(A)CASI and other self-reported data collection literature review: focused on cost-effectiveness, acceptability, cognitive testing and reliability.

Compiled by Jeffrey Schouten

October 11, 2012

Contents
Overview and Summary ........................................................................................................................................... 1
Review Articles/Recommendations ............................................................................................................................ 3
Cost Effectiveness ......................................................................................................................................................... 11
Acceptability/Social Desirability Bias Reduction ..................................................................................................... 13
Cognitive Testing/Development .................................................................................................................................. 19
Reliability/Data Quality ............................................................................................................................................... 22
References .................................................................................................................................................................... 30

Overview and Summary

Most of the literature on the use of (A)CASI and other self-reported data collection is focused on acceptability and reliability. Very little literature exists on the cost-effectiveness. There are two major review articles and one recommendations paper published in 2008 by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) electronic patient reported outcomes (ePRO) Good Research Practices Task Force. One review paper published in 2006 compared the effectiveness of hand-held computers versus paper methods. They reviewed nine randomized studies and concluded that “handhelds are an effective alternative to paper and pencil modes of data collection; they are faster and were preferred by most users.” A subsequent review paper published in 2010 was a systematic review and meta-analysis of quantitative interviewing tools to investigate self-reported HIV and STI associated behaviours in low- and middle-income countries. They observed that “differences between FTFI and non-interviewer-administered interview methods for the reported sensitive behaviour investigated were not uniform.” Additionally they “observed trends and variations in the level of reporting according to the outcome, study and population characteristics. FTFI may not always be inferior to innovative interview tools depending on the sensitivity of the question as well as the population assessed.” The key issues identified included 1) the determination of the extent of modification required to administer the PRO on the ePRO device and 2) the selection and implementation of an effective strategy for testing the measurement equivalence of the two modes of administration.

There are only a few articles evaluating the cost effectiveness of computer or web-based assisted data collection. One is a theoretical model which found high initial costs, which could be offset if the (A)CASI software were used in multiple studies. Another study evaluated the use of computer-assisted child mental health assessments and in the clinic setting found them to be less costly than a traditional clinical assessment. The third paper studied the cost of paper versus web-based collection. They concluded:
These findings highlight the advantages of using a web-based computerized data collection and management system. First, such a system ensures data integrity and compliance with Health Insurance Portability and Accountability Act of 1996 (HIPAA) regulations. Second, it increases the accuracy and reliability of the data by reducing the opportunities for human error. Third, as competition for funding continues to increase, strategies that maximize research dollars by reducing operating costs are becoming more important. Web-based computerized data collection and management save time in all phases of research studies. Initial startup costs associated with computerized data collection may be high, but with repeated use, large samples, or large data sets, computerized data collection becomes less expensive, more efficient, and more reliable than human data collection and management. Lastly, web-based computer data collection may reach research participants more effectively through availability 24 hour/day and 7 days/week and may reduce attrition and increase compliance with electronically delivered reminders and incentives.¹

In the general category of acceptability and/or social desirability bias reduction there are many more publications. While in general these reports found that (A)CASI resulted in increased reporting of sensitive data, but the results are not universally consistent. Many of the studies focused on adolescent populations and found (A)CASI to be an acceptable instrument. But there were differences noted between men and women, types of reported behavior, geography and educational level. Less sensitive data was not observed to have a difference in reporting by method of collection. Overall, most studies conclude that (A)CASI does reduce social desirability bias.

The next group of papers discuss cognitive testing/interviewing and other design parameters. “Researchers should consider cognitive interviewing when developing survey questionnaires to investigate new or poorly described health concepts, for researching and translating questionnaires for culturally diverse groups and when developing questionnaires for samples where questionnaire completion may pose particular problems.”²

Farnik and Pierzchala published a very helpful methodology paper for the development of patient reported outcome assessments. They conclude:

> For successful transferring of the concept of research to new instrument development and implementation, investigators should start with a strong definition of specific study objectives and follow a methodology of instrument development. The process of designing a new tool should involve a panel of experts, including clinicians, psychologists (preliminary phase), and statisticians (scale development, scoring), as well as patients (cognitive debriefing). Patient-related outcomes measures could provide important data for the current state of the art in medical care and even have an impact on macrodecisions.³

Several other papers argue that cognitive testing should be an important part of questionnaire design, in addition to the traditional standardizing and validation of questionnaires. A few other articles included discuss the issue of translation, questionnaire design and length. Simple word translation does not assure literal and cultural adaptation. Additional design issues include whether to include skip patterns and other checks for internal consistency. One study recommended surveys no longer than 20 minutes.

The last group of papers focus on the reliability of self-reported data collected by electronic means and the data quality. On October 2, 2012, AIDS and Behavior published a paper on the effect of ACASI in the MTN 035 trial from Gorbach at al.⁴ This paper highlights one of the issues noted in prior studies, that ACASI resulted in increased reporting of sensitive behaviors, but also results in more incomplete or inconsistent data as well. Electronic data capture though reduces the errors associated with data entry and allows in the design for increased accuracy, if errors are identified immediately and subjects have the opportunity to change their responses. One study did note that ACASI may not be appropriate for vulnerable older respondents in developing countries.⁵ Test-retest reliability on life time sexual history was found to reliable in a couple of studies in different populations.⁶ ⁷

This review is not exhaustive, but is representative of the range of publications and findings reported in the field. Below are the abstracts from the 49 articles collected in this compilation.
Review Articles/Recommendations

A review of randomized controlled trials comparing the effectiveness of hand held computers with paper methods for data collection

Background: Handheld computers are increasingly favoured over paper and pencil methods to capture data in clinical research.

Methods: This study systematically identified and reviewed randomized controlled trials (RCTs) that compared the two methods for self-recording and reporting data, and where at least one of the following outcomes was assessed: data accuracy; timeliness of data capture; and adherence to protocols for data collection.

Results: A comprehensive key word search of NLM Gateway’s database yielded 9 studies fitting the criteria for inclusion. Data extraction was performed and checked by two of the authors. None of the studies included all outcomes. The results overall, favor handheld computers over paper and pencil for data collection among study participants but the data are not uniform for the different outcomes. Handheld computers appear superior in timeliness of receipt and data handling (four of four studies) and are preferred by most subjects (three of four studies). On the other hand, only one of the trials adequately compared adherence to instructions for recording and submission of data (handheld computers were superior), and comparisons of accuracy were inconsistent between five studies.

Conclusion: Handhelds are an effective alternative to paper and pencil modes of data collection; they are faster and were preferred by most users.

8
<table>
<thead>
<tr>
<th>Study</th>
<th>Purpose</th>
<th>Design</th>
<th>Duration of Follow Up</th>
<th>Location</th>
<th>Patient Population</th>
<th>Number of Subjects</th>
<th>Instrument and Mode of Entry</th>
<th>Outcome Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quinn P et al. 2003</td>
<td>To assess the effectiveness of a portable electronic diary as a data collection device for symptoms of an overactive bladder (OAB)</td>
<td>Randomized crossover study</td>
<td>7 days/arm, 14 days total</td>
<td>Patient's residence</td>
<td>Patients with a diagnosis of over-active bladder.</td>
<td>35 patients were recruited, 2 were excluded post randomization.</td>
<td>Intervention: Customized version of MiniDoc Daily diary Visual Analogue Scale (VAS) Control: Paper diary Visual Analogue Scale (VAS) Effectiveness of the electronic diary Acceptability to patients</td>
<td></td>
</tr>
<tr>
<td>Jamison RN et al. 2002</td>
<td>To compare the e-VAS (electronic) with p-VAS (paper) for cognitive and sensory stimuli.</td>
<td>Single centre randomized crossover study</td>
<td>Data collected in a 1 hour session</td>
<td>Institution</td>
<td>Healthy volunteers</td>
<td>24 subjects</td>
<td>Intervention: Palm Pilot IIIxe Visual Analogue Scale (VAS) Control: Paper Visual</td>
<td>Validity and equivalence of methods</td>
</tr>
<tr>
<td>Study</td>
<td>Author(s)</td>
<td>Year</td>
<td>Intervention Description</td>
<td>Randomized Design</td>
<td>Chart Data Collection</td>
<td>Medical Student Assignment</td>
<td>Control Assignment</td>
<td>Speed/Time</td>
</tr>
<tr>
<td>-------</td>
<td>-----------</td>
<td>------</td>
<td>--------------------------</td>
<td>-------------------</td>
<td>----------------------</td>
<td>--------------------------</td>
<td>-------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Lal SO et al.</td>
<td>2000</td>
<td>35</td>
<td>To determine whether electronic data collection and downloading to a personal computer spreadsheet is faster and more accurate than written data.</td>
<td>Randomized crossover design</td>
<td>Chart data collected within a 96 hour window period</td>
<td>3 medical students retrieving data from 110 medical charts</td>
<td>Shriners Burns Hospital Medical student volunteering for data collection</td>
<td>3Com Palm I1x Data entry into an Excel spreadsheet</td>
</tr>
<tr>
<td>McBride JS et al.</td>
<td>1999</td>
<td>36</td>
<td>To examine how data can be collected at point of care. Comparison of electronic and paper versions of a standard quality survey.</td>
<td>Randomized design</td>
<td>Data collected in one session</td>
<td>Patients visiting an orthopedic clinic</td>
<td>Wake Forest PhysiciansOrthopedics Department Clinics</td>
<td>Mini Doc, a portable electronic data capturing device.</td>
</tr>
<tr>
<td>Stratton RJ et al.</td>
<td>1998</td>
<td>37</td>
<td>To assess an electronic visual analogue scale with a paper method for appetite rating To examine test-retest reliability.</td>
<td>Randomized crossover study design</td>
<td>4 day study Test-retest over 2 additional days</td>
<td>Healthy free-living volunteers</td>
<td>Subject’s residence</td>
<td>Apple Newton Message Pad Visual analogue scale questionnaire</td>
</tr>
<tr>
<td>Study</td>
<td>Design</td>
<td>Intervention</td>
<td>Control</td>
<td>Outcome Measures</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>--------</td>
<td>--------------</td>
<td>---------</td>
<td>------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tiplady B et al. 1997 (study 1) 38</td>
<td>Randomized two period crossover design</td>
<td>Apple Message Pad</td>
<td>Paper and pencil</td>
<td>To assess the suitability of PDAs compared to paper diaries, for daily collection of data on lung function.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tiplady B et al. 1997 (study 2)</td>
<td>Observational study</td>
<td>Completed electronic diary for 1 month</td>
<td>Paper and pencil</td>
<td>To assess the suitability of electronic diary for home use, transmitting respirology data.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drummond HE et al. 1995 39</td>
<td>Randomized, open, two period crossover</td>
<td>Apple Newton Message Pad</td>
<td>Paper and pencil</td>
<td>To compare the responses obtained from a quality of life (QOL) questionnaires using electronic (PDA) and conventional (paper).</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Objective</td>
<td>Design</td>
<td>Time</td>
<td>Control</td>
<td>Intervention</td>
<td>Outcome Measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>-------------------------------</td>
<td>-------------</td>
<td>--------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rivellesse AA et al. 1991</td>
<td>To evaluate an electronic (Food-Meter) method for recording 7-day food intake.</td>
<td>Randomized cross-over design repeated once.</td>
<td>4 weeks</td>
<td>patient's residence</td>
<td>Insulin-dependent diabetic patients (IDDM)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Walker I et al. 2004</td>
<td>To compare handheld computers and paper diaries for recording intravenous infusions of hemophilic clotting factor concentrates.</td>
<td>Randomized controlled trial, parallel design.</td>
<td>6 months</td>
<td>patients' residence</td>
<td>Patients with hemophilia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rivellesse AA et al. 1991</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Intervention:**
"Food-Meter" (Miles, Elkhart, IN)

