



Published in final edited form as:

*AIDS Behav.* 2010 December ; 14(6): 1320–1329. doi:10.1007/s10461-010-9761-5.

## Low numeracy predicts reduced accuracy of retrospective reports of frequency of sexual behavior

Timothy L. McAuliffe<sup>1</sup>, Wayne DiFranceisco<sup>1</sup>, and Barbara R. Reed<sup>2</sup>

<sup>1</sup>Center for AIDS Intervention Research (CAIR) Department of Psychiatry and Behavioral Medicine Medical College of Wisconsin Milwaukee, Wisconsin

<sup>2</sup>Falls Church, Virginia (formerly with the Medical College of Wisconsin)

### Abstract

Assessment of the frequency of sexual behavior relies on participants' ability to arithmetically aggregate information over time and across partners. This study examines the effect of numeracy (arithmetic skills) on the accuracy of retrospective reports of sexual behavior. For 91 days, participants completed daily reports about their sexual activity. Participants then completed a survey on sexual behavior over the same period. The discrepancies between the survey-based and the diary-based measures of frequency of vaginal and anal intercourse were evaluated. Multiple regression analysis showed that the discrepancy between retrospective and diary measurements of sexual intercourse increased with lower numeracy ( $p=.026$ ), lower education ( $p=.001$ ), aggregate question format compared to partner-by-partner format ( $p=.031$ ) and higher frequency of intercourse occasions ( $p<.001$ ). Lower numeracy led to a 1.5-fold increase (adjusted-mean=14.1 to 20.9) in the discrepancy for those using the aggregate question format and a 2.0-fold increase (adjusted-mean=3.7 to 7.6) for those using the partner-by-partner format.

### Keywords

Numeracy; accuracy; surveys; sexual behavior; frequency

## INTRODUCTION

Accurate self-report data on individual sexual risk behavior are essential to the evaluation of HIV prevention interventions. Evaluation of HIV prevention interventions often relies on the measurement of reduction in sexual risk behaviors, including frequencies and types of sexual activities and the contexts in which these behaviors occur. The most common source for risk behavior data is the retrospective survey.

The development of computer-administered self-interview (CASI) methodologies has improved the way sexual behavior data are collected retrospectively. CASI methodology affords participants greater privacy than person-to-person interviews for reporting socially

---

Correspondence: Timothy L. McAuliffe, Ph.D., Center for AIDS Intervention Research (CAIR), Department of Psychiatry and Behavioral Medicine, Medical College of Wisconsin, 2071 N Summit Avenue, Milwaukee, WI 53202 USA., Fax: 414-287-4206, Voice: 414-955-7700, mcauliff@mcw.edu.

sensitive health behaviors, like sexual activity; reduces the literacy demands when audio-enhanced (ACASI) technology is employed; and simplifies the completion of complex surveys. Question formats that may cue the participant's recall of sexual activities by providing a context and focus for recall of their past experiences can also reduce inaccuracies in self-report data brought about by incomplete recall.

However, even after refinements in survey question format and use of CASI techniques, substantial error in participants' retrospective reports of sexual activity remains. [1] While much of this inaccuracy—both underestimation and overestimation—may be due to recall bias or self-presentation (e.g., social desirability) bias, an additional source of error may arise from the difficulties that some participants face when working with numbers that are implicit in both the recall and integration of information required to answer questions on sexual risk behavior. Thus, low arithmetic ability may further affect the accuracy of self-reports of sexual behavior.

Numeracy is defined as the knowledge and skills to apply arithmetic operations using numbers embedded in printed materials (e.g., balancing a checkbook, computing change or a tip, or completing an order form). [2] Evidence suggests that many persons have a poor sense of numbers or difficulty working with numbers to cope with the practical demands of everyday life. In a large survey of adult literacy, half were either unable to perform calculations typical of everyday life that involve the use of basic arithmetic skills (primarily addition) or more than a single basic calculation. Persons who performed low in arithmetic skills were more likely to have less than a high school education, be of nonwhite race, and be unemployed. [3-4]

Low numeracy has been recognized as a factor associated with poor adherence with medication regimens, poor perception of cancer risk and poor quality of life assessment. [5-11] Similarly, numeracy problems appear to be associated with inaccuracies that occur when individuals report about their health behaviors, including their sexual activities. Those who have difficulty working with numbers may face difficulty in recall and integration of information on their behavior and revert to guessing. For infrequent or irregularly-occurring behaviors, a recall-and-count strategy is most often employed by respondents. [12] However, when the frequency of activity exceeds 7 to 10 occurrences, or when recalling behavior that occurred over a longer time period, research has shown that estimation strategies are often used for recalled activity. [13-14] Respondents may employ highly idiosyncratic strategies, of varying accuracy, when reporting details about their sexual behavior. [15-17]

Several measures have been used to assess health numeracy. Generally, these measures consist of 3 to 7 items and have focused on the use of probabilities and percentages, [9-10, 18] or the number of pills to be taken per day. [7, 19-20] Longer instruments have been designed to assess a broader range of skills. [21]

This study examines the effect of numeracy on the accuracy of retrospective self-reports of sexual behavior. For 91 days, participants completed a daily report (diary) of their sexual activity and returned the diary by mail the same day. At the end of the 91 days, participants

completed a retrospective survey on their sexual activity during the diary period and a measure of numeracy. The accuracy of their retrospective report was judged against data in their daily-completed reports (diaries). We hypothesized that participants with greater numeracy would more accurately report in retrospective survey the frequency of their sexual behavior.

