TITLE: How to ask sensitive questions in conservation: A review of specialized questioning techniques

RUNNING TITLE: "Asking sensitive questions"

NUNO, Ana [a,b,*], ST JOHN, Freya A. V. [c,*]
[a] Department of Life Sciences, Imperial College London, Silwood Park, Buckhurst Road, Ascot, SL5 7PY, UK
[b] Centre for Ecology and Conservation, College of Life and Environmental Sciences, University of Exeter Cornwall Campus, Penryn, Cornwall TR10 9EZ, UK
[c] Durrell Institute of Conservation and Ecology, School of Anthropology and Conservation, University of Kent, Canterbury, Kent, CT2 7NZ, UK
[*] Co-corresponding authors. E-mail: a.m.g.nuno@exeter.ac.uk. Telephone: +44 756 4011 550; f.a.v.stjohn@kent.ac.uk. Telephone: +44 1227 82 7139

Word count: 12008 (inclusive of text, references, figures and tables)
Number of tables: 1
Number of figures: 8

ABSTRACT

Tools for social research are critical for developing an understanding of conservation problems and assessing the feasibility of conservation actions. Social surveys are an essential tool frequently applied in conservation to assess both people's behaviour and to understand its drivers. However, little attention has been given to the weaknesses and strengths of different survey tools. When topics of conservation concern are illegal or otherwise sensitive, data collected using direct questions are likely to be affected by non-response and social desirability biases, reducing their validity. These sources of bias associated with using direct questions on sensitive topics have long been recognised in the social sciences but have been poorly considered in conservation and natural resource management.

We reviewed specialized questioning techniques developed in a number of disciplines specifically for investigating sensitive topics. These methods ensure respondent anonymity, increase willingness to answer, and critically, make it impossible to directly link incriminating data to an individual. We describe each method and report their main characteristics, such as data requirements, possible data outputs, availability of evidence that they can be adapted for use in illiterate communities, and summarize their main advantages and disadvantages. Recommendations for their application in conservation are given. We suggest that the conservation toolbox should be expanded by incorporating specialized questioning techniques, developed specifically to increase response accuracy. By considering the limitations of each survey technique, we will ultimately contribute to more effective evaluations of conservation interventions and more robust policy decisions.

Keywords: bias; decision-making; illegal; measurement error; survey methods; uncertainty

1. INTRODUCTION

Effective conservation and natural resource management require the identification of the underlying causes of multiple threats to biodiversity such as overexploitation, habitat fragmentation and climate change (Lande, 1998; Thomas et al., 2004). Processes of human decision-making play a key role in understanding how humans use natural resources (Agrawal and Gibson, 1999), protect certain species while persecuting others (Treves and Karanth, 2003), support policy (Treves, 2009), and allocate research investments (Martín-López et al., 2009). Understanding the drivers and impacts of human behaviour is thus at the core of several disciplines and increasingly more attention has been given to their study in conservation.

Many human activities undermining the success of conservation and natural resource management strategies are illegal or otherwise sensitive (e.g. they are taboo; Jones et al., 2008; Keane et al., 2008). Examples of the consequences of illegal natural resource exploitation include extensive deforestation in Indonesia (Jepson et al., 2001); reproductive collapse in the saiga antelope (*Saiga tatarica*) (Milner-Gulland et al., 2003); and "fish wars" between and among user groups and managers in Southeast Asia fisheries (Pomeroy et al., 2007). Whilst indirect approaches for measuring the extent of illegal resource extraction exist (e.g. remote sensing of deforestation rates (Linkie et al., 2004); and analysing ivory seizures data (Underwood et al., 2013)), such techniques tell us little about the characteristics of rules breakers or what drives their behaviour. Yet effective conservation and informed policy decisions require an understanding of the drivers and impacts of human behaviour (St. John et al., 2013). Illegal or sensitive behaviour is thus a frequent source of uncertainty affecting management decisions and compromising evaluations of conservation interventions.

1.1 Assessing human behaviour

Among the methods used to assess human behaviour, for example indirect observation as applied in market surveys, self-reporting through diaries, or the consultation of law-enforcement records (Gavin et al., 2010; Knapp et al., 2010), questionnaires, delivered through face-to-face interviews or self-completed, are the most commonly applied. Questionnaires frequently assess behaviour through direct questions (e.g. *"Have you done X" Yes / No)*. However, when the topic under investigation is illegal or otherwise sensitive, both non-response and social desirability biases can reduce the validity of data. For example, a non-random proportion of respondents may refuse to participate partly or wholly in the survey creating non-response bias (Groves, 2006); or respondents may provide dishonest answers in order to conform with prevailing social norms, introducing social desirability bias (Fisher, 1993). This tendency of respondents to answer questions in a manner that will be viewed favourably by others may result in under-reporting of undesirable behaviour, such as rule breaking, or over-reporting of desirable behaviour, such as rule compliance (Fisher, 1993).

These sources of bias associated with using direct questions on sensitive topics have long been recognised in the social sciences (e.g. Barton, 1958; Warner, 1965). A number of approaches have been applied in an attempt to identify and correct for these biases, such as relating self-reported behaviours to social-desirability scales (Lee and Sargeant, 2011); measuring comfort with answering sensitive questions (Zink et al., 2006); and analysing mood ratings before and after sensitive questions (Jackson et al., 2012). In addition, question wording or presentation has been manipulated in an attempt to increase reporting of sensitive information. For example, Näher and Krumpal (2011) used forgiving wording, whilst Acquisti et al. (2012) included dummy information on how others responded. Further, by convincing respondents that researchers can discern truthful answers despite what they say, for example, through biological validation, the bogus pipe line procedure seeks to encourage truthful reporting (Adams et al., 2008). The order of questions has also been considered; whilst it is generally recommended that sensitive questions are asked towards the end of questionnaires (Brace, 2008), Acquisti et al. (2012) provide some evidence that respondents are more likely to divulge sensitive information when questions are presented in decreasing order of intrusiveness.

Different modes of survey administration have also been adopted based upon the premise that increased privacy increases data validity. For example, anonymous self-complete answer sheets were posted into a ballot box to reduce bias in sexual behaviour surveys in Zimbabwe (Langhaug et al., 2011); Makkai and Mcallister (1992) assessed drug use by using a "sealed booklet", in which both questions and answers were coded; and Lindstrom et al. (2012) developed a "nonverbal response card" to assess sexual coercion amongst youth in Ethiopia. In addition, advances in technology have led to increased use of computers to deliver surveys, which are not necessarily restricted by literacy as Audio Computer-Assisted Self-Administered Interview (ACASI) systems exist. Highly portable tools such as personal digital assistants (PDAs) have also made an important contribution to investigating sensitive topics. For example, Langhaug et al. (2010) provide evidence that PDAs reduced reporting bias by respondents in developing countries when compared to asking questions about sexual behaviour face-to-face. Other modes of administration that may encourage more honest reporting by increasing respondents' perceived level of protection include video-enhanced self-administrated computer interviews, computer-assisted telephone interviews, internet-based surveys and interactive voice response (Tourangeau and Yan, 2007).

Interview setting and the presence of an interviewer or of other people whilst a questionnaire is being administered are also important factors that may affect people's responses, particularly when the topic is sensitive (Tourangeau and Yan, 2007). The behaviour and characteristics of the person delivering a questionnaire to a respondent can contribute to misreporting, for example survey responses may be influenced by the way in which a question is read out (interviewer behaviour), or the gender of the interviewer (interviewer characteristic). Catania et al. (1996) found that matching respondents and interviewers on gender or allowing respondents to select their interviewer's gender reduced the discrepancies in self-reported sexual behaviour, but that men and women were not equally affected by these interview conditions and also that these effects varied between topics. Interviewer gender effects have been suggested to occur even for recorded voices using ACASI (Dykema et al., 2012). Because the presence of a third party also affects reporting on sensitive topics, ideally, no one but the interviewer and respondent should be present during the administration of the questions (Tourangeau and Yan, 2007), particularly if that third person is not familiar with the information the respondent has been asked to provide and if the respondent fears any repercussions from revealing it to the bystander (Aquilino et al., 2000).

