Responding to sensitive questions in surveys: A comparison of results from Online panels, face to face, and self-completion interviews.

Yehuda Dayan, Ipsos MORI
Carina Schofield Paine, PhD, Research Executive, Ipsos MORI
AJ Johnson, Global Director and Head of Online Research, UK
Ipsos Interactive Services

Abstract

It is generally accepted that conducting surveys online is both faster and cheaper than other, more traditional, survey methodologies. Such advantages have helped to drive the growth of online surveys over recent years. In addition to the growth of online surveys, respondents are increasingly required to answer more personal and sensitive questions online. It is therefore important to investigate and understand the behaviour of respondents to sensitive questions in surveys in order to ensure the most effective methodology is employed.

A salient issue in online survey research is the removal of an interviewer. This is particularly relevant when dealing with sensitive topics - when the lack of interviewer presence can remove response bias. Much research has demonstrated that surveys administered online, without an interviewer being present, are characterised by higher levels of self disclosure (Weisband and Kiesler 1996), an increased willingness to answer sensitive questions (Tourangeau 2004) and reductions in socially desirable responding (Frick et al. 2001; Joinson 1999). Furthermore, survey methodologies that reduce the level of question administration by human interviewers (e.g. via computer-aided self interviews) also increase responses to sensitive personal questions and yield more honest, candid answers.

As part of the ongoing experimental work at Ipsos MORI we are investigating the affect of different survey methodologies on respondents' behaviour to sensitive questions.

In the present paper we present a two part study. Part 1 searches evidence of survey mode effect on disclosure levels and examines data consisting of participants interviewed in one of three conditions. In condition one, 1,645 members of the Ipsos Online Panel completed an online survey. In condition two, 902 were interviewed offline, face-to-face using Computer Assisted Personal Interviewing (CAPI) interviewing. Finally, in condition three, 1028 participants were again interviewed offline, using Computer Assisted Self Interviewing (CASI). Direct comparison were possible between the two offline samples. Allocation to the online sample, on the other hand, was not randomized thus propensity score adjustment was applied to control for possible confounding of online/offline comparisons. Respondents were asked more than 50 questions about a variety of topics from

politics to media consumption. Within these questions respondents were asked five which were deemed as sensitive. The topics for the sensitive items covered: immigration, adultery, drink driving, abortion, and attitudes toward debt.

Part 2 examined the association between the level of sensitivity and level of disclosure, and specifically any differences between the three survey modes. To estimate the social sensitivity an ad hoc panel of five experienced independent social researchers sampled from a larger pool of experts and were asked to rank levels of sensitivity of each of the five questions. After passing reliability tests of agreement between raters the estimated sensitivity was correlated with item disclosure level by mode.

Finally, implications for the handling of sensitive questions in survey research are discussed.

1 Introduction

Online research emerged in the early 1990s, but increased significantly during the beginning of the new millennium, with a significant boost in technical development and an increasing number of people connected to the internet via dial-up or broadband connections. Online methods have transformed survey data collection, and some experts estimate that in 2010, 30% of all market research will be conducted online (Wilson and Laskey 2003).

For Ipsos MORI, a global research company, online research plays an important and increasingly major role. Alongside online methods, face to face data collection is also widely used, particularly Computer Assisted Personal Interviewing (CAPI) and Computer Assisted Self Interviewing (CASI).

In addition to changes in the form of data collection over recent years, surveys are also increasingly asking for sensitive information (Tourangeau and Smith 1996). When asking sensitive questions, many studies have investigated i) the presence, versus the absence, of an interviewer and ii) the administration of interview (interviewer administration versus self administration).

Earlier studies compared the administration of the interview. For example, Waterton and Duffy (1984) compared self administered interviews to interviewer administered interviews. In their local survey, their results showed that respondents reported more alcohol consumption when interviewers were self administered. Following this, Mosher and Duffy (1994) compared Audio CASI (ACASI) with CAPI. In a National Survey of Family Growth, they found that participants reported more abortions when interviews were self administered using ACASI. Similarly, Locke et al. (1992) found that CASI elicited significantly more HIV risk factors than a conventional face to face interview when surveying potential blood donors. Respondents also rated the CASI interview as more private.