## METHODS

### Participants and Procedures

Participants were recruited over an 11-month period in 2003, during an ongoing study of the effects of question format and mode of administration on the accuracy of retrospective sexual behavior reports. Within 5 days of taking the follow-up survey for the main study, volunteers returned for personal interviews on recall strategies used in completing their retrospective reports. Eighty-one (32%) of the 254 subjects invited agreed to participate. The most common reason given for not participating was time availability (44%).

Details of the recruitment and enrollment procedures of the main study are described elsewhere. [1] Visitors at two Milwaukee, Wisconsin urban healthcare clinics and an anonymous HIV testing clinic were invited to complete a one-page questionnaire to determine their eligibility to participate in a study of the reliability of sexual behavior self-reports. In the main study, 493 sexually active men and women, 18 years or older, and not in an exclusive relationship with one partner at enrollment completed a daily diary of sexual activities for 91 days and returned each report by mail the same day. At the end of the 91-day period, participants then completed a retrospective survey on their sexual behavior occurring during the same 91-day period.

### Measures

**Sexual Activity Diary**—A structured daily diary form allowed the separate recording of all sexual activities for up to two sexual partners, with supplementary forms available for recording activities with each additional sexual partner on the same day. For days when the participant had sex, they recorded the date of the activity, and for each partner with whom they had sex with that day, the number of times that they had vaginal intercourse and anal intercourse, and the number of times that condoms were used for each type of intercourse. Overall, 71% of the diaries were postmarked within one mail-collection day of the activity date and 94% within two days.

**Retrospective Sexual Behavior Surveys**—After submitting their final diaries for the 3-month period, participants then returned to complete a retrospective survey on their sexual behavior occurring over the 3-month diary period. Participants completed the retrospective survey within five days of their final diary—94% completed the survey within two days of submitting their final diary.

In the retrospective surveys, participants reported their total number of sexual partners, the number of times that they engaged in each type of sexual activity (anal and vaginal intercourse), and the number of times that a condom was used during each type of activity

Participants were randomly assigned to one of six experimental retrospective sexual behavior survey conditions, representing a combination of one survey question format (aggregate or partner-by-partner) and one mode of administration (paper questionnaire, non-audio CASI, audio-enhanced CASI). Respondents provided complete sexual-activity information for the 3-month period in terms of actual frequencies (i.e., the number of times that they engaged in vaginal intercourse, or in anal intercourse, and the number of occasions for which a condom was used). Participants randomized to the “aggregate” survey question format condition reported their total number of different sexual partners, the total number of times that they engaged in each type of sexual intercourse, and the total number of times that a condom was used during each activity summed across all sexual partners or a specific subgroup of these partners over a 3-month time period. Participants in the “partner-by-partner” question format condition also reported the number of different partners with whom they had sex and then, individually for each partner (up to 12), described their sexual activity (e.g., the number of times that they engaged in vaginal intercourse with partner “X” in the past 3-month period, and the number of occasions for which a condom was used with partner “X” in the past 3-month period). Sexual activities with any remaining partners beyond 12 were reported in the aggregate, separately.

Surveys were administered privately. A facilitator read a set of general instructions to the participant before starting the survey and was available for further explanation of instructions or to read questions, if the participant asked for assistance. Among the participants in the current study, 21 completed the retrospective survey as a self-administered paper questionnaire, 37 as an audio-enhanced CASI, and 23 as a non-audio CASI; also, 38 were asked about their sexual behavior in a partner-by-partner survey question format and 43 in an aggregate question format.

**Numeracy**—To assess numeracy, three questions were asked that invoke the use of arithmetic skills typical of those involved when answering questions eliciting frequency information about recalled sexual behavior. To ensure the content validity of the numeracy scale, the construction of these questions was guided by a 5-level quantitative literacy model used in national adult literacy surveys. [3] In a large national population survey, three of the quantitative literacy levels characterized 78% (22%, 25% and 31%, respectively) of adults. [3] The questions used to assess numeracy in this study target these levels of quantitative skills. Figure 1 shows the numeracy questions and the corresponding level of the quantitative literacy criterion that each addresses.

The first question asked the respondent to assess their ability to add numbers in their mind. A 5-point response scale (1=poor, 2=fair, 3=moderate, 4=good, and 5=excellent) was used. A rating of moderate to excellent ability was the targeted response. The second question describes a task, to calculate a percentage, for which the numbers are stated in the question. The response to this question was given as a percentage (%), using an open format. The third question describes a task that involves performing two or more arithmetic operations (multiplication and addition) sequentially for which the numerical information and the operations are inferred from the question text. The answer to the third question was given as a number, using an open format. The numeracy score was calculated by summing the number of correct or targeted responses for the three numeracy questions—Question 1

adding numbers in their mind (1=moderate-to-excellent ability, 0=poor-to-fair ability); Question 2, calculate a percentage (1=correct, 0=not correct); and Question 3, perform a calculation involving multiple additions and multiplications (1=correct, 0=not correct).

## Data Analysis

For each participant, the difference in total numbers of vaginal or anal intercourse occasions (times) reported in retrospective survey and collectively in their diaries was calculated. The absolute value of the difference (i.e., ignoring the sign) was calculated to reflect the magnitude of the difference in either direction. The means and standard deviations for the magnitude of the difference (discrepancy) between the retrospective and diary measurements are presented. Analysis of variance was used to examine the univariate effect of numeracy on the magnitude of the discrepancy between retrospective and diary reports of number of vaginal or anal intercourse occasions (times) over the 3-month period.