Whilst these approaches may, to varying degrees, encourage reporting of sensitive information, evidence suggests that data validity may be increased by applying methods specifically developed for investigating sensitive topics. Such methods, which we refer to as 'specialized questioning techniques' (also known as 'indirect questioning techniques'), developed in disciplines including political and health sciences, ensure respondent anonymity, increase willingness to answer honestly, and critically, make it impossible to directly link incriminating data to an individual (Warner, 1965; Chaudhuri and Christofides, 2013). Despite some recent applications (Solomon et al. 2007; Blank et al. 2009; Razafimanahaka et al. 2012; St. John et al., 2012; Nuno et al. 2013b), most of these techniques have not been applied within a conservation and natural resource management context suggesting unaddressed potential to ask about illegal or otherwise sensitive topics using novel survey techniques. In this study we review methods specifically developed for investigating sensitive topics, providing examples and recommendations for their potential application in conservation.

2. METHODS

To identify methods specifically developed for investigating sensitive topics we searched both ISI (Web of Knowledge) and Google Scholar with the following keywords: "sensitive question*", "indirect question*", "sensitive topic*" and "social desirability bias". We read abstracts for all publications and selected those that mentioned theoretical or empirical applications of methods developed to ask survey participants about sensitive topics. We also considered relevant studies cited by articles found via keyword searches. We did not aim to compile an exhaustive list of papers using each of the specialized questioning techniques found, but rather to identify: a) the different types of specialized questioning techniques described in peer-reviewed literature and; b) the different versions of each of the techniques found.

We described each method and recorded their main characteristics, such as data requirements (e.g. need for data on a non-sensitive characteristic), possible data outputs (e.g. estimate of behaviour prevalence, link to explanatory variables associated with behaviour), availability of evidence that they can be adapted for use in illiterate communities, and summarized their advantages and disadvantages. When available, we recorded information when researchers compared different techniques (e.g. in terms of accuracy, efficiency, perceptions, etc.). When a certain technique had not been used in illiterate communities and/or a developing country context, we considered that the following requirements would have to be met for its potential use under such conditions: place minimal cognitive demands on respondents; being highly portable; and inexpensive. Several methods reported in different studies were adaptations or variants of a previously described method so we grouped them accordingly.

3. RESULTS

We identified seven types of method developed specifically for investigating sensitive topics, particularly for estimating the proportion of respondents involved in sensitive activities: randomised response techniques; nominative technique; unmatched-count technique; grouped answer method; crosswise, triangular, diagonal and hidden sensitivity models; surveys with negative questions; and the bean method (Table 1).

Table 1. Summary of methods reported in this study and a non-exhaustive list of studies in which these techniques were used

Technique	Previously used in conservation or natural resource management?	Methods comparison studies completed	Evidence that method can be adapted for use in illiterate community?	Possible data outputs
Randomised response technique (RRT; Warner et al. 1965)	Yes (Solomon et al., 2007; Blank et al. 2009; St. John et al., 2010, 2012)	RRT with direct questions (Solomon et al. 2007); RRT with UCT (Coutts and Jann, 2011); RRT with nominative (St. John et al., 2010)	Yes	Proportion of sample population engaging in sensitive behaviour + link to explanatory variables associated with behaviour
Nominative technique (Miller, 1985)	Yes (St. John et al., 2010)	Nominative with RRT and direct questions (St. John et al., 2010)	Yes	Proportion of sample population engaging in sensitive behaviour
Unmatched-count technique (UCT; Droitcour et al., 1991)	Yes (Nuno et al., 2013b)	UCT with direct questions (Tsuchiya et al., 2007); UCT with RRT (Coutts and Jann, 2011)	Yes	Proportion of sample population engaging in sensitive behaviour + link to explanatory variables associated with behaviour
Grouped answer method (Droitcour and Larson, 2002)	No	None	Yes	Proportion of sample population engaging in sensitive behaviour

Crosswise model (CM; Yu et al. 2008), Triangular model (TM; Yu et al. 2008), Diagonal model (DM; Groenitz 2014) Hidden sensitivity model (HSM; Tian et al., 2007)	No	CM with direct questions (Jann et al. 2012)	Maybe	Proportion of sample population engaging in sensitive behaviour + link to explanatory variables associated with behaviour
Surveys with negative questions (Esponda and Guerrero, 2009)	No	None	Maybe	Proportion of sample population engaging in sensitive behaviour
Bean method (BM; Lau et al. 2011)	No	BM with direct questions (Lau et al. 2011)	Yes	Proportion of sample population engaging in sensitive behaviour

3.1 Randomized response techniques

First described by Warner (1965), the randomised response technique (RRT) uses a randomising device (e.g. dice or a spinner) to introduce an element of chance into the question-answer process. RRT has been subject to considerable methodological development aimed at increasing statistical efficiency whilst maintaining respondent protection (Lensvelt-Mulders et al., 2005). Various RRT designs have been applied across a range of sensitive topics including illegal abortion (Silva and Vieira, 2009); social security fraud (Böckenholt and van der Heijden, 2007); and illegal drugs use (Simon et al., 2006). RRT has also been applied to rule-breaking in conservation (Blank et al. 2009; St. John et al., 2010, 2012) where there is evidence that it can be adapted for completion by people with low literacy levels (Solomon et al. 2007; Razafimanahaka et al., 2012). Due to the randomization of questions, there is an added source of variability and RRT requires larger sample sizes than direct questions; the forced-response randomised response technique is one of the more statistically efficient designs (Lensvelt-Mulders et al., 2005). Forced-response RRT instructs (rather than forces) respondents to either: respond to a sensitive question truthfully (answering ‘yes’ or ‘no’); or to give a prescribed ‘yes’ or ‘no’ answer. For example, rolling a pair of dice, respondents may be instructed to: answer a sensitive question truthfully when the dice sum five through to ten (probability = 0.75); give a fixed answer ‘yes’ when the dice sum two, three or four (probability = 0.167); and a fixed answer ‘no’ when the dice sum 11 or 12 (probability = 0.083) (Figure 1). Respondents never reveal the result of the dice roll so it is impossible to distinguish truthful from prescribed responses. Following Hox and Lensvelt-Mulders (2004), prevalence of sensitive behaviours are calculated by:

$$\pi = \frac{\lambda - \theta}{s} \qquad \text{eqn 1}$$

where π is the estimated proportion of the sample who have undertaken the behaviour, λ is the proportion of all responses in the sample that are ‘yes’, θ is the probability of the answer being a ‘forced yes’, s is the probability of having to answer the sensitive question truthfully.

By adapting the standard logistic regression model (van den Hout et al., 2007), it is possible to explore how covariates relate to people’s involvement in sensitive behaviours. For example, St. John et al. (2012) investigated how innocuous indicators of behaviour, such as farmers’ attitudes towards carnivores, relate to illegal carnivore killing reported via RRT. Further, the development of a sum score proportional odds model for RRT data offers an opportunity to reveal associations that remain undetected when data are analysed in a univariate way (Cruyff et al., 2008). Such studies pave the way for using RRT to identify drivers of illicit behaviour.

Typically, RRT estimates the proportion of a population engaged in stigmatizing or illegal behaviours. However, in addition to knowing the proportion of the population involved in such behaviours, we often want to understand the quantitative nature of the behaviour. For example, we may want to simultaneously know the proportion of a population illegally killing a species, and the quantity that they kill. Crude estimates of quantity can be made by using a randomising device (e.g. a spinner with blank and numbered segments) and instructing respondents to: respond truthfully by ticking one of several discrete categorical response options when the spinner lands on a blank segment (e.g. *‘1 = killed zero leopards’*, *‘2 = killed between one and five leopards’*, *3 =* etc.); or ‘forcing’ them to tick the corresponding category when the spinner lands on a numbered segment (Peeters et al., 2010) (see also Conteh et al., this issue). However, more refined estimates become possible when respondents ‘scramble’ their answers. For example, by adding a number from a known distribution to their numeric response (‘additive’ RRT) (Pollock and Bek, 1976) (Figure 2); or by multiplying their numeric response by a number chosen at random from a known distribution and reporting the product (‘multiplicative’ RRT) (Eichhorn and Hayre, 1983). A major advantage of both additive and multiplicative RRT is that they allow sensitive data to be gathered from every respondent. However, RRT designs such as these place considerable demand upon respondents and may therefore not be viable where literacy and numeracy are low. The application of these types of RRT in a conservation context is in its infancy as such their utility still remains to be explored.

