, scanning, verification, as well as computer processing time to commit a batch, and time for the file transfer to upload into the secure website. A paper survey could have three levels of OCR complexity: messy, average/normal, or easy. OCR processing time for a messy survey took about 0.5 minute longer than for an average survey, and more than a minute longer than for an easy survey. The time spent in OCR processing was the highest cost for the mailed paper-mode approach ($1.52/mailed survey), and was completely eliminated for the web-mode (Tables 3 and 4).

Variable Costs of Double Data Entry and Filing. DDE for quality assurance monitoring and filing were added expenses. DDE and filing was more costly for paper-

mode ($1.01/survey) compared with web-mode surveys ($0.27/survey). Web-based methods required half of the need for DDE than did paper processing, 10% of web-mode versus 20% of paper-mode, due to higher accuracy. While filing was reduced substantially with the web-mode, it was not completely eliminated. All participants were sent a medical request consent form that was signed and returned. Additionally, hospital audits were requested on 13% of paper and 14% of web surveys to verify patient reported data. All of this content required filing regardless of mode (Tables 3 and 4).

Variable Cost of Storage. Storage was a cost category that was eliminated entirely by the web-based system ($2.51/paper-mode survey) and resulted in substantial savings for web-mode. It should be noted that the web-mode surveys were stored electronically and costs were subsumed in the contracting fee associated with the processing the web-mode surveys. These costs contributed little to that overall cost. Subsequently, due to enhancements in the OCR software, a digital image of the paper survey was captured during the batch export process thereby eliminating much of this cost category. However, at the time of the study, these costs were significant in terms of rental and storage cabinet costs. The paper-mode is now digitally archived which requires 3 gigabytes of storage space for a single mailing. Archiving digital images adds costs which may vary (i.e., monthly per gigabyte storage fee), such as the ongoing costs of content management software to organize and curate the asset, and electronic security costs to safeguard personal health information (Tables 3 and 4).

Benefits or Effectiveness. Two non-monetary benefits of the mixed-mode approach were examined: response rate and error rate. Although only a minority of participants chose to use the web-mode, that mode had a substantially higher response rate (88%) compared to paper-mode (76%) (Z-statistic 284.69; p <0.01). In addition,

when reviewed for errors, the web-mode had only a 9% error rate compared to a 14% rate for paper-mode (not statistically significant).

Cost-effectiveness. Finally mixed-mode costs were compared in four ways: Cost per survey, total costs, costs per response received, and costs per error-free response received. Table 5 reports actual costs which were highly volume dependent. At the current baseline volume, paper-mode costs were lower by $3.81 for sent surveys; $1.72 for received surveys; and only 65 cents cheaper for error-free response surveys.

To illustrate the effects of the volume and proportion who used each mode on cost, total costs were estimated using a hypothetical sample of 5,000 units (i.e., all paper, all web, or varying proportions) to estimate the total costs per ratio of paper-mode to web-mode surveys (Figure 2). As the proportion of web-mode use was increased to 100%, the total costs declined by approximately 50%.

Substantial value difference in mixed-mode adoption (volume) and the benefits of response rate and error-free response rate were used as additional denominators when comparing the variable costs of mixed-mode when the volume of surveys was scaled to a higher number. Initially both paper and web-modes are quite expensive per unit. The volume at which costs between modes became equivalent was at 200 surveys in each mode beyond which web continues to decline in cost. At 5,000 units in each mode group, paper-mode costs 2.6 times more per unit than web-mode ($22.50 versus $8.50, respectively) at which point the cost difference stabilizes when scaled to a hypothetical number of 99,999 units (Figure 3).

Discussion

This is the first study to examine the costs of performing mix-mode PRO survey administration within a longitudinal registry of men with prostate cancer. Findings from

this study suggest that at the current volume and mixture with only nine percent web-mode, the paper-mode surveys were less expensive per unit when compared to the cost for the mixed-mode surveys. These cost differences were due to the small proportion of participants who opted to use the web-mode which suggests that it was more cost effective at the current volume mix to administer the paper-mode. The low rate of web-mode use by this sample reduced the impact of mixed-mode on overall response rate and costs, despite web-mode's positive impact on data quality.

An accurate estimation of the population mix that will use a web-mode approach is an important consideration when determining the resources required in building a parallel business process for web-mode administration. Additionally, this estimate will enable the researcher to forecast the potential gain in efficiency of mixed-mode survey offering before initiating large scale longitudinal studies that incorporate PRO surveys. Because of the high fixed costs of contract services for processing, storing, and providing security for web-mode data, estimating the volume of participants that are willing to utilize the web-mode approach needs to be determined when estimating the efficiency of the overall survey collection approach used. These estimates may be difficult to derive. In an earlier survey of CaPSURE participants, 30% expressed willingness to use the web but in practice only 10% opted for web-mode [35]. The experience of the CaPSURE study is similar to findings reported by Couper and colleagues in the Health and Retirement Study (HRS) with regards to Internet survey usage. Noncoverage, due to lack of access to the Internet, was of greater concern for study participation than nonresponse (unwillingness to participate given Internet access) among older Americans who were HRS participants [36]. Finally, if contract services become more efficient, these costs may decrease, making the web-mode approach more cost effective.

This study has five important findings. First, based on the mix of mailed paper-mode (91%) versus web-based (9%) surveys, mailed surveys were less expensive per survey, per response received, and per error-free response received. However, these cost differences decreased when the mix of paper-mode versus web-mode was varied. Second, web-mode participants were significantly younger; better educated, reported higher annual incomes, had more low risk disease, and were healthier than paper-mode participants. These characteristics may explain some of the increased accuracy of the study data using the web-mode administration. Third, in this study, the response rate for web-mode was higher than for paper. However, given the small proportion of participants who chose web-mode, the overall response rate was unchanged (88% web, 76% paper, and 77% combined) which did little to improve overall response rate as reported in other national studies [8, 10-12, 19]. Fourth, the proportion of persons who requested a switch from web-mode back to paper-mode was low (2%) but was not zero. Data on mode switch has not been reported in the literature. Mode switches during this study were due to technical issues with navigating the web, or human-computer interaction factors such as survey length. Requests for mode switch from web-mode back to paper-mode may increase as a person ages due to changes in cognitive, visual, and motor skills that impact an older person's ability to interact with computing technology (see Supplemental Table 3). However, the provision of both modes minimizes the potential for sample coverage bias. Findings from this study suggest that few health surveys can be 100% web-based, if they are to reflect the general population of men with prostate cancer. Finally, a clearer understanding of the business processes for mixed-mode administration emerged that allowed the data coordinating center to identify areas to improve efficiency by mode.

This study has several limitations. Inherent difficulties exist in the ability to cost some of the items involved in each survey process, especially for the web-mode surveys. Many of these technology and contracting costs will decrease with time, making the web-based approach more cost efficient. In addition, web-based processes are likely to improve and become more cost efficient. For example, ability to flag errors in the web-mode approach will decrease the number of errors that need post-survey editing. In addition, mailed-based approaches are continually being combined with more automated methods which will further reduce their costs. In this study, our mailed survey approach that used OCR scanning instead of manual entry had demonstrated improved data processing efficiency [32]. Scanning, digitized filing, and storage of paper surveys would result in additional cost savings for the paper-mode. Finally, some costs were not included in this study such as survey design, development of separate workflows for each mode, training, updating equipment, and software costs (i.e., encryption and electronic security costs) for either approach. All of these factors may have different costs between modes.

Conclusions

These analyses demonstrated that with our current mixture and volume, the mailed paper-mode approach with computerized scanning, was cheaper per survey, per response, and per error-free response. However, as the survey volume increases and/or the mixture of methods increases for the web-mode approach, the web-mode approach quickly becomes much more cost-efficient. These considerations can be used to plan for new longitudinal PRO studies which are so important for CER studies when comparing two or more different health care treatments for prostate cancer, their outcomes, and costs [5].

Table 1. CaPSURE Questionnaire Field Descriptions

Instrument Name	Number of Items	Complexity
Administrative fields	11	Moderated – Date or write-in fields
Short Form – 36 (SF-36) –version 1	36	Low – Choice fields only
UCLA-Prostate Cancer Index (PCI)	20	Low – Choice fields only
Satisfaction with treatment	3	Low – Choice fields only
Signs & Symptoms Check List	63	Moderate
Health Care Resource Utilization	214	High – 51 open text fields; eight that link to look-up lists for standardized coding
Worker Productivity	19	Moderate
Health Insurance Status	21	Moderate – Six open text fields that link to master look-up table
Satisfaction with Health Care	12	Low – Choice fields only
Proxy Administration	12	Low – Choice fields only
Total Number of Response Fields	411	

Table 2. Socio-demographic, Clinical, and Survey Characteristics of Patient-reported Outcome Survey Responders

Study Characteristic	Value	Paper N = 3426†	(%)	Online N = 410	(%)	Pearson †† χ^2 *p*-value
Age at completion (*years*)	43-64	906	26	179	44	<.01
	65-69	650	19	78	19	.
	70-74	714	21	75	18	
	75-96	1156	34	78	19	.
Education	High school graduate	1115	36	54	14	<.01
	Some college	633	20	75	19	.
	College graduate	1370	44	268	68	.
	Unknown	308		13		.
Household Income	$<50,000	1476	52	104	28	<.01
	$50-75,000	590	21	82	22	.
	>$75,000	769	27	187	50	.
	Unknown	591		37		.
Race						<.01
	White	2884	92	382	96	
	Other	243	8	15	4	.
	Unknown	299		13		.
Risk Group‡	Low	1396	44	191	51	<.01
	Intermediate	1084	34	131	35	.
	High	693	22	54	14	.
	Unknown	253		34		.
Primary Treatment	Radical Prostatectomy	1876	56	298	75	<.01
	Radiation	802	24	58	15	
	Hormones	320	10	15	4	.
	Other	194	6	14	4	.
	Watchful Waiting	157	5	15	4	.
	Unknown	77		10		.
Health Status	Excellent	526	16	104	25	<.01
	Very Good	1387	41	180	44	.
	Good	946	28	93	23	.
	Fair	415	12	30	7	.
	Poor	81	2	3	1	.
	Unknown	71		0		.
Prescription Medications	None (0)	379	11	28	7	<.01
	1-3	1149	34	170	41	.
	4-6	1049	31	120	29	.
	7-23	849	25	92	22	.
Hospitalization	Yes	462	13	53	13	0.75
	No	2964	87	357	87	.
Hospitalization Count	0	2968	87	357	87	0.97
	1	391	11	45	11	.
	2-4	67	2	8	2	.
Outpatient Procedure	Yes	435	13	61	15	0.21
		2991	87	349	85	.
Outpatient Procedure Count	None (0)	2991	87	349	85	0.34
	1	304	9	40	10	.
	2-6	131	4	21	5	.

† N does not equal 3446 due to missing data.

††Pearson chi-square statistics tested for differences between observed and expected proportions in any single category. For statistically significant results at the 5% level, we followed these with tests of all pairs; the only categories that were not statistically significantly different between the modes were low and intermediate risk, medications counts 0 and 7-23, 4-6 and 7-23, pairs with treatment WW or hormones due to low cell counts, and the health status categories between good, fair, and poor due to low cell count.

‡D'Amico risk classification was calculated using clincal tumor stage, prostate-specific antigen value and Gleason tumor grade at diagnosis [37].

Table 3. Fixed and Variable Costs by Each Processing Step for Paper-mode

MAILED SURVEYS					
		Probability	**Unit Cost†**		**Total Cost**
FIXED COSTS					
	OCR Use Costs				$ 3,000
VARIABLE COSTS †					
Distribution	N = 4541	**Probability**	**Unit Cost**		**Total Cost**
Printing	Qx Packet	1.00	$ 4.60		$20,889
	Reminder Postcard	0.15	$ 0.30		$208
Postage	Postage Sent	1.00	$ 1.41		$6,403
	Postage Received	0.76	$ 1.00		$3,446
	Reminder	0.16	$ 0.20		$141
	Sub-total		**$ 6.85**		**$31,086**
	Response rate: 0.76; n = 3446				
Receiving / TS / Triage		**Probability**	**Unit Wage‡**	**Unit Time**	**Total Cost**
	Phone TS	0.25	$ 0.55	1.86	$ 883
	Receiving	1.00	$ 0.52	0.52	$ 933
	Triage	1.00	$ 0.52	0.52	$ 933
	Sub-total		**$ 0.60**		**$ 2,749**
Editing	n = 3446	**Probability**	**Unit Wage‡**	**Unit Time**	**Total Cost**
	Hospital Audit	0.14	$ 0.55	6.03	$ 1,600
	New Medication	0.49	$ 0.55	4.83	$ 4,486
	Uncomplicated Meds	0.40	$ 0.52	2.88	$ 2,064
	Uncomplicated Qx	0.11	$ 0.52	3.12	$ 615
	Outlier	0.13	$ 0.55	3.22	$ 793
	Sub-total			**2.12**	**$ 9,558**
OCR Processing	n = 3446	**Probability**	**Unit Wage‡**	**Unit Time**	**Total Cost**
	Preparation	1.00	$ 0.57	0.80	$1,571
	Scanning	1.00	$ 0.52	0.80	$1,434
	Messy	0.15	$ 0.55	3.38	$ 961
	Normal	0.50	$ 0.55	2.87	$ 2,085
	Easy	0.35	$ 0.52	2.06	$ 389
	Upload	1.00	$ 0.55	0.27	$ 512
	Sub-total			**1.52**	**$ 6,951**
DDE/Filing	n = 3446	**Probability**	**Unit Wage‡**	**Unit Time**	**Total Cost**
	Double Data Entry (DDE)	0.20	$ 0.57	2.58	$ 1,014
	Filing	1.00	$ 0.52	2	$ 3,584
	Sub-total				**$ 4,592**
Storage	n = 3446				
	Storage	1.00	$ 2.51		$ 8,649
	Sub-total				**$ 8,649**
Totals	**Total / Mode**				**$ 66,591**
	Grand Total				**$ 75,216**