Multiple regression analysis was performed to test for independent predictors of the magnitude of the discrepancy between prospective diary-based and retrospective survey-based measurements of total number of vaginal or anal intercourse occasions over the same 3-month reporting period. The regression model includes indicator variables for partner-by-partner question format (versus aggregate), for audio-enhanced CASI and non-audio CASI (versus self-administered paper questionnaire) modes of collection used in the retrospective surveys, and for interactions. Ordinal covariates included in the model are the numeracy score, level of education (1=less than high school degree, 2=high school degree or GED or skills training certificate, and 3=some college), and the number of vaginal and anal intercourse occasions reported in retrospective survey and in daily diaries over the 3-month period. To reduce skewness in the magnitude of the discrepancy, the data are square-root transformed for regression analysis. Adjusted group estimates of the magnitude of the discrepancy are evaluated at the mean values for other covariates in the regression equation, and then back transformed.

## RESULTS

### Participant Characteristics

The sample (60 men and 21 women) had an average age of 38.3 years, and was 58% (n=47) African-American, 38% (n=31) Caucasian, and 4% (n=3) of other or mixed racial backgrounds. Fifty-eight percent (n=47) of participants were full-or part-time employed and 27% (n=22) were unemployed but actively seeking job opportunities. The median annual income was between \$10,000 and \$20,000. Fifteen percent (n=12) of the sample had less than a high school education, 36% (n=29) completed high school, GED or technical training, 36% (n=29) had some college education, and 14% (n=11) had completed a college baccalaureate. The Marlowe-Crowne Social Desirability scale [22] was used to measure the tendency to reply in surveys in a manner that would be viewed favorably by others. The mean social desirability score was 17.2 (SD=6.4 on the scale 0 to 33).

## Numeracy

Table 1 presents descriptive statistics for the numeracy items and the calculated numeracy score. Three numeracy questions ask respondents about counting and working with numbers in their heads. Twenty-two percent of participants judged their ability to perform basic addition of numbers in their heads as poor to fair (5% poor, 17% fair) and 78 percent as moderate to excellent (31% moderate, 32% good and 15% excellent). Thirty-one percent of subjects determined the correct percentage when asked the question “A raffle awards ten prizes. Two hundred tickets are entered. What percentage of tickets will win a prize?”. Twenty-six percent of subjects answered a *question that* involved multiple additions and multiplications correctly (“During the 13 weeks of summer, it rained on 3 days during 1 week, on 2 days each week during 5 weeks, on 1 day each week during 3 weeks, and not at all during 4 weeks. How many days did it rain during the summer?”). These tasks involve the use of quantitative skills implicit when answering questions elicit reporting about the frequency of sexual activity over some time period. Eighteen (22%) participants reported difficulty adding numbers in their heads; 34 (42%) rated their ability to add numbers in their heads as moderate-to-excellent (able) but were unable to correctly answer either of the other numeracy questions; 20 (25%) were able to add numbers extemporaneously and, also, correctly answered one of the two other numeracy questions; and 9 (11%) were able to add numbers in their heads and, also, answered both of the other numeracy questions correctly.

The mean numeracy score (number of correct or targeted—moderate-to-excellent rating—responses) was 1.35 (SD=.87, range=0 to 3). The numeracy scores increased with level of education completed (means, 0.83 with less than high school, 1.24 with high school diploma or GED certificate, and 1.53 with some college; Spearman’s  $\rho=.29$ ), were higher for Caucasians compared to other races (means, 1.71 versus 1.12;  $\rho=.30$ ), and higher for men than for women (means, 1.50 versus .90;  $\rho=.29$ ). However, no demographic characteristic was found to be an independent predictor of the numeracy score.

## Comparison of retrospective and prospective (diary) measures of frequency of sexual activity

Overall, the magnitude of the difference (ignoring the sign) between the retrospective and the diary-based measurements of total number of vaginal and anal intercourse occasions during the same 3-month period had a mean of 19.3 (SD=25.4). Figure 2 presents a Bland-Altman scatter plot [23] of the difference between the retrospective and prospective measurements and the average of the measurements for the total number (frequency) of vaginal and anal intercourse occasions. The 95% limits of agreement between the two measurements range from a negative -71.2 to 46.9. Figure 2 shows that the differences trends toward larger negative values (i.e., fewer occasions are reported in retrospective survey). Figure 2, also shows that the variance of the differences increases with frequency of activity, suggesting that other factors could account for some of the variation in the difference between the paired measurements of frequency of intercourse. Table 2 gives the means, standard deviations and univariate tests of significance of the magnitude of the difference between participants’ retrospective self-report and their diary (prospective) measurements of total number of vaginal and anal intercourse occasions over the 3-month period by experimental survey groups, numeracy score and level of education. The

unadjusted difference between the retrospective survey and the diary-based measurements of the number of intercourse occasions was 13 times larger for respondents with lower numeracy compared to those with high scores (means, 40.3 for numeracy score=0, 18.5 for numeracy score=1, 16.0 for numeracy score=2, and 2.9 for numeracy score=3,  $r$ -squared=.15,  $p<.001$ ). The magnitude of the discrepancy was 2 times larger for respondents with less than high school education compared to those with some college education (means, 33.6 for less than high school education, 18.9 for high school degree/GED/skills certificate, and 15.4 for some college education,  $r$ -squared=.12,  $p=.001$ ). The relationships of the magnitude of discrepancy with numeracy and level of education were found to be linear (tests of non-linearity,  $p=.74$  and  $p=.62$ , respectively). The discrepancy between the retrospective survey and the diary-based measurements of the number of intercourse occasions over the same 3-month period was larger for respondents who were asked about their sexual behavior in an aggregate survey question format than those who were asked in a partner-by-partner question format (means, 23.0 for the aggregate format group versus 5.2 for the partner-by-partner format group), and larger for respondents who completed the retrospective survey as a self-administered paper questionnaire than those who completed the retrospective CASI survey (means, 22.8 for paper questionnaire group, 18.6 for audio CASI group, and 17.2 for non-audio CASI group), but these differences were not statistically significant.