	**INSTRUCTIONS** **Please shake the two dice – do not let me see what they land on** Remember the rules, add together the numbers on the two dice: **2, 3, 4** = say **‘Yes’** **5 – 10** answer the question below truthfully **‘Yes’** or **‘No’** **11, 12** = say **‘No’**
QUESTION	**In the last 12 months did you kill any leopards?**

Figure 1. An example instruction card for the forced response randomized response technique. Respondents are provided with an opaque beaker, two dice and a set of question cards each displaying the instructions. The dice are rolled and the instructions followed. Depending upon how the survey is administered, respondents provide their answers either by saying ‘yes’ or ‘no’ out load to an interviewer, or by personally recording their answer. The respondent never reveals the result of the dice role. Killing a leopard is used here (and in Figures 2 and 3) as an example of an activity of conservation concern that may be illegal in some study systems.

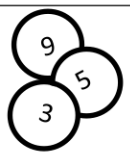	**INSTRUCTIONS** **Please take one numbered ball out of the sack – do not show it to me** Remember the rules: **Add the number on the ball to your truthful answer**
QUESTION	**In the last 12 months how many leopards did you kill?**

Figure 2. An example instruction card for the additive randomized response technique. Respondents are provided with a cloth sack containing numbered balls with a known distribution. Respondents select one ball from the sack and add the number shown on the ball to their numeric response to the question. The respondent never reveals the number displayed on the ball they select. Respondent may call their answers out load to an interviewer or record them personally.

3.2 Nominative technique

The nominative technique (NT) is a variant of multiplicity sampling (sometimes called network sampling) (Sirken, 1972; Sudman et al., 1988) and was developed expressly to investigate heroin use (Miller, 1985). The NT requires respondents to report on the deviant behaviour of close friends. With correction for duplication, to account for multiple respondents reporting the same person, the number of people doing the deviant behaviour can be estimated (Miller, 1985). On three occasions the NT was used to investigate heroin use in the American National Survey on Drug Abuse. On each occasion the

NT estimated higher lifetime prevalence use of heroin compared to anonymous self-complete questionnaire data. Despite this apparent advantage, the NT does not appear to have been applied beyond the Miller (1985) studies before St. John et al. (2010) applied it to rule-breaking in conservation; although this may be due to researchers' reluctance to publish unfavourable findings. The NT is easy to use: respondents are asked to report the number of close friends that they know for certain have done a certain behaviour (e.g. broken a hunting rule); and how many other people they believe know about the nominated friend's behaviour (Figure 3). Based on this information, prevalence rates can be calculated by:

$$T_X = \sum_{j=1}^{n} \frac{A_j}{1+B_j} \qquad \text{eqn 2}$$

where T_x is the number of people breaking a rule in a sample of size n, A_j is the number of rule breakers known to individual j and B_j is the number of friends (other than j) that know of the nominated friend's rule-breaking (Miller, 1985; St. John et al., 2010). Before using the NT, familiarity of respondents with their friend's behaviour in respect of the topic under investigation must be considered. Where respondents' knowledge of their friend's behaviour is weak, NT reveals little about the prevalence of sensitive behaviours (St. John et al., 2010).

1. Most of us know many people. But usually only a few, if any, of these are close friends. **About how many of your close friends go hunting?**

If answer to is 0, *end interview here.*

2. This question is about those close friends. Keep their names to yourself. We want to know about them, but we do not want to know who they are. **How many of your close friends who go hunting can you say for certain have broken hunting rules in the last year?**

If answer is 0, *end interview here.* ***If only 1 friend is reported*** *go to next question directly****.*** *If* ***more than 1 friend reported*** *apply randomised selection*.*

3. Please answer the following question with respect to your close friend that you are thinking of. **As far as you know in the last 12 months did your friend kill any leopards?**

4. Now we would like you to think about this friend's other close friends, besides yourself. **As far as you know, how many of this person's close friends, besides yourself, know for sure that this person has broken hunting rules in the last year?**

Figure 3. Example questions for the nominative technique. This method could be administered through a face-to-face interview or self-administered using pen-and-paper, or computer. *Randomised selection requires respondents to write down the initials of each friend and number them from 1 to the end of the list; predefined instructions (e.g. if the number of close friends reported in question 1 is 5, ask about friend number 2 on the list) in order to identify which friend they should think about when answering the sensitive question(s).

3.3 Unmatched-count technique

The unmatched-count technique (UCT), also known as the list experiment or item count technique, has been used in the last three decades to ask about sensitive topics such as sexual risk behaviours (Hubbard et al., 1989), dangerous driving (Sheppard and Earleywine, 2013), racial prejudice (Blair and Imai, 2012) and illegal bushmeat hunting (Nuno et al., 2013b). Survey respondents are randomly allocated into baseline and treatment groups. Baseline group members receive a list of non-sensitive items while the treatment group members are shown this same list with an additional sensitive item added to it (Figure 4). All respondents are asked to indicate how many, but not which, items apply to them. Differences in the means between baseline and treatment groups are used to estimate the prevalence of the sensitive behaviour (Droitcour et al., 1991).

If the respondents are engaged in all or none of the listed activities, answer secrecy is removed and they may deflate (to avoid association with a socially undesirable item) or inflate (to avoid dissociation with a socially desirable item) their true answers, causing ceiling and floor effects (Zigerell, 2011). To minimize these issues, non-sensitive items should include at least one item whose prevalence is extremely low and one item with very high prevalence (Tsuchiya et al., 2007). Also, non-sensitive items completely different from the target item may cause suspicion (Hubbard et al., 1989); a common theme should be used (e.g. include the sensitive item “poaching” together with non-sensitive livelihood strategies, such as herding and farming). Tsuchiya et al. (2007) suggested that lists should include two or three non-sensitive items in order to ensure answer secrecy while allowing easy mental counting. To analyse UCT data, UCT answers can be analysed in function of the explanatory variables, card type (i.e. treatment or baseline) and interactions of the card type variable with each predictor; the interactions between predictor variables and treatment status indicate differences between the reported number of behaviours in the two conditions for each predictor (Holbrook and Krosnick, 2010).

There is some evidence that the UCT is more effective than direct questions for estimating prevalence of sensitive behaviours (Tsuchiya et al., 2007; Sheppard and Earleywine, 2013) and produces similar or higher estimates than RRT (Coutts and Jann, 2011). In addition, UCT has been reported as less troublesome and easier to understand than RRT (Hubbard et al., 1989). Its simplicity and ease of use in areas of high illiteracy are two main advantages (Nuno et al., 2013). However, UCT has been shown to have limited use for very rare behaviours given the wide standard errors around estimates (Tsuchiya et al., 2007). Further, UCT requires large sample sizes; more than 1000 respondents completed UCT questions administered to determine household participation in bushmeat hunting in western Serengeti returning an estimate with a ±5% standard error (Nuno et al., 2013b), suggesting potential trade-offs between accuracy and precision.