Table 4. Fixed and Variable Costs by Each Processing Step for Web-mode

WEB-BASED SURVEYS				
Probability		**Unit Cost**		**Total Cost**
FIXED COSTS				
Web Software & Hosting				$ 6,000
VARIABLE COSTS†				
Distribution	N = 467	**Unit Cost**		**Total Cost**
Production Cost	1.00	$ 1.55		$ 724
Postage	1.00	$ 0.83		$ 388
Sub-total		**$ 2.38**		**$ 1,112**
Response rate: 0.88; n = 410				
Receiving / TS / Triage	**Probability**	**Unit Wage‡**	**Unit Time**	**Total Cost**
Phone TS	0.15	$ 0.55	3.855	$ 131
Receiving				$ -
Triage				$ -
Sub-total			**0.28**	**$ 131**
Editing	N = 410	**Unit Wage‡**	**Unit Time**	**Total Cost**
Hospital Audit	0.13	$ 0.55	6.70	$196
New Medication	0.47	$ 0.55	4.10	$ 435
Uncomplicated Meds	0.40	$ 0.52	2.30	$ 196
Outpatient ICD-9 code	0.15	$ 0.52	6.70	$ 227
Outlier	0.13	$ 0.55	3.22	$ 202
Sub-total			**2.69**	**$ 1,256**
	N = 410	**Unit Wage‡**	**Unit Time**	**Total Cost**
OCR Processing				$ -
				$ -
				$ -
				$ -
				$ -
				$ -
Sub-total				$ -
DDE/ Filing	N = 410	**Unit Wage‡**	**Unit Time**	**Total Cost**
Double Data Entry	0.10	$ 0.55	2.77	$ 62
Filing	0.15	$ 0.52	2.00	$ 64
		0.27		**$ 127**
Storage				
Storage	0.00	$ -		$ -
Totals	**Total / Mode**			**$ 8,625**
	Grand Total			**$ 75,216**

Table 5. Costs for CaPSURE Mixed-mode Survey Administration

Outcome	**Paper-mode (n = 4541)**	**Web-mode (n = 467)**	**Total (N = 5008)**
Cost / survey sent	$14.66	$18.47	$15.02†
Response rate*	3,446 (76%)	410 (88%)	
Cost / response received	$19.32	$21.04	
Error-rate by DDE‡	14%	9%	
Cost / error-free survey	$22.47	$23.12	
Total cost by mode	$66,591	$8,625	
Total cost			$75,216

† Average cost per survey for both modes

*Response rate was calculated as a ratio = (the number of surveys returned divided number surveys sent *100); expressed in percent (%) [38].

‡DDE is double data entry

Figure 1. CaPSURE Mixed-mode Sample

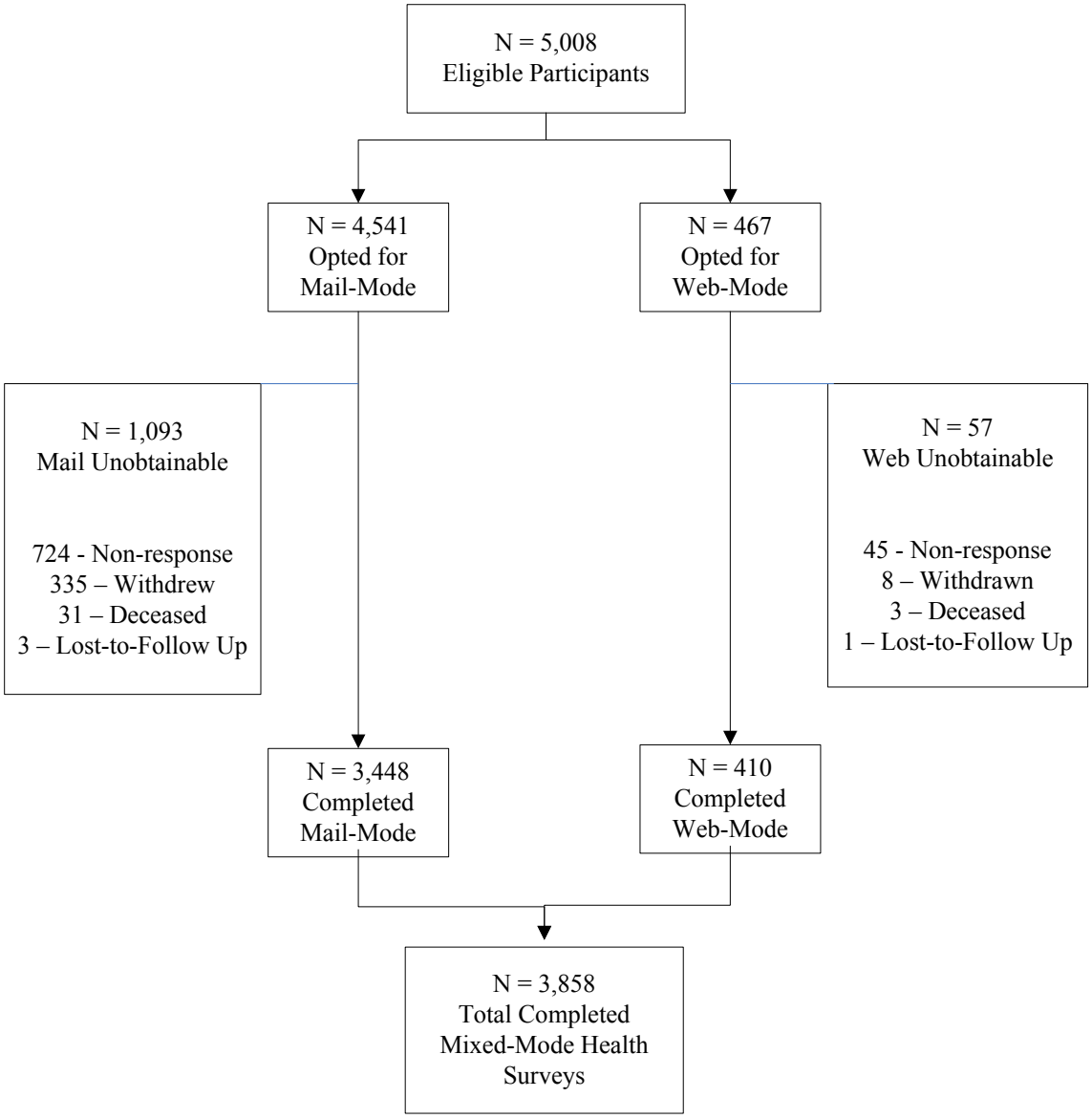
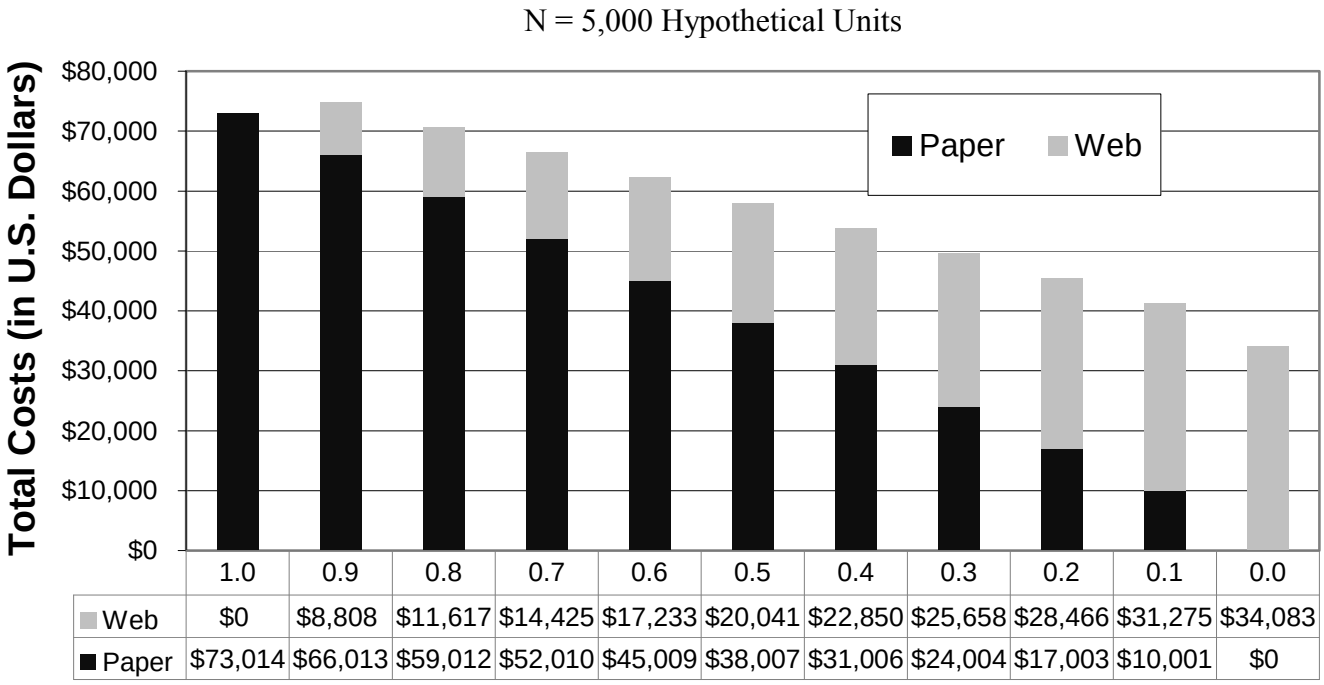
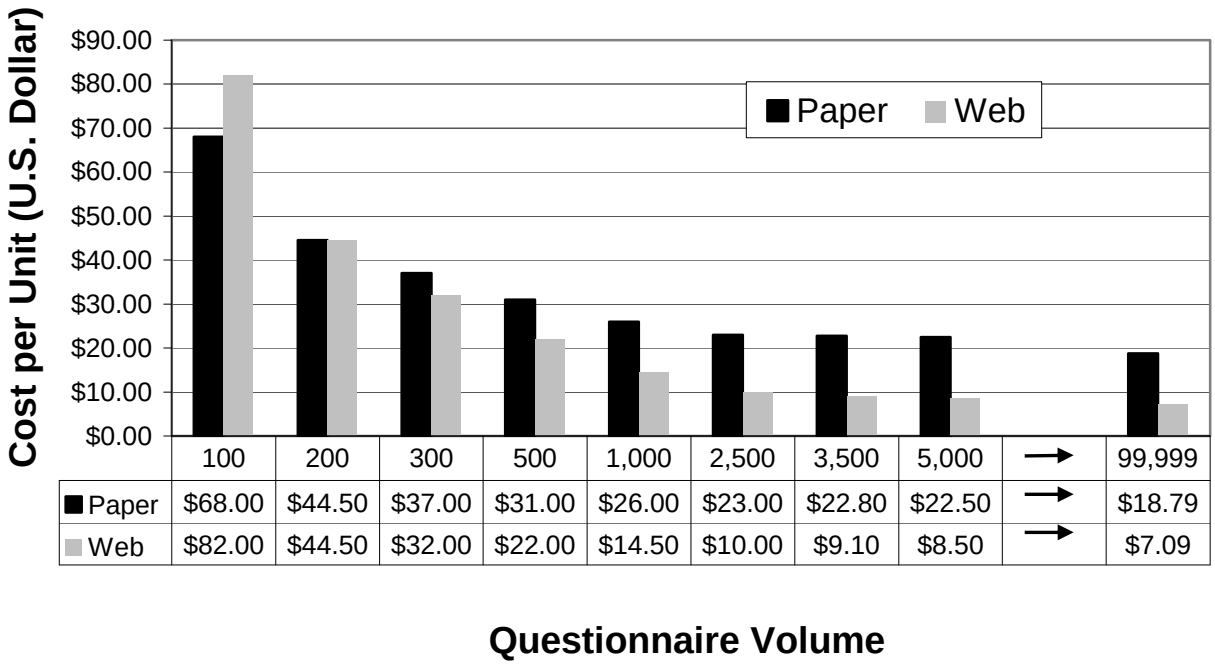


Figure 2. Cost Comparison Varying Proportion of Paper to Web Mixture



	1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.0
Web	$0	$8,808	$11,617	$14,425	$17,233	$20,041	$22,850	$25,658	$28,466	$31,275	$34,083
Paper	$73,014	$66,013	$59,012	$52,010	$45,009	$38,007	$31,006	$24,004	$17,003	$10,001	$0

Figure 3. Cost Comparison per Accurate Questionnaire by Volume and Mode



	100	200	300	500	1,000	2,500	3,500	5,000	→	99,999
Paper	$68.00	$44.50	$37.00	$31.00	$26.00	$23.00	$22.80	$22.50	→	$18.79
Web	$82.00	$44.50	$32.00	$22.00	$14.50	$10.00	$9.10	$8.50	→	$7.09

Supplemental Table 1. Comparison of Business Process Steps Used In Developing Time and Motion Study