Figure 3 shows the median and interquartile range for the magnitude of the difference in measurements of total frequency of vaginal and anal intercourse. This figure shows that the magnitude of the median difference decreased from 32 for those with numeracy score of 0 to 1 for those with a numeracy score of 3.

### **Effect of numeracy on the discrepancy between retrospective survey and diary-based measurements**

In multiple regression analysis, we examined the effect of numeracy on the magnitude of the discrepancy between the number of vaginal and anal intercourse occasions reported in retrospective surveys and in daily diaries over the same 3-month period. The dependent variable, magnitude of the difference, was square-root transformed for analysis in minimizing skewness in the data. In addition to the numeracy score, our model included indicator variables for the partner-by-partner question format and for the ACASI and non-audio CASI modes of administration of retrospective survey, and interaction terms. The model also included the two ordinal covariates level of education and the average of the subject's total frequency of vaginal and anal intercourse recorded in the retrospective survey and in the participant's diaries. Limited by sample size and multicollinearity, interactions between numeracy and partner-by-partner format and between partner-by-partner format and modes of administration were included in the model.

Table 3 presents the results of the multivariate analysis for discrepancy between measurements of frequency of vaginal and anal intercourse over the 3-month period. Numeracy ( $t = -2.21$ ,  $p=.031$ ), partner-by-partner question format ( $t = -2.20$ ,  $p=.031$ ), level of education ( $t = -3.60$ ,  $p=.001$ ), and number of vaginal and anal intercourse occasions reported in survey and diaries ( $t=13.21$ ,  $p<.001$ ) were found to be independent predictors of the magnitude of the discrepancy between the survey-based and diary-based measurements.

Larger discrepancies were associated with lower numeracy scores, aggregate question format, lower education, and higher frequency of sexual activity. Neither mode of survey administration nor any interactions terms were found to be significant. The multiple regression model explained 75% of the variation (adjusted R-squared=.75,  $F [9, 71]=27.3$ ,  $p<.001$ ) in the magnitude of the discrepancy between retrospective and diary-based measures of number of sexual intercourse occasions. Table 4 presents the adjusted mean and 95% confidence interval estimates of the magnitude of the discrepancy by numeracy level and question format. The magnitude of the discrepancy between measurements of the number of intercourse occasions over the 3-month period increased with lower levels of numeracy for respondents in both question format groups. For respondents in the aggregate question format group, the adjusted mean increased from 14.06 for respondents with numeracy score 3 to mean 20.86 for respondents with numeracy score 0. For respondents in the partner-by-partner question format group, the adjusted mean increased from 3.72 to 7.55 with lower levels of numeracy.

## DISCUSSION

Efforts to improve how participants answer questions about their behavior have focused primarily on issues of literacy and comprehension, and on reduction in presentation and recall biases. The use of ACASI methodology, together with the framing of questions about sexual risk behavior in a “partner-by-partner” format, can improve the accuracy of retrospective reports compared to traditional self-administered paper questionnaires, which use question formats that require participants to aggregate their behavior across all partners. However, even with these refinements, there can be a substantial degree of error in the data. [1]

The accuracy of a participant’s response to a question on frequency (number of acts) of sexual activity relies on the ability to both recall the behavior and to count the number of occurrences. The response may involve counting the number of distinct sexual encounters with each different partner during the recall period. Thus, the accuracy of response to questions about frequency of activity relies on the participant’s arithmetic or quantitative skills (numeracy). No published research has explored the role of numeracy in the accuracy of retrospective self-reports of sexual behavior.

For 91 consecutive days, participants completed diaries of their sexual activity and returned the diaries by mail. At the end of the 91-day period, participants then completed a retrospective survey on sexual activity over the same period, along with a measure of numeracy. We hypothesized closer agreement between retrospective survey-based and prospective diary-based measurements of sexual activity among participants with greater numeracy skills.

To measure numeracy, three questions were asked to assess the type of quantitative skills used when answering questions about frequency of sexual behavior. Construction of these questions was guided by a quantitative literacy model that has been used in national adult literacy surveys. [3] Questions used to assess numeracy targeted three quantitative skill levels. Because formulation of an answer for a question about the frequency of sexual



activity is done primarily in one's mind, the numeracy questions focused on their ability to add numbers extemporaneously. One question involved calculation of a percentage. Another question involved calculation of a number where a series of additions and multiplications were required. The latter type of scenario arises when aggregating activity counts across all sexual partners. In our sample, 22% of participants rated their ability to add numbers in their heads as poor to fair (78% moderate-to-excellent), 31% correctly answered a question involving the calculation of a percentage, and 26% correctly calculated a number involving a series of inferred additions and multiplications. In comparison, a national survey found that the three targeted levels of quantitative literacy characterized 78% (22% Level 1, 25% Level 2, and 31% Level 3) of adults surveyed. [3] The numeracy score was calculated based on the answers for the three numeracy questions: moderate-to-excellent ability to add numbers in their heads, correctly calculate a percentage, and correctly calculating a number involving multiple additions and multiplications. In our sample, numeracy scores increased with level of education and were higher for Caucasians and men. However, no single demographic characteristic was found as an independent predictor of the numeracy score. The numeracy score accounted for 15% of the variance of the magnitude of the difference between retrospective-based and diary-based measures of the number of sexual intercourse occasions for the same period. Though numeracy and level of education are correlated in our sample, in multivariate analysis, we found both to be independent predictors of the magnitude of the discrepancy.