Ongoing UCT developments have focused on increasing its statistical efficiency by improving the estimation process (Corstange, 2009; Blair and Imai, 2012) and the survey administration design (Droitcour et al., 1991; Petróczi et al., 2011; Glynn, 2013). For example, Imai (2011) proposed nonlinear least squares and maximum likelihood estimators for a multivariate analysis. Instead of using a standard design, a double UCT presents the sensitive item to all respondents by using two baseline lists; both experiments provide estimators of the sensitive behaviour that can be averaged (Droitcour et al., 1991). Recently described by Petróczi et al. (2011), a simplified and more efficient version of the UCT, the single sample count (SSC), also asks respondents how many items apply to them without revealing which ones but embeds the sensitive question among four unrelated innocuous questions with known population distributions (e.g. phone numbers ending in odd numbers or birthdays in the first half of the year). This avoids the need to allocate respondents to control groups, since all participants see the same questions. The prevalence estimate from SSC data is then calculated as:

$$\pi = (\lambda/n) - b \qquad \text{eqn 3}$$

where π is the estimated population distribution of the 'yes' answers to the sensitive question, λ is the observed number of 'yes' answers, n is the sample size, and b is the expected value of responses for the baseline non-sensitive questions. Another recent adaptation of UCT, the item sum technique (IST; Trappmann et al., 2014), quantifies sensitive behaviours (e.g. how much time people spend poaching instead of only how many people poach). The IST is administrated similarly to the UCT but it incorporates sensitive and innocuous items that can be measured on a quantitative scale (preferably the same scale, such as hours or monetary units). Respondents are asked to report the sum of the answers to all the activities they engage in (e.g. how many hours they spend per month herding, farming and hunting). However, because respondents in the baseline group only report the sum from non-sensitive activities, the extent of the sensitive behaviour can be calculated from the mean difference of answers between the two subsamples (Trappmann et al., 2014).

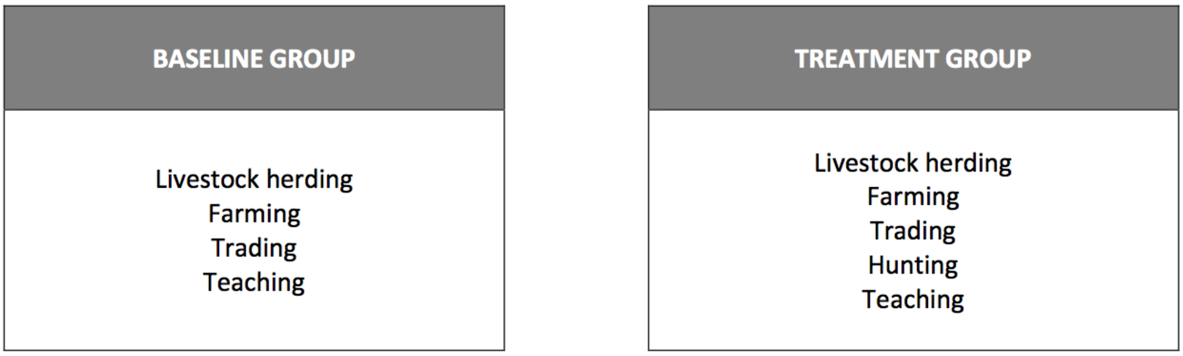

Figure 4. An example of baseline and treatment unmatched-count technique (UCT) lists viewed by survey respondents randomly allocated to either baseline or treatment groups. Respondents are required to report the total number of items that apply to them without identifying any individual item. “Hunting” is used here (and all figures thereafter) as an example of an activity of conservation concern that may be conducted illegally in some study systems and/or under certain conditions.

3.4 Grouped answer method

The grouped answer method, also known as the 2- or 3-card method, was developed in the late 1990s to estimate irregular migration, including illegal or undocumented status (GAO, 1999; Droitcour and Larson, 2002). A list of mutually exclusive items including the sensitive item (e.g. the person’s main occupation) is divided into three groups. The respondents are randomly allocated to one of two treatments (e.g. Card 1 or Card 2, Figure 5), which differ only in the grouping of non-sensitive items with the sensitive item (e.g. hunting); i.e. in Figure 5, the sensitive item remains in Box B for both cards but non-sensitive activities swap between Box A and B. The respondents are then asked to indicate which group they belong to (e.g. A, B or C of Card 1, Figure 2), but not which actual item within the group applies to them. The prevalence of the sensitive item is then estimated by comparing the proportion of people from each of the two treatments who picked the answer group containing the sensitive item, while variance of the sensitive behaviour is estimated by adding the variances from the groups incorporated in the calculations (Droitcour and Larson, 2002). For example, a simple estimate of the sensitive behaviour can be obtained by subtracting the proportion of people that choose Box A in Card 1 from those who choose Box B when shown Card 2 (Figure 5). If the mutually exclusive items are also exhaustive, then the prevalence of the sensitive behaviour can be estimated by subtracting the Box C (averaged from Card 1 and 2) and Box A (summed from Card 1 and 2) percentages from a total of 100%.

GAO (2007) recommended using follow-up questions for respondents who did not pick a group with the sensitive item. These follow-up questions would aim to identify the specific category that applied to the respondents by obtaining direct information on all non-sensitive items for validity checking through comparison with other data sources. If respondents are asked other sociodemographic characteristics during follow-up, then correlates for each non-sensitive category may be obtained directly.

Respondent acceptability and understanding of this technique were considered by GAO (2006) and Larson and Droitcour (2012) who described this technique as promising, although still requiring further testing. To date, this method has only been recommended to produce group-level estimates, without any attempt to conduct univariate or multivariate analysis. For example, to link predictor variables with engagement in the sensitive activity, one could split the analyses according to main variables of interest. Additionally, to our knowledge, estimates from this method have never been compared with direct questioning. Main limitations of this technique are thus its current lack of evidence that it can be subjected to efficient multivariate analysis, large sample size requirements, and the current lack of comparison and validation studies. Nevertheless, its simplicity in administration and ease of use mean that further investigation into this technique may be worthwhile.

CARD 1	
A	Farming Livestock herding
B	Trading Remittances Hunting
C	Other

CARD 2	
A	Trading Remittances
B	Farming Livestock herding Hunting
C	Other

Figure 5. An example of cards used for the grouped answer method. Depending upon the treatment group they are assigned to, respondents are required to report which group on Card 1 or 2 they belong to without identifying which items apply to them.

3.5 Crosswise, triangular, diagonal and hidden sensitivity models

Developed to address concerns that asking respondents to use randomizing devices can create confusion (Chaudhuri and Christofides, 2013), the techniques that follow do not depend on a randomizing device. However, randomization occurs implicitly (Tian and Tang, 2013).

The crosswise (CM) and triangular (TM) models, first described by Yu et al. (2008), expose respondents to two questions, only one of which is sensitive, and respondents then provide a joint answer to both questions. For both techniques, the probability distribution of the non-sensitive question must be known (e.g. month of birth) and it should be unrelated to the sensitive behaviour. However, these techniques differ in their specific response rules. In the CM, respondents are told to choose option *A* if the answer is the same for both questions (i.e. 'yes' to both questions or 'no' to both questions) and option *B* if one answer is 'yes' and the other is 'no'. In the TM, respondents are asked to choose option *A* if the answer is 'no' to both questions and option *B* if at least one answer is 'yes' (Figure 6).

While both the TM and CM ask one sensitive question at a time, the hidden sensitivity model (HSM) has been developed to analyse the association between several sensitive questions by asking them simultaneously (Tian et al., 2007). To ask two sensitive questions simultaneously, e.g. about illegal hunting and corruption, HSM requires a non-sensitive question with four mutually exclusive response categories each with a known probability distribution (e.g. A, B, C and D corresponding to different quarters in a year). Respondents who do not engage in any of the sensitive behaviours, should reply truthfully to the non-sensitive question (A, B, C or D) while the other respondents should choose B if they are only engaged in the second sensitive behaviour, C if only the first and D if both, hiding the sensitive attribute of respondents (Figure 7).