Finally, Jobe et al. (1996) manipulated both the presentation (computer versus paper) and the administration of a survey (self versus interviewer). In a study of 1000 women they found that computerisation had very little effect of level of reporting. However, self-administration (computer assisted or not) had a clear impact on the level of reporting, especially on the reporting of sexual behaviour: with respondents in the self administered interview condition disclosing more information.

More recent studies have compared the presence of an interviewer on disclosure to sensitive questions in surveys, by comparing online and offline surveys. Several experiments have demonstrated advantage of online over other interview methods for asking sensitive questions, in that surveys administered via the Internet have been associated with higher levels of self disclosure to such questions. For example, Weisband and Kiesler (1996) report a statistical meta-analysis of literature from 1969 to 1994 (39 studies using 100 measures). All of these studies showed that people disclosed more on a computer form than in an interview or on a paper form. Effect sizes were larger comparing computer administration with face to face interviews, when forms solicited sensitive information, and when medical or psychiatric patients were the subjects.

Reasons for increased levels of disclosure to online surveys (where an interviewer is removed) have described how computer mediated communication promoted visual anonymity, heightened private self-awareness combined with reduced public selfawareness (Joinson 2001). Surveys administered via the Internet have also been associated with reductions in socially desirable responding. For example, Frick et al. (2001) conducted a study in which participants completed measures of self-consciousness, social anxiety, self esteem and social desirability on-line or on paper. They were also assigned to either an anonymous or non-anonymous condition. Results showed that people reported lower social anxiety and social desirability, and higher self-esteem, when they were anonymous as opposed to when non-anonymous. Participants also reported lower social anxiety and social desirability when using the Internet compared to when using paper-based methods.

In summary, studies have consistently shown that survey methodology techniques that tend to reduce human involvement in question administration demonstrate higher levels of self disclosure and an increased willingness to answer sensitive questions. Furthermore, survey methodologies that reduce the level of question administration by human interviewers (e.g. via CASI) also increase responses to sensitive personal questions and yield more honest, candid answers.

Following the aims of the study below, the paper goes on to detail aspects of our survey design (Sections 2.1 to 2.4), after that Section 3 discusses the conceptual frame work of part 1 of our study including sample imbalances, ways of correcting them (Section 3.1 and 3.2), where subsequently Section 3.3 displays our results of average and item disclosure across the alternative survey modes examined; Section 4 summarizes the second part of our study which inspects the link between expected sensitivity and level of disclosure. Finally the paper concludes with a short discussion in Section 5.

1.1 Aims

Part 1 of the present study aims to investigate the effect of different survey methodologies on respondents' behaviour to sensitive questions in surveys. It is hypothesized that average disclosure levels will be significantly higher for surveys administered to an online panel than offline. It is also hypothesized that average disclosure levels for CASI will be significantly higher than for CAPI .

Part 2 aims to investigate the effect of sensitivity of a question on respondents' behavior. It is hypothesized that item specific disclosure is dependent on the sensitive topic addressed in the question. We also aim to test whether some of this dependency can be removed by mode of survey. We expect that the dependency of disclosure on sensitivity is smaller in the online panel compared to CASI, and that this dependency is smaller in the CAPI environment.

Both parts of our study rely on online data originating from a random sample of the Ipsos Interactive Services (IIS) panel- a large pre-selected volunteer web panel. This fact requires some consideration as panel populations and specifically internet panel populations are known to be different from the general offline population to a degree that may confuse inference. Web panels are usually effected by (internet) coverage error, self selection bias and one possibly might speculate a general higher tendency of panel members to reveal sensitive details in surveys[1]. It is possible that observed (offline-online) differences in disclosure levels is driven then by these web panel factors and not necessarily by survey technology factors (interviewer presence, interviewer administration).

We do not ignore this issue and in Sections 3.1 and 3.2 we outline our approach for controlling sample profile imbalances and to some extent the possible panel-disclosure association. However,assuming that web panel error terms might remain we primarily focus on estimating the *online-panel effect* on disclosure behavior without isolating the unique *online effect*. Some results in our study, though, support such isolation.