Our sample of 81 adults was racially diverse, was broadly aged (mean=38 years, range=23 to 73) and included both men and women. Our findings show that the difference between retrospective and diary (prospective) measurements of frequency of sexual intercourse, even when adjusted for frequency of sexual activity, education, question format and mode of administration used in retrospective survey, is greater for participants with low numeracy. Since disadvantaged populations are frequently the target of HIV prevention programs, low numeracy may adversely affect efforts to evaluate the efficacy of these programs.

We acknowledge the limitations of this study and the need for caution when interpreting our findings. With the modest sample size, it was not possible to investigate potential effects of demographics factors beyond numeracy and education on the discrepancy between retrospective and diary-based measurements. Additional research with larger samples is necessary to validate these results. Another limitation of these findings is that both retrospective- and diary-based measurements are based on self-report data. We are unable to determine the extent to which these self-report data represent the respondents' actual behaviors. Although diaries—near-contemporaneous reports of sexual behavior—were completed daily and returned by mailed, paper diaries lack completion-time information to support the prospective validity of the data. Further research utilizing electronic diary technology could be used to validate the findings. This study investigated the impact of numeracy on the accuracy of self- and computer-administered retrospective sexual behavior survey data collected using two commonly used question formats (aggregate and partner-by-partner).

Retrospective self-reports of sexual behavior commonly used in the HIV research field can misrepresent the sexual activities of participants of low numeracy. Much of this inaccuracy

may be due to the need for participants to aggregate frequencies of behavioral events. Sexual behavior questions that require participants to aggregate frequency information about their behavior across partners and over time put a heavy burden on participants with low numeracy skills. Based on our findings that the framing of questions about the frequency of sexual behavior in an aggregate format and low numeracy among respondents are associated with reduced accuracy of retrospective sexual behavior self-report, we recommend that researchers frame questions about sexual risk behavior in a “partner-by-partner” format to mitigate the impact of low numeracy in the evaluation of HIV sexual risk behavior reduction programs. We further recommend that HIV prevention program evaluation take into consideration numeracy among the participants. With computer-administered interviewing techniques together with framing of sexual behavior frequency questions in a “partner-by-partner” format, participants are only asked as many questions as they have different sexual partners. Further inquiry is needed to see if numeracy impacts data collected in an event format (e.g., timeline-followback) similarly. Alternatively, diary methodology may avoid the impact of numeracy in the collection of data on frequency of behavior by eliminating the need for a participant to aggregate numbers of events across days or weeks. The findings of this study may also be relevant for other research fields in which behavioral data are collected in a retrospective manner, as the numbers of times that behavior events occurred.

## Acknowledgments

This research was supported by grants R01-MH62961 and P30-MH52776 from the National Institute of Mental Health.