MAILED SURVEYS	WEB-BASED SURVEYS
Fixed Costs **N = 4541** Cardiff TeleForms® Optical Character Recognition (OCR) software use costs	**Fixed Costs**: **N = 467** Software web hosting
Variable Costs **N = 3446 (76%)**	**Variable Costs** **N = 410 (88%)**
Pre-Processing Phase	
Distribution • Paper, sales tax, shipping & handling (S&H), and printing. • Collate packet to contain booklet survey, newsletter and 9 x 12 business reply envelope • Sort mailed packet by zip code • Scan out inventory and mail • Print reminder card and mail to non-responder.	**Distribution** • Paper, sales tax, shipping & handling (S&H), and printing. • Collate mailed. Mail packet contained cover letter, medical release and newsletter, and 4 x 9 business reply envelope. • Sort packet by zip code, scan out inventory, and mail. • Reminder for non-response sent by e-mail
Processing Phase	
Receiving • Open envelopes and scan as received • Troubleshooting o Disposition out of phase o Disposition study status changes • Article requests • Process site-specific VA medical release form (per IRB mandate)	**Receiving** • Check online email account daily (Research Analyst (RA) of week) • Troubleshooting o Disposition study status changes o Respond to technical issues i.e. e-mail delivery issues (failures), calls from patient about IT / browser access issues o Mode change from online to paper administration – mail out paper • Article requests • Process site-specific VA medical release form (per IRB mandate) •
Editing • Triage into levels of complexity: • Flag by complexity • Create Call Sheets • Distribute into appropriate bins o Low complexity: Uncomplicated or medications that were unchanged o Moderate complexity: New medication look up	**Editing** • Online triage into levels of complexity: • Create Call Sheets o All editing completed by RA of the week o Automated triage (new meds, hospitalization audit (HA) or, outpatient procedure (OP) procedure) • Edit: o Edit survey online (RA of week)

MAILED SURVEYS	WEB-BASED SURVEYS
o High Complexity: Hospital audit, outpatient procedure, or patient call etc. • Edit: o Editing o Logging as edited into web application • Outlier events to editing: o Respond to comments o Patient call o Discuss at data meeting review	• Outlier events to editing: o Respond to comments o Patient call o Discuss at data meeting review
OCR Processing • Preparation of batch: o Clipping into batches of 5 o Staple removal o OCR batch routing sheet • Scan: o Scan into Reader o Evaluation by Verifier • Verification (TeleForms®) by staff RA • Upload to Application: o Committing batch o Processing by Secure Website o Error Correction	
Post-Processing Phase	
Double Data Entry (DDE) per batch and Filing • Distribution to batch for DDE by editor • Review of errors **Filing - 100% of returns** • Sort by research ID • Drill two-hole punch into document • File survey in research chart	**Double Data Entry (DDE) per batch and Filing** • Print out 5 surveys before begin editing for DDE • Review of errors **Filing**: • File single page medical release form only for those who returned signed release and for filing HA for the 13% who reported inpatient event
Storage • Survey stored in paper research chart in file cabinets	**Storage** • Release form stored in paper research chart in file cabinets

Supplemental Table 2. Average Time (*in minutes*) to Process a Batch of 15 Questionnaires†*

Task	**Paper (minutes)**	**Online (minutes)‡**
Receiving	7.75	n/a
Triage	7.75	n/a
Editing (time for outlier events not included)	54.33*	54.55
Batch Preparation	11.93	n/a
Scan	11.80	n/a
Verification in TeleForms®	32.93	n/a
Upload	4.00	n/a
File	3.22	n/a
TOTAL Process Time/15 Surveys	133.71 **(8.9 min ea.)**	54.55 **(3.6 min ea.)**

† Average time data obtained from internal time and motion study.

* Initial time trial result for online editing was 69.38 minutes (8 batches of 15/excluding Uncomplicated); redid with additional 3 batches of Uncomplicated @ 1 minute/each for result of 54.55 minutes average.

‡Online result may be low because did not include time to fill out DDE tracking form. DDE was performed on all open text fields for the web-mode survey.

Supplemental Table 3. Reasons for Mode Switch From Web to Paper, N = 8

Reason	Category
• Technical difficulties with America Online (AOL) browser	Technical
• Letter from participant "I prefer to get a handwritten survey – I have too much stuff on email to handle".	Human-computer interaction
• Did online pilot, thinks questionnaire is too long to fill out online	Human-computer interaction
• Participant was unable to access online survey requested "snail mail"	Human-computer interaction
• Email failed and postal address 'return to sender'; spoke to participant and confirmed postal address correct - wants to do paper instead of online b/c of computer problems	Technical
• No reason stated (n=3)	Unclassified

References

1. Dubois, R.W. and J.S. Graff, *Setting priorities for comparative effectiveness research: from assessing public health benefits to being open with the public.* Health Affairs, 2011. **30**(12): p. 2235-42.
2. Fleurence, R.L., H. Naci, and J.P. Jansen, *The critical role of observational evidence in comparative effectiveness research.* Health Affairs, 2010. **29**(10): p. 1826-33.
3. Sox, H.C., *Defining comparative effectiveness research: the importance of getting it right.* Medical Care, 2010. **48**(6 Suppl): p. S7-8.
4. Tunis, S.R., J. Benner, and M. McClellan, *Comparative effectiveness research: Policy context, methods development and research infrastructure.* Statistics in Medicine, 2010. **29**(19): p. 1963-76.
5. Garber, A.M. and H.C. Sox, *The role of costs in comparative effectiveness research.* Health Affairs, 2010. **29**(10): p. 1805-11.
6. Dreyer, N.A., et al., *GRACE principles: recognizing high-quality observational studies of comparative effectiveness.* American Journal of Managed Care, 2010. **16**(6): p. 467-71.
7. Dreyer, N.A., et al., *Why observational studies should be among the tools used in comparative effectiveness research.* Health Affairs, 2010. **29**(10): p. 1818-25.
8. Gliklich, R.E. and N.A. Dreyer, *Registries for evaluating patient outcomes: A users guide, Second Edition*, in *Outcome DEcIDE Center [Outcome Sciences, Inc. d/b/a/ Outcome]*, R.E. Gliklich and N.A. Dreyer, Editors. 2010, Agency for Healthcare Research and Quality AHRQ Publication No. 10-EHC049: Rockville, MD. p. 1-347.

9. Dillman, D., J. Smyth, and L. Christian, *Internet, mail and mixed-mode surveys: The tailored design method*. 3rd ed. 2009, Hoboken, NJ: John Wiley & Sons, Inc.

10. Groves, R.M., et al., *Survey methodology*. 2nd ed. Wiley Series in Survey Methodology. 2009, Hoboken, New Jersey: A. John Wiley & Sons, Inc. 461.

11. Gwaltney, C.J., A.L. Shields, and S. Shiffman, *Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: a meta-analytic review.* Value in Health, 2008. **11**(2): p. 322-33.

12. deLeeuw, E., J. Hox, and D. Dillman, *Mixed-mode surveys: When and why*, in *International handbook of survey methodology*, E. deLeeuw, J. Hox, and D. Dillman, Editors. 2008, Psychology Press Taylor and Francis Group: New York. p. 299-316.

13. deLeeuw, E.D., *To mix or not to mix data collection modes in surveys.* Journal of Official Statistics, 2005. **21**(2): p. 233-255.

14. van Gelder, M.M., R.W. Bretveld, and N. Roeleveld, *Web-based questionnaires: the future in epidemiology?* American Journal of Epidemiology, 2010. **172**(11): p. 1292-8.

15. Kypri, K., S.J. Gallagher, and M.L. Cashell-Smith, *An internet-based survey method for college student drinking research.* Drug and Alcohol Dependence, 2004. **76**(1): p. 45-53.

16. McAlindon, T., et al., *Conducting clinical trials over the internet: feasibility study.* BMJ, 2003. **327**(7413): p. 484-7.

17. Ekman, A. and J.E. Litton, *New times, new needs; e-epidemiology.* European Journal of Epidemiology, 2007. **22**(5): p. 285-92.

18. Rodriguez, H.P., et al., *Evaluating patients' experiences with individual physicians: a randomized trial of mail, internet, and interactive voice response telephone administration of surveys.* Medical Care, 2006. **44**(2): p. 167-74.

19. Greenlaw, C. and S. Brown-Welty, *A comparison of web-based and paper-based survey methods: testing assumptions of survey mode and response cost.* Evaluation Review, 2009. **33**(5): p. 464-80.

20. McHorney, C.A., M. Kosinski, and J.E. Ware, Jr., *Comparisons of the costs and quality of norms for the SF-36 health survey collected by mail versus telephone interview: results from a national survey.* Medical Care, 1994. **32**(6): p. 551-67.

21. Lubeck, D.P., et al., *The CaPSURE database: a methodology for clinical practice and research in prostate cancer. CaPSURE Research Panel. Cancer of the Prostate Strategic Urologic Research Endeavor.* Urology, 1996. **48**(5): p. 773-7.

22. Cooperberg, M.R., et al., *The contemporary management of prostate cancer in the United States: lessons from the cancer of the prostate strategic urologic research endeavor (CapSURE), a national disease registry.* Journal of Urology, 2004. **171**(4): p. 1393-401.

23. Broering, J.M., et al., *Validation of an Internet based patient health-related quality of life questionnaire: Data from CaPSURE [abstract].* Value in Health, 2006. **9**(3): p. A114.

24. Litwin, M.S., et al., *The UCLA Prostate Cancer Index: development, reliability, and validity of a health-related quality of life measure.* Medical Care, 1998. **36**(7): p. 1002-12.

25. Litwin, M.S., et al., *Quality-of-life outcomes in men treated for localized prostate cancer.* Journal of the American Medical Association, 1995. **273**(2): p. 129-35.

26. Lubeck, D.P., et al., *Measurement of health-related quality of life in men with prostate cancer: the CaPSURE database.* Quality of Life Research, 1997. **6**(5): p. 385-92.

27. Lubeck, D.P., et al., *An instrument to measure patient satisfaction with healthcare in an observational database: results of a validation study using data from CaPSURE.* American Journal of Managed Care, 2000. **6**(1): p. 70-6.

28. Lubeck, D.P., P.W. Spitz, and J.F. Fries, *The health assessment questionnaire (HAQ): Assessment of personal economic costs [abstract].* Arthritis and Rheumatism, 1982. **25**: p. S24.

29. Lubeck, D.P., et al., *A multicenter study of annual health service utilization and costs in rheumatoid arthritis.* Arthritis and Rheumatism, 1986. **29**(4): p. 488-93.

30. Ware, J.E., Jr., M. Kosinski, and B. Gandek, *SF-36® health survey: Manual & Interpretation guide*. 2004, Lincoln, RI: QualityMetric Incorporated.

31. *TeleForm® Version 9.2*. Autonomy CARDIFF 2011 [cited 2011 July 30]; Available from: http://www.cardiff.com/products/teleform/.

32. Broering, J.M., et al., *CaPSURE.net e-Tools: Comprehensive business process management for research questionnaire adminstration [abstract]*, in *National Cancer Institute, Critical Issues in e-Health Research Conference*. 2005: Bethesda, MD.

33. *Microsoft.NET Framework*. [cited 2011 July 30]; Available from: http://www.microsoft.com/net.

34. Torrance, G.W., J.E. Siegel, and B.R. Luce, *Framing and designing the cost-effectiveness analysis*, in *Cost-effectiveness in health and medicine*, M.R. Gold, et al., Editors. 1996, Oxford University Press: New York. p. 54-80.

35. Broering, J.M., et al., *Predictors of subject interest to complete a web based health-related quality of life questionnaire: Data from CaPSURE [abstract]*, in *National Cancer Institute, Critical Issues in e-Health Research Conference*. 2005: Bethesda, MD.

36. Couper, M.P., et al., *Noncoverage and nonresponse in an Internet survey.* Social Science Research, 2007. **36**: p. 131-148.

37. D'Amico, A.V., et al., *Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer.* Journal of the American Medical Association, 1998. **280**(11): p. 969-74.

38. AAPOR, *Standard definitions: Final dispositions of case codes and outcome rates for surveys*. 2009, American Association of Public Opinion Research (AAPOR): Ann Arbor, MI.

Chapter 4

Data Quality and Mixed-Mode Administration of Patient-reported Outcome Surveys

Jeanette M. Broering[1†], Alan Paciorek[1], Peter R. Carroll[1], Leslie S. Wilson[2], Mark S. Litwin[3], and Christine Miaskowski[4]

Jeanette M. Broering, RN, PhD, MPH [1†] - Corresponding Author
Department of Urology, School of Medicine
University of California San Francisco (UCSF)
San Francisco, CA 94143-1355
Tel (415) 514-0203
Fax (415) 514-0210
jbroering@urology.ucsf.edu

[1] Alan Paciorek, BS
Biostatistical Programmer Analyst
Department of Urology, School of Medicine
University of California San Francisco (UCSF)
apaciorek@urology.ucsf.edu

[1] Peter R. Carroll, MD, MPH
Chairperson and Professor
Department of Urology, School of Medicine
University of California San Francisco (UCSF)
pcarroll@urology.ucsf.edu

[2] Leslie S. Wilson, PhD
Adjunct Professor
Department of Clinical Pharmacy, School of Pharmacy
University of California San Francisco (UCSF)
wilsonl@pharmacy.ucsf.edu

[3] Mark S. Litwin, MD, MPH
Chairperson and Professor of Urology and Health Services Research
Department of Urology, David Geffen School of Medicine and Fielding School of Public Health
University of California Los Angeles (UCLA)
mlitwin@mednet.ucla.edu

[4] Christine Miaskowski, RN, PhD, FAAN
Professor of Nursing
Department of Physiological Nursing, School of Nursing
University of California San Francisco (UCSF)
Chris.miaskowski@nursing.ucsf.edu

Abstract

Objective: To examine the effects of mixed-mode questionnaire administration (i.e., paper-mode versus web-mode) on data quality parameters that can affect external validity.