Finally, results from both studies will be used to provide 'best practice' information to be used for handling sensitive questions in future surveys.

2 Part I Method

2.1 Design

A between participants design was used with three survey modes. Each mode is summarized in Table 1. As noted above, we do not isolate the panel and online effects and explicitly define the *online-panel* survey mode only.

[1] Possibly due to monetary incentive or a link between panel participation and disclosure.

Table 1: Summary of three survey modes

Survey Mode	Presentation of Survey	Interview Setting	Interviewer present?	Administration of survey
Offline I: CAPI	Verbal	Home	Yes	Interviewer
Offline II: CASI	Computerised	Home	Yes	Self
Online Panel	Computerised	Home	No	Self

The fieldwork periods for all surveys were the same: Surveys were carried out concurrently over five days in June 2006. Generally, we expect disclosure to follow a hierarchical structure such that online panel will have a higher disclosure level than both offline administered modes, and also CASI will have a higher disclosure level than CAPI. We further assume that this relationship can be mainly explained by the gradual removal of interview intervention in the process of data collection of sensitive information. The online panel will be referred to as 'online' from now on.

2.2 Participants

Online Participants

Online participants were members of the IIS Panel. This is a invitation only panel of over 200,000 adults in the UK. Panel members volunteer to complete surveys regularly throughout the year. In total 5500 members of the research panel were invited by e-mail to complete the web-based questionnaire. 1,645 panel members responded (response rate: 30%).

Offline Participants

A total of 1,930 respondents of the adult (15+) UK population were sampled offline where 902 were CAPI respondents and 1,028 were CASI respondents.
The offline survey was conducted in a randomized trial format where half the participating respondents were randomly selected to be surveyed by CAPI and the other by CASI, a design that leads to an (approximately) balanced profile on all research covariates, observed and not observed alike. For this reason when describing the Offline sample profile we sometime ignore the CAPI and CASI sub samples.

2.3 Materials

Survey Questions:

Respondents were asked more than 50 questions on a variety of topics (from politics to media consumption). Within the 50 questions were five 'sensitive' questions. Each of these

questions contained multiple choice response options. Topics for sensitive items covered: immigration, adultery, drink driving, abortion, attitudes toward debt. For example, respondents were asked "Have you ever been unfaithful to a partner / someone you have been in a relationship with?".

Two of the sensitive questions required participants to disclose their behaviour (regarding adultery and drink driving) The remaining three questions required participants to disclose their attitudes (regarding immigration, abortion and debt).

Respondents could 'disclose' by selecting a response option (in the previous example the response options were: "Frequently; Occasionally; Once or twice; Never"). Alternatively, respondents could select the response option "don't know", or not answer a question. Question wordings in each version were kept as similar as possible.

2.4 Procedure

The online sample was provided by IpsosŠ global online research business, Ipsos Interactive Services (IIS). IIS recruits and manages panels in 14 countries and has access to more than 1.6 million respondents worldwide through its partners. The Ipsos GB panel of 200,000 individuals is recruited by invitation only and is validated during a double opt-in registration process. Panelists are recruited from both online and off-line resources.

The Online Access Panel has strict panel usage rules to avoid respondent conditioning and eliminate professional respondents. A points based incentive programme is in place to appropriately reward panelists.
Sample is drawn to match the requirements of individual projects, and quotas are controlled through survey invitation and during fieldwork. Quotas are set to be representative of the GB population using standard demographics including gender, age, social class, region and working status.

The offline sample were recruited using Capibus, a weekly sample of 2000 adults' aged 15+ in GB. All fieldwork is conducted in home by trained interviewers usually using CAPI, for our experiment CASI also. No incentives are offered to respondents and one interview is completed per household. The sample is designed to represent the UK population at a national and regional level. To achieve such representativeness a two stage random location design is applied. Ipsos MORI also use a geodemographic system in the selection process to increase representativeness.

The Sample collected is weighted to correct for any minor deficiencies or bias in the sample. Capibus uses the common raking weighting algorithm (Deming and Stephan 1940) to NRS defined profiles for age, social grade, region and working status - within sex along with household tenure and ethnicity.