**Control:**
Paper and pencil

**Agreement between methods**

**Walker I et al. 2004**

**Intervention:**
Palm III with bar code reader

**Control:**
Touch-sensitive paper diary.

**Outcome Measures**
Compliance, Timeliness, Accuracy, Preference

---

### Table 2

Summary of the results of data accuracy assessed in six randomized controlled trials

<table>
<thead>
<tr>
<th>Study</th>
<th>Results related to data Accuracy</th>
<th>Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hand Held Computers</td>
<td>Paper</td>
</tr>
<tr>
<td>Tiplady B et al. 1997</td>
<td>Missing data 8.91%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Problematic data** 5.64%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Missing data 0.16%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Problematic data** 0.24 %</td>
<td></td>
</tr>
<tr>
<td>McBride JS et al. 1999</td>
<td>No difference in missing item</td>
<td></td>
</tr>
<tr>
<td></td>
<td>responses between PDA and paper</td>
<td></td>
</tr>
<tr>
<td></td>
<td>in 4/5* subscales, (p &lt; 0.05).</td>
<td></td>
</tr>
<tr>
<td></td>
<td>No differences in missing item</td>
<td></td>
</tr>
<tr>
<td></td>
<td>responses between PDA and paper</td>
<td></td>
</tr>
<tr>
<td></td>
<td>in 4/5* subscales, (p &lt; 0.05).</td>
<td></td>
</tr>
<tr>
<td>Lal SO et al. 2000</td>
<td>2.8% error frequency.</td>
<td>6.7% error frequency</td>
</tr>
<tr>
<td>Jamison RN et al. 2002</td>
<td>Of 503 paired verbal stimuli in</td>
<td></td>
</tr>
<tr>
<td></td>
<td>24 subjects the correlation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>between paper and PDA ratings</td>
<td></td>
</tr>
<tr>
<td></td>
<td>was $r = 0.97$ (range 0.95–0.98),</td>
<td></td>
</tr>
<tr>
<td></td>
<td>for sensory stimuli $r = 0.86$ (range 0.81–0.92).</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Correlation between group</td>
<td></td>
</tr>
<tr>
<td></td>
<td>electronic VAS and paper VAS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ratings to common verbal stimuli $r^2 = 0.997$, for the common sensory</td>
<td></td>
</tr>
<tr>
<td></td>
<td>stimuli group correlation was $r^2 = 0.99$.</td>
<td></td>
</tr>
<tr>
<td>Quinn P et al. 2003</td>
<td>Errors not possible in electronic</td>
<td>Errors &quot;detected&quot; in 80% of paper</td>
</tr>
<tr>
<td></td>
<td>diaries due to prompt and format</td>
<td></td>
</tr>
<tr>
<td></td>
<td>of questions and responses.</td>
<td></td>
</tr>
<tr>
<td>Walker I et al. 2004</td>
<td>3 vials/patient not accounted</td>
<td>5 vials/patient not accounted for (P = 0.45).</td>
</tr>
<tr>
<td></td>
<td>for 15 patients with errors.</td>
<td>13 patients with errors (P = 1.00).</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*The subscale in which differences in missing item responses were found was the 1\textsuperscript{st} response choice on PDA, authors suggest this could be attributed to learning effect.


A systematic review and meta-analysis of quantitative interviewing tools to investigate self-reported HIV and STI associated behaviours in low- and middle-income countries.
Studies identifying risks and evaluating interventions for human immunodeficiency virus (HIV) and other sexually transmitted infections often rely on self-reported measures of sensitive behaviours. Such self-reports can be subject to social desirability bias. Concerns over the accuracy of these measures have prompted efforts to improve the level of privacy and anonymity of the interview setting. This study aims to determine whether such novel tools minimize misreporting of sensitive information.

Methods: Systematic review and meta-analysis of studies in low- and middle-income countries comparing traditional face-to-face interview (FTFI) with innovative tools for reporting HIV risk behaviour. Crude odds ratios (ORs) and 95% confidence intervals (CIs) were calculated. Cochran’s chi-squared test of heterogeneity was performed to explore differences between estimates. Pooled estimates were determined by gender, region, education, setting and question time frame using a random effects model.

Results: We found and included 15 data sets in the meta-analysis. Most studies compared audio computer-assisted self interview (ACASI) with FTFI. There was significant heterogeneity across studies for three outcomes of interest: ‘ever had sex’ (I²=93.4%, P<0.001), non-condom use (I²=89.3%, P<0.001), and number of partners (I²=75.3%, P<0.001). For the fourth outcome, ‘forced sex’, there was homogenous increased reporting by non-FTFI methods (OR 1.47; 95% CI 1.11–1.94). Overall, non-FTFI methods were not consistently associated with a significant increase in the reporting of all outcomes. However, there was increased reporting associated with non-FTFI with region (Asia), setting (urban), education (460% had secondary education) and a shorter question time frame.

Conclusion: Contrary to expectation, differences between FTFI and non-interviewer-administered interview methods for the reported sensitive behaviour investigated were not uniform. However, we observed trends and variations in the level of reporting according to the outcome, study and population characteristics. FTFI may not always be inferior to innovative interview tools depending on the sensitivity of the question as well as the population assessed.
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.

BACKGROUND: Patient-reported outcomes (PROs) are the consequences of disease and/or its treatment as reported by the patient. The importance of PRO measures in clinical trials for new drugs, biological agents, and devices was underscored by the release of the US Food and Drug Administration’s draft guidance for industry titled “Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims.” The intent of the guidance was to describe how the FDA will evaluate the appropriateness and adequacy of PRO measures used as effectiveness end points in clinical trials. In response to the expressed need of ISPOR members for further clarification of several aspects of the draft guidance, ISPOR’s Health Science Policy Council created three task forces, one of which was charged with addressing the implications of the draft guidance for the collection of PRO data using electronic data capture modes of administration (ePRO). The objective of this report is to present recommendations from ISPOR’s ePRO Good Research Practices Task Force regarding the evidence necessary to support the comparability, or measurement equivalence, of ePROs to the paper-based PRO measures from which they were adapted.

METHODS: The task force was composed of the leadership team of ISPOR’s ePRO Working Group and members of another group (i.e., ePRO Consensus Development Working Group) that had already begun to develop recommendations regarding ePRO good research practices. The resulting task force membership reflected a broad array of backgrounds, perspectives, and expertise that enriched the development of this report. The prior work became the starting point for the Task Force report. A subset of the task force members became the writing team that prepared subsequent iterations of the report that were distributed to the full task force for review and feedback. In addition, review beyond the task force was sought and obtained. Along with a presentation and discussion period at an ISPOR meeting, a draft version of the full report was distributed to roughly 220 members of a reviewer group. The reviewer group comprised individuals who had responded to an emailed invitation to the full membership of ISPOR. This Task Force report reflects the extensive internal and external input received during the 16-month good research practices development process.
RESULTS/RECOMMENDATIONS: An ePRO questionnaire that has been adapted from a paper-based questionnaire ought to produce data that are equivalent or superior (e.g., higher reliability) to the data produced from the original paper version. Measurement equivalence is a function of the comparability of the psychometric properties of the data obtained via the original and adapted administration mode. This comparability is driven by the amount of modification to the content and format of the original paper PRO questionnaire required during the migration process. The magnitude of a particular modification is defined with reference to its potential effect on the content, meaning, or interpretation of the measure's items and/or scales. Based on the magnitude of the modification, evidence for measurement equivalence can be generated through combinations of the following: cognitive debriefing/testing, usability testing, equivalence testing, or, if substantial modifications have been made, full psychometric testing. As long as only minor modifications were made to the measure during the migration process, a substantial body of existing evidence suggests that the psychometric properties of the original measure will still hold for the ePRO version. Hence, an evaluation limited to cognitive debriefing and usability testing only may be sufficient. However, where more substantive changes in the migration process has occurred, confirming that the adaptation to the ePRO format did not introduce significant response bias and that the two modes of administration produce essentially equivalent results is necessary. Recommendations regarding the study designs and statistical approaches for assessing measurement equivalence are provided.