Methods: The Medical Outcomes Short-Form 36 (SF-36) and the UCLA Prostate Cancer Index were administered to participants in the Cancer of the Prostate Strategic Urologic Research Endeavor (CaPSURE), a prostate cancer registry. Testing metrics included comparison of response rates; non-respondent characteristics; data completeness of items and computable scale scores; comparison of mean scale scores and ceiling and floor response effects; and comparison of physical and mental component summary scores against US norms.

Results: The sample included 4,836 participants of whom 4,376 (90%) opted for paper-mode and 460 (10%) for the web-mode. Overall response rate was 77%, paper-mode 76% versus 88% web-mode (p <0.01). Predictors of overall response were older age, being Caucasian, having attained college education, and living in a significant relationship. Paper-mode participants were less likely to respond (OR 0.69, 95%CI .68–.70) and had significantly more missing data on the SF-36 (mean missing .67 versus .10, p < .0001) and the UCLA-PCI (mean missing .67 versus .16, p <0 .000) and fewer computable scale scores. All mean scale scores were lower among paper-mode participants but effect sizes for clinically meaningful differences between modes were small. Comparison to US based age-stratified norms for males suggests CaPSURE participants had higher levels of physical and mental health functioning.

Conclusions: Significant differences were found between mode of administration and data quality for respondents but differences were small. These findings suggest that the use of mixed-mode administration does not introduce significant measurement differences rather it allowed for participation by older men.

Introduction

The incorporation of a patient-reported outcome (PRO) surveys to measure the effects of treatment-related toxicities associated with therapies for cancer has grown in importance for use in clinical trials, observational registries, and clinical care [1-4]. A variety of methods can be used to contact participants and to deliver and administer PRO instruments. When answering PRO questions, considerable cognitive demand is placed on the respondent to comprehend the meaning of the questions, to recall relevant information from memory, to link the information retrieved with the questions asked, and to communicate a response. The communication channel that is used to present information to a respondent (e.g., auditory, oral, visual) may place additional literacy demands on the respondent especially for self-administered PRO surveys [5]. The use of mixed-mode administration, or more than one approach to contact and collect PRO data, is not a new phenomenon [6]. While the paper-mode and web-mode employ a visual channel of communication, the web-mode adds the burden of computer literacy skills and Internet access.

Only five studies have evaluated the impact of mixed-mode administration (paper-mode versus computer-mode) on data quality within oncology settings [7-11]. The purposes of these studies were to pilot test the feasibility of data capture using an e-Tablet [7, 10], or to assess the technology platform for feasibility and acceptability within clinical care settings [8, 11]. All of these mixed-mode PRO studies were performed at a single academic institution; included relatively small samples (i.e., 56 to 149 participants); and were conducted within a clinical environment where technical assistance was readily available [7-11]. In a study that evaluated home log-in access to

self-administered PRO instruments, a large disparity in response rates was found in terms of the patients' gender and cancer diagnosis. Younger female patients with gynecologic cancers had the highest remote access rate when compared to men with lung cancer (66% versus 15%, respectively) [8].

Assessment of data quality within the context of mixed-mode administration can be evaluated by a number of parameters including survey response rates, item response rates, the accuracy of response rates, absence of bias (e.g., social desirability bias, acquiescence bias, or interviewer bias), and the completeness of information. Data quality were not consistently evaluated in the five oncology mixed-mode studies. In three of these studies, only response rates and study attrition were evaluated [7, 9, 11]. Only two studies evaluated item non-responses [9, 11].

Cross-sectional data from the 1990 National Survey of Functional Health Status, a population-based study, was used to evaluate for differences in cost and data quality when the Short Form -36 (SF-36) Health Survey was administered by paper-mode or telephone-mode [12-13]. This study provides guidance on the metrics that can be used to evaluate data quality including: overall response rates, predictors of non-responses, and assessment of responses to individual items.

Given the importance of PRO outcomes in oncology [2, 14-17] and the paucity of research on mixed-mode administration, the purpose of this study was to evaluate for differences in data quality between paper-mode versus web-mode administration of two PRO questionnaires (i.e., Medical Outcomes Study (MOS) SF-36 version 1 and the UCLA Prostate Cancer Index (UCLA-PCI) in patients with prostate cancer [18-20]. Differences in data quality between modes were compared while accounting for

differences in response rates, predictors of response, data completeness, differences in mean scale scores and response effects (i.e., ceiling and floor effects), and an anchor-based comparison of the sample data against United States (US) normative data for men stratified by age [21].

Methods

Participants

Participants were derived from a single cross-sectional wave of questionnaires administered to all active registry participants in the Cancer of the Prostate Strategic Urologic Research Endeavor (CaPSURE) database that contained 5,008 eligible individuals [22]. CaPSURE participants completed a baseline PRO survey at enrollment by paper-mode and were asked to complete a follow-up PRO survey every 6 months thereafter. In subsequent administrations, participants were given the option to migrate to a web-mode version of the PRO questionnaire that was designed to be visually equivalent to the paper-mode [23]. Two exclusion criteria were applied for the creation of the analytic dataset. One-hundred and twenty-seven participants were excluded from this analysis because PRO data were reported by a proxy [24]. An additional 45 persons had died. Of the 4,836 eligible participants, 4,376 (90%) opted for the paper-mode and 460 (10%) opted for the web-mode. Participants were not compensated for their participation. The Committee on Human Research at University of California San Francisco (UCSF) approved this study.

Instruments

The MOS SF-36 version 1 consists of 36 items that evaluate overall physical and mental health status. The SF-36 consists of eight health status scales that measure

physical functioning (PF), role limitations due to physical health (RP), bodily pain (BP), social functioning (SF), general mental health (MH), role limitations due to emotional problems (RE), vitality, energy, or fatigue (VT), general health perceptions (GH) and a single item that compares transition in health status over the past year (HT) [18]. The eight scales, excluding the HT item, are scored to create a summary component measures for physical and mental health outcomes. The PF, RP, BP, and GH scales correlate most highly with the physical component summary (PCS) score. The VT, SF, RE, and MH scales correlate most highly with the mental component summary (MCS) score [21].

The UCLA-PCI consists of 20 items with six dimensions that assess urinary, bowel, and sexual function and bother [19-20]. Both the SF-36 and the UCLA-PCI instruments are scored from zero to 100, with a higher score indicating better HRQOL. Both instruments used a standard 4–week recall period.

Data Collection

The paper-mode questionnaire was returned by pre-paid business reply mail for processing by the data coordinating center. The paper-mode questionnaire was designed using TeleForms® version 9.2 optical character recognition (OCR) software in a booklet format that allowed for scanning, verification, and batch process submission into the secure web application, using the OCR technology [25]. In CaPSURE, OCR is a more efficient and accurate method than manual data entry [26]. The web-mode was a digital version of the same survey that was completed on a personal computer and stored on a secure server. A web-mode survey option was offered to all active CaPSURE participants after measurement equivalence between the two modes was established [27].

A letter of invitation was inserted into the participant's questionnaire packet that offered them the opportunity to complete their semi-annual questionnaire using the Internet. Interested participants were instructed to navigate to a secure web site where they registered for this option. Inclusion criteria for the web-mode were: English language literacy; access to broadband Internet connectivity using Internet Explorer or Mozilla Firefox; and willingness to self-administer through the secure web site. The paper-mode questionnaire was available in English and Spanish while the web-mode offering was available only in English.

Statistical Analyses

Response rate

Response rates were calculated as the ratio of returned PRO surveys to the total number of eligible units for the total sample and by each mode [12]. Partial completion of a survey was included as a responder and incorporated into the calculation of the response rate [24].

Nonresponse bias

Nonresponse bias was defined as differences between participants who did or did not respond in pre-existing characteristics that might bias response estimates [12]. To evaluate nonresponse bias, Chi-square test of correlation evaluated differences in demographic and clinical characteristics by response status (responded vs. nonresponse) and mode (paper vs. web). A multivariable logistic regression analyses evaluated predictors of PRO response. Predictor variables were added in the order of strongest association from the univariate analyses. Each predictor was retained in the final model if it had a Wald p-value of $\leq .05$. At each step of adding a predictor, categories were

redefined to obtain the strongest associations (e.g., participants with advanced stage tumor stage 3 and 4 were combined into a single category). Odds ratios (OR) and their 95% confidence intervals (CI) were used to quantify the relative risk of response associated with each of the explanatory variables. The predictive performance of the logistic regression was assessed by the *C* statistic [28]. The Hosmer-Lemeshow Chi-square goodness of fit test determined the extent to which the data fit the model. A likelihood ratio test evaluated competing models for final model selection.

Data quality

Data quality parameters were defined as the completeness of data at the item and scale levels [5, 12]. Data completeness was evaluated by three methods. For each item on the SF-36 and the UCLA-PCI, the proportion of each item completed and the number of scales with computable scores were calculated. The Fisher's Exact statistical test was used to evaluate for differences in completeness between modes of administration [29]. Mean, standard deviation (SD), minimum and maximum, and the percentages of ceiling and floor scores of each scale were computed. A *t*-test was used to compare mean scores between modes. When statistically significant differences in means scores were found, a distribution-based method was used to calculate an effect size (ES) to determine if these findings represented clinically meaningful differences [30-32].

Data Comparability

Comparison of group means of scale scores for the CaPSURE sample and US national normative data was done using an anchor-based approach [31-32]. A *t*-test was used to compare means in the PCS and the MCS scores between CaPSURE and US

norms for males by age category [21]. When statistical differences were found, an ES was calculated.

Analyses were generated using SAS® software Version 9.2 (SAS Institute, Cary, NC, USA) and the multivariable logistic regression models were generated using STATA version 11. A *p*-value of ≤ .05 for two-sided test of correlation was considered statistically significant.

Results

Participant characteristics: Prior to this PRO questionnaire administration, the average duration of participation in CaPSURE was 4.4 years (SD 3.0). The mean number of months since diagnosis with prostate cancer was 46 months. On average, a CaPSURE participant had completed eight surveys (SD 6.9). However, 296 participants (6.1%) had never completed a prior PRO survey. On average paper-mode respondents were older (70.7 years, SD 8.8) than web-mode respondents (66.8 years, SD 8.5, Table 1).

Response rate: The overall response rate was 76.7%. A significantly lower response rate was observed for the paper-mode at 75.6% compared to 88.1% for the web-mode (Pearson $p < 0.01$).

Differences between response status and modes: A series of preliminary univariate analyses were performed between response status (responded versus nonresponse) and mode (paper-mode versus web-mode). Statistically significant differences were found by response status for all demographic, health status, and clinical characteristics (all Pearson $p < 0.001$) with the exception of annual income (Pearson $p = 0.28$) and the number of comorbid health conditions at the time of study enrollment (Pearson $p = 0.20$). Statistically significant differences were found by mode of

administration on all demographic, health status, and clinical characteristics with the exception of tumor-stage (Pearson $p = 0.09$) (Data not shown).

Differences between respondents by mode: Differences between responders who used the paper-mode or the web-mode of administration are summarized in Table 1 (columns 1 versus 3). Significant differences were observed for all demographic, health status and clinical characteristics (all p values < .05). Paper-mode respondents were significantly more likely to be older (35% vs. 21% for ≥75 years old); non-White (7% vs. 4%); to have less education (≤ high school diploma, 34% vs. 13%); to be widowed or not be in a significant relationship; to have a lower annual income (<$50K 51% vs. 28%); and to have opted for primary treatments other than radical prostatectomy (i.e., other treatments 42% vs. 25%). In contrast, web-mode respondents were more likely to be from the youngest age category (41% vs. 25%, for the 41–64 year old group) to be Caucasian (96% vs. 92%); to report college education (68% vs. 45%); to report an annual income of > $75K (50% vs. 28%); to report managed care or preferred provider organization as their source of health insurance (57% vs. 39%); to have low risk prostate cancer disease (51% vs. 45%); to report their health status as excellent to very good (69% vs. 58%); and to have opted for radical prostatectomy (RP) as their primary treatment (75% vs. 57%).

Differences between non-respondents by mode. Differences between non-responders who used the paper-mode or the web-mode of administration are summarized in Table 1 (columns 2 versus 4). For non-respondents by mode, no differences were observed by age category ($p = 0.68$), health insurance status ($p = 0.11$), and risk group ($p = 0.47$). Paper-mode non-respondents were significantly more likely to be non-White

(21% vs. 6%); to report ≤ high school diploma (43% vs. 12%); to not be living alone (9% vs. 0%); to have an annual income of less than $50K (51% vs. 33%); and were less likely to have opted for RP as a primary treatment (49% vs. 36%).

Predictors of PRO response: The result of the multiple logistic regression analysis of predictors of PRO response category for all 4,836 participants is shown in Table 2. The following potential predictors were not significant and were not retained in the model: disease risk status, prostate specific antigen (PSA) value at diagnosis, tumor stage at diagnosis, Gleason grade, and insurance status. Except for living alone (e.g., widowed) all of the predictor variables were associated with being a respondent. Holding all predictors constant in the model, younger participants were 21% less likely to respond (OR .79, 95%CI .67–.94). Non-white participants were half as likely to respond (OR .50, 95%CI .41–.62). Those participants with less than a college education or whose education status was unknown; were less likely to respond. Participants with a primary treatment other than radical prostatectomy or whose treatment was unknown; were less likely to respond. Completion of a prior survey was positively associated with completion of the current survey (OR 1.10, 95%CI 1.08–1.12).