The allocation of respondents between Online and Offline was not random and strictly speaking nor from the same population, and so it was not surprising that the profiles of the two sample groups, Online and Offline, are somewhat different. We attempt to controll for these imbalances through propensity score adjustment tools as will be described further in section 3.2.

3 Part I Results: Mode of survey and Disclosure

We wish to measure the effect of survey mode[2] on respondents behaviour to sensitive questions, or in other words to inspect whether there is a causal relationship between mode of survey and disclosure.

3.1 Sample Imbalances

Here we denote mode effect as τ_{mode} and define it as the simple difference between the theoretical discloser levels by mode of survey. In the randomized offline setting, estimation of the CAPI-CASI mode effect is straightforward so that $\widehat{\tau}_{CAPI-CASI} = d_{CAPI} - d_{CASI}$ is an unbiased estimate[3] where d is average sample disclosure.

Online respondents, however, were sampled separately through the IIS panel, and fearing sample profile discrepancy, population disclosure may not be readily derived from their survey estimates d_{on} and we must proceed with some care.

For example, when comparing[4] to the offline sample, the online data collected has a higher proportion of female respondents (57% female online vs. 53% female offline) and also displays an age distribution with an overrepresentation of the tails to the degree that the age band 25-64 is almost 10% underrepresented. Not surprisingly, this segment of the population (i.e. female, 25 younger or 65 older) are significantly more likely to disclose sensitive information than the rest of the population.
Another example (table 2, below) is the almost linear association between lower social grade and lower levels of disclosure. In light of this, the fact that our online data has a distinctly higher social grade distribution with 7% more respondents of social grade A, B or C1 relative to the offline data is problematic. left unchecked, any difference of disclosure between the online and offline groups that we find can be explained, at least partially, by the observed differences in social grade and as such attributing the average disclosure difference to the mode of survey would be false.

So, all this implies that directly estimating mode effect τ_{mode} through unadjusted survey

[2]Which includes as stated previously both mechanism and possible sample source.
[3]Of the true effect $\tau_{CAPI-CASI}$. Thus assuming $\tau_{CAPI-CASI}$=$E(\widehat{\tau}_{CAPI-CASI})$.
[4]Both Online and Offline figures in this section are of unweighted samples.

Table 2: Association of SG to Disclosure detail

Social Class	I do not believe in abortion
A	3.7%
B	8.9%
C1	9.0%
C2	12.5%
D	19.5%
E	24.4%

disclosure difference is expected to be biased[5], an artifact arising from the unbalanced samples.

3.2 Post Survey Adjustment: Estimating Propensity score and converting to weights

Data imbalances such as described above are common in non randomized observational studies. Conceptually, we would like to adjust the data in such a way that the distribution of all (relevant) covariates will be identical both in the online and offline samples. However, this is technically difficult. One common method of such post survey adjustment is the propensity score adjustment (PSA).

The propensity score (Rosenbaum and Rubin 1983) is the probability that an individual is assigned to one of the samples as a function of observed covariates. Propensity score (PS) is commonly employed to adjust for confounding in large non randomized studies Rubin (1997) ; and Dehejia and Wahba (1999) and has been applied also onto the web panel context (see Lee 2006)- mainly in the form of post sampling weights proportional to the inverse of the propensity score estimators.The key consequence of Rosenbaum and Rubin (1983) is that for valid inference it is sufficient to control for the propensity score rather than controlling covariates themselves[6]

It must be noted, however, that this adjustment cannot control for unobserved covariates to the degree they do not correlate with any of those modeled in PSA. Particularly this is relevant to the speculated higher tendency (independent of data collection mode) of a pre-recruited panelist to reveal sensitive details- due to financial incentive, and perhaps also, a link between participation in panel and feeling obligation to disclose. In this study we attempted to treat this issue by controlling (through PSA) on attitudinal and behavioral covariates that could be considered proxies of such tendencies, such as stated sensitivity to personal information

[5]meaning that $E(d_{on}) - E(d_{off}) = \tau_{On-Off} + (\overline{u}_{on} - \overline{u}_{off})$ where $\overline{u}_{on}$ and $\overline{u}_{off}$ are some function of the background variables for the respective groups, and represent a bias from our target statistic τ_{On-Off}.