The diagonal model (DM) recently developed by Groenitz (2014) expands upon CM, TM and HSM by allowing researchers to investigate multichotomous sensitive questions, such as levels of income (which is often considered sensitive). Again, respondents are asked a sensitive and a non-sensitive question with known distribution, each with multiple categories (e.g. four in the example below). Respondents give the answer:

$$A = [(W - X) * mod\ k] + 1 \qquad \text{eqn 4}$$

where *W* is the number (1 to 4) corresponding to their categorical answer to the non-sensitive question, X is the number (1 to 4) corresponding to their categorical answer to the sensitive question, and *k* is the number of categories in the non-sensitive question. However, respondents are not provided with this formula but simply with a table from which they can select their answer to the sensitive and non-sensitive questions simultaneously (Figure 8). Using the table, respondents report only the number in the table which provides the required answer A depending on X and W. Because it is not possible to identify the X value from their answer A, answer secrecy is guaranteed. When asking a respondent multiple sensitive questions (e.g. how many leopards did you kill in the last 12 months?; how many lions did you kill in the last 12 months?) where responses may fall within the same category (e.g. category 1 equals none, category 2 equals between 1 and 3 etc.), the non-sensitive question posed simultaneously must also be changed in order to ensure that respondents do not reveal truthful responses to either X or W.

To our knowledge, only the CM and HSM have been empirically explored (Tian et al., 2007; Jann et al., 2012; Vakilian et al., 2014). Given this, and the similarity between these four techniques, we will now focus on the CM. For CM, prevalence estimates are calculated by:

$$\pi = \frac{\lambda+p-1}{2p-1}, p \neq 0.5 \qquad \text{eqn 5}$$

where π is the estimated proportion of the sample who have undertaken the sensitive behaviour, λ is the observed proportion of all responses in the sample that choose option A (i.e., ‘yes’ to both questions or ‘no’ to both questions), and p is the known population prevalence of the non-sensitive item (Jann et al., 2012). To analyse the effects of multiple covariates, modified logistic regression models and modified linear probability models may be used. For example, Jann et al. (2012) used this technique to investigate plagiarism by students, linking to several predictors, and found that CM produced higher prevalence rates than direct questioning. Although no comparative analysis is available, Jann et al. (2012) also suggest that the CM may be better than RRT and UCT due to its statistical efficiency and lack of an obvious self-protective answering strategy.

Q1: Is your birthday in January, February or March?
Q2: Did you hunt without a license last year?

<u>How are your answers to these two questions? Pick A or B</u>

CROSSWISE MODEL	
A	NO to both questions **OR** YES to both questions
B	YES to one of the questions **AND** NO to the other

TRIANGULAR MODEL	
A	NO to **BOTH** questions
B	YES to **ONE** of the questions

Figure 6. An example of a question card to be used in studies applying either the crosswise model or the triangular model. Respondents are asked to provide a joint answer to both questions following different rules according to specific technique.

According to the table below please <u>pick the option (A, B, C or D) that corresponds to your answers</u>:

W: When is your birthday?

X: Did you hunt without a license last year?

Y: Did you pay a bribe to a park ranger last year?

	W=Jan-Mar	W=Apr-Jun	W=Jul-Sep	W=Oct-Dec
X=No, Y=No	A	B	C	D
X=No, Y=Yes	Please tick option B			
X=Yes, Y=No	Please tick option C			
X=Yes, Y=Yes	Please tick option D			

Figure 7. An example of a question card to be used in studies applying the hidden sensitivity model. Respondents are asked to answer A, B, C or D according to the card instructions; people that have done any of the sensitive activities are required to answer irrespectively of their actual birthday, protecting their answers.

Using the table below please <u>pick the number (1, 2, 3 or 4) that corresponds to your answers to both questions simultaneously</u>:

- When is your birthday? *(= W)*
- How many times did you go hunting inside the park last year? *(= X)*

	January February March	April May June	July August September	October November December
0	1	2	3	4
1, 2 or 3	4	1	2	3
4, 5 or 6	3	4	1	2
6, 7, 8….	2	3	4	1

Figure 8. An example of a question card to be used in studies applying the diagonal model. After being read or shown two questions (one sensitive and the other non-sensitive), respondents should report the number (1, 2, 3 or 4) in the table that provides the required answer depending on both questions simultaneously.

3.6 Surveys with negative questions

In conventional closed check-list questions (Newing, 2011) respondents are required to answer questions or statements phrased in a positive direction (e.g. *'I earn…'*) by selecting the response category that applies to them. However, 'negative questions' ensure respondent privacy by phrasing questions in a negative direction (e.g. *'I do not earn…*) and asking respondents to select a response category to which they do not belong (Esponda and Guerrero, 2009). For example, a negative question for assessing annual income may look like this (Esponda and Guerrero, 2009):

I do *not* earn:
Less than 30 000 dollars a year
Between 30 000 and 60 000 dollars a year
More than 60 000 dollars a year

The number of respondents *e* that belong to a certain category *j* is estimated using:

$$e_j = n - (c - 1) \times r_j \quad \text{eqn 6}$$

where *n* is the total number of participants, *c* is the number of categories and r_j is the number of respondents who report category *j* (Xie et al., 2011).

This technique requires both that questions be phrased in the negative (e.g. 'I do not earn…'), and that multiple true options are available for respondents to choose from. For example, if a respondent earns more than 60,000 dollars a year they could choose either option a) or b) as their answer to the question above because both answer the negative question truthfully. However, in order to reduce the chance of bias in respondents' selection of response categories, a randomizing device with *c-1* options is used in private by the respondent to obtain a value *m*, they then choose the m^{th} true alternative from the list accordingly (Esponda and Guerrero, 2009). Rather than using a randomising device with known probabilities drawn from a uniform distribution, Xie et al. (2011) proposed that the probability of selecting response categories should follow a Gaussian distribution centred at the positive category. This approach achieves higher accuracy but reduces respondent privacy. Bao et al. (2013) has also suggested improvements to this method that ensure that estimates of the number of people selecting each category are always positive (negative estimates can unrealistically occur with a standard estimation process with low sample sizes).

'Negative questions' is a relatively recent survey technique still under development, with the few empirical applications currently limited to communications and technology. For example, Horey et al. (2007) used this approach to implement anonymous data collection on sensor network platforms. Easy to administer, it seems a promising method although its validity and how it compares to other questioning techniques still remain to be investigated.

3.7 Bean method

The "bean method" was recently developed to collect information on health risk behaviours (Lau et al., 2011). This method presents respondents with one large and one small jar of beans, both containing mixed-up beans of different colours. The number of beans should be large enough so that addition or removal of a single bean from either jar is not noticeable. Respondents are instructed to move a black bean from the smaller jar to the large jar if the answer to a sensitive question is 'no' and to move a bean of another specified colour from the small jar to the large jar if the answer is 'yes'. Respondents do this in private, without being watched by the interviewer. After multiple respondents have completed the exercise, changes in the bean composition in the jars are used to estimate the prevalence of a sensitive behaviour.

This method is technologically simple, very easy to administer and Lau et al. (2011) reported that it was well received by respondents. Further, it generally produced similar or higher estimates of the sensitive behaviour compared to face-to-face direct questions (Lau et al. 2011). However, if administered as described here, the bean method only produces group-level estimates.

4. DISCUSSION

Increasing emphasis is being placed upon the social dimensions of conservation (Sandbrook et al., 2013) and this may present challenges to scientists trained largely in the natural sciences. However, social science techniques must be applied with the same rigour demanded of methods used to monitor ecological factors (St. John et al., 2013). Tools for social research are essential for understanding the feasibility of conservation actions and identifying the scope of conservation problems (Raymond and Knight, 2013). Social surveys are an essential tool often used in conservation both to assess people's behaviour and to understand its drivers (White et al., 2005). However, the weaknesses and strengths of different tools must be considered. When topics of conservation concern are illegal or otherwise sensitive, inferences drawn from survey data must be interpreted and used very carefully due to potential influences of non-response and social-desirability bias (St. John et al., 2010). We suggest that the conservation toolbox should be expanded by incorporating specialized questioning techniques that have been developed in a range of disciplines specifically to reduce these biases and improve response accuracy.