CONCLUSIONS: The electronic administration of PRO measures offers many advantages over paper administration. We provide a general framework for decisions regarding the level of evidence needed to support modifications that are made to PRO measures when they are migrated from paper to ePRO devices. The key issues include: 1) the determination of the extent of modification required to administer the PRO on the ePRO device and 2) the selection and implementation of an effective strategy for testing the measurement equivalence of the two modes of administration. We hope that these good research practice recommendations provide a path forward for researchers interested in migrating PRO measures to electronic data collection platforms.  

Electronic Patient Diaries and Questionnaires - ePRO Now Delivering on the Promise? (Editorial)

Those considering implementing ePRO methods will need to consider cost and the impact on study processes. While the up-front costs for software development and purchase of devices are significantly more than the process for paper, the cost of implementing ePRO represents only a modest increase in the overall cost of running a paper, diary-based study given the gains in efficiency and the quality of the data. The need for setup, testing, and staff training before systems are used for live data imposes a ‘front-end load’ on the study process, even though it may not increase total workload. This needs to be taken into account in planning. In some cases, for example, the use of unsupervised patient diaries for regulatory submissions, the improvement in data quality and documented compliance may be sufficient on its own to justify the use of ePRO. ePRO have been a promising new technology for quite a while, but have now come of age. They should be in the toolbox of all those who need to get the patient’s point of view accurately, quickly, and in ways that they are comfortable using.  

Cost Effectiveness

Computer-assisted self-interviews: a cost effectiveness analysis.

Computer-assisted self-interview (CASI) questionnaires are being used with increased frequency to deliver surveys that previously were administered via self-administered paper-and-pencil questionnaires (SAQs). Although CASI may offer a number of advantages, an important consideration for researchers is the assessment modality's immediate and long-term costs. To facilitate researchers' choice between CASI and SAQ, this article provides theoretical cost models with specific parameters for comparing the costs for each assessment type. Utilizing these cost models, this study compared the cost effectiveness in a health behavior study in which both CASI (n = 100) and SAQ (n = 100) questionnaires were administered. Given the high initial costs, CASI was found to be less cost effective than SAQ for a single study. However, for studies with large sample sizes or when
CASI software is to be used for multiple studies, CASI would be more cost effective and should be the assessment mode of choice.\textsuperscript{12}

**Cost effectiveness of three child mental health assessment methods: computer-assisted assessment is effective and inexpensive.**

In order to survive severe funding reductions, community mental health centers (CMHCs) have implemented a number of systems-level interventions that attempt to minimize the impact of budget cuts on treatment effectiveness. The present study focused on ways to maintain the effectiveness of clinical assessment while lowering the assessment cost. The present study evaluated the relative cost effectiveness of three methods for collecting information and developing clinical assessment reports on children at a CMHC: (a) a traditional narrative clinical assessment report; (b) a form-style clinical assessment; and (c) a computer-assisted clinical assessment. The results revealed that the computer-assisted assessments were at least as effective as the two alternative assessment methods and only 20 percent to 45 percent as costly. The effect of using the computer-assisted assessments was reported to be favorable by therapists. While computer technology can be used to cut service delivery costs, the use of computers in CMHCs has generally been limited to administrative tasks, and clinical applications have been ignored.\textsuperscript{13}

**A comparison study: paper-based versus web-based data collection and management.**

The purpose of this pilot study was to compare the cost, accuracy, and efficiency of a web-enhanced handheld computer data collection system with those of the traditional paper-based data collection and management system and to increase awareness/knowledge of researchers on these two data collection and management methods. This is an important topic because funding resources are diminishing and high startup costs associated with automated data collection systems may give researchers pause when faced with these financial expenditures. Hence, this information will position grant writers and funders to make intelligent decisions regarding the feasibility and advantage of web-enhanced electronic data collection strategies.\textsuperscript{1}

\begin{table}[h]
\centering
\begin{tabular}{|l|c|c|c|c|c|c|}
\hline
\textbf{Category} & \textbf{Paper-based (n = 10)} & \textbf{Computer-based (n = 10)} & \textbf{Errors (frequency percent)} & \textbf{Errors (frequency percent)} \\
\hline
\hline
\text{Data collection} & Mean time per case (min) & Mean cost per case (dollars) & – & Mean time per case (min) & Mean cost per case (dollars) & – \\
\hline
\text{Coding} & 7.7 & 1.03 & – & 4.4 & 0.59 & – \\
\hline
\text{Entry} & 2.6 & 0.35 & 17 (8.5) & 0.00 & 0.00 & 0.0 \\
\hline
\text{Mean total score} & 17.3 & 2.31 & 19% & 4.4 & 0.59 & 0.0 \\
\hline
\end{tabular}
\caption{Average time, variable cost, and error: paper-based data collection and management system versus computer-based system.}
\end{table}

* Includes the cost of cellular connectivity that would be omitted if internet access were available.
Acceptability/Social Desirability Bias Reduction

The reporting of sensitive behavior by adolescents: a methodological experiment in Kenya.

Does audio computer-assisted self-interviewing (ACASI) produce more valid reporting of sexual activity and related behaviors than face-to-face interviews or self-administered interviews? This analysis, based on data collected from over 6,000 unmarried adolescents in two districts of Kenya--Nyeri and Kisumu--indicates substantial and significant differences in reported rates of premarital sex across interview modes, although not always in the expected direction. Our assumption that girls underreport sexual activity in face-to-face interviews by comparison with ACASI is not confirmed by the Nyeri data, but our results from Kisumu are considerably more promising. As for boys, who we believe exaggerate their level of sexual activity in face-to-face interviews, a more nuanced set of expectations regarding the reporting of sensitive behaviors was offered; our results from Kisumu, although not always significant, by and large conform to expectations.14

A comparison of audio computer-assisted self-interviews to face-to-face interviews of sexual behavior among perinatally HIV-exposed youth.

Computer-assisted interview methods are increasingly popular in the assessment of sensitive behaviors (e.g., substance abuse and sexual behaviors). It has been suggested that the effect of social desirability is diminished when answering via computer, as compared to an interviewer-administered face-to-face (FTF) interview, although studies exploring this hypothesis among adolescents are rare and yield inconsistent findings. This study compared two interview modes among a sample of urban, ethnic-minority, perinatally HIV-exposed U.S. youth (baseline = 148 HIV+, 126 HIV-, ages 9-16 years; follow-up = 120 HIV+, 110 HIV-, ages 10-19 years). Participants were randomly assigned to receive a sexual behavior interview via either Audio Computer-Assisted Self-Interview (ACASI) or FTF interview. The prevalence of several sexual behaviors and participants’ reactions to the interviews were compared. Although higher rates of sexual behaviors were typically reported in the ACASI condition, the differences rarely reached statistical significance, even when limited to demographic subgroups--except for gender. Boys were significantly more likely to report several sexual behaviors in the ACASI condition compared to FTF, whereas among girls no significant differences were found between the two conditions. ACASI-assigned youth rated the interview process as easier and
more enjoyable than did FTF-assigned youth, and this was fairly consistent across subgroup analyses as well. We conclude that these more positive reactions to the ACASI interview give that methodology a slight advantage, and boys may disclose more sexual behavior when using computer-assisted interviews.\textsuperscript{15}

Assessment of an electronic daily diary in patients with overactive bladder.

OBJECTIVE: To assess the effectiveness of a portable electronic diary as a data collection device for overactive bladder symptoms, and to evaluate its level of patient acceptability compared with a conventional paper-based voiding diary.

PATIENTS AND METHODS: Patients were identified through USA and UK hospital incontinence clinics. Patients were trained in the use of paper and electronic diaries before randomization, to complete either diary for 7 days. The diaries were then collected and, after a further training session, patients completed the other diary type for 7 days.

RESULTS: In all, 35 patients were recruited into the trial; overall, patients using the paper diaries (35) and electronic diaries (33) recorded similar data for the median number of incontinent episodes per week (8.2 and 7.0, respectively) and for the median number of significant leaks per day (0.4 and 0.5, respectively). However, the number of daily micturitions was slightly lower for the electronic than for the paper diary (7.3 vs 8.5, respectively). The frequency of urgency recorded in the electronic diary was higher than that recorded in the paper diary (5.8 vs 4.7). As 94\% of patients found the electronic diary easy to use, and the electronic diary reflects real-time data entries, the electronic diary data may provide a more accurate reflection of patient symptoms.

CONCLUSION: We confirmed that the electronic diary is a novel method of collecting clinically relevant symptom data from patients with an overactive bladder. In addition, the ease-of-use ratings support the use of the electronic diary as a superior alternative to paper diaries, providing real-time data which can be rapidly analysed, and thus allowing a speedy review of data during ongoing clinical studies.\textsuperscript{16}

Using Touch Screen Audio-CASI to Obtain Data on Sensitive Topics

This paper describes a new interview data collection system that uses a laptop personal computer equipped with a touch-sensitive video monitor. The touch-screen-based audio computer-assisted self-interviewing system, or touch screen audio-CASI, enhances the ease of use of conventional audio CASI systems while simultaneously providing the privacy of self-administered questionnaires. We describe touch screen audio-CASI design features and operational characteristics. In addition, we present data from a recent clinic-based experiment indicating that the touch audio-CASI system is stable, robust, and suitable for administering relatively long and complex questionnaires on sensitive topics, including drug use and sexual behaviors associated with HIV and other sexually transmitted diseases.\textsuperscript{17}

Adolescents' perceptions of a health survey using multimedia computer-assisted self-administered interview.

OBJECTIVE: To ascertain young people's perceptions of an adolescent health survey when administered by multimedia computer assisted self-administered Interview (M-CASI) through analysis of (1) questionnaire item responses and (2) focus group interviews.

METHODOLOGY: Setting: Auckland, New Zealand, 1999. Study type: Pilot testing of a 488-item branching questionnaire delivered using a youth-oriented and user-friendly M-CASI interface in a variety of settings using both desktop and laptop computers. Post pilot focus groups of participants identifying their perceptions and experiences of the survey. Sample: 110 school students aged 12 to 18 years.