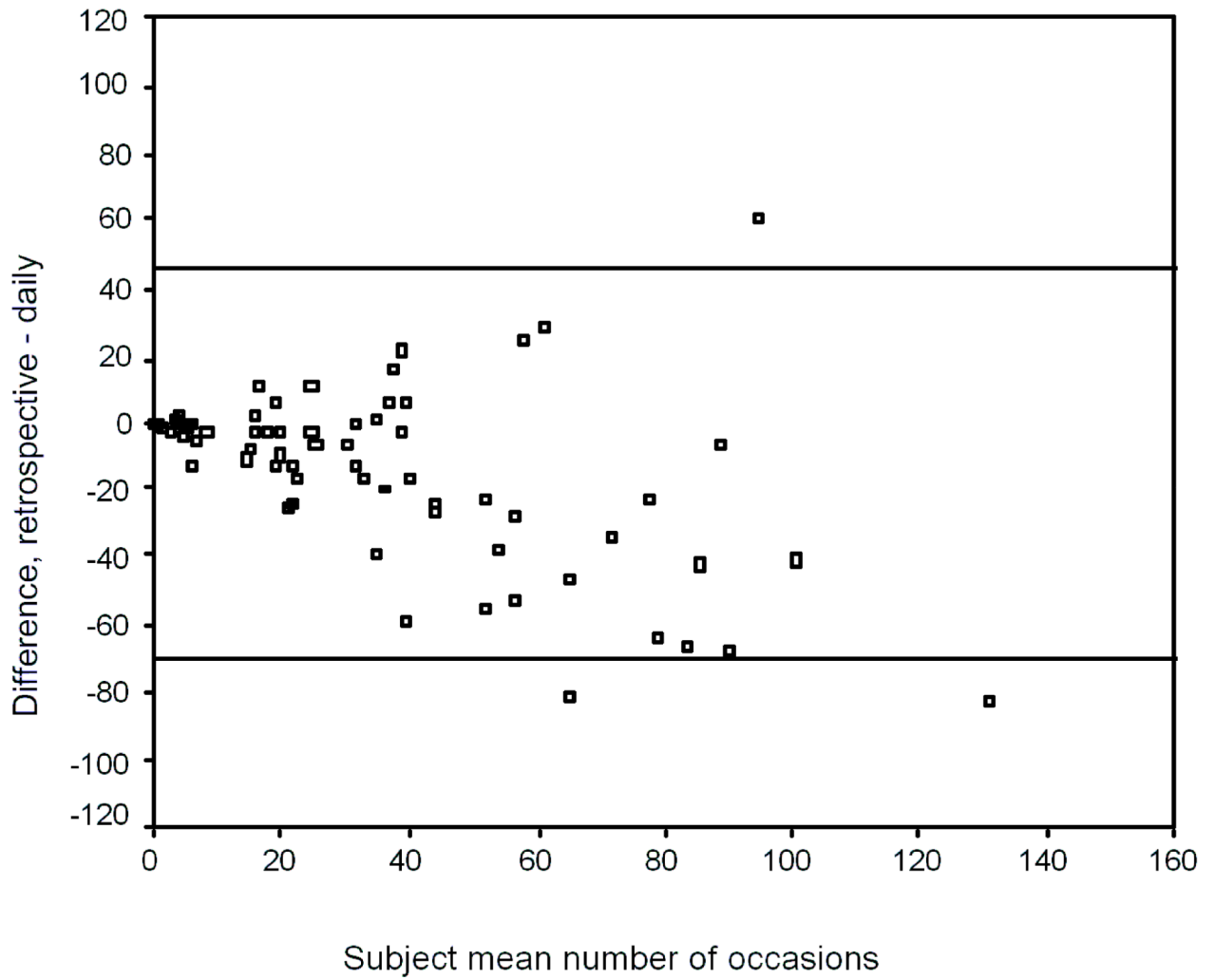
## REFERENCES

- 1). McAuliffe TL, DiFranceisco WJ, Reed BR. Effects of question format and collection mode on the accuracy of retrospective surveys of health risk behavior: a comparison with daily sexual activity diaries. *Health Psychol.* 2007; 26(1):60–7. [PubMed: 17209698]
- 2). National Center for Education Statistics. NCE 199909. NCE, US Dept of Education; Washington, DC: 1992. National Adult Literacy Survey. 1993. Available at <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=199909>
- 3). Kirsch, IS.; Jungeblut, A.; Jenkins, L.; Kolstad, A. NCE 93275. NCE, US Dept of Education; Washington, DC: 1993. Adult literacy in America: A first look at the results of the National Adult Literacy Survey. Available at <http://nces.ed.gov/pubs93/93275.pdf>
- 4). Kutner, M.; Greenberg, E.; Jin, Y.; Boyle, B.; Hsu, Y.; Dunleavy, E. NCE 2007480. NCE, US Dept of Education; Washington, DC: 2007. Literacy in everyday life: Results from the 2003 National Assessment of Adult Literacy. Available at <http://nces.ed.gov/Pubs2007/2007480.pdf>
- 5). Schwartz SR, McDowell J, Yueh B. Numeracy and the shortcomings of utility assessment in head and neck cancer patients. *Head Neck.* 2004; 26:401–7. [PubMed: 15122656]
- 6). Woloshin S, Schwartz LM, Moncur M, Gabriel S, Tosteson AN. Assessing values for health: numeracy matters. *Med Decis Making.* 2001; 21:382–90. [PubMed: 11575488]
- 7). Estrada CA, Martin-Hryniewicz M, Peek BT, Collins C, Byrd JC. Literacy and numeracy skills and anticoagulation control. *Am J Med Sci.* 2004; 328(2):88–93. [PubMed: 15311167]
- 8). Cavanaugh K, Huizinga MM, Wallston KA, Gebretsadik T, Shintani A, Davis D, et al. Association of numeracy and diabetes control. *Ann Intern Med.* 2008; 148:737–46. [PubMed: 18490687]
- 9). Schwartz LM, Woloshin S, Black WC, Welch HG. The role of numeracy in understanding the benefits of screening mammography. *Ann Intern Med.* 1997; 127:966–72. [PubMed: 9412301]

- 10). David SL, Schapira MM, McAuliffe TL, Nattinger AB. Predictors of pessimistic breast cancer risk perceptions in a primary care population. *J Gen Intern Med.* 2004; 19:310–5. [PubMed: 15061739]
- 11). Black WC, Nease RF, Tosteson AN. Perceptions of breast cancer risk and screening effectiveness in women younger than 50 years of age. *J Natl Cancer Inst.* 1995; 87(10):720–30. [PubMed: 7563149]
- 12). Sudman, S.; Bradburn, NM.; Schwarz, N. Thinking about answers: The application of cognitive processes in survey methodology. Jossey-Bass; San Francisco: 1996.
- 13). Burton S, Blair E. Task conditions, response formulation process, and response accuracy for behavioral frequency questions in surveys. *Public Opin Q.* 1991; 55:50–79. Spring.
- 14). Blair E, Burton S. Cognitive processes used by survey respondents to answer behavioral frequency questions. *J Consumer Res.* Sep.1987 14:280–8.
- 15). Bradburn NM, Rips L, Shevell S. Answering autobiographical questions: The impact of memory and inference on surveys. *Science.* 1987; 236:157–61. [PubMed: 3563494]
- 16). Croyle, RT.; Loftus, EF.; Klinger, MR.; Smith, KD. Reducing errors in health related memory: Progress and prospects. In: Schement, JR.; Ruben, BD., editors. *Between communication and information: Information and behavior.* Vol. 4. Transaction; Piscataway (NJ): 1993. p. 255-68.
- 17). Menon G. The effects of accessibility of information in memory on judgments of behavioral frequencies. *J Consumer Res.* 1993; 20:431–40.
- 18). Lipkus IM, Samsa G, Rimer BK. General performance on a numeracy scale among highly educated samples. *Med Decis Making.* 2001; 21:37–44. [PubMed: 11206945]
- 19). Parker RM, Baker DW, Williams MV, Nurss JR. The test of functional health literacy in adults: A new instrument for measuring patient's literacy skills. *J Gen Intern Med.* 1995; 10(10):537–41. [PubMed: 8576769]
- 20). Williams MV, Parker RM, Baker DW, Parikh NS, Pitkin K, Coates WC, et al. Inadequate functional health literacy among patients at two public hospitals. *JAMA.* 1995; 274(21):1677–82. [PubMed: 7474271]
- 21). Huizinga MM, Elasy TA, Wallston KA, Cavanaugh K, Davis D, Gregory RP, et al. Development and validation of the diabetes numeracy test (DNT). *BMC Health Serv Res.* 2008; 8:96+. Available from: <http://www.biomedcentral.com/1472-6963/8/96>. [PubMed: 18452617]
- 22). Crowne D, Marlowe D. A new scale of social desirability independent of psychopathology. *J Consult Psychol.* 1960; 24(4):349–54. [PubMed: 13813058]
- 23). Bland JM, Altman DG. Statistical methods for assessing agreement between two methods if clinical measurement. *Lancet.* 1986; 327(8476):307–10. [PubMed: 2868172]