[6]Effectively, online data, after PSA adjustment, is similar on all available covariates that may potentially confound comparison to the offline data and so now $E(d_{on}|\, e_{obs}) - E(d_{off}) = E(d_{On}^{PS}) - E(d_{off}) = \tau_{On-Off}$., an unbiased estimate. further technical detail in 3.2.1

usage, political views, respondent financial indicators and more. However, echoing results in (Lee 2006) non social demographic covariates showed a negligible effect on estimates while adversely effecting the data quality through increasing variance[7].

3.2.1 Technical Details of PSA

Briefly, and largely following (Lee 2006), we denote the IIS sample as s^{On} with n^{On} units and w_j^{On} is the base weight for unit j; j=1, ..., n^{On}. Similarly, s^{Off}, n^{Off} and w_k^{Off} to be the equivalent Offline denotations. Let $s = (s^{On} \cup s^{Off})$ of size $n = n^{On} + n^{Off}$ units.

Also, define the propensity score of unit i as the likelihood of the unit in participating in the web survey rather than the Offline one, given auxiliary variables. We denote $e(x_i)$ the propensity score of unit i and this is equal to $P(i \in s^{On}/x_i,\ i = 1, ..., n)$ where x being an $n \times p$ matrix and p is the number of covariates used in the model. These scores are estimated by logistic regression and are denoted $\hat{e}_{iobs}$. PSA then proceeds by re-arranging the distribution of s^{On} on $\hat{e}_{obs}$ so that resembles that of s^{Off}. Mechanically this is done by the sorting of s on $\hat{e}_{iobs}$ and partitioning to C classes of approximately equal sizes (Normally $C = 5$, following Cochran (1968)) so that $n_C = n_C^{On} + n_C^{Off}$ where $n_1 = n_2 = ... = n_C$.

Given this outline the adjustment factor is

$$f_c \ = \frac{\sum_{k \in (s_c^{Off})} w_k^{Off} / \sum_{k \in (s^{Off})} w_k^{Off}}{\sum_{j \in (s_c^{On})} w_j^{On} / \sum_{j \in (s^{On})} w_j^{On}}$$

and so the final adjustment weight for any unit j of class c is

$$w_j^{On.PSA} = f_c \cdot w_j^{On}$$

so that the estimated average disclosure d_{On} is then

$$d_{On}^{PS} = \sum_c \sum_{j \in s_c^{On}} w_j^{On.PSA} d_j / \sum_c \sum_{j \in s_c^{On}} w_j^{On.PSA}$$

3.3 Results of Disclosure

3.3.1 Differences in Overall disclosure Levels

Following data adjustment we proceed with calculating survey disclosure. Table 3 below shows results of average disclosure over all five sensitive items studied. Table 3 indicated that the majority of respondents answered all of the five sensitive items.

Also, the discloser behavior across the three modes of survey follow our expectations of hierarchy as presented in Table 1 where $d_{On} > d_{CASI} > d_{CAPI}$, However, interestingly the CAPI-CASI difference is not significant[8]. That said, even significant differences (CAPI-Online,

[7] PSA estimates' standard errors were approximated by estimating the design effect (see Kish 1995).

[8] Even though a one sided significance test ($\alpha = 0.05$) was applied- given our expectation of the direction of the relationship. Significant results are indicated with *.

Table 3: Average Disclosure by mode of survey

Disclosure total	Offline CAPI	Offline CASI	Offline both	Online	Online (PS)
Mean	4.46	4.51	4.49	4.72	4.68
STDV	(0.92)	(0.92)	(0.92)	(0.60)	(0.66)

CASI-Online) are low and attest more to the large sample size of the study and less to the difference in average discloser. In real terms the Online-Offline survey estimated mode effect is 0.2 ($\widehat{\tau}_{[On-Off]} = 0.2$) which in percentage is only a 5% difference in disclosure levels online.