\textsuperscript{15} More enjoyable than FTF-assigned youth, and this was fairly consistent across subgroup analyses.\textsuperscript{16} Assessment of an electronic daily diary in patients with overactive bladder.\textsuperscript{17} Using Touch Screen Audio-CASI to Obtain Data on Sensitive Topics.
RESULTS: The mean number of questions answered by participants was 316 with the median time to completion being 48 minutes. On average 65% of the total number of questions were seen and of these 1.5% were deliberately not answered. A high level of acceptability and enjoyment of M-CASI was found in the analysis of focus group responses and agreed with the item responses relating to M-CASI within the questionnaire itself. Participants identified privacy and confidentiality as being particularly important for the honesty of their responses. The passive matrix screens of the computers were popular as they could only be viewed from in front.

CONCLUSIONS: M-CASI is an acceptable instrument for the administration of a youth health survey. Laptop computers with passive matrix screens are able to enhance perceptions of privacy and confidentiality, which may improve honesty of responses.

IMPLICATIONS: M-CASI is now feasible and offers advantages in health surveying.

Consistency in the reporting of sexual behaviour by adolescent girls in Kenya: a comparison of interviewing methods.

OBJECTIVES: To investigate in a district in Kenya the level and consistency of reporting of sexual behaviour among adolescent girls randomly assigned to two modes of survey interview: face to face interview and audio computer assisted self-interview (ACASI).

METHODS: The analysis is based on a subsample of over 700 never married girls aged 15-21 years in Kisumu, Kenya, drawn from a population based survey of over 2100 respondents. A questionnaire with 69 questions was used, two thirds of which were considered sensitive, including questions about risky sexual behaviour, alcohol and drug use, contraceptive practice, pregnancy, induced abortions, and births.

RESULTS: ACASI produced significantly higher reporting of sex with a relative, stranger, or older man, and higher reporting of coerced sex. However, differences by mode for ever had sex and sex with a boyfriend were not significant. Relative to ACASI, the interviewer administered mode produced highly consistent reporting of sexual activity, both within the main interview and between the main and exit interviews.

CONCLUSIONS: Both the mode of survey administration and the probing for various behaviours significantly affect the observed prevalence of sexual activity. The ACASI results suggest that adolescent girls in Kenya have more complex and perilous sex lives than traditional face to face surveys of sexual activity indicate. The level of consistency in the interviewer mode is argued to be suspect, particularly given the much lower levels of reporting, relative to ACASI, for types of sexual partners and coerced sexual activity.

Does Audio-CASI Improve Reports of Risky Behavior? Evidence from a Randomized Field Trial Among Young Urban Men in India

This study compares the effectiveness of audio computer-assisted self-interviewing (Audio-CASI) with face-to-face interviews and self-administered questionnaires in collecting sensitive information on risky sexual and other behaviors among young men in urban India. A randomized study design compared data collected from 900 male college students using the three data-collection approaches and from 600 young men residing in slums using Audio-CASI and face-to-face interviews. Among the college students, the reported prevalence of risky behaviors was generally higher for young men interviewed through the Audio-CASI approach than with face-to-face interviews; self-administered questionnaires failed to yield significantly higher estimates than face-to-face interviews. Among the slum residents, the results were more mixed; the Audio-CASI approach failed to yield consistently higher responses for many risky behaviors compared with the face-to-face interview mode. The results demonstrate that although Audio-CASI appears to yield higher estimates of risky behavior among college-educated, computer-literate populations of young men, the efficacy of this approach among less-educated and less computer-literate populations appears more doubtful.
T-ACASI Reduces Bias in STD Measurements: The National STD and Behavior Measurement Experiment

**Background:** Although telephone surveys provide an economical method for assessing patterns of diagnosed sexually transmitted diseases (STDs) and STD-related behaviors in populations, the requirement that respondents report such information to human telephone interviewers introduces an opportunity for substantial reporting bias. Telephone computer-assisted self-interviewing (T-ACASI) surveys substitute a computer for human interviewers when asking sensitive questions.

**Methods:** A randomized experiment was embedded in a telephone survey that drew probability samples of the populations of the United States (N 1543) and Baltimore city (N 744). Respondents were randomly assigned to have sensitive questions asked either by a TACASI computer or by a human telephone interviewer.

**Results:** Respondents interviewed by a T-ACASI computer were more likely to report STD symptoms [dysuria, genital sores, genital discharge, and genital warts; adjusted odds ratios (ORs) 1.5–2.8] and a diagnosis of gonococcal or chlamydial infection during the past year (adjusted ORs 3.6 and 6.1). T-ACASI respondents with a main sex partner in the past year were more likely to report that their partner has had an STD (adjusted OR 2.4). For some measurements, the impact of T-ACASI was strongest among younger and less-educated respondents. When sampling weights were applied to project National STD and Behavior Measurement Experiment results to the populations of the United States and Baltimore, we found that reliance on data obtained by human interviewers would underestimate the annual incidence of chlamydial and gonococcal infections in these populations by factors of 2.4 to 9.7.

**Conclusions:** Compared with human telephone interviewers, T-ACASI surveys obtain increased reporting of STD symptoms, infections, and STD-related behaviors.21

A randomized trial of audio computer and in-person interview to assess HIV risk among drug and alcohol users in Rio De Janeiro, Brazil

This study compares drug patterns and prevalence of risk behaviors in a randomized trial using two methods of administration, Audio Computer-Assisted Self-Interview (ACASI) and Interviewer-Administered Questionnaire (IAQ), among drug users seeking treatment in a drug treatment center. We randomized 735 participants: 367 to ACASI and 368 to IAQ. No significant difference in sociodemographic variables were found between subjects in the two arms of the study. Those interviewed by ACASI were more likely to report use on 7 of 10 substances assessed. Rates of reporting of sexual risk behaviors (male-to-male and commercial sex) were higher among participants in the ACASI arm. ACASI seems to be a key resource in improving the reporting of sensitive data in Brazil, as it has been in prior international studies.22

Acceptability of audio computer-assisted self-interview (ACASI) among substance abusers seeking treatment in Rio de Janeiro, Brazil.

This study aimed to determine the acceptability of the ACASI approach to risk assessment and the impact of personal preference regarding mode of interview on reporting risk behaviors among drug users entering treatment in Rio de Janeiro, Brazil. We assessed 268 substance users who completed the ACASI arm in a randomized trial comparing the ACASI with the Interviewer-Administered Questionnaire (IAQ). The vast majority of interviewees (90.7%) reported no problem using the computer, and 37.3% felt that their privacy was best protected by the ACASI (vs. 16.4% who preferred the IAQ). Nearly half (45.5%) reported that the computer interview would produce more “honest” answers, whereas 30.6% selected the IAQ. In the adjusted regression analysis, problems using the computer were associated only with lower educational level (p<0.05). We found no evidence that preference had an impact on reporting risk behaviors or drug use. Our study showed both good feasibility and acceptability of the ACASI for interviewing drug users in Brazil. The findings extend our understanding of the role of the ACASI method by suggesting the utility of this approach in assessing HIV risk among low-to middle-income drug users in a cultural setting quite different from previous studies.23
A Clinical Trial Comparing Interviewer and Computer-Assisted Assessment Among Clients With Severe Mental Illness

Objective: Demographic, behavioral, and diagnostic information should routinely be collected from clients with severe mental illness, and data gathering should employ the most efficient techniques available. Surveys are increasingly conducted via Web-based computer-assisted interviewing (CAI), but this technique is not well validated for patients with severe mental illness. A randomized clinical trial of 245 clients was carried out to compare face-to-face and computer-assisted interviewing (233 clients completed two surveys).

Methods: Self-report data were collected on demographic characteristics, substance abuse, risk behaviors for bloodborne diseases, trauma history, and posttraumatic stress disorder. Each client was assessed twice and randomly assigned to one of the four possible combinations of interviewer and computer (computer and computer, N=53; computer and interviewer, N=56; interviewer and computer, N=59; and interviewer and interviewer, N=65). The two formats were compared on feasibility, client preference, cost, reliability, convergent validity, and criterion validity.

Results: This study demonstrated the feasibility of CAI across a variety of inpatient and outpatient settings. All participants who began the CAI process completed the interview and responded to over 95% of the survey items. Participants liked using the computers as well as they liked face-to-face interviews, and they completed the CAI as quickly. CAI produced data as reliable and valid as face-to-face interviews produced and was less expensive, and results were available more quickly. The two formats were similar in criterion validity.

Conclusions: CAI appears to be a viable technology for gathering clinical data from the population with severe mental illness and for transforming such information into a useful, quickly accessible form to aid in clinical decision making.

Comparison of 3 data collection methods for gathering sensitive and less sensitive information.

OBJECTIVE: When gathering sensitive information about personal experiences such as child abuse, drug and alcohol use, and intimate partner violence (IPV), it is especially important for both research and clinical purposes to use optimal methods to limit socially desirable responses. The purpose of this paper is to determine which of the following 3 methods is optimal for gathering data: 1) face-to-face interviews, 2) self-administered paper and pencil questionnaires, or 3) audio computer-assisted self-interviews (ACASI).

METHODS: The sample consisted of 514 parents bringing their preschoolers (0-5 years) to a pediatric primary care clinic for a checkup. The parent screening questionnaire (PSQ) addressing psychosocial problems was completed by participants themselves. Participants completed the PSQ in 1 of 3 ways: paper and pencil, face-to-face interview, or directly onto a computer (ACASI).

RESULTS: In general, ACASI yielded the highest rates for sensitive problems such as social isolation and parental stress, with face-to-face interviews occupying an intermediate position. The differences between ACASI and self-administered paper and pencil questionnaires were significant for many items. The differences between ACASI and face-to-face interviews, however, were modest. There were no significant group differences among the 3 methods in the prevalence rates of the neutral, less sensitive items.

CONCLUSION: ACASI resulted in greater disclosure of sensitive information than did a paper and pencil approach. No significant differences were observed between the computer-assisted interview and the face-to-face interview, both done in a research setting. The 3 methods appeared similar when gathering less sensitive data.