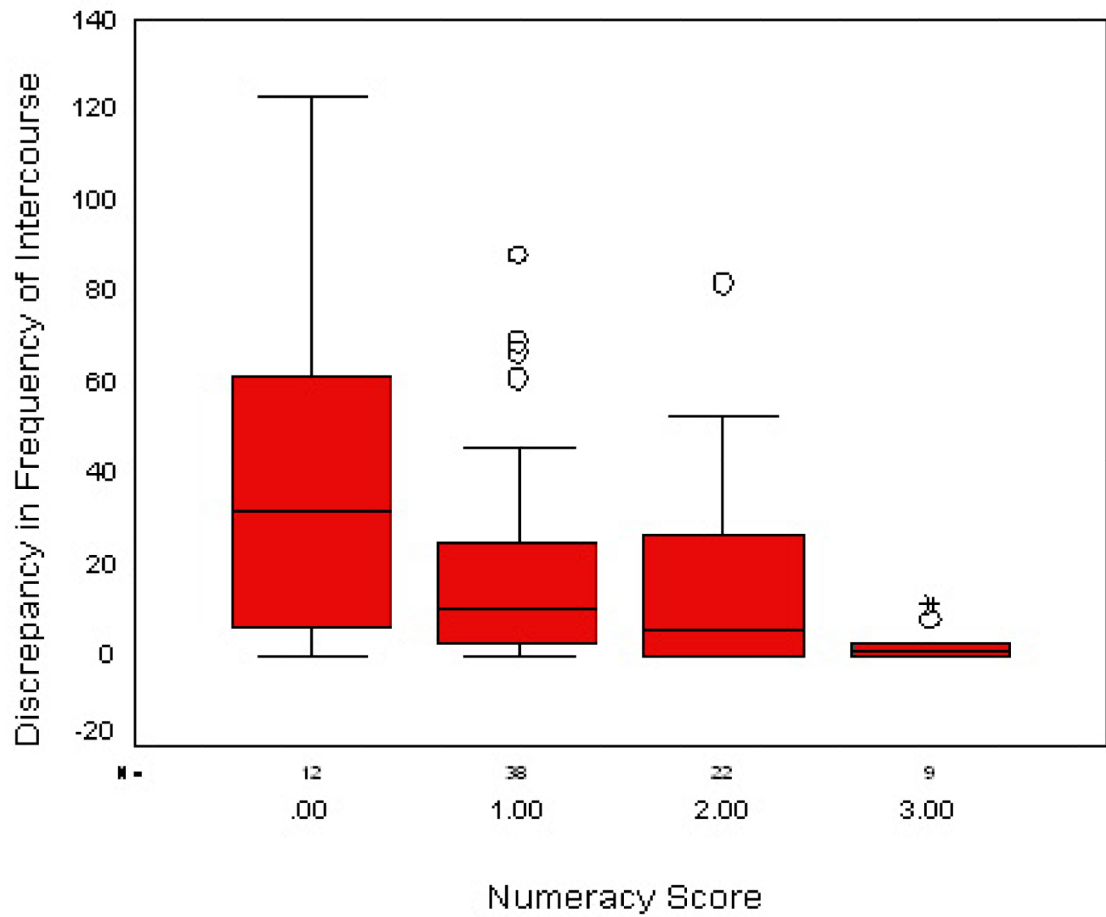
Level	Numeracy Question	Description of Quantitative Literacy Level <sup>a</sup>
1	Using one of the terms, 1=poor, 2=fair, 3=moderate, 4=good or 5=excellent, what best indicates your ability to add numbers in your head? Response: Categorical	"Tasks in this level require readers to perform single, relatively simple arithmetic operation, such as addition. The numbers to be used are provided and the arithmetic operation to be performed are specified."
2	A raffle awards ten prizes. Two hundred tickets are entered. What percent of the tickets will win a prize? Response: Percent (open format)	"Tasks in this level typically require readers to perform a single operation using numbers that are either stated in the task or easily located in the material. The operation to be performed may be stated in the question or easily determined from the format of the material (for example, an order form)."
3	During the 13 weeks of summer, it rained on 3 days during 1 week, on 2 days each week during 5 weeks, on 1 day during 3 weeks, and not at all during 5 weeks. How many days did it rain during the summer? Response: Number (open format)	"In tasks in this level, two or more numbers are typically needed to solve the problem, and these must be found in the material. The operation(s) needed can be determined from the arithmetic relation terms used in the question or directive."
<sup>a</sup> Levels 1, 2 and 3 characterized the quantitative literacy for 78% of adults study in the 1992 National Adult Literacy Survey <sup>3</sup> .		