As shown in our study, a variety of specialized questioning techniques have been developed to protect respondent confidentiality and increase response accuracy. Whilst these techniques represent promising and useful developments in the field of quantitative social science, they should not be seen as a panacea. Their limitations should be considered and evaluated against other criteria. The general disadvantages in using these specialized techniques rather than direct questioning include the increased complexity of data analysis, higher sample size requirements and the more limited form that the questions can take. Nevertheless, a number of recent studies have presented improved designs and analysis for these techniques (e.g. Bullock et al., 2011; Petróczi et al., 2011; Blair and Imai, 2012). In some cases, given the larger sample size required for some of the techniques, it is not cost-efficient to use them for non-sensitive topics. Also, regardless of the survey technique, some participants may still give evasive responses; as such estimates are still likely to be conservative. A key consideration is the limited availability of studies comparing different techniques and their estimates' accuracy. Ground-truthing estimates from social surveys is rarely conducted (White et al., 2005) and validation studies in which the reliability of responses is verified (e.g. by surveying convicted criminals and comparing their answers to their criminal records) are particularly difficult when dealing with sensitive topics. The use of complementary methods for triangulation may help overcome the constraints inherent to each individual research tool.

Although these specialized questioning techniques have been applied in a number of socio-demographic and cultural contexts (e.g. Solomon et al. (2007) in villages in Uganda and St. John et al. (2010) with fishers in the UK), relatively little attention has been given to the trade-offs between technique complexity and intellectual demand, perceived feelings of anonymity and trust. For example, while the UCT was developed to address some of the criticisms of RRT (that it may be constrained by belief in trickery or by respondents' feelings of confusion and education level (Hubbard et al., 1989; Landsheer et al., 1999)), little attention has been given to exploring respondents' perceptions towards these techniques. In a small pilot study conducted to investigate the feasibility of using specialized questioning techniques to assess poaching in the Serengeti, Nuno (2013) found that respondents found the UCT easier to understand than the RRT. Pilot studies testing the feasibility of multiple techniques before conducting the main data collection can thus provide essential information about the adequacy of different survey instruments and the importance of such pilots cannot be overemphasized. Additional studies that robustly consider the appropriate use of each of these techniques in terms of costs, suitability in low literacy populations and efficiency of statistical estimators would provide much needed information that could be used to compare their feasibility, advantages and potential problems in a single framework.

Novel applications of existing methodologies may also contribute to our understanding of involvement in illicit behaviours. For example, Moro et al. (2013) used choice experiments to elicit a household's intention to hunt illegally in the Serengeti under different conditions by embedding hunting as one option across a range of livelihood strategies. Nielsen et al. (2013) also suggested that the use of hypothetical scenarios in choice experiments is likely to make the elicitation of preferences about illegal activities less sensitive. Choice experiments may then be used to obtain essential information on sensitive behaviours by providing information on preferences and trade-offs in relation to several attributes of the choice to engage in those activities. Other techniques developed in the economic sciences that may be useful to investigate decisions about engagement into sensitive behaviours include, for example, willingness-to-pay studies (e.g. asking willingness to accept compensation for forgoing illegal harvest) and economic experiments using lotteries to investigate relations between income and wildlife harvest (Sirén et al., 2006).

Advances in technology also present opportunities; for example, smartphones have been used to obtain information about illegal activities which has been collected by local communities in developing countries (Vitos et al., 2013). Additionally, occupancy modelling has been suggested as a potential tool to determine more accurate illegal wildlife trade estimates from market data by taking detectability into account (Barber-Meyer, 2010), and capture-recapture methods have been used to estimate the size of difficult-to-count human populations (e.g. clients of prostitution; Roberts and Brewer, 2006) through overlap between different datasets or subsequent arrest records.

While our study focused on describing specialized questioning techniques that have been developed to investigate sensitive topics, and mainly focuses on techniques used to reduce non-response and under-reporting due to social desirability biases, there are a number of other factors to be considered. For example, despite being generally unaddressed in conservation, it is likely that people over-report involvement in pro-conservation behaviours, as already observed for other pro-social behaviours such as charitable giving (Lee and Sargeant, 2011). Moreover, acquiescence bias (tendency to agree or disagree with all or most of the questions asked) and extremity bias (tendency to choose extreme ratings in response-scale formats) are frequent problems affecting social surveys. For example, Javeline (1999) showed the magnitude of the acquiescence problem in societies where norms of civility and respect distort attitude reports, and suggested that forced-choice questions (offering two opposing views and instructing respondent to select one of them) are more effective than traditional Likert scales in addressing this problem. Identifying, reducing and/or accounting for these multiple sources and types of bias in social surveys in conservation is thus essential and deserves further attention and research.

5. CONCLUSION

Given the promising ongoing developments in survey techniques and the well-known limitations of asking sensitive questions directly, we suggest that specialized questioning techniques developed specifically to investigate sensitive topics should be further explored. When evaluating conservation interventions and making policy decisions, observation uncertainty related to the measurement process and its implications should be made explicit, and should be fully considered (Nuno et al., 2013a). By identifying and acknowledging the limitations of each survey technique, we can incorporate this information into wider conceptual and methodological frameworks aimed at supporting decision-making, such as the management strategy evaluation (Bunnefeld et al., 2011). Only by guaranteeing that decisions are evaluated in a comprehensive, robust and transparent manner can we plan for effective conservation.

ACKNOWLEDGEMENTS

AN was supported by the Portuguese Foundation for Science and Technology (FCT; doctoral grant SFRH/BD/43186/2008). The funding source had no involvement in the conduct of the research and preparation of the article. Thank you to Dr David Roberts of DICE, School of Anthropology and Conservation for comments on an earlier draft of this manuscript. Two anonymous reviewers gave valuable comments on the manuscript.

REFERENCES

Acquisti, A., John, L.K., Loewenstein, G., 2012. The impact of relative standards on the propensity to disclose. J. Mark. Res. 49, 160–174.

Adams, J., Parkinson, L., Sanson-Fisher, R.W., Walsh, R.A., 2008. Enhancing self-report of adolescent smoking: the effects of bogus pipeline and anonymity. Addict. Behav. 33, 1291–6.

Agrawal, A., Gibson, C.C., 1999. Enchantment and disenchantment: the role of community in natural resource conservation. World Dev. 27, 21.

Aquilino, W. S., Wright, D. L., Supple, A. J., 2000. Response effects due to bystander presence in CASI and paper-and-pencil surveys of drug use and alcohol use. Substance Use and Misuse 35, 845–867.

Bao, Y., Luo, W., Zhang, X., 2013. Estimating positive surveys from negative surveys. Stat. Probab. Lett. 83, 551-558.

Barber-Meyer, S.M., 2010. Dealing with the clandestine nature of wildlife-trade market surveys. Conserv. Biol. 24, 918–923.

Barton, A.H., 1958. Asking the embarrassing question. Public Opin. Q. 22, 67.

Blair, G., Imai, K., 2012. Statistical analysis of list experiments. Polit. Anal. 20, 47–77.

Blank, S.G., Gavin, M.C., 2009. The randomized response technique as a tool for estimating non-compliance rates in fisheries: a case study of illegal red abalone (*Haliotis rufescens*) fishing in Northern California. Environmental Conservation 36, 112-119.

Böckenholt, U., van der Heijden, P.G.M., 2007. Item randomized-response models for measuring noncompliance: risk-return perceptions, social influences, and self-protective responses. Psychometrika 72, 245–262.

Brace, I., 2008. Questionnaire design: how to plan, structure and write survey material for effective market research, 2nd ed. Kogan Page Publishers.

Bullock, W., Imai, K., Shapiro, J.N., 2011. Statistical analysis of endorsement experiments: measuring support for militant groups in Pakistan. Polit. Anal. 19, 363–384.

Bunnefeld, N., Hoshino, E., Milner-Gulland, E.J., 2011. Management strategy evaluation: a powerful tool for conservation? Trends Ecol. Evol. 26, 441–447.

Catania, J. A., Binson, D., Canchola, J., Pollack, L. M., Hauck, W., Coates, T. J., 1996. Effects of interviewer gender, interviewer choice, and item wording on responses to questions concerning sexual behavior. Public Opin Q. 60, 345–375.

Chaudhuri, A., Christofides, T., 2013. Indirect questioning in sample surveys. Springer.