Given these results, a second overall statistic that measures the proportion of respondents answering a specific number of items per mode (see Table 4 below), may be more insightful .

Table 4: Overall Disclosure Distribution by mode of survey

No. items Disclosed to	Offline CAPI	Offline CASI	Offline both	Online	Online (PS)
0	1.2	1.2	1.2	0.0	0.3
1	0.9	0.9	0.9	0.1	0.4
2	1.7	2.2	1.9	0.7	0.9
3	7.8	5.9	6.9	2.4	2.8
4	23.6	21.0	22.4	19.2	20.5
5	64.9	68.8	66.8	77.3	75.1

Table 4 does reinforce results of Table 3, that is a large majority of respondents had disclosed information on all five items, but illuminates further the variation between the three modes of survey. Specifically, we note that 75% online (adjusted) have answered all five items online, over 10% more than the same CAPI group. Also, for the first time, we find empirical evidence that CASI respondents are more likely to disclose sensitive information compared to CAPI- as our expected hierarchy outlined. CASI levels of complete disclosure are nearly 4% higher than the CAPI sample, a statistically significant result [9].

3.3.2 Results of Item disclosure

Whereas Tables 3 and 4 estimate overall disclosure from the sensitive item sample, in Table 5 we inspect the specific average-item disclosure. Unfaithfulness is the item with strongest mode effect, with average disclosure differences of over 6% between CAPI and CASI and over 16% difference between online and offline. Both Drinking and driving and Finance items are also significantly different (Online-Offline) but with disclosure differences that in a commercial research context can easily be ignored.

[9] Although please note again that estimated standard error are only approximation.

It is worth noting also that three out of the five items show a significant Online-Offline mode effect. Furthermore, it is the three behavioural items (Unfaithfulness, Drink Driving and Personal Finances) that in which this mode effect is detected, whereas both attitudinal statements do not show such effect. If we test this result as a 2×2 contingency table of detected mode effect (yes/no) against type of sensitive item (behavioural/attitudinal) this is not surprisingly a statistically significant result[10]. That said, this is a post study examination and so strictly speaking can not be considered a valid hypothesis testing. Regardless, this classification does make sense when considering causes of sensitivity and discloser and also implies that the more sensitive the item (here, as a function of type of question) the bigger the mode effect is, something we look at in Section 4.

Table 5: Average Disclosure of Item by survey mode

Sensitive Item	Offline CAPI	Offline CASI	Offline both	Online	Online (PS)
Abortion	82.3	81.0	81.7	83.8	81.6
Unfaithful	78.2	84.8*	81.3	96.6*	96.6*
Immigration	92.6	92.1	92.4	94.4*	93.6
Drink Driving	96.9	97.8	97.3	99.2*	99.1*
Finance	96.6	95.2	96.0	97.9*	97.4*

It is reassuring to see that most items show similar relative distributions across survey modes, in particular the answer to abortion with a low average disclosure of approximately 82% across total sample. Considering that even the online sample has such a low level of disclosure comes to challenge, to some extent, the risk that a sample of panelists (independent of mode of survey) universally answer any item in high levels whether for exogenous or endogenous reasons as discussed previously in last paragraph of Section 3.2.

[10]Both Chi square test ($\chi^2_{1,1} = 5$) and the, possibely more appropriate, Fisher exact test are significant.

4 Part II: Associating Item Sensitivity to Mode Effect

4.1 Introduction

Our second examination tests whether there is a correlation between degree of item sensitivity and item level of disclosure, and of more consequence is to inspect the variation of this relationship across mode of survey- is this an association that is constant across modes or rather diminishes as interviewer intervention is removed. In this framework we expect two things (a) In an uncontrolled environment sensitivity has a positive correlation with disclosure, and (b) this correlation will become smaller and smaller as isolation increases.

4.2 Design

A small panel of expert social researchers were asked to independently rate the five items studied. From a pool of approximately forty possible raters a simple random sample of 5 were chosen. These panelists were sent a simple questionnaire via email requesting to rank the five items in order of social sensitivity. ties were allowed. None of the researchers on the expert panel had any involvement with the study. To estimate the sensitivity-discloser association, we correlate the average disclosure level in the survey with the sensitivity ranking derived from this panel of experts.