Differences in Young People’s Reports of Sexual Behaviors According to Interview Methodology: A Randomized Trial in India

Objectives. We compared reports of sexual behaviors given in standard face-to-face interviews with reports given in audio computer-assisted self-interviews (ACASIs) and culturally specific interactive interviews among adolescents in India. We sought
to determine which of the interview methods leads to higher reporting of sexual behaviors among economically disadvantaged 15-19-year-olds in urban India.

Methods. We conducted a randomized trial in which each participant (583 boys and 475 girls) was assigned to 2 interview methods: face-to-face interview and ACASI or interactive interview. We used matched case–control analyses to assess differences in the individual's reporting on the 2 methods.

Results. Female participants consistently reported fewer sexual behaviors in ACASIs than in face-to-face interviews, whereas male participants’ reports differed according to type of sexual behavior and interview mode. Both male and female participants reported more sexual behaviors during interactive interviews than during face-to-face interviews. Twenty-eight percent of male participants reported having engaged in heterosexual intercourse in interactive interviews, as compared with 20% in face-to-face interviews (P<.01); the corresponding percentages for female participants were 7% and 2% (P<.01).

Conclusions. Our results showed that young people were more likely to report sexual behaviors in culturally specific interactive interviews than in face-to-face interviews. By contrast, ACASIs did not uniformly lead to higher reporting levels than did face-to-face interviews.26

Is Audio Computer-Assisted Self-Interview (ACASI) Useful in Risk Behaviour Assessment of Female and Male Sex Workers, Mombasa, Kenya?

Background: Audio computer-assisted self-interview (ACASI) may elicit more frequent reporting of socially sensitive behaviours than face-to-face (FtF)-interview. However, no study compared responses to both methods in female and male sex workers (FSW; MSW) in Africa.

Methodology/Principal Findings: We sequentially enrolled adults recruited for an HIV-1 intervention trial into a comparative study of ACASI and FtF-interview, in a clinic near Mombasa, Kenya. Feasibility and acceptability of ACASI, and a comparative analysis of enrolment responses between ACASI and FtF on an identical risk assessment questionnaire were evaluated. In total, 139 women and 259 men, 81% of eligible cohort participants, completed both interviews. ACASI captured a higher median number of regular (2 vs. 1, p<0.001, both genders) and casual partners in the last week (3 vs. 2, p = 0.04 in women; 2 vs. 1, p=0.001 in men). Group sex (21.6 vs. 13.5%, p<0.001, in men), intravenous drug use (IDU; 10.8 vs. 2.3%, p=0.001 in men; 4.4 vs. 0%, p = 0.03 in women), and rape (8.9 vs. 3.9%, p = 0.002, in men) were reported more frequently in ACASI. A surprisingly high number of women reported in ACASI that they had paid for sex (49.3 vs. 5.8%, p<0.001). Behaviours for recruitment (i.e. anal sex, sex work, sex between males) were reported less frequently in ACASI. The majority of women (79.2%) and men (69.7%) felt that answers given in ACASI were more honest. Volunteers who were not able to take ACASI (84 men, and 37 women) mostly lacked reading skills.

Conclusions/Significance: About 1 in 5 cohort participants was not able to complete ACASI, mostly for lack of reading skills. Participants who completed ACASI were more likely to report IDU, rape, group sex, and payment for sex by women than when asked in FtF interview. ACASI appears to be a useful tool for high risk behaviour assessments in the African context.27

Consistency in Women’s Reports of Sensitive Behavior In an Interview Mode Experiment, São Paulo, Brazil

CONTEXT: Inaccurate reporting of sexual behavior creates a misleading picture of individuals’ risk for STI infection. Despite a substantial body of U.S. research on the consistency of self-reports of sensitive behavior, only a few such studies have been conducted in developing countries.

METHODS: Consistency in the reporting of sexual activity and other sensitive behaviors was assessed among 818 women aged 18–40 who enrolled in 2004 in a study examining STI screening and diagnosis in São Paulo, Brazil. Participants were randomized into face-to-face interview and audio computer-assisted self-interview (audio-CASI) groups, and a six-week follow-up interview was conducted using audio-CASI for all participants. Differences between groups were assessed using t tests, and logistic
regression analyses were used to estimate the likelihood of inconsistency within the enrollment interview and between the enrollment and follow-up interviews.

RESULTS: Consistency in reporting at the enrollment interview was higher in the face-to-face group than in the audio-CASI group, likely because interviewers prompted women to reconcile discrepant responses, whereas the audio-CASI program did not enforce logical consistency. However, consistency between enrollment and follow-up was significantly lower in the face-to-face group for abortion, marijuana use, transactional sex, coerced sex and number of lifetime sexual partners, because of increased reporting at follow-up using audio-CASI.

CONCLUSION: Although the analysis of internal consistency at enrollment suggests that computerized interviewing may increase random measurement error, it appears to reduce social desirability bias and encourage higher reporting of sensitive behaviors.

Cognitive Testing/Development

Instrument development and evaluation for patient-related outcomes assessments.

Abstract: Patient-related outcomes measures could provide important information for the current state of the art in medical care and even have an impact on macrodecisions in the health care system. Patient-related outcomes were initially defined as subjective health indicators that allow disability and illness to be assessed, based on patient, caregiver, or physician self-reports. As illness involves psychological and behavioral complex processes of care, a multidisciplinary approach in measuring patient-reported outcomes should be recommended, such as quality of life questionnaires. Patient-related outcomes measures should correspond to specific clinical situations and bring opportunities to improve quality of care. Objective measurements enable quantitative data to be collected and analyzed. Depending on the aim of the research, investigators can use existing methods or develop new tools. This publication presents a methodology for developing patient-related outcomes measures, based on a multistage procedure. The proper definition of specific study objectives and the methodology of instrument development are crucial for successfully transferring the study concept. The model of instrument development is the process of starting from the preliminary phase and includes questionnaire design and scaling, pilot testing (cognitive debriefing), revision of the preliminary version, evaluation of the new tool, and implementation. Validation of the new instrument includes reliability, reproducibility, internal consistency, and responsiveness. The process of designing the new tool should involve a panel of experts, including clinicians, psychologists (preliminary phase), and statisticians (scale development and scoring), and patients (cognitive debriefing). Implementation of a new tool should be followed by evaluation study - assessment of the tool's usefulness in clinical practice. An instrument must show not only the expected methodological properties and performance but also a positive contribution to care. The necessity of implementation of direct patient-reporting methods has been highlighted by both the Food and Drug Administration and the European Medicines Agency.

Cognitive psychology and self-reports: Models and methods

This article describes the models and methods that cognitive psychologists and survey researchers use to evaluate and experimentally test cognitive issues in questionnaire design and subsequently improve self-report instruments. These models and methods assess the cognitive processes underlying how respondents comprehend and generate answers to self-report questions. Cognitive processing models are briefly described. Non-experimental methods – expert cognitive review, cognitive task analysis, focus groups, and cognitive interviews – are described. Examples are provided of how these methods were effectively used to identify cognitive self-report issues. Experimental methods – cognitive laboratory experiments, field tests, and experiments embedded in field surveys – are described. Examples are provided of: (a) how laboratory experiments were designed to test the capability and accuracy of respondents in performing the cognitive tasks required to answer self-report questions, (b) how a field experiment was conducted in which a cognitively designed questionnaire was effectively tested against the original questionnaire, and (c) how a cognitive experiment embedded in a field survey was conducted to test cognitive predictions.
Pretesting survey instruments: an overview of cognitive methods.

Abstract: This article puts forward the case that survey questionnaires, which are a type of measuring instrument, can and should be tested to ensure they meet their purpose. Traditionally survey researchers have been pre-occupied with 'standardising' data collection instruments and procedures such as question wording and have assumed that experience in questionnaire design, coupled with pilot testing of questionnaires, will then ensure valid and reliable results. However, implicit in the notion of standardisation are the assumptions that respondents are able to understand the questions being asked, that questions are understood in the same way by all respondents, and that respondents are willing and able to answer such questions. The development of cognitive question testing methods has provided social researchers with a number of theories and tools to test these assumptions, and to develop better survey instruments and questionnaires. This paper describes some of these theories and tools, and argues that cognitive testing should be a standard part of the development process of any survey instrument.

Cognitive interviewing: verbal data in the design and pretesting of questionnaires.

PURPOSE: The purpose of this paper is to discuss problems that occur in questionnaire responses and how cognitive interviewing can be used to identify problematic questions prior to using the questionnaire in the field.

BACKGROUND: Questionnaire design involves developing wording that is clear, unambiguous and permits respondents successfully to answer the question that is asked. However, a number of problems in relation to respondents' understanding and successfully completing questionnaires have been identified. Cognitive interviewing, an amalgamation of cognitive psychology and survey methodology, has been developed to identify problematic questions that may elicit response error. The overall aim is to use cognitive theory to understand how respondents perceive and interpret questions and to identify potential problems that may arise in prospective survey questionnaires.

METHODS: A literature review is used to examine the process of questionnaire design and how cognitive interviewing can be used to reduce sampling error and increase questionnaire response rates.

FINDINGS: Cognitive interviewing involves interviewers asking survey respondents to think out loud as they go through a survey questionnaire and tell them everything they are thinking. This allows understanding of the questionnaire from the respondents' perspective rather than that of the researchers. Cognitive interviews have been used in a number of areas in health care research to pretest and validate questionnaires and to ensure high response rates. Interviewing has been found to be highly effective in developing questionnaires for age specific groups (children and adolescents) and in ascertaining respondents' understanding in health surveys prior to distribution. However, cognitive interviews have been criticized for being overly subjective and artificial.

CONCLUSION: Cognitive interviews are a positive addition to current methods of pretesting questionnaires prior to distribution to the sample. They are most valuable in pretesting questions that are complex, where questions are sensitive and intrusive and for specific groups for whom questionnaire completion may pose particular difficulties.

Issues and challenges of instrument translation.