The last predictor entered into the model was mode of administration. After adjusting for all the other predictors, participants who elected the paper-mode of administration were 31% less likely to respond (OR 0.69, 95% CI .68–.07). The full model had a *C*-statistic of 0.736, which fits moderately well and suggests acceptable discrimination. Although some unmeasured predictor/s were not accounted for in the model for the likelihood of response (Hosmer-Lemeshow test with 8 *df* showed that a good fit was not achieved (p <0.01)). Comparison of the full model to a nested model

that excluded mode suggested that mode was moderately associated with the likelihood of being a respondent (Likelihood Ratio χ^2, $p = 0.08$).

To evaluate predictors of response within mode, separate multivariable logistic regression models were generated for the paper-mode and web-mode of administration. For the paper-mode logistic nested model, all explanatory variables were significant and the ORs were similar to the full model. The *C* statistic for the paper-mode model was 0.745. In contrast, for the web-mode logistic nested model, only one explanatory variable remained significant (i.e., living alone compared to living with a spouse or partner) remained significant ($p = .01$). The *C* statistic for web-mode model was 0.639 (data not shown).

Data quality

Data completeness: For both PRO instruments, the average number of missing items was significantly higher for the paper-mode of administration. For the SF-36, the mean number of missing items was .67 (SD 2.5) for the paper-mode and .10 (SD .54) for the web-mode (*t*-test, $p < .01$). For the UCLA-PCI, the mean number of missing items was .67 (SD 1.95) for the paper-mode and .16 (SD .72) for the web-mode (*t*-test, $p < .01$).

Table 3 presents the proportion of respondents by mode with complete items and computable scale scores for the SF-36 and the UCLA-PCI. For the SF-36, the proportion of scales with missing data ranged from 92% to 99% for the paper-mode and 97% to 100% for the web-mode. Item-level differences for missing data were statistically significantly different for PF, BP, GH, SF, MH and HT scales. The SF scale experienced the most missing data at the item-level and computable scale score level (92.3%) for the paper-mode. Computable scale scores were achieved for 100% of five of the eight scales

for the web-mode. For computable scale scores, the PF, RP and RE did not differ by mode whereas the other five scales were statistically significant (i.e., BP, GH, VT, SF, and MH).

For the UCLA-PCI, four of the six scales had significantly more missing data at both the item-level and computable scale score level (UF, SF, SB, and BB). For the paper-mode the UF and the SF scales had the most missing data (89% and 90%, respectively) compared to the web-mode (96% and 95%, respectively). No differences were found between mode for the UB and BF scales.

Tables 4 and 5 report the mean scale scores for the SF-36 and the UCLA-PCI, respectively. Web-mode respondents had significantly higher mean scale scores, as well as PCS and MCS scores (all $p < 0.05$). Differences in effect sizes by mode were the largest for the PCS, PF, and VT scores (*ES* = .30, .36, and .32, respectively). The inability to compute the PCS and MCS scores due to missing data was most apparent for the paper-mode respondents compared to the web-mode (i.e., 11% vs. 1%, respectively). A similar trend was observed for the UCLA-PCI, with web-mode respondents reporting significantly higher mean scores. The sexual function domain had the largest observed effect size (*ES* = .34).

Tables 4 and 5 report the minimum and maximum scale scores, and the percentage of the sample with the lowest or highest score for each scale and the physical and mental component summary scores for each PRO instrument by mode. A PRO instrument should be able to reliably assess a range of health states in a specific population [33]. Established standards for minimal floor and ceiling effects (i.e., <15%

of a specific population with the lowest or highest possible score) have been published [34].

For the SF-36, the VT, MH, BP, and GH scale scores did not achieve the floor value of zero for the web-mode (i.e., 10, 28, 22.5, and 15, respectively). The PF, RP, RE, and SF scales did achieved the lowest value (i.e., 0 score and < 15% of each scale) for both modes. The highest value (ceiling of 100) was reported for all eight SF-36 scales for both modes. A ceiling effect that exceeded the 15% rule; was found for the PF, RP, RE, and BP scales for both modes (Table 4).

In terms of the UCLA-PCI, for the web-mode group, floor effects were not achieved for urinary function. For both modes, a floor effect on the bowel function scale was not achieved. For both modes, all scales of the UCLA-PCI achieved a ceiling value (100) and exceeded the 15% standard for the percent of participants who had the highest possible scores for urinary function, urinary bother, sexual function, bowel function, and bowel bother (Table 5).

Comparison to US normative data: Tables 6 and 7 report the PCS and the MCS scores for the entire sample as well as by mode. The means scores for the entire sample were compared against age stratified US norms for men using a *t*-test [21]. For the PCS, CaPSURE participants reported significantly higher mean scores for each age group when compared to US norms (all $p < 0.05$). The largest effect size (i.e., 0.47) was for the 55–64–year old age group. For the MCS, no differences in MCS scores were found between the youngest age group. However, significant differences in MCS scores were found between the 55–64 and the 65+ age groups, although the effects sizes were small (i.e., 0.22 to 0.30, respectively).

Discussion

This study is the first to evaluate a number of data quality parameters for a generic and a disease-specific PRO measures that were administered by mixed-mode (i.e., paper-mode and web-mode) in older age men diagnosed with prostate cancer. Overall response rates in this study were comparable to response rates of 50% to 70% of the general population [23, 35]. While paper-mode participants in this study had a significantly lower response than web-mode, both response rates were above the accepted average of 60%. These higher response rates are consistent with prior studies of cancer survivors who may be more likely to respond to surveys because of their personal interest in survivorship issues [36]. Additionally, the use of mixed-mode offering has been associated with an increase in overall response rates [37]. However, compared to the CaPSURE survey response rate for this administration (76% paper-mode and 88 % web-mode), the overall response rate did not increase dramatically (77%). This finding may be attributed to the relatively small number of participants who chose the web-mode of administration and the number of participants who switched back to paper-mode. Mode-switch included technical issues with navigating the web or human-computer interaction factors such as survey length. Request for mode-switch may increase as a person ages due to changes in cognitive, visual, and motor skills that impact an older person's ability to interact with computing technology.

The multivariable logistic model was able to predict the likelihood of being a respondent. This model that contained explanatory variables such as demographic characteristics (e.g., age, race, education, and income), treatment selection, prior survey participation, and mode selection has reasonable discrimination between those who

responded and those who did not. Findings from two mode-specific nested models suggest that the explanatory variables were more relevant in their predictive abilities when applied to the paper-mode group. The fact that younger age was associated with a higher risk of not being a respondent was an unanticipated finding given that Internet access is higher among younger persons [38]. The differential participation by younger men may be explained by the leverage-salience theory which posits that people differ in the importance that they attach to a request to participate in a survey [39]. Additional research would be required to understand the levers that are associated with the attenuation of participation by younger men.

As sample size increases the size of the differences that can be detected decreases. Prior PRO research has found that data analysis that incorporates large sample size (e.g., > 400 persons) can result in the detection of statistically significant differences of mean scale scores between groups which differ by a relatively small amount (e.g., ± 3 points on a 100–point scale) [31]. For both the SF-36 and the UCLA-PCI, mean scale scores were always higher for the web-mode participants. Differences in SF-36 mean scale scores for web-mode participants were largest for the PF, RP, RE, and VT scales (i.e., 6.4 to 9.5 points higher). For the UCLA-PCI, a similar trend was found with a difference in mean scale scores of 2 to 4 points with web-mode participants reporting higher mean scale scores. The only exception was for the sexual function and bother scales which were higher for the web-mode participants (i.e., 9 to 7 points, respectively). Post hoc calculation of an effect sizes were performed for each scale of the SF-36 and UCLA-PCI to assess clinical significance. It has been noted that few scales achieve the degree of reliability to detect meaningful change clinically (i,e., ±14 points on the SF-36) to make

decisions about clinical care at the individual level [34]. Only two scales on the SF-36, PF and MH, met the lower-bound standard of .90 reliability and none of the 8 scales meet the .95 degree of reliability required for decision making at the individual level [40].

Significant differences between mean scale scores were observed, with web-mode participants' generally reporting higher quality of life scores. However, the introduction of the web-mode did not appear to negatively affect overall data quality when formalized testing for minimal important differences was assessed. Rather the inclusion of mixed-mode, did allow for inclusion of a broader range of PRO scale scores, self-reported health status, and participation by older men.

Additionally, ceiling effects were observed on a number of scales in this study. Ceiling effects are problematic for individual-patient assessment for two reasons. First, measurement of change within longitudinal monitoring, such as improvement from a previous baseline, is impossible. Second, a false-negative finding may be reported by an instrument which lacks the precision to measure a full range of health states. A ceiling effect value implies that the person has perfect functioning. In reality, the instrument may only be able to measure or discriminate between the most severe case of dysfunction [34].

Using an anchor-based comparison, the CaPSURE sample did report higher PRO scores for the PCS and MCS component summary scores when compared to the national norms for men stratified by age. This suggests that the distribution of PRO scores were skewed and is consistent with other findings that those who continue to participate in PRO studies are generally the best performers in terms of outcomes [41].

This study has several strengths. It was conducted among a sample of older men from multiple community-based urology practices throughout the US. Since 1999, the CaPSURE study has prospectively recruited men who are newly diagnosed with prostate cancer. Initially the study offered only a single paper-mode of PRO administration. The addition of a mixed-mode strategy suggests that providing both modes allows for a broader representation of participants based on important health characteristics such as age, health status, and treatment selection.

This study has a several limitations. The CaPSURE sample was derived from a consecutive sampling strategy of men newly diagnosed with prostate cancer for study recruitment within an observational registry. Men who declined to participate may differ systematically from those who participated in CaPSURE. This study did not evaluate factors related to computer literacy, Internet access, or regional variation in access to the Internet. Examination of participant zip codes could provide some additional insight into Internet access since variation between urban and rural locations still exist with regards to connectivity [38]. Participants in CaPSURE appear to have higher quality of life, were older, and more likely to be Caucasian. This limits the generalizability of our findings since men who are non-Caucasian, have less education, and lower incomes were under represented in the CaPSURE. The study did not control for any confounding effects such as second treatments which may impact PRO scores.

Conclusions

This study found significant differences in data quality by mode of administration in response rates, missing data, and mean scale scores between paper-mode and web-mode participants. These differences did not rise to the level of meaningful difference

but does illustrate that the incorporation of both modes capturers a broader range of PRO levels, health status, and participation by older participants. This study extends the work by McHorney and colleagues and contributes to the literature by providing a contemporary comparison of the effects of mixed-mode on data quality by including comparisons of more traditional paper-mode administration to a newer Internet web-mode technology when the SF-36 was used within an observational longitudinal disease-specific registry [12]. No prior studies have evaluated mixed-mode PRO administration and data quality with the UCLA-PCI.

Table 1. CaPSURE Demographic, Health Status, and Clinical Characteristics by Mode and Response Category ((N = 4,836)

	Paper-mode N = 4,376		Web-mode N = 460		*P*-value†	*P*-value‡
Study Characteristic	Respond	Non-Respond	Respond	Non-Respond	**Mode**	**Response**
Column Number	(1)	(2)	(3)	(4)		
Age at Questionnaire	N = 3305 (SD)	N = 1071 (SD)	N = 405 (SD)	N = 55 (SD)		
Age Categorical	n (%)	n (%)	n (%)	n (%)		
41-64 years old	835 (25)	397 (37)	167 (41)	23 (42)	Responded	Web
65-74 years old	1318 (40)	372 (35)	154 (38)	18 (33)	**<0.001**	0.82
75-84 years old	980 (30)	248 (23)	79 (20)	13 (24)	Nonresponse	Paper
85-92 years old	172 (5)	54 (5)	5 (1)	1 (2)	0.68	**<0.001**
Mean Age (*years*)	70.7 (8.8)	68.4 (9.7)	66.8 (8.5)	67.4 (9.0)		
Race						
Caucasian	3041 (92)	845 (79)	388 (96)	52 (95)	Responded	Web
African-American	180 (5)	181 (17)	10 (2)	1 (2)	**0.02**	0.61
Other	78 (2)	40 (4)	7 (2)	2 (4)	Nonresponse	Paper
Unknown	6	5	0	0	**0.01**	**<0.001**
Education						
< High school diploma	279 (9)	100 (14)	6 (2)	0 (0)	Responded	Web
High school graduate	752 (25)	203(29)	44 (11)	6 (12)	**<0.001**	0.80
Some college	626 (21)	136 (19)	74 (19)	11 (22)	Nonresponse	Paper
College graduate	1352 (45)	267 (38)	268 (68)	34 (67)	**<0.001**	**<0.001**
Unknown	296	365	13	4		
Relationship status						
Living w/ spouse or partner	2628 (88)	594 (85)	348 (90)	43 (86)	Responded	Web
In a significant relationship	93 (3)	37 (5)	15 (4)	7 (14)	**0.003**	N/A*
No significant relationship	154 (5)	37 (5)	25 (6)	0 (0)	Nonresponse	Paper
Widowed	96 (3)	29 (4)	0 (0)	0 (0)	**0.01**	**0.02**
Unknown	334	374	17	5		
Annual income						
<$50,000	1402 (51)	321 (51)	103 (28)	16 (33)	Responded	Web
$50,000-75,000	576 (21)	141 (23)	81 (22)	12 (24)	**<0.001**	0.63
>$75,000	760 (28)	164 (26)	185 (50)	21 (43)	Nonresponse	Paper
Unknown	567	445	36	6	**0.02**	0.62
Health insurance status						
Medicare supplement	1154 (35)	266 (25)	80 (20)	12 (22)	Responded	Web
Medicare only	431 (13)	172 (16)	29 (7)	5 (9)	**<0.001**	N/A*
HMO/PPO	1279 (39)	455 (42)	231 (57)	30 (55)	Nonresponse	Paper
Fee for service	240 (7)	54 (5)	40 (10)	6 (11)	0.11	**<.0001**
Other	112 (3)	78 (7)	18 (4)	2 (4)		
VA	80 (2)	34 (3)	7 (2)	0 (0)		
Medicaid	9 (<1)	12 (1)	0 (0)	0 (0)		
Prostate Cancer Risk Group [42]						
Low	1382 (45)	425 (43)	191 (51)	28 (51)	Responded	Web
Intermediate	1033 (34)	316 (32)	129 (35)	16 (29)	**0.003**	0.47
High	662 (22)	251 (25)	53 (14)	11 (20)	Nonresponse	Paper
Number Unknown	228	79	32	0	0.47	**0.05**
Health Status						
Excellent	518 (16)	0	103 (25)	0	Responded	Web
Very Good	1357 (42)	0	177 (44)	0	**<0.001**	N/A*
Good	906 (28)	0	93 (23)	0	Nonresponse	Paper
Fair	387 (12)	0	30 (7)	0	N/A*	N/A*
Poor	66 (2)	0	2 (< 1)	0		
Number Unknown	71	1071	0	55		
Primary Treatment						
Radical Prostatectomy	1840 (57)	490 (49)	294 (75)	34 (64)	Responded	Web
Other	143 (4)	50 (5)	10 (3)	2 (4)	**<0.001**	**0.08**
Radiation	782 (24)	235 (24)	58 (15)	15 (28)	Nonresponse	Paper
Hormones	290 (9)	151 (15)	15 (4)	2 (4)	**0.02**	**<0.001**
Watchful Waiting	151 (5)	74 (7)	15 (4)	0		
Number Unknown	99	71	13	2		