4.3 Panel Ranking and Reliability of Results

The results of the panel of raters on item sensitivity are displayed below in table 6 where 5-highest relative sensitivity and 1-lowest item relative sensitivity- a 5×5 table of items by raters. Note that although some between rater variation is apparent, results seem very consistent and produce a relatively clear ranking.

Table 6: Expert Ranking of Sensitivity

Question	Item	Panel Member				
ID	Title	A	B	C	D	E
Q29	Unfaithful	5	4	5	5	5
Q31	Drink driving	4	5	4	4	4
Q30	Immigration	2	2	3	2	2
Q28	Abortion	1	3	2	3	1
Q32	Finances	3	1	1	1	3

To confirm this and get an indication of how much we can trust the raters ranking, we use the intraclass correlation coefficient (ICC) which is a common statistic (for example, Shrout and Fleiss 1979) measuring the consistency of inter-rater reliability, the idea being that the more homogeneous the ranking of the judges (i.e. raters giving same responses) the more confident we are in summarizing the separate answers into one estimate of the true (relative) sensitivity of an item.

This is an ANOVA-type model[11] and compares the variability of different ratings of the same expert to the total variation across all ratings and all experts.

Table 7: Reliability Test of Expert Panel

Intraclass Correlation	95% Confidence Interval		F Test
	Lower Bound	Upper Bound	Sig
0.928	0.731	0.992	0.000

As we assume the sensitive items to be a random sample from a large pool of possible sensitive items and also the experts (the other source of variation) are considered sample from a large pool of trained researchers, a two-way random effects model was applied[12].
The estimated IIC is 0.928 (table 7) supports our initial thoughts that there is a strong consistency between ranks across the panel members, and so we can summarize the expert individual responses into one estimated ranking of item sensitivity.

4.4 Correlation Between Sensitivity and Disclosure

Finally, we correlate this estimated relative item sensitivity with their respective disclosure levels across the three modes of survey, Table 8 includes only the Pearson correlation statistic estimate but both Kendall and Spearman estimates were also run with very consistent results. The Pearson correlation estimate between CAPI disclosure level and sensitivity is 0.599, but this correlation decreases to 0.338 in the CASI setting and is in effect 0 for the online dataset.

Table 8: Correlation of Expected Sensitivity and Levels of Disclosure by Mode

		Disclosure		
		CAPI	CASI	Online
Sensitivity	Pearson	0.599	0.338	0.088
	Sig. (1-tailed)	0.143	0.289	0.444
	N	5	5	5

Despite not being 95% significant they clearly indicate that there is a strong correlation between disclosure and sensativity but that this relationship gradually diminishes as levels of interview presence and intervention are reduced. This removal of correlation nicely summarizes the idea that disclosure is a function of both sensitivity and environment parameters and by controlling for these components, the dependency between sensitivity and disclosure can be drastically reduced.

[11] This assumes continuous data, the possibly more appropriate test for rank data, Kendall's coefficient of concordance was run also, resulting with Kendall's W= 0.776 (asymptotic significance=0.004) significantly rejecting the null hypothesis of no agreement among the raters.

[12] Moreover, we assume that the raters have similar patterns of scores, so we check for consistency rather than absolute agreement.

5 Discussion

Most research surveys have at least some elements considered sensitive, and with the availability of an increasing number of alternative data collection tools there is a real need to better understand how each of survey mode effect respondents behavior and consequently survey inference. The study aimed to examine the behavior of respondents to sensitive questions in three main environments common in commercial research- Online panel, CASI and CAPI . Overall we have found sufficient evidence that disclosure levels to sensitive questions significantly vary across the three modes of survey, where online panel disclosure level were higher than both offline modes and also we found that CASI respondents on average tend to disclose more than CAPI administered respondents. Also, it is clear that levels of disclosure are dependent on level of sensitivity. Interestingly, it seems that this discloser sensitivity can be partially explained by item classifications such as attitudinal and actual respondent behavior type where the latter being a more sensitive research area.