Abstract: The purpose of this article was to discuss the challenges of instrument translation, using the translation of the Medical Outcomes Study Social Support Survey into Chinese as an example. Brislin's model of translation, which highlights the need for forward and backward translation, was used. Major considerations in conducting translation, and the strategies used to overcome the challenges arising from cultural and linguistic differences between the source and target languages, were discussed. Examples were used to illustrate how difficulties, such as maintenance of the original intent of the questionnaire, maximization of the cultural relevancy of the concept in question, and enhancement of the comprehensibility of the translated
questionnaire, were handled. The importance of literal and cultural adaptation of a developed instrument, rather than its simple word translation in the maintenance of an equivalent translation is highlighted. 31

So you want to do research? 5: Questionnaire design.

Abstract: This article describes the key aspects in the design, construction and adaptation of survey questionnaires. There are different types of questionnaire, each of which has its advantages and disadvantages. Aspects of constructing the questionnaire are discussed in detail; choosing the mode of administration; the objectives of the survey; availability of resources; characteristics of the target population; and quality of data. Issues concerning the identification of the questionnaire’s content, wording and sequencing of the questions through to the overall appearance and layout of the questionnaire are also considered. Differences in the role of open-ended and closed questions, together with their strengths and weaknesses, are outlined, and the need to undertake pre-testing and piloting as an integral part of questionnaire development is highlighted. Finally, issues around the adaptation of existing questionnaires are discussed with particular emphasis on their use in different language and cultural groups, and the need to achieve conceptual, content, semantic, operational and functional equivalence is described. An overview of the translation process is provided. 32

Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey

This paper investigates how expected and actual questionnaire length affects cooperation rates and a variety of indicators of data quality in web surveys. We hypothesized that the expected length of a web-based questionnaire is negatively related to the initial willingness to participate. Moreover, the serial position of questions was predicted to influence four indicators of data quality. We hypothesized that questions asked later in a web-based questionnaire will, compared to those asked earlier, be associated with (a) shorter response times, (b) higher item-nonresponse rates, (c) shorter answers to open-ended questions, and (d) less variability to items arranged in grids. To test these assumptions, we manipulated the stated length (10, 20, and 30 minutes) and the position of questions in an online questionnaire consisting of randomly ordered blocks of thematically related questions. As expected, the longer the stated length, the fewer respondents started and completed the questionnaire. In addition, answers to questions positioned later in the questionnaire were faster, shorter, and more uniform than answers to questions positioned near the beginning. 33

Editing data: what difference do consistency checks make?

Abstract: In 1998, the Florida Department of Health undertook a self-administered school-based survey of tobacco use, attitudes, and behaviors among nearly 23,000 public school students in grades 6-12. The survey design did not use skip patterns; therefore, students had multiple opportunities to contradict themselves. By using examples from the high school portion (grades 9-12) of the survey, the authors examined five possible approaches to handling data inconsistencies and the effect that each has on point estimates. Use of these approaches resulted in point estimates of current cigarette use ranging from 25.6% to 29.7%. The number of missing respondents varied from 33 (less than 1%) to 1,374 (13%), depending on which approach was used. After stratification by gender and race, the prevalence estimates changed marginally for girls but strikingly for boys. Non-Hispanic White students were substantially more likely than non-Hispanic Black students to report current cigarette use, but the magnitude of this difference varied significantly according to the analytical approach used. The approach used to check data consistency may influence point estimates and comparability with other studies. Therefore, this issue should be addressed when findings are reported. 34
Survey Length Can Impact Concept Measures, Data Quality

This study demonstrates that concept test results are impacted starting around 20 minutes. Therefore, Lightspeed Research has determined that optimum questionnaire length is 20 minutes or less. This length maximizes respondent engagement and survey satisfaction. Increased satisfaction leads to less attrition from the panel and higher quality responses.

To create a survey of optimum length, follow these guidelines:

1. Include only questions that directly relate to the research objectives and success criteria. Exclude questions designed to gather “nice to know,” but not essential data.
2. Minimize the number of questions asked prior to concept exposure.
3. Ensure that questions are easy to answer from the respondent’s point of view. For example, it can be hard for respondents to remember the specific brand SKU they last purchased or answer specific questions about a product when all they have seen is a concept.
4. Use technology to make questionnaires more respondent friendly. For example, program the questionnaire to automatically skip questions that don’t apply.
5. Eliminate questions that can seem repetitious. Respondents, in general, do not detect subtle nuances in questions. For example, respondents may view the question, “Is this product healthy,” as exactly the same as “Is this product good for you.”
6. Consider using a split questionnaire design to break a long questionnaire into manageable segments. A second questionnaire can be fielded to those returning the first. This approach makes the task less daunting for respondents while still gathering all the information from the same respondents. (http://www.lightspeedaheadnewsletter.com/?p=292, accessed 10/4/2012)

Effect of Computer-Assisted Interviewing on Self-Reported Sexual Behavior Data in a Microbicide Clinical Trial

Abstract In a microbicide safety and effectiveness trial (HPTN 035) in Malawi, 585 women completed the same questionnaire through a face-to-face interview (FTFI) and an audio computer-assisted self-interview (ACASI). Concordance between FTFI and ACASI responses ranged from 72.0 % for frequency of sex in the past week to 95.2 % for anal intercourse (AI) in the past 3 months. Reported gel and condom use at last sex act were marginally lower with ACASI than FTFI (73.5 % vs. 77.2 %, p = 0.11 and 60.9 % vs. 65.5 %, p = 0.05, respectively). More women reported AI with ACASI than FTFI (5.0 % vs. 0.2 %, p < 0.001). Analyses of consistency of responses within ACASI revealed that 15.0 % of participants in the condom only arm and 28.7 % in the gel arm provided at least one discrepant answer regarding total sex acts and sex acts where condom and gel were used (19.2 % reported one inconsistent answer, 8.1 % reported two inconsistent answers, and 1.4 % reported three inconsistent answers). While ACASI may provide more accurate assessments of sensitive behaviors in HIV prevention trials, it also results in a high level of internally inconsistent responses.

Evaluation of microbicide gel adherence monitoring methods.

BACKGROUND: An objective and accurate method that measures adherence to vaginal microbicide gel regimens during clinical trials could provide more accurate estimates of microbicide efficacy, aid in targeting adherence promotion resources, and enable objective assessment of adherence promotion strategies.

METHODS: We evaluated 4 methods to assess whether or not gel applicators had been vaginally inserted. At the study site, 50 women inserted hydroxyethylcellulose universal placebo gel through a polypropylene vaginal applicator and handled, but did not insert a second "sham-inserted” applicator. Applicators were discarded into a container capped with a medical event monitor system (MEMS) that recorded the time and date of opening. Fifteen additional participants did likewise at 2 study site visits, and administered gel on 6 intervening days at home. Applicators were scored as inserted, or not, by direct inspection
under ambient light, ultraviolet (UV) light, staining with Alcian blue, and microscopic detection of vaginal cells stained with iodine.

RESULTS: Mean sensitivity/specificity of 2 readings each by 3 test readers for UV, Alcian blue, ambient light, and iodine methods were 84/83, 79/83, 76/63, and 65/80%, respectively. Sensitivity of all methods was significantly higher in applicators inserted after one or more prior insertions of gel, with the highest sensitivity (95%) obtained with UV. MEMS caps accurately recorded applicator disposal time.

CONCLUSIONS: The modest accuracy of all 4 methods for applicator insertions without prior gel applications may limit their accuracy in monitoring coital regimens. However, for daily dosing regimens, MEMS monitoring and UV inspection should provide a rapid, reliable, and quantitative assessment of adherence. 

The reliability of sensitive information provided by injecting drug users in a clinical setting: Clinician-administered versus audio computer-assisted self-interviewing (ACASI).

Abstract Research with injecting drug users (IDUs) suggests greater willingness to report sensitive and stigmatised behaviour via audio computer-assisted self-interviewing (ACASI) methods than during face-to-face interviews (FFIs); however, previous studies were limited in verifying this within the same individuals at the same time point. This study examines the relative willingness of IDUs to report sensitive information via ACASI and during a face-to-face clinical assessment administered in health services for IDUs. During recruitment for a randomised controlled trial undertaken at two IDU-targeted health services, assessments were undertaken as per clinical protocols, followed by referral of eligible clients to the trial, in which baseline self-report data were collected via ACASI. Five questions about sensitive injecting and sexual risk behaviours were administered to participants during both clinical interviews and baseline research data collection. “Percentage agreement” determined the magnitude of concordance/discordance in responses across interview methods, while tests appropriate to data format assessed the statistical significance of this variation. Results for all five variables suggest that, relative to ACASI, FFI elicited responses that may be perceived as more socially desirable. Discordance was statistically significant for four of the five variables examined. Participants who reported a history of sex work were more likely to provide discordant responses to at least one socially sensitive item. In health services for IDUs, information collection via ACASI may elicit more reliable and valid responses than FFI. Adoption of a universal precautionary approach to complement individually tailored assessment of and advice regarding health risk behaviours for IDUs may address this issue.

Electronic quality of life questionnaires: a comparison of pen-based electronic questionnaires with conventional paper in a gastrointestinal study.

The use of pen-based electronic questionnaires and conventional paper questionnaires was compared in a randomized crossover study. Forty-six patients, aged 17-81 years, suffering from gastro-intestinal disorders, initially filled in a paper quality of life questionnaire for familiarization purposes, then on two subsequent visits completed electronic and paper questionnaires in randomized order. At the last visit they completed a preference survey. The results showed a high degree of acceptability of the electronic questionnaire, with 57% of patient preferring electronic and 13% preferring paper, while the remaining 30% expressed no preference. Neither age, gender nor familiarity with technology showed any marked association with patients’ preferences. All patients found both paper and electronic questionnaires easy to use. Data were more complete on the electronic questionnaire (100%) than on the paper (99.1%). Data handling procedures were greatly simplified. These results show that major benefits in completeness of data, speed of data flow, and data handling workload can be obtained from the use of pen-based electronic questionnaires.
Evaluation of new computerized method for recording 7-day food intake in IDDM patients.