† Post hoc Pearson Chi-square *P*-values between **mode** within either responder or non-responder (i.e., responders comparisons are between columns 1 and 3; non-responders comparisons are between columns 2 and 4).

‡ Post hoc Pearson Chi-square *P*-values between **responder and non-responder** within either web-mode or paper-mode (i.e., comparisons are between paper-mode columns 1 and 2; web-mode comparisons are between columns 3 and 4).

* Not Applicable (NA) - Post-hoc comparisons not calculated due to small cell size or no records in the category.

Table 2. Multivariate Logistic Regression Predicting the Likelihood of PRO Response (N = 4,836)

Parameter		Coefficient (β)	Standard Error	Wald χ^2	OR† (95% CI)	*p*-value*
Intercept		1.53	.18	8.65		.000
Age	41–64–years	-.24	.09	-2.72	.79 (.67 - .94)	**.007**
	65–92–years				1.0	
Race	Other	-.69	.11	-6.30	.50 (.41 - .62)	**.000**
	Caucasian				1.0	
Education	No College	-.28	.09	-3.21	.75 (.64 - .90)	**.001**
	College				1.0	
Education	Unknown	-1.18	.16	-7.48	.31 (.23 - .42)	**.000**
	College				1.0	
Relationship†	Living alone	-.59	.20	-3.02	.55(.37 - .81)	**.003**
Living w/ spouse or partner					1.0	
Relationship‡	Other	-.05	.14	-.35	.95 (.73 – 1.25)	.73
Living w/ spouse or partner					1.0	
Treatment	Other	-.26	.08	-3.14	.77 (.65 - .91)	**.002**
Radical Prostatectomy					1.0	
Treatment	Unknown	-.67	.18	-3.75	.51 (.36 - .73)	**.000**
Radical Prostatectomy					1.0	
Previous Questionnaire Count (Range 0 to 28)		.10	.01	12.20	1.10 (1.08 – 1.12)	**.000**
Mode	Paper	-.37	.16	-2.38	.69 (.68 – .07)	**.02**
	Web				1.0	

† OR is Odds Ratio

* Probability of Wald χ^2 statistic significant at $p \leq 0.05$.

The concordance index or *C*-statistic for this model was 0.736

‡Relationship status defined as "**In**" is defined as in a significant relationship buy not living with person; "**living w**/" is defined as living with spouse or partner; or "**Othe**r" is defined as widowed or unknown relationship status.

Table 3. Proportion of Respondents with Complete Short Form-36 and UCLA-Prostate Cancer Index Items and Computable Scale Scores by Mode

SF-36 Scales	Items	% with Complete Items			% with Computable Scale Scores		
		Paper	Web	*p*-value†	Paper	Web	*p*-value†
Physical Functioning	10	96.4	98.5	**0.03**	99.3	99.7	0.50
Role Physical	4	98.7	99.0	1.00‡	99.3	100.0	0.16
Bodily Pain	2	97.9	99.5	**0.03**	97.9	99.5	**0.03**
General Health	5	95.7	99.2	**0.00**	97.9	100.0	**0.00**
Vitality	4	96.6	97.8	0.44	97.9	100.0	**0.00**
Social Functioning	2	92.3	99.5	**0.00**	92.3	99.5	**0.00**
Role Emotional	3	99.2	99.0	0.55‡	99.3	100.0	0.16
Mental Health	5	96.6	99.2	**0.00**	98.0	100.0	**0.00**
Health Transition	1	97.8	100.0	**0.00**	97.8	100.0	**0.00**
UCLA-PCI Scales							
Urinary function	5	88.6	95.5	**0.00**	97.7	100.0	**0.00**
Urinary bother	1	98.9	99.7	0.17‡	98.9	99.7	0.17‡
Sexual function	8	90.4	95.3	**0.00**	96.4	99.5	**0.00**
Sexual bother	1	94.5	99.0	**0.00**	94.5	99.0	**0.00**
Bowel function	4	98.7	99.7	0.08‡	99.2	100.0	0.07‡
Bowel bother	1	99.0	100.0	**0.04**‡	99.0	100.0	**0.04**

† Fisher's Exact Test two sided P-value significant at ≤0.05

‡ 25% of the cells have expected counts less than 5 for the web-mode. Chi-Square may not be a valid test.

Table 4. Mean Score, Missing Data, and Ceiling and Floor Effects by Survey Mode - All SF-36 Domains and Composite Scores

Scale/ Score	Mode	N*	Mean Scale Score	SD	Minimum Floor 0-100 (%)	Maximum Ceiling 0-100 (%)	*t*-test	*p*-value	Effect Size*
PCS†	Paper	2938	47.5	10.2	7.2 (.03)	68.6 (.07)	5.15	**<0.001**	.30
PCS	Web	400	50.2	8.9	17.6 (.25)	61.4 (.25)			
MCS‡	Paper	2938	54.3	8.2	15.3 (.03)	72.3 (.03)	2.21	**.03**	.12
MCS	Web	400	55.3	7.1	23.9 (.25)	68.3 (.25)			
PF	Paper	3284	80.2	23.1	0 (.52)	100 (26)	6.77	**<0.001**	.36
PF	Web	404	88.2	17.2	0 (.50)	100 (41)			
RP	Paper	3285	71.7	39.4	0 (17)	100 (60)	4.66	**<0.001**	.23
RP	Web	405	81.2	34.1	0 (10)	100 (73)			
RE	Paper	3284	86.3	29.9	0 (8)	100 (80)	1.89	**<0.001**	.22
RE	Web	405	92.7	21.8	0 (3)	100 (88)			
VT	Paper	3238	64.8	20.2	0 (.37)	100 (2)	6.12	**<0.001**	.32
VT	Web	405	71.2	17.1	10 (.25)	100 (4)			
MH	Paper	3239	82.2	14.6	4 (.06)	100 (7)	2.25	**.02**	.11
MH	Web	405	83.9	12.9	28 (.49)	100 (7)			
SF	Paper	3051	88.0	19.9	0 (22)	100 (.16)	2.71	**0.001**	.15
SF	Web	403	90.9	17.7	0 (11)	100 (.25)			
BP	Paper	3238	79.2	21.4	0 (.15)	100 (31)	3.34	**.001**	.18
BP	Web	403	82.9	18.91	22.5 (.99)	100 (35)			
GH	Paper	3198	70.6	19.9	0 (.06)	100 (9)	3.53	**.001**	.19
GH	Web	405	74.3	18.8	15 (.25)	100 (11)			

PF = Physical Functioning; RP = Role=physical; RE = Role-Emotional; VT = Vitality; MH = Mental Health; SF = Social-Functioning; BP – Bodily Pain; GH = General Health

†Physical Component Score (PCS) summary measure of physical health
‡Mental Component Score (MCS) summary measure of mental health
* Sample size for paper-mode respondents was 3,305. Sample sized for web-mode respondents was 405.
**Effect size calculated using Cohen's D [30].

Table 5. Mean Score, Missing Data, and Ceiling and Floor Effects by Survey Mode - All UCLA-PCI Scales

Scale†	Mode	N*	Mean	SD	Minimum Floor 0-100 (%)	Maximum Ceiling 0-100 (%)	*t*-test	*p*-value	Effect Size**
UF	Paper	3229	78.9	22.5	0 (.25)	100 (34)	2.25	**.02**	.12
UF	Web	405	81.2	19.9	5 (.49)	100 (35)			
UB	Paper	3270	78.6	26.1	0 (2)	100 (49)	2.98	**.003**	.16
UB	Web	404	82.7	22.8	0 (.50)	100 (54)			
SF	Paper	3186	26.8	26.7	0 (22)	100 (.16)	6.36	**<.0001**	.34
SF	Web	403	35.8	28.0	0 (11)	100 (.25)			
SB	Paper	3125	44.3	39.1	0 (33)	100 (22)	3.52	**.0004**	.19
SB	Web	401	51.6	36.3	0 (19)	100 (25)			
BF	Paper	3277	87.3	14.2	10 (.03)	100 (25)	2.69	**.0072**	.14
BF	Web	405	89.3	12.4	37 (.49)	100 (29)			
BB	Paper	3274	86.3	22.6	0 (1)	100 (65)	1.94	**.05**	.10
BB	Web	405	88.6	21.4	0 (.49)	100 (72)			

† UF = Urinary Function; UB = Urinary Bother; SF = Sexual Function; SB = Sexual Bother; BF = Bowel Function; BB = Bowel Bother.

* Sample size for paper-mode respondents was 3,305. Sample size for web-mode respondents was 405.

**Effect size calculated using Cohen's D [30].

Table 6. Physical Component Summary Scores United States Norms for Males by Age Group Compared to CaPSURE

PCS	US Norms By Age	CaPSURE All N = 3,710	*p*-value/ ES*	CaPSURE Paper-mode N = 3,305	CaPSURE Web-mode N = 405
Age 40-54	n = 145	n = 136		n = 119	n = 28
Mean	50.40	53.61	.002	53.65	53.44
25th Percentile	48.14	52.70	ES .36	52.67	52.78
50th Percentile	53.36	55.83		56.05	55.62
75th Percentile	56.13	57.62		57.87	56.76
SD	9.68	7.74		8.09	6.21
Range	13-67	23-69			
Age 55-64	n = 105	n = 799		n = 716	n = 139
Mean	46.90	51.19	.000	50.76	53.25
25th Percentile	40.57	48.94	ES .47	48.43	51.01
50th Percentile	49.50	54.11		53.88	55.34
75th Percentile	54.99	56.90		56.68	57.41
SD	10.82	8.70		9.05	6.45
Range	16-58	14-67			
Age 65+	n = 293	n = 2,403		n = 2,470	n = 238
Mean	41.95	46.34	.000	46.15	48.11
25th Percentile	33.48	39.28	ES .42	39.10	42.02
50th Percentile	43.84	49.26		49.03	51.70
75th Percentile	51.64	54.70		54.59	55.44
SD	11.35	10.26		10.30	9.70
Range	9-59	7-69			

US norms for males by age derived from Appendix D, Table 5 pp. 188-189 [21].
*Effect size calculated using Cohen's D [30].