Not all significant discloser differences were indeed substantial and to most sensitive items examined disclosure levels were high enough that observed mode effects were of no real consequences. It was insightful then to find that an experienced enough group of social researchers can go quite long in predicting the level of sensitivity of an item. In case of perceived high sensitivity, disclosure levels can be substantial and researchers are advised to move survey to a more controlled environment.

Of course these results serve as a platform for more research. After finding sufficient evidence of disclosure level variation, a likely study to follow is to isolate the factors of the online panel effect we have identified- the interviewer, environment and the separate panel effects such as monetary incentive and predisposition to disclosure. We are currently working on a number of experimental designs to reach this.

Further we would recommend examination of the different types of non-discloser. All non-discloser definition such as 'Don't Know', 'Not Answered' and 'Skip' were combined under one umbrella. Is there a identifiable mechanism the governs an active rather than passive non-discloser? In our data, it was interesting to find that the relationship discloser-sensitivity conditional on mode is not constant and reaches very different levels across surprisingly basic segments of the population. Of course, we do not know that what a participant has responded with is accurate.

References

Cochran, W. G. (1968). The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics 24*, 205–213.

Dehejia, R. H. and S. Wahba (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association 94*, 1053–1062.

Deming, W. E. and F. F. Stephan (1940). On a least squares adjustment of a sampled frequency table when the expected marginal tables are known. *The Annals of Mathematical Statistics 11*, 427–444.

Frick, A., M. T. Bachtiger, and U. D. Reips (2001). Financial incentives, personal information and drop-out in online studies. In U.-D. Reaps and M. Bosnjak (Eds.), *Dimensions of Internet Science*, pp. 209–219. Lengerich, Germany: Pabst Science Publishers.

Jobe, J. B., W. F. Pratt, R. Tourangeau, A. Baldwin, and K. Rasinski (1996). Effects of interview mode on sensitive questions in a fertility survey. In L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwartz, and D. Trewin (Eds.), *Survey measurement and process quality*. New York: Wiley.

Joinson, A. N. (1999). Anonymity, disinhibition and social desirability on the internet. *Behaviour Research Methods, Instruments and Computers 31*, 433–438.

Joinson, A. N. (2001). Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity. *European Journal of Social Psychology 31*, 177–192.

Kish, L. (1995). Methods for design effects. *Journal of Ocial Statistics 11*(1), 55–77.

Lee, S. (2006). Propensity score adjustment as a weighting scheme for volunteer panel web surveys. *Journal of Official Statistics 22*(2), 329–349.

Locke, S. E., H. B. Kowaloff, R. G. Hoff, C. Safran, M. A. Popovsky, D. J. Cotton, D. M. Finkelstein, P. P. L., and W. V. Slack (1992). Computer based interview for screening blood donors for risk of hiv transmission. *Journal of the American Medical Association 29*, 1301–1305.

Mosher, W. and A. Duffy (1994). Experiments in survey data collection: the national survey of family growth pretest. In *Paper presented at the meeting of the Population Association of America*, Miami. Sponsoring Organization.

Rosenbaum, P. R. and D. B. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika 7*(1), 41–55.

Rubin, D. B. (1997). Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine 127*, 757–763.

Shrout, P. and J. L. Fleiss (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin 86*, 420–428.

Tourangeau, R. (2004). Survey research and societal change. *Annual Review of Psychology 55*, 775–801.

Tourangeau, R. and T. Smith (1996). Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opinion Quarterly 60*, 275–304.

Waterton, J. and J. Duffy (1984). A comparison of computer interviewing techniques and traditional methods for the collection of self report alcohol consumption in a field study. *International statistical review 52*, 173–183.

Weisband, S. and S. Kiesler (1996). Self-disclosure on computer forms: Meta-analysis and implications. In *Proceedings of CHI96*. Available online at http://www.acm.org/sigchi/chi96/proceedings/papers/Weisband/sw$_{t}xt.htm$.

Wilson, A. M. and N. Laskey (2003). Internet based marketing research: A serious alternative to traditional research methods? *Marketing Intelligence Planning 21*(2), 79–84.