OBJECTIVE: To evaluate a new computerized method for recording 7-day food intake.

RESEARCH DESIGN AND METHODS: Randomized crossover trial was conducted with patients recording the amount and type of every food and drink consumed during a week by either a computerized device (Food-meter) or recording the data in a diary. Each method was applied twice. Twenty-one insulin-dependent diabetic patients (mean +/- SD age 25 +/- 9 yr) were studied.

RESULTS: The two methods showed very good agreement in the evaluation of the patients’ diets (1792 +/- 408 vs. 1764 +/- 436 kcal/day, 84 +/- 19 vs. 82 +/- 21 g/day protein, 68 +/- 22 vs. 67 +/- 23 g/day fat, 210 +/- 60 vs. 207 +/- 58 g/day carbohydrate with the conventional and computerized methods, respectively). The variability between the methods and the variability within each method were of similar magnitude.

CONCLUSIONS: The Food-meter represents a useful tool for computerizing the 7-day food record. The method is easy, reliable, and time saving. Moreover, it minimizes the risk of transcriptional errors.38

Comparing hand-held computers and paper diaries for haemophilia home therapy: a randomized trial.

Treatment of severe haemophilia with factor concentrates is by self-infusion in the home. Adherence to record keeping on paper diaries is poor. A randomized-controlled trial compared adherence with record keeping of paper diaries with hand-held computers. Forty-one individuals with severe haemophilia, were randomized to hand-held computers (n = 22) or paper diaries (n = 19) and followed for 6 months. About 86.2% (679 of 788) of infusions by patients in the computer group were in compliance with the data submission schedule compared with only 48.3% (358 of 741) of infusions by patients using paper diaries (P < 0.0001). The time intervals between infusions and the receipt of data were shorter in the computer group (median 0.25 vs. 25 days respectively, P < 0.0001). Reminder phone calls by the clinic were made less frequently to users of hand-held computers than to users of paper diaries (median one vs. five times, P < 0.0001). Accuracy of data was similar for both methods. Compliance with hand-held computers was superior to paper diaries. The clinic received data from hand-held computers mostly on the same day, and nurses could thereby provide clinical advice more effectively. Although hand-held computers did not result in increased accuracy, errors could be detected and corrected more rapidly. Electronic data can more easily be verified, analysed and summarized than that from paper diaries.39

Comparative study of electronic vs. paper VAS ratings: a randomized, crossover trial using healthy volunteers.

The visual analogue scale (VAS) is an established, validated, self-report measure usually consisting of a 10 cm line on paper with verbal anchors labeling the ends. Palmtop computers (PTCs also known as personal digital appliances) have incorporated VAS entry by use of a touch screen. However, the validity and psychophysical properties of the electronic VAS have never been formally compared with the conventional paper VAS. The aim of this study is to determine the agreement between the electronic (eVAS) and paper (pVAS) modes. Twenty-four healthy volunteers were recruited for this study. Each study participant provided input using both measurement methods by marking the eVAS and pVAS in response to two kinds of stimuli, cognitive and sensory. A verbal rating scale of seven descriptors of intensity represented the cognitive stimuli. Participants were asked to mark the location that best corresponded to the pain intensity described by each word on scales from ‘no pain’ to ‘worst possible pain’. The sensory stimuli used were a set of test weights consisting of plastic containers ranging from 7 to 129 g. The VAS for sensory stimuli ranged from 0 (no weight) to ‘reference weight’ (the heaviest weight outside the range of test weights). There were two types of input stimuli and two modes for recording responses for a total of four experimental conditions. Two evaluators independently measured and recorded all the pVAS forms to the nearest millimeter. A total of 2016 stimuli were rated. The overall correlation for ratings of both sensory and cognitive stimuli on eVAS and pVAS was r = 0.91. For paired verbal stimuli the correlation was r = 0.97. For paired sensory stimuli the correlation was r = 0.86. The correlation between group eVAS and pVAS ratings to common verbal stimuli was r = 0.99. For common sensory stimuli the group correlation was r = 0.99. The median of correlations comparing eVAS and pVAS ratings was 0.99 for verbal stimuli and 0.98 for sensory stimuli. Multivariate
analyses showed equivalent stimuli to be rated much the same whether entered on paper VAS or PTC touch screen VAS (P < 0.0001). Support was found for the validity of the computer version of the VAS scale.  

**Palm computer demonstrates a fast and accurate means of burn data collection.**

Manual biomedical data collection and entry of the data into a personal computer is time-consuming and can be prone to errors. The purpose of this study was to compare data entry into a hand-held computer versus hand written data followed by entry of the data into a personal computer. A Palm (3Com Palm IIIx, Santa Clara, Calif) computer with a custom menu-driven program was used for the entry and retrieval of burn-related variables. These variables were also used to create an identical sheet that was filled in by hand. Identical data were retrieved twice from 110 charts 48 hours apart and then used to create an Excel (Microsoft, Redmond, Wash) spreadsheet. One time data were recorded by the Palm entry method, and the other time the data were handwritten. The method of retrieval was alternated between the Palm system and handwritten system every 10 charts. The total time required to log data and to generate an Excel spreadsheet was recorded and used as a study endpoint. The total time for the Palm method of data collection and downloading to a personal computer was 23% faster than hand recording with the personal computer entry method (P < 0.05), and 58% fewer errors were generated with the Palm method. The Palm is a faster and more accurate means of data collection than a handwritten technique.  

**Using a hand-held computer to collect data in an orthopedic outpatient clinic: a randomized trial of two survey methods.**

OBJECTIVES: In a randomized study, the authors examine how data can be collected at the point of care. Specifically, examining to what extent handheld computer data collection systems introduce bias or increase respondent difficulty.

METHODS: Volunteers were randomized to 1 of 2 survey methods: the hand-held computer or a paper and pencil form of similar content. Differences between group scale scores were compared using the Wilcoxon (rank sum) test.

RESULTS: The hand-held computer system produced comparable scores to paper and pencil surveys. However, there was evidence of lower internal consistency reliability with the handheld computer.

CONCLUSIONS: This study demonstrated the comparability of the hand-held computer methodology to the paper and pencil methodology in obtaining survey information in an ambulatory clinic. The hand-held computer method of survey data collection offers an alternative to paper methods when point-of-care administration is acceptable. Preliminary evidence shows that this method produces comparable results to paper forms.  

**Comparison of the traditional paper visual analogue scale questionnaire with an Apple Newton electronic appetite rating system (EARS) in free living subjects feeding ad libitum.**

OBJECTIVE: Assessing the value of a newly developed electronic visual analogue scale questionnaire (Apple Newton Message Pad) with the traditional paper method for appetite rating.

DESIGN: In a random, crossover design, subjects completed both electronic and paper questionnaires to compare results obtained by the two methods; individual methods were completed consecutively to assess test-retest reliability; preference was established using a questionnaire.

SETTING/SUBJECTS: Healthy, free-living adults were studied for comparison of methods (n = 12), test-retest reliability (n = 8) and preference (n = 13).

INTERVENTION: Visual analogue scales were completed each waking hour to assess appetite. Preference was assessed after both methods were completed.
RESULTS: There was no significant difference in the hourly results obtained by the paper and electronic methods for 'desire to eat', 'how much can you eat now', 'urge to eat' and 'preoccupation with thoughts of food'. Small differences in 'hunger' and 'fullness' ratings were noted (approximately 5% mean difference between methods, P < 0.05), but patterns of change and sensitivity for these and all other parameters remained similar for both methods across the visual analogue scale. Test-retest reliability demonstrated was similar for both methods. Seven (54%) subjects preferred to use the paper questionnaire, five (38%) the electronic method and one (8%) had no preference.

CONCLUSIONS: The electronic Apple Newton questionnaire is as sensitive and reliable as the paper method, has the advantage that it automatically records the time of data acquisition and data collection and processing are more efficient for the researcher. The two methods should not be used interchangeably.


Purpose: To evaluate the test–retest reliability of self-reported sexuality-related data in a sample of African American adolescents residing in four U.S. cities.

Methods: Using audio computer-assisted self-interviewing (ACASI), 156 African American teens (mean age = 15.5 years) provided data on lifetime and recent sexual behavior, HIV/sexually transmitted disease (STD) testing, and theoretical antecedents of sexual risk behavior on two occasions separated by 2 weeks.

Results: Most self-reports of lifetime sexual behavior and STD/HIV testing were stable across the two assessment points. Test–retest agreement was substantial for dichotomous indices of lifetime sexual behaviors and STD testing (kappas ranging from .61–.87), and for 3-month recall of vaginal sex (kappa¼ .72) and number of sexual partners (intraclass correlation coefficient [ICC] = .68). Lower reliability estimates emerged for count data of unprotected vaginal sex occasions (ICC = .44). Test–retest reliability estimates for antecedents of sexual risk behavior were highest for a validated measure of HIV-related knowledge (r=.73), but somewhat lower for peer norms(r=.58) and condom use self-efficacy (r=.50).

Conclusions: Although variability in the stability of self-report data was observed, findings confirm that most sexual behavior, STD and HIV testing history, and psychosocial measures can be assessed reliably among adolescents. Research should continue to identify strategies to enhance the reliability of self-report sexual behavior data from youth at risk for HIV and other sexually transmitted infections.

Does modality of survey administration impact data quality: audio computer assisted self interview (ACASI) versus self-administered pen and paper?

BACKGROUND: In the context of a randomized controlled trial (RCT) on HIV testing in the emergency department (ED) setting, we evaluated preferences for survey modality and data quality arising from each modality.

METHODS: Enrolled participants were offered the choice of answering a survey via audio computer assisted self-interview (ACASI) or pen and paper self-administered questionnaire (SAQ). We evaluated factors influencing choice of survey modality. We defined unusable data for a particular survey domain as answering fewer than 75% of the questions in the domain. We then compared ACASI and SAQ with respect to unusable data for domains that address sensitive topics.