Table 7. Mental Component Summary Scores United States Norms for Males by Age Compared to CaPSURE

MCS	US Norms By Age	CaPSURE All N = 3,710	*p*-value / ES*	CaPSURE Paper-mode N = 3,305	CaPSURE Web-mode N = 405
Age 40-54	n = 145	n = 147		n = 119	n = 28
Mean	51.03	51.19	.89	50.82	52.72
25th Percentile	48.10	47.16	ES n/a	46.05	52.43
50th Percentile	53.94	54.85		54.06	55.66
75th Percentile	57.51	57.92		57.84	58.84
SD	9.86	10.11		10.19	9.83
Range	17-67	20-68			
Age 55-64	n = 105	n = 855		n = 716	n = 139
Mean	51.60	53.49	.04	53.29	54.46
25th Percentile	48.46	51.81	ES .22	51.49	52.87
50th Percentile	54.63	56.22		56.18	56.31
75th Percentile	57.58	58.75		58.72	58.89
SD	9.11	8.56		8.77	7.44
Range	25-63	15-69			
Age 65+	n = 293	n = 2,708		n = 2,470	n = 238
Mean	52.51	54.96	.000	54.84	56.07
25th Percentile	47.95	52.45	ES .30	52.28	53.53
50th Percentile	54.83	57.30		57.29	57.57
75th Percentile	59.44	59.81		59.79	59.85
SD	9.78	7.76		7.89	6.37
Range	19-74	19-72			

US norms for males by age derived from Appendix D, Table 5 pp. 188-189 [21].
*Effect size calculated using Cohen's D [30].
n/a is not applicable

References

1. Greenhalgh, J., *The applications of PROs in clinical practice: what are they, do they work, and why?* Quality of Life Research, 2009. **18**(1): p. 115-23.
2. Lipscomb, J., et al., *Patient-reported outcomes assessment in cancer trials: taking stock, moving forward.* Journal of Clinical Oncology, 2007. **25**(32): p. 5133-40.
3. Lohr, K.N. and B.J. Zebrack, *Using patient-reported outcomes in clinical practice: challenges and opportunities.* Quality of Life Research, 2009. **18**(1): p. 99-107.
4. Rothman, M.L., et al., *Patient-reported outcomes: conceptual issues.* Value in Health, 2007. **10 Suppl 2**: p. S66-75.
5. Bowling, A., *Mode of questionnaire administration can have serious effects on data quality.* Journal of Public Health 2005. **27**(3): p. 281-91.
6. deLeeuw, E.D., *To mix or not to mix data collection modes in surveys.* Journal of Official Statistics, 2005. **21**(2): p. 233-255.
7. Abernethy, A.P., et al., *Improving health care efficiency and quality using tablet personal computers to collect research-quality, patient-reported data.* Health Services Research, 2008. **43**(6): p. 1975-91.
8. Basch, E., et al., *Evaluation of an online platform for cancer patient self-reporting of chemotherapy toxicities.* Journal of the American Medical Informatics Association, 2007. **14**(3): p. 264-8.
9. Dupont, A., et al., *Use of tablet personal computers for sensitive patient-reported information.* Journal of Supportive Oncology, 2009. **7**(3): p. 91-7.

10. Farnell, D.J., et al., *Efficacy of data capture for patient-reported toxicity following radiotherapy for prostate or cervical cancer.* European Journal of Cancer, 2010. **46**(3): p. 534-40.

11. Velikova, G., et al., *Automated collection of quality-of-life data: a comparison of paper and computer touch-screen questionnaires.* Journal of Clinical Oncology, 1999. **17**(3): p. 998-1007.

12. McHorney, C.A., M. Kosinski, and J.E. Ware, Jr., *Comparisons of the costs and quality of norms for the SF-36 health survey collected by mail versus telephone interview: results from a national survey.* Medical Care, 1994. **32**(6): p. 551-67.

13. Ware, J.E., Jr. and C.D. Sherbourne, *The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection.* Medical Care, 1992. **30**(6): p. 473-83.

14. Lipscomb, J., et al., *Cancer Outcomes Research.* Journal of the National Cancer Institute Monographs., 2004. **2004**(33): p. 178-197.

15. Lipscomb, J., M. Donaldson, and R. Hiatt, *Cancer Outcomes Research and the Arenas of Application.* Journal of the National Cancer Institute Monographs, 2004. **2004**(33): p. 1-7.

16. Lipscomb, J., C.C. Gotay, and C.F. Snyder, *Patient-reported outcomes in cancer: a review of recent research and policy initiatives.* CA: A Cancer Journal for Clinicians, 2007. **57**(5): p. 278-300.

17. Lipscomb, J., C. Snyder, and C. Gotay, *Cancer outcomes measurement: Through the lens of the Medical Outcomes Trust framework.* Quality of Life Research, 2007. **16**(1): p. 143-64.

18. Hays, R.D., C.D. Sherbourne, and R.M. Mazel, *The RAND 36-Item Health Survey 1.0.* Health Economics, 1993. **2**(3): p. 217-27.

19. Litwin, M.S., et al., *The UCLA Prostate Cancer Index: development, reliability, and validity of a health-related quality of life measure.* Medical Care, 1998. **36**(7): p. 1002-12.

20. Litwin, M.S., et al., *Quality-of-life outcomes in men treated for localized prostate cancer.* Journal of the American Medical Association, 1995. **273**(2): p. 129-35.

21. Ware, J.E., Jr. and M. Kosinski, *SF-36® physical and mental health summary scales: A manual for users of version 1.0*. 2nd ed. 2001, Lincoln, RI. : QualityMetric Incorporated.

22. Lubeck, D.P., et al., *The CaPSURE database: a methodology for clinical practice and research in prostate cancer. CaPSURE Research Panel. Cancer of the Prostate Strategic Urologic Research Endeavor.* Urology, 1996. **48**(5): p. 773-7.

23. Dillman, D., J. Smyth, and L. Christian, *Internet, mail and mixed-mode surveys: The tailored design method*. 3rd ed. 2009, Hoboken, NJ: John Wiley & Sons, Inc.

24. AAPOR, *Standard definitions: Final dispositions of case codes and outcome rates for surveys*. 2009, American Association of Public Opinion Research (AAPOR): Ann Arbor, MI.

25. *TeleForm® Version 9.2*. Autonomy CARDIFF 2011 [cited 2011 July 30]; Available from: http://www.cardiff.com/products/teleform/.

26. Broering, J.M., et al., *CaPSURE.net e-Tools: Comprehensive business process management for research questionnaire adminstration [abstract]*, in *National*

Cancer Institute, Critical Issues in e-Health Research Conference. 2005: Bethesda, MD.

27. Broering, J.M., et al., *Validation of an Internet based patient health-related quality of life questionnaire: Data from CaPSURE [abstract]*. Value in Health, 2006. **9**(3): p. A114.

28. Hermansen, S.W. *Evaluating predictive models: Computing and interpreting the c statistic*. SAS Global Forum 2008 [cited 2012 May 1]; Paper 143-2008:[1-9]. Available from: http://www2.sas.com/proceedings/forum2008/143-2008.pdf

29. Rosner, B., *Fisher's Exact Test*, in *Fundamentals of Biostatistics, 6th Ed.* 2006, Thomson Higher Education Belmont, CA. p. 402-407.

30. Cohen, J., *Calculation of effect size*, in *Statistical power analysis for the behavioral sciences*, 2nd, Editor. 1988, Lawrence Erlbaum Associates: Hillsdale, New Jersey. p. 24-27.

31. Sloan, J.A., et al., *The clinical significance of quality of life assessments in oncology: a summary for clinicians.* Supportive Care in Cancer, 2006. **14**(10): p. 988-98.

32. Norman, G.R., J.A. Sloan, and K.W. Wyrwich, *Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation.* Medical Care, 2003. **41**(5): p. 582-92.

33. Wyrwich, K.W., et al., *Estimating clinically significant differences in quality of life outcomes.* Quality of Life Research, 2005. **14**(2): p. 285-95.

34. McHorney, C.A. and A.R. Tarlov, *Individual-patient monitoring in clinical practice: are available health status surveys adequate?* Quality of Life Research, 1995. **4**(4): p. 293-307.

35. Dillman, D.A., *The design and administration of mailed surveys.* Am Rev Sociol, 1991. **17**: p. 225-249.

36. Edwards, P., et al., *Increasing response rates to postal questionnaires: systematic review.* BMJ, 2002. **324**(7347): p. 1183.

37. Greenlaw, C. and S. Brown-Welty, *A comparison of web-based and paper-based survey methods: testing assumptions of survey mode and response cost.* Evaluation Review, 2009. **33**(5): p. 464-80.

38. Jones, S. and S. Fox. *Generations online in 2009*. 2009 November 22, 2009]; Available from: http://www.pewinternet.org/Reports/2009/Generations-Online-in-2009.aspx.

39. Groves, R.M., E. Singer, and A. Corning, *Leverage-saliency theory of survey participation: description and an illustration.* Public Opinion Quarterly, 2000. **64**(3): p. 299-308.

40. McHorney, C.A., et al., *The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups.* Medical Care, 1994. **32**(1): p. 40-66.

41. Sloan, J.A. and A. Dueck, *Issues for statisticians in conducting analyses and translating results for quality of life end points in clinical trials.* Journal of Biopharmaceutical Statistics, 2004. **14**(1): p. 73-96.

42. D'Amico, A.V., et al., *Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer*. Journal of the American Medical Association, 1998. **280**(11): p. 969-74.

Chapter 5

Conclusions

The overall purpose of this dissertation research was to examine phenomenon associated with mixed-mode administration of patient reported outcome (PRO) surveys for measurement equivalence, cost, and data quality when surveys were administered using paper-mode or web-mode. Most PRO instruments were originally developed and psychometrically evaluated for administration by a paper-mode and later adapted to web-mode. Systematic evaluation of the reliability and validity of PRO instruments is recommended when moderate to significant changes are introduced with the web-mode offering (Coons, Gwaltney, Hays, et al., 2009).

Few studies have examined measurement equivalence within oncology (Abernethy, Herndon, Wheeler, et al., 2008; Basch, Artz, Iasonos, et al., 2007; Dupont, Wheeler, Herndon, et al., 2009; Farnell,Routledge, Hannon, et al., 2010; Velikova, Wright, Smith, et al., 1999). The purposes of these studies were to pilot test the feasibility of data capture using an e-Tablet (Abernethy et al., 2008; Farnell et al., 2010) or to assess the technology platform for feasibility and acceptability within clinical care settings (Basch et al., 2007; Velikova et al., 1999). All of these mixed-mode PRO studies were performed at a single academic institution; included relatively small sample (i.e., 56 to 149 participants); and were conducted within a clinical environment where technical assistance was readily available (Abernethy et al., 2008; Basch et al., 2007; Dupont et al., 2009; Farnell et al., 2010; Velikova et al., 1999).

A meta-analysis that assessed measurement equivalence between paper-and-pencil and electronic modes of administration demonstrated small differences that were

not statistically or clinically significant (Gwaltney, Shields, & Shiffman, 2008). However, this meta-analysis included few studies of older persons who are less likely to have computer literacy; and are known to have lower rates of Internet access which vary by geographic region (i.e., urban versus rural), ethnicity, personal income, and employment status (Jones & Fox, 2009; Strickling & Gomez, 2010). Additionally, older persons may require special accommodations to support their use of computerized devices (e.g., augmentation to the visual screen display).

The *Total Survey Error Paradigm* theoretical model (Groves, Fowler, Couper, et al., 2009) was used to guide the analytic approaches in the three data-based papers in this dissertation. Specifically, measurement issues that were addressed included: sample coverage, non-response error, and data quality metrics. Published recommendations provided guidance for the study design and statistical analyses that were used to evaluate measurement equivalence between electronic and paper-based PRO measures (Coons et al., 2009). The first data-based paper in this dissertation describes the findings from a cross-over design study that evaluated measurement equivalence among a convenience sample of 209 men with prostate cancer enrolled in the Cancer of the Prostate Strategic Urologic Research Endeavor (CaPSURE), a multi-institutional prostate cancer registry (Lubeck, Litwin, Henning, et al., 1996). Once measurement equivalence was established, a logical extension of inquiry, were research questions related to the effects of providing a mixed-mode offering within the full registry focusing on cost and data quality. The second data-based paper reports finding on the total costs, the cost per survey, the cost per response, and the cost per error-free response to administer a mixed-mode survey offering within a cross-sectional timeframe of a semi-annual PRO administration to 5,088

participants in CaPSURE. Conceptual guidance was derived from the work by Martha Gold and colleagues on *Cost-effectiveness in Health and Medicine* (Gold, Siegel, Russel, & Weinstein, 1996) for the design and framing of the costs and cost-effectiveness analysis to reflect the complexity of the workflow process required to support both modes for a cross-sectional administration (Torrance, Siegel, & Luce, 1996). The third data-based paper reports on a comparison of data quality metrics by mode for the same cross-sectional timeframe on 4,836 eligible participants. This third paper reports on a subset of participants from the cross-sectional timeframe since deceased participants and those PRO surveys reported by a proxy were excluded. Prior research, that assessed data quality between mixed-mode administration of the Medical Outcomes Study-Short Form (SF-36) health survey comparing paper mailed and telephone administered modes, provided guidance on the metrics to assess data quality and with the creation of the analysis plan (McHorney, Kosinski, & Ware, 1994). Findings from this dissertation are summarized by each of these specific areas of inquiry. Relevant clinical and research implications are discussed for each of these areas of mixed-mode administration.

Summary of Findings and Clinical Implications

Measurement equivalence

Comparison of test-retest data using the intraclass correlation coefficients (ICC) that compared paper-mode to web-mode on all 56–items in the SF-36 (Hays, Sherbourne, & Mazel, 1993) and University of California Los Angeles-Prostate Cancer Index (UCLA-PCI) (Litwin, Hays, Fink, et al., 1998; Litwin, Hays, Fink, et al., 1995) were for the most part above the standard cut point of 0.75 for excellent correlation. Significant differences were observed when mean scale scores were compared by mode for the mental

component summary (MCS) and role emotional scales. However, the effect sizes for these differences were small (<.20). No differences were observed when mean scale scores were compared by mode with the UCLA-PCI scale scores. A linear mixed-methods approach revealed ordering effects for the physical component summary (PCS) score and the physical function scale. Interactions between mode and order of administration were observed with the MCS score and the role emotional, social function, vitality, and mental health scale scores. No mode, ordering, or the interaction of mode and order of administration effects were noted for the UCLA-PCI instrument. Paper-mode participants had more missing data which suggests differential data quality issues by mode. The majority of participants (70%) in the measurement equivalence study expressed a strong preference for the web-mode while the remainder of participants was agnostic about mode of administration or continued to prefer paper-mode (9%).