**Figure 1.**  
Numeracy Scale Items and Target Levels of Quantitative Literacy.



**Figure 2.**

Bland-Altman plot of the difference between retrospective and diary-based measures of the frequency of sexual intercourse occasions over a 91-day period and the average of the two measures, and the 95% limits of agreement between the two measures.



**Figure 3.**

Box plots of the magnitude of errors in retrospective surveys of frequency of sexual intercourse occasions over a 91-day period when judged against activity recorded in prospective diaries by numeracy score. Box plots show the median, interquartile range, outliers, and extreme cases.

**Table 1**

Numeracy scale items and summary response characteristics.

	N	%
Using one of the terms, 1=poor, 2=fair, 3=moderate, 4=good or 5=excellent, what best indicates your ability to add numbers in your head?		
Poor	4	5
Fair	14	17
Moderate	25	31
Good	26	32
Excellent	12	15
A raffle awards ten prizes. Two hundred tickets are entered. What percent of the tickets will win a prize?		
Incorrect	56	69
Correct	25	31
During the 13 weeks of summer, it rained on 3 days during 1 week, on 2 days each week during 5 weeks, on 1 day during 3 weeks, and not at all during 5 weeks. How many days did it rain during the summer?		
Incorrect	60	74
Correct	21	26
Numeracy Score <sup>a</sup>		
0	12	15
1	38	47
2	22	27
3	9	11

<sup>a</sup>Score is the sum of the number of correct or targeted responses: Moderate-to-excellent ability to add numbers in your head, correct percentage of winning tickets, and correct number of days it rained.

**Table 2**

Univariate analyses of the magnitude of the discrepancy between retrospective survey- and diary-based measurements of number of vaginal and anal intercourse occasions during the past 3-month period.

	Difference in number of vaginal and anal intercourse occasions		
	Mean (SD)	F <sup>a</sup>	p
Overall (n=81)	19.3 (25.4)		
Question format		.86	.36
Aggregate (n=43)	23.0 (30.1)		
Partner-by-Partner (n=38)	15.2 (18.2)		
Mode of Collection		.20	.82
Paper self-administered questionnaire (n=21)	22.8 (31.4)		
Audio-enhanced Computer-administered (n=37)	18.6 (23.0)		
Non-audio Computer-administered (n=23)	17.2 (23.7)		
Numeracy score		13.60 <sup>b</sup>	<.001
0 (n=12)	40.3 (37.6)		
1 (n=38)	18.5 (22.0)		
2 (n=22)	16.0 (21.6)		
3 (n=9)	2.9 (4.3)		
Education		10.95 <sup>b</sup>	.001
Less than a high school education (n=12)	33.6 (24.0)		
High school degree, GED or skills certificate (n=29)	18.9 (21.1)		
Some college (n=40)	15.4 (27.6)		

<sup>a</sup> Analysis of variance F-test on the square-root of the magnitude of the discrepancy .

<sup>b</sup> Linear regression.



**Table 3**

Multiple regression analysis of the magnitude of discrepancy between the surveybased and the diary-based measures of number of vaginal and anal sexual intercourse occasions during the past 3-month period.

Parameter	Coefficient <sup>a,b,c</sup>	t	Significance
Partner-by-Partner (PxP) format indicator	-.326	-2.20	.031
Non-audio CASI mode indicator	-.081	-.80	.423
Audio CASI mode indicator	-.041	-.41	.682
PxP Non-audio CASI interaction	.132	1.23	.221
PxP Audio CASI interaction	.102	.89	.375
Number of intercourse occasions	.781	13.21	.000
Numeracy <sup>d</sup>	-.185	-2.21	.031
PxP Numeracy interaction	.235	1.80	.076
Education <sup>d</sup>	-.216	-3.60	.001

<sup>a</sup> Square-root of the magnitude of the discrepancy.

<sup>b</sup> Adjusted  $R^2 = .75$ ; Kolmogorov-Smirnov Test for Normal distribution of residuals with Lilliefors significance level,  $p = .20$ ; Collinearity: Condition Index=14.7, minimum tolerance=.14.

<sup>c</sup> Standardized coefficient.

<sup>d</sup> Numeracy (ordinal): 0, 1, 2, 3; Level of education achieved (ordinal): 1=less than high school education, 2=high school degree, GED or Skills training certificate, 3=some college.

**Table 4**

Adjusted mean magnitude of discrepancy between the survey-based and the diary-based measures of number of vaginal and anal sexual intercourse occasions during the past 3-month period.

Question Format	Numeracy Score	Mean <sup>a, b</sup>	95% Confidence Interval	
			Lower Bound <sup>b</sup>	Upper Bound <sup>b</sup>
Aggregate	0	20.86	14.04	29.03
Aggregate	1	18.44	12.56	25.46
Aggregate	2	16.18	10.00	23.83
Aggregate	3	14.06	7.15	23.28
Partner by Partner	0	7.55	2.68	14.90
Partner by Partner	1	6.13	2.34	11.69
Partner by Partner	2	4.85	1.72	9.56
Partner by Partner	3	3.72	0.90	8.46

<sup>a</sup> Covariates in the model are evaluated at the mean values: Number of intercourse occasions in past 3-month period = 34.6975, Audio-enhanced CASI indicator = .4568, Non-audio CASI indicator = .2840, Partner by Partner (PxP) with Non-audio CASI interaction = .1358, PxP with Audio-enhanced CASI interaction = .1975, and Education (1=Less than high school (HS) degree, 2=HS degree / GED / Skills certificate, 3=Some college) = 2.3457.

<sup>b</sup> Back transformed to the original scale.