RESULTS: Of 758 enrolled ED patients, 218 (29%) chose ACASI, 343 chose SAQ (45%) and 197 (26%) opted not to complete either. Results of the log-binomial regression indicated that older (RR = 1.08 per decade) and less educated participants (RR = 1.25) were more likely to choose SAQ over ACASI. ACASI yielded substantially less unusable data than SAQ.

CONCLUSIONS: In the ED setting there may be a tradeoff between increased participation with SAQ versus better data quality with ACASI. Future studies of novel approaches to maximize the use of ACASI in the ED setting are needed.
The effect of mode of administration on Medical Outcomes Study health ratings and EuroQol scores in AIDS

Brief measures of health-related quality of life are being used with increased frequency in AIDS clinical trials. Self-administration of questionnaires can reduce costs in this setting because they require little time. However, the equivalence between self and interview-administered responses in clinical trials is not known. We evaluated patient and proxy responses to the Medical Outcomes Study HIV Health Survey (MOS-HIV) and the EuroQol. We randomized 68 patients with advanced HIV disease on (1) mode of administration (self vs. interview); (2) type of interview (face-to-face vs. telephone); (3) questionnaire order (MOS-first vs. EuroQol-first); and (4) 2- vs. 3-item response categories for physical limitations. There were few differences in scores between self and interview administration and type of interview. Proxy respondents viewed patients as more impaired than did patients themselves on subjective aspects of health including mental health (63.8 vs. 75.7, p < 0.001), health distress (67.3 vs. 77.1, p = 0.007), pain (64.4 vs. 70.0, p = 0.04), and vitality (48.4 vs. 55.5, p = 0.04). Results concerning questionnaire order and number of response categories were not conclusive. Our results suggest that for patients with advanced HIV disease, data from the MOS-HIV and the EuroQol collected using different modes may be pooled, but that proxy responses should be calibrated.45

Comparing the Reliability of Responses to Telephone-Administered versus Self-Administered Web-Based Surveys in a Case-Control Study of Adult Malignant Brain Cancer

Introduction: To determine whether a Web-based survey was an acceptable method of data collection for a clinic-based case-control study of adult brain cancer, the authors compared the reliability of paired responses to a main and resurvey for participants completing surveys by telephone (n = 74) or self-administered on the Web (n = 465) between 2003 and 2006.

Methods: Recruitment of cases was done at the Evanston Northwestern Healthcare Kellogg Cancer Care Center and the Duke University Medical Center Cancer Control division, and controls were friends and siblings of cases. Twenty-five variables were examined, including smoking, oral contraceptive and residential histories, water sources, meat preparation, fruit and vegetable consumption, and pesticide use. Weighted and simple k’s were estimated for categorical and binary variables, respectively.

Results: The number of concordant paired responses was summed for use in linear regression. Respondents were 97% White and 85% had postsecondary education. Kappa’s for individual questions ranged from 0.31 (duration of residence in a single family house) to 0.96 (ever smoked), with a median of 0.57 (95% confidence interval, 0.47-0.64). The median number of concordant responses was 16.2 (range, 9-22). Reliability was greater for controls than cases, Web-based versus telephone responders, females, and higher-income responders. Frequency of e-mail and Internet use was not associated with reliability.

Conclusions: A self-administered, Web-based survey was a feasible and appropriate mode of interview in this study. The comparable reliability of Web compared with telephone responses suggest that Web-based self-interviews could be a cost-effective alternative to traditional modes of interview.46

Differential response effects of data collection mode in a cancer screening study of unmarried women ages 40–75 years: A randomized trial

Background: Little is known about the impact of data collection method on self-reported cancer screening behaviours, particularly among hard-to-reach populations. The purpose of this study is to examine the effects of data collection mode on response to indicators of cancer screenings by unmarried middle-aged and older women.

Methods: Three survey methods were evaluated for collecting data about mammography and Papanicolaou (hereafter, Pap) testing among heterosexual and sexual minority (e.g., lesbian and bisexual) women. Women ages 40–75 were recruited from June 2003 – June 2005 in Rhode Island. They were randomly assigned to receive: Self-Administered Mailed Questionnaire [SAMQ; N = 202], Computer-Assisted Telephone Interview [CAT; N = 200], or Computer-Assisted Self-Interview [CASI; N = 197]. Logistic regression models were computed to assess survey mode differences for 13 self-reported items related to cancer screenings, adjusting for age, education, income, race, marital status, partner gender, and recruitment source.
Results: Compared to women assigned to CATI, women assigned to SAMQ were less likely to report two or more years between most recent mammograms (CATI = 23.2% vs. SAMQ = 17.7%; AOR = 0.5, 95% CI = 0.3 – 0.8) and women assigned to CASI were slightly less likely to report being overdue for mammography (CATI = 16.5% vs. CASI = 11.8%; AOR = 0.5, 95% CI = 0.3 – 1.0) and Pap testing (CATI = 14.9% vs. CASI = 10.0%; AOR = 0.5, 95% CI = 0.2 – 1.0). There were no other consistent mode effects.

Conclusion: Among participants in this sample, mode of data collection had little effect on the reporting of mammography and Pap testing behaviours. Other measures such as efficiency and cost-effectiveness of the mode should also be considered when determining the most appropriate form of data collection for use in monitoring indicators of cancer detection and control.

Handheld computers for self-administered sensitive data collection: A comparative study in Peru

Background: Low-cost handheld computers (PDA) potentially represent an efficient tool for collecting sensitive data in surveys. The goal of this study is to evaluate the quality of sexual behavior data collected with handheld computers in comparison with paper-based questionnaires.

Methods: A PDA-based program for data collection was developed using Open-Source tools. In two cross-sectional studies, we compared data concerning sexual behavior collected with paper forms to data collected with PDA-based forms in Ancon (Lima).

Results: The first study enrolled 200 participants (18–29 years). General agreement between data collected with paper format and handheld computers was 86%. Categorical variables agreement was between 70.5% and 98.5% (Kappa: 0.43–0.86) while numeric variables agreement was between 57.1% and 79.8% (Spearman: 0.76–0.95). Agreement and correlation were higher in those who had completed at least high school than those with less education. The second study enrolled 198 participants. Rates of responses to sensitive questions were similar between both kinds of questionnaires. However, the number of inconsistencies (p = 0.0001) and missing values (p = 0.001) were significantly higher in paper questionnaires.

Conclusion: This study showed the value of the use of handheld computers for collecting sensitive data, since a high level of agreement between paper and PDA responses was reached. In addition, a lower number of inconsistencies and missing values were found with the PDA-based system. This study has demonstrated that it is feasible to develop a low-cost application for handheld computers, and that PDAs are feasible alternatives for collecting field data in a developing country.

Fear, hope and social desirability bias among women at high risk for HIV in West Africa

Background Self-reports are widely used for measuring behaviour in HIV research and prevention, yet the accuracy of these measures has been shown to be questionable in many cases. Social desirability bias (SDB) is one of the key factors identified as affecting self-report accuracy.

Methods Using in-depth interviews, we examined SDB from the perspective of 60 women at high risk for HIV in two West African countries: Ghana and Nigeria. We solicited suggestions for reducing SDB in the context of HIV research and prevention, and asked for feedback regarding methods currently being employed to reduce SDB.

Results Themes pertaining to fear and a desire to have a better life were pervasive throughout the data. Thematic structure was similar between sites and age groups, although younger women tended to be more concerned about the interview context.

Conclusions Vulnerability of a population should be considered when asking sensitive questions. Audiocomputer-assisted self-interviews may not be appropriate for vulnerable populations in developing countries, particularly for older respondents.
The donor health assessment questionnaire: potential for format change and computer-assisted self-interviews to improve donor attention

BACKGROUND: The Canadian donor health assessment questionnaire (DHAQ) has developed gradually over many years. The purpose of this study was to determine whether the format and method of administration of the DHAQ influences donor attentiveness.

STUDY DESIGN AND METHODS: Between May 2004 and September 2005, a total of 1397 donors participated in the study by completing the DHAQ by one of three methods: the current Canadian Blood Services (CBS) format and method of administration, which is partly self-administered and partly interviewer administered (Method 1); the DHAQ reformatted to the AABB Uniform Donor Health Questionnaire format and self-administered (Method 2); and an audiovisual computer-assisted self-interview (CASI; Method 3). This was followed by a short, scripted interview assessing recall of 17 specific items queried on the DHAQ. Time to completion of the DHAQ and degree of familiarity with computer use were also assessed.

RESULTS: The percentages of donors identifying all 17 items correctly were 9.4, 20.9, and 34.8 percent and the mean percentages recall of items were 53.9, 56.0, and 69.8 percent with DHAQ administration Methods 1, 2, and 3, respectively (p < 0.0001). This difference was largely attributable to the poor recall of items queried as part of a list in Method 1. Mean times to complete the DHAQ were 3.1, 3.8, and 8.1 minutes for Methods 1, 2, and 3, respectively. More than 95 percent of donors had used a personal computer in the past year.

CONCLUSION: The current format of the Canadian DHAQ is not optimal for donor attention to specific questions asked as part of a list. Attention was improved by use of AABB uniform donor history questionnaire format and was best with use of a CASI format.

Test-Retest Reliability of a Sexual Behavior Interview for Men Residing in Brazil, Mexico, and the United States

Understanding the natural history of sexually transmitted infections requires the collection of data on sexual behavior. However, there is concern that self-reported information on sexual behavior may not be valid, especially if study participants are culturally and linguistically distinct. The authors completed a test-retest reliability study of 1,069 men recruited in Brazil, Mexico, and the United States in 2005 and 2006. All of the men completed the same computer-assisted self-interview approximately 3 weeks apart. Refusal rates, kappa coefficients, and intraclass correlation coefficients were calculated for the full sample and by country, age, and lifetime number of female sex partners. Reliability coefficients for each study site and the combined population were high for almost all questions. With few exceptions, the authors found high test-retest reliability with a computer-assisted self-interview on sexual behavior used in 3 culturally and linguistically distinct countries.
References