This study added to the body of scientific literature in that it included an older sample of men with prostate cancer. The study incorporated a rigorous study design and analytic methods (i.e., an a priori power calculation, randomized cross-over design, and ICC with the linear mixed methods modeling). This study illustrated the highly motivated and selective participant characteristics that have been describe elsewhere as the digital divide effect (Jones & Fox, 2009; Strickling & Gomez, 2010).

Cost of Mixed-mode

A time and motion study was conducted to assist in building the cost model equations for the decision analysis. The cost model used tasks associated with pre-processing, processing, and post-processing of the PRO survey. Costs were defined as fixed (i.e., the cost of software and web hosting) or variable (i.e., the cost associated with

producing a single survey). The model assumed that not all surveys required each step as outlined in the processing model. Costs were framed in 2006 United States (US) dollars and assumed maximal efficiency of the data coordinating center staff at the University of California, San Francisco (UCSF). The response rates for the paper-mode survey were significantly lower compared to the web-mode (76% vs. 88%, respectively). Improved data quality was noted for the web-mode by examination of the double data entry (DDE) error rate (9% vs. 14%, respectively). However, these differences in DDE error rates were not statistically significant. The total cost to process this cross-sectional timeframe was $75,216 with the average cost per survey calculated to be $15.02. Costs per paper-mode were lower at $14.66/survey and web-mode costs were higher at $18.47/survey. The efficiency of web-mode surveys depends on several factors such as the proportion of participants that opt to use web-mode, the potential for higher response rates, and fewer errors. A cost-effectiveness model found that the web-mode became cheaper as the volume increased and becomes a higher proportion of the survey mix. Although web-mode has to potential to be more cost-efficient, participants' often required both methods to mitigate coverage bias and to accommodate personal preferences or circumstances.

Implementation of electronic modes for data capture is often adopted within clinical care settings without a full examination of the cost drivers associated with implementing such a system. The costs to support a mixed-mode approach for the administration of PRO instruments could be more expensive. These costs need to be balanced against the potential to improve coverage for the target population or to alter the cancer care treatment plan. The business process model used in the cost study could be

adopted, modified, and replicated to allow clinical care environments to gain a better understanding of the cost to provide a mixed-mode approach.

Data Quality and Mixed-mode

Examination of data quality between paper-mode and web-mode were examined using aggregate group data derived from the CaPSURE sample. A number of metrics were used to assess data quality. Specifically, response rates, modeling to describe predictors of response, data completeness, and differences in mean scale scores, ceiling and floor effects observed within scales, and comparison of mean scale scores for the PCS and MCS scores reported by CaPSURE participants against US norms for men stratified by age. Paper-mode participants had lower a response rate, were older, had more missing data, and reported significantly lower mean scale scores for all scales. However, a comparison of effect sizes, suggested that these differences were small.

A total of 127 participants were excluded from the data quality analysis due to proxy administration. The majority of those excluded (98%) had opted for paper-mode. An inherent assumption with this select sub-group of paper-mode participants was that they were too ill to self-administer the survey. The provision of a paper-mode offering within a clinical care or clinical research setting allows for participation by those persons, indirectly through their proxy, without the constraints of monitoring a personal e-mail account, or knowledge of any personal security password. For this group, the paper-mode facilitated being a respondent. Additionally, migration to other electronic modes such as smart phones, e-tablets, or talking e-tablets may require usability testing as well testing for psychometric equivalence (Nielsen, 1993).

Implications for Research

Measurement Equivalence

The measurement equivalence study evaluated the psychometric equivalence that was restricted to two PRO instruments contained within the CaPSURE survey using a distinct set of assumptions. The specific assumptions for these mixed-mode studies included 1) self-administration of the survey; 2) conducted outside of a clinical setting; 3) using a personal computer or paper-and-pencil and; 4) technical assistance that was available through a toll-free telephone support and/or by e-mail correspondence with the data coordinating center at UCSF.

The 56–items that were included in the SF-36 and the UCLA-PCI contained only structured Likert scale choice field options. Additional analyses could be performed to determine if measurement equivalence is observed for the remainder of the items and scales contained in the semi-annual CaPSURE survey, such as the satisfaction with health care and the health care utilization portions of the survey. These portions of the survey were not analyzed and may demonstrate more variability in the test-retest comparison since they required reporting in a number of open text fields with written responses. Findings from this measurement equivalence study cannot be generalized to other situations if there are moderate or substantial modifications made to an instrument (Coons et al., 2009). For example, the use of a mobile smart phone application or of a tablet device which may alter how items are presented (e.g., a single item presentation versus multiple items, the demand for scrolling up and down the screen, or changes to the font size) may require test-retest evaluation of measurement equivalence.

The methods used in this measurement equivalence study addressed only psychometric equivalence within a predominately monocultural group. Validation of measurement equivalence with mode of administration for these instruments may vary when they are administered to other minority and cultural groups. The need to incorporate additional measurement techniques and conceptual models to assess shared meaning across diverse cultural groups has the ability to extend the relevance of the use of PRO instruments, delivered by a variety of electronic modes within a more culturally diverse group of prostate cancer survivors (Johnson, 2006).

Cost of Mixed-mode

No studies have evaluated the cost of deploying a mixed-mode approach to administer PRO instruments within a large, longitudinal, disease-specific registry. Cost estimates in this study were derived from a primary data source and provide for benchmark cost estimates that are anchored in 2006 US dollars to support mixed-mode administration. Future research could include cost trends over time. Using the current study parameters and applying an estimate for inflation or deflation of costs with the fixed and variable costs in 2012 US dollars, new cost estimates could be calculated and compared against the historical benchmark.

The cost study provides a model for identifying costs in the pre-processing, processing, and post-processing phases of survey administration. In the new study design proposal phase, researchers will be confronted with selecting the mode of administration which provides for the inclusion of the target population (coverage), balanced with proposed response rate, and most efficient use of resources (study costs) to collect PRO data. The model used for identifying areas of cost could be adapted and applied to

costing mixed-mode approaches where emerging electronic technologies are employed. Finally, this study did not perform a formal sensitivity analyses which could evaluate the range of uncertainty for each univariate predictor in the model (Manning, Fryback, & Weinstein, 1996).

Data Quality and Mixed-mode

Areas for future research that emerged from the data quality analysis relate to participant age, the impact of proxy administration, and the direct affect that web-mode administration has on data quality. Research issues associated with each of these factors are discussed below.

Men who were in the younger age category were less likely to complete a PRO survey regardless of mode. Different theoretical approaches have been suggested to examine survey participation. Leverage-salience theory posits that people vary in the importance that they assign to a request to respond to a survey (i.e., topical relevance, incentive awards, or sponsor credibility) (Groves, Presser, & Dipko, 2004; Groves, Singer, & Corning, 2000). Additional research is required to better understand non-response among this group of younger men.

Historically, the CaPSURE study has collected information about the proxy administration of the PRO survey. Three percent of the analytic cross-sectional sample reported assistance with survey completion by a family member or close associate. The frequency of proxy administration was extraordinarily low by the web-mode participants (n=2). Therefore, any analysis would, in all likelihood, be restricted to an examination of proxy administration for the paper-mode. No analyses of proxy reported data has ever

been performed within CaPSURE. Future analysis could evaluate this group for variations in data quality.

Systematic evaluation of data quality by mode suggests that administration by web-mode resulted in superior data quality as measured by the higher response rate and more data completeness. However, an alternate explanation for improved data quality could be the favorable characteristics of those participants. Since CaPSURE is a longitudinal study and all of the current web-mode participants in all likelihood completed a paper-mode PRO, a comparison of data quality across time periods could be performed using this group of web-mode participants to examine this hypothesis of mode effects versus favorable participant characteristic effects.

Conclusions

Analytic results were presented in three data-based papers that examined the effects of mixed-mode administration for PRO surveys on measurement equivalence, cost, and data quality. Data from the measurement equivalence study supported psychometric equivalence for the SF-36 and the UCLA-PCI components of the CaPSURE survey. Costs associated with mixed-mode administration were lower per paper-mode survey and higher for the web-mode survey due to the low proportion (10%) that chose to adopt the web-mode. Response rates in the cross-sectional examination of a single CaPSURE administration phase were significantly lower for the paper-mode when compared to the web-mode participants. Different participant characteristics were observed between the two modes. Variation in data quality was observed by mode of PRO survey administration. Paper-mode participants had more missing data in both the measurement equivalence and the data quality studies. Mixed-mode administration was

not able to fully mitigate coverage bias issues related to age, ethnic diversity, and poorer health status.

References

Abernethy, A. P., Herndon, J. E., 2nd, Wheeler, J. L., Patwardhan, M., Shaw, H., Lyerly, H. K., & Weinfurt, K. (2008). Improving health care efficiency and quality using tablet personal computers to collect research-quality, patient-reported data. *Health Services Research, 43*, 1975-1991.

Basch, E., Artz, D., Iasonos, A., Speakman, J., Shannon, K., Lin, K., et al. (2007). Evaluation of an online platform for cancer patient self-reporting of chemotherapy toxicities. *Journal of the American Medical Informatics Association, 14*, 264-268.

Coons, S. J., Gwaltney, C. J., Hays, R. D., Lundy, J. J., Sloan, J. A., Revicki, D. A., et al. (2009). Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcomes (PRO) measures: ISPOR ePRO good research practices task force report. *Value in Health, 12*, 1-11.

Dupont, A., Wheeler, J., Herndon, J. E., 2nd, Coan, A., Zafar, S. Y., Hood, L., et al. (2009). Use of tablet personal computers for sensitive patient-reported information. *Journal of Supportive Oncology, 7*, 91-97.

Farnell, D. J., Routledge, J., Hannon, R., Logue, J. P., Cowan, R. A., Wylie, J. P., et al. (2010). Efficacy of data capture for patient-reported toxicity following radiotherapy for prostate or cervical cancer. *European Journal of Cancer, 46*(3), 534-540.

Gold, M. R., Siegel, J. E., Russel, L. B., & Weinstein, M. C. (1996). *Cost-effectiveness in health and medicine*. New York: Oxford University Press.

Groves, R. M., Fowler, J., F.J, Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). Hoboken, New Jersey: A. John Wiley & Sons, Inc.

Groves, R. M., Presser, S., & Dipko, S. (2004). The role of topic interest in survey participation decisions. *Public Opinion Quarterly, 68*, 2-31.

Groves, R. M., Singer, E., & Corning, A. (2000). Leverage-saliency theory of survey participation - Description and an illustration. *Public Opinion Quarterly, 64*, 299-308.

Gwaltney, C. J., Shields, A. L., & Shiffman, S. (2008). Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: a meta-analytic review. *Value in Health, 11*, 322-333.

Hays, R. D., Sherbourne, C. D., & Mazel, R. M. (1993). The RAND 36-Item Health Survey 1.0. *Health Economics, 2*, 217-227.

Johnson, T. P. (2006). Methods and frameworks for crosscultural measurement. *Medical Care, 44*(11 Suppl 3), S17-20.

Jones, S., & Fox, S. (2009). Generations online in 2009 Retrieved November 22, 2009, from http://www.pewinternet.org/Reports/2009/Generations-Online-in-2009.aspx .

Litwin, M. S., Hays, R. D., Fink, A., Ganz, P. A., Leake, B., & Brook, R. H. (1998). The UCLA Prostate Cancer Index: development, reliability, and validity of a health-related quality of life measure. *Medical Care, 36*, 1002-1012.

Litwin, M. S., Hays, R. D., Fink, A., Ganz, P. A., Leake, B., Leach, G. E., & Brook, R. H. (1995). Quality-of-life outcomes in men treated for localized prostate cancer. *Journal of the American Medical Association, 273*, 129-135.

Lubeck, D. P., Litwin, M. S., Henning, J. M., Stier, D. M., Mazonson, P., Fisk, R., & Carroll, P. R. (1996). The CaPSURE database: a methodology for clinical practice and research in prostate cancer. CaPSURE Research Panel. Cancer of the Prostate Strategic Urologic Research Endeavor. *Urology, 48*, 773-777.

Manning, W. G., Fryback, D. G., & Weinstein, M. C. (1996). Reflecting uncertainty in cost-effectiveness analysis. In M. R. Gold, J. E. Siegel, L. B. Russel & M. C. Weinstein (Eds.), Cost-effectiveness in health and medicine (pp. 176-213). New York: Oxford University Press.

McHorney, C. A., Kosinski, M., & Ware, J. E., Jr. (1994). Comparisons of the costs and quality of norms for the SF-36 health survey collected by mail versus telephone interview: results from a national survey. *Medical Care, 32*, 551-567.

Nielsen, J. (1993). *Usability engineering*. New York: Academic Press.

Strickling, L. E., & Gomez, A. (2010). Digital Nation: 21st Century America's Progress Toward Universal Broadband Internet Access. *An NTIA Research Preview* Retrieved June 1, 2010, from http://www.ntia.doc.gov/reports/2010/NTIA_internet_use_report_Feb2010.pdf .

Torrance, G. W., Siegel, J. E., & Luce, B. R. (1996). Framing and designing the cost-effectiveness analysis. In M. R. Gold, J. E. Siegel, L. B. Russell & W. C. Weinstein (Eds.), *Cost-effectiveness in health and medicine* (pp. 54-80). New York: Oxford University Press.

Velikova, G., Wright, E. P., Smith, A. B., Cull, A., Gould, A., Forman, D., et al. (1999). Automated collection of quality-of-life data: a comparison of paper and computer touch-screen questionnaires. *Journal of Clinical Oncology, 17*, 998-1007.

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Jeanette M. Burkey

Author Signature

7/16/2012

Date