Corstange, D., 2009. Sensitive questions, truthful Answers? Modeling the list experiment with LISTIT. Polit. Anal. 17, 45–63.

Coutts, E., Jann, B., 2011. Sensitive questions in online surveys: experimental results for the randomized response technique (RRT) and the unmatched count technique (UCT). Sociol. Methods Res. 40, 169–193.

Cruyff, M.J.L.F., van den Hout, A., van der Heijden, P.G.M., 2008. The analysis of randomized response sum score variables. J. R. Stat. Soc. Ser. B (Statistical Methodol. 70, 21–30.

Droitcour, J., Caspar, R.A., Hubbard, M.L., Parsley, T.L., Visscher, W., Ezzati, T.M., 1991. The item count technique as a method of indirect questioning: A review of its development and a case study application, in: Biemer, P.P., Groves, R.M., Lyberg, L.E., Mathiowetz, N.A., Sudman, S. (Eds.), Measurement Errors in Surveys. New York, NY, pp. 185–210.

Droitcour, J., Larson, E.M., 2002. An innovative technique for asking sensitive questions: the three-card method. Bull. Méthodologie Sociol. 75, 5–23.

Dykema, J., Diloreto, K., Price, J., White, E., Schaeffer, N., 2012. ACASI Gender-of-interviewer voice effects on reports to questions about sensitive behaviors among young adults. Public Opin Q. 76, 311-325.

Eichhorn, B.H., Hayre, L.S., 1983. Scrambled randomized response methods for obtaining sensitive quantitative data. J. Stat. Plan. Inference 7, 307–316.

Esponda, F., Guerrero, V.M., 2009. Surveys with negative questions for sensitive items. Stat. Probab. Lett.

Fisher, R., 1993. Social desirability bias and the validity of indirect questioning 20, 303–315.

GAO, 1999. Survey methodology: an innovative technique for estimating sensitive survey items. Washington, DC.

GAO, 2006. Estimating the undocumented population: a “grouped answers” approach to surveying foreign-born respondents. DIANE Publishing.

GAO, 2007. Estimating irregular migration in a survey: the “two-card follow–up” method. U.N. Sixth coordination meeting on international migration. New York.

Gavin, M.C., Solomon, J.N., Blank, S.G., 2010. Measuring and monitoring illegal use of natural resources. Conserv. Biol. 24, 89–100.

Glynn, A.N., 2013. What can we learn with statistical truth serum?: design and analysis of the list experiment. Public Opin. Q. 77, 159–172.

Groenitz, H., 2014. A new privacy-protecting survey design for multichotomous sensitive variables. Metrika 77, 211–224.

Groves, R.M., 2006. Nonresponse rates and nonresponse bias in household surveys. Public Opin. Q. 70, 646–675.

Holbrook, A.L., Krosnick, J.A., 2010. Social desirability bias in voter turnout reports: tests using the item count technique. Public Opin. Q. 74, 37–67.

Horey, J., Groat, M.M., Forrest, S., Esponda, F., 2007. Anonymous data collection in sensor networks, in: 2007 Fourth Annual International Conference on Mobile and Ubiquitous Systems: Networking & Services (MobiQuitous). IEEE, pp. 1–8.

Hox, J., Lensvelt-Mulders, G., 2004. Randomized response analysis in Mplus. Struct. Equ. Model. A Multidiscip. J. 11, 615–620.

Hubbard, M.L., Caspar, R.A., Lessler, J.T., 1989. Respondent reactions to item count lists and randomized response, in: Proceedings of the American Statistical Association, Section for Survey Research Methods. pp. 544–548.

Imai, K., 2011. Multivariate regression analysis for the item count technique. J. Am. Stat. Assoc. 106, 407–416.

Jackson, Y., Gabrielli, J., Tunno, A.M., Hambrick, E.P., 2012. Strategies for longitudinal research with youth in foster care: a demonstration of methods, barriers, and innovations. Child. Youth Serv. Rev. 34, 1208–1213.

Jann, B., Jerke, J., Krumpal, I., 2012. Asking sensitive questions using the crosswise model: an experimental survey measuring plagiarism. Public Opin. Q. 76, 32–49.

Javeline, D., 1999. Response effects in polite cultures - a test of acquiescente in Kazakhstan. Public Opin. Q. 63, 1–28.

Jepson, P., Jarvie, J.K., MacKinnon, K., Monk, K.A., 2001. The end for Indonesia's lowland forests? Science 292, 859–861.

Jones, J.P.G., Andriamarovololona, M.M., Hockley, N., 2008. The importance of taboos and social norms to conservation in Madagascar. Conserv. Biol. 22, 976–86.

Keane, A., Jones, J.P.G., Edwards-Jones, G., Milner-Gulland, E.J., 2008. The sleeping policeman: understanding issues of enforcement and compliance in conservation. Anim. Conserv. 11, 75–82.

Knapp, E.J., Rentsch, D., Schmitt, J., Lewis, C., Polasky, S., 2010. A tale of three villages: choosing an effective method for assessing poaching levels in western Serengeti, Tanzania. Oryx 44, 178–184.

Lande, R., 1998. Anthropogenic, ecological and genetic factors in extinction and conservation. Res. Popul. Ecol. (Kyoto). 40, 259–269.

Landsheer, J.A., Heijden, P. Van Der, Gils, G. Van., 1999. Trust and understanding, two psychological aspects of randomized response. Quality & Quantity 33, 1–12.

Langhaug, L.F., Cheung, Y.B., Pascoe, S.J.S., Chirawu, P., Woelk, G., Hayes, R.J., Cowan, F.M., 2011. How you ask really matters: randomised comparison of four sexual behaviour questionnaire delivery modes in Zimbabwean youth. Sex. Transm. Infect. 87, 165–73.

Langhaug, L.F., Sherr, L., Cowan, F.M., 2010. How to improve the validity of sexual behaviour reporting: systematic review of questionnaire delivery modes in developing countries. Trop. Med. Int. Health 15, 362–81.

Larson, E.M., Droitcour, J.A., 2012. The grouped answer method for estimating immigration status: analysis of data from the 2004 general social survey, in: Hoque, N., Swanson, D.A. (Eds.), Opportunities and Challenges for Applied Demography in the 21st Century. Springer Netherlands, Dordrecht, pp. 311–334.

Lau, J.T.F., Yeung, N.C.Y., Mui, L.W.H., Tsui, H.Y., Gu, J., 2011. A simple new method to triangulate self-reported risk behavior data - the bean method. Sex. Transm. Dis. 38, 788–92.

Lee, Z., Sargeant, A., 2011. Dealing with social desirability bias: an application to charitable giving. Eur. J. Mark. 45, 703–719.

Lensvelt-Mulders, G.J.L.M., Hox, J.J., Heijden, P.G.M. van der, 2005. How to improve the efficiency of randomised response designs. Quality & Quantity 39, 253–265.

Lindstrom, D.P., Hattori, M.K., Belachew, T., Tessema, F., 2012. Lifting the curtain on the conditions of sexual initiation among youth in Ethiopia. J. Adolesc. Health 50, 614–20.

Linkie, M., Smith, R.J., Leader-Williams, N., 2004. Mapping and predicting deforestation patterns in the lowlands of Sumatra. Biodivers. Conserv. 13, 1809–1818.

Makkai, T., Mcallister, I., 1992. Measuring social indicators in opinion surveys: a method to improve accuracy on sensitive questions. Soc. Indic. Res. 27, 169–186.

Martín-López, B., Montes, C., Ramírez, L., Benayas, J., 2009. What drives policy decision-making related to species conservation? Biol. Conserv. 142, 1370–1380.

Miller, J.D., 1985. The nominative technique: a new method of estimating heroin prevalence., in: Rouse, B.A., Kozel, N.J., Richards, L.G. (Eds.), Self-Report Methods of Estimating Drug Use: Meeting Current Challenges to Validity. NIDA - National Institute on Drug Abuse, pp. 104–24.

Milner-Gulland, E.J., Bukreeva, O.M., Coulson, T., Lushchekina, A.A., Kholodova, M. V, Bekenov, A.B., Grachev, I.A., 2003. Reproductive collapse in saiga antelope harems. Nature 422, 135.

Moro, M., Fischer, A., Czajkowski, M., Brennan, D., Lowassa, A., Naiman, L.C., Hanley, N., 2013. An investigation using the choice experiment method into options for reducing illegal bushmeat hunting in western Serengeti. Conserv. Lett. 6, 37–45.

Näher, A.-F., Krumpal, I., 2011. Asking sensitive questions: the impact of forgiving wording and question context on social desirability bias. Quality & Quantity 46, 1601–1616.

Newing, H., 2011. Conducting research in conservation: social science methods and practice. Routledge.

Nielsen, M.R., Jacobsen, J.B., Thorsen, B.J., 2013. Factors determining the choice of hunting and trading bushmeat in the Kilombero Valley, Tanzania. Conserv. Biol.

Nuno, A., 2013. *Managing social-ecological systems under uncertainty: implications for conservation*. Imperial College London. PhD thesis.

Nuno, A., Bunnefeld, N., Milner-Gulland, E.J., 2013a. Matching observations and reality: using simulation models to improve monitoring under uncertainty in the Serengeti. J. Appl. Ecol. 50, 488–498.

Nuno, A., Bunnefeld, N., Naiman, L.C., Milner-Gulland, E.J., 2013b. A novel approach to assessing the prevalence and drivers of illegal bushmeat hunting in the Serengeti. Conserv. Biol. 27, 1355–1365.

Peeters, C.F.W., Lensvelt-Mulders, G.J.L.M., Lasthuizen, K., 2010. A note on a simple and practical randomized response framework for eliciting sensitive dichotomous and quantitative information. Sociol. Methods Res. 39, 283–296.

Petróczi, A., Nepusz, T., Cross, P., Taft, H., Shah, S., Deshmukh, N., Schaffer, J., Shane, M., Adesanwo, C., Barker, J., Naughton, D.P., 2011. New non-randomised model to assess the prevalence of discriminating behaviour: a pilot study on mephedrone. Subst. Abuse Treat. Prev. Policy 6, 20.

Pollock, K.H., Bek, Y., 1976. A comparison of three randomized response models for quantitative data. J. Am. Stat. Assoc. 71, 884–886.

Pomeroy, R., Parks, J., Pollnac, R., Campson, T., Genio, E., Marlessy, C., Holle, E., Pido, M., Nissapa, A., Boromthanarat, S., 2007. Fish wars: conflict and collaboration in fisheries management in Southeast Asia. Mar. Policy 31, 645–656.

Raymond, C.M., Knight, A.T., 2013. Applying social research techniques to improve the effectiveness of conservation planning. Bioscience 63, 320–321.

Razafimanahaka, J.H., Jenkins, R.K.B., Andriafidison, D., Randrianandrianina, F., Rakotomboavonjy, V., Keane, A., Jones, J.P.G., 2012. Novel approach for quantifying illegal bushmeat consumption reveals high consumption of protected species in Madagascar. Oryx 46, 584–592.

Roberts, J.M., Brewer, D.D., 2006. Estimating the prevalence of male clients of prostitute women in Vancouver with a simple capture-recapture method. J. R. Stat. Soc. Ser. A (Statistics Soc. 169, 745–756.

Sandbrook, C., Adams, W.M., Büscher, B., Vira, B., 2013. Social research and biodiversity conservation. Conserv. Biol. 27, 1487–90.

Sheppard, S.C., Earleywine, M., 2013. Using the unmatched count technique to improve base rate estimates of risky driving behaviours among veterans of the wars in Iraq and Afghanistan. Inj. Prev. 19, 382–386.

Silva, R. de S. e, Vieira, E.M., 2009. Frequency and characteristics of induced abortion among married and single women in São Paulo, Brazil. Cad. Saude Publica 25, 179–187.

Simon, P., Striegel, H., Aust, F., Dietz, K., Ulrich, R., 2006. Doping in fitness sports: estimated number of unreported cases and individual probability of doping. Addiction 101, 1640–4.

Sirén, A. H., J. C. Cardenas, J.C., Machoa, J.D., 2006. The relation between income and hunting in tropical forests: an economic experiment in the field. Ecology and Society 11, 44.

Sirken, M.G., 1972. Stratified sample surveys with Multiplicity. J. Am. Stat. Assoc. 67, 224–227.

Solomon, J.N., Jacobson, S., Wald, K.D., Gavin, M.C., 2007. Estimating illegal resource use at a Ugandan park with the randomized response technique. Hum. Dimens. Wildl. 12, 75–88.

St. John, F.A.V., Edwards-Jones, G., Gibbons, J.M., Jones, J.P.G., 2010. Testing novel methods for assessing rule breaking in conservation. Biol. Conserv. 143, 1025–1030.

St. John, F.A.V., Keane, A., Edwards-Jones, G., Jones, L., Yarnell, R.W., Jones, J.P.G., 2012. Identifying indicators of illegal behaviour: carnivore killing in human-managed landscapes. Proc. R. Soc. B Biol. Sci. 279, 804–812.

St. John, F.A.V., Keane, A., Milner-Gulland, E.J., 2013. Effective conservation depends upon understanding human behaviour, in: Macdonald, D.W., Willis, K.J. (Eds.), Key Topics in Conservation Biology 2. John Wiley & Sons, Oxford, U.K., pp. 344–361.

Sudman, S., Sirken, M.G., Cowan, C.D., 1988. Sampling rare and elusive populations. Science 240, 991–6.

Thomas, C.D., Cameron, A., Green, R.E., Bakkenes, M., Beaumont, L.J., Collingham, Y.C., Erasmus, B.F.N., De Siqueira, M.F., Grainger, A., Hannah, L., Hughes, L., Huntley, B., Van Jaarsveld, A.S., Midgley, G.F., Miles, L., Ortega-Huerta, M.A., Peterson, A.T., Phillips, O.L., Williams, S.E., 2004. Extinction risk from climate change. Nature 427, 145–8.

Tian, G.-L., Tang, M.-L., 2013. Incomplete categorical data design: non-randomized response techniques for sensitive questions in surveys. CRC Press.

Tian, G.-L., Yu, J.-W., Tang, M.-L., Geng, Z., 2007. A new non-randomized model for analysing sensitive questions with binary outcomes. Stat. Med. 26, 4238–4252.

Tourangeau, R., Yan, T., 2007. Sensitive questions in surveys. Psychol. Bull. 133, 859–883.

Trappmann, M., Krumpal, I., Kirchner, A., Jann, B., 2014. Item sum: a new techniques for asking quantitative sensitive questions. J. Surv. Stat. Methodol. 2, 58-77.

Treves, A., 2009. Hunting for large carnivore conservation. J. Appl. Ecol. 46, 1350–1356.

Treves, A., Karanth, K.U., 2003. Human-carnivore conflict and perspectives on carnivore management worldwide. Conserv. Biol. 17, 1491–1499.

Tsuchiya, T., Hirai, Y., Ono, S., 2007. A study of the properties of the item count technique. Public Opin. Q. 71, 253–272.

Underwood, F.M., Burn, R.W., Milliken, T., 2013. Dissecting the illegal ivory trade: an analysis of ivory seizures data. PLoS One 8, e76539.

Vakilian, K., Mousavi, S.A., Keramat, A., 2014. Estimation of sexual behavior in the 18-to-24-years-old Iranian youth based on a crosswise model study. BMC Res. Notes 7, 28.

Van den Hout, A., van der Heijden, P.G.M., Gilchrist, R., 2007. The logistic regression model with response variables subject to randomized response. Comput. Stat. Data Anal. 51, 6060–6069.

Vitos, M., Lewis, J., Stevens, M., Haklay, M., 2013. Making local knowledge matter, in: Proceedings of the 3rd ACM Symposium on Computing for Development - ACM DEV '13. ACM Press, New York, USA.

Warner, S.L., 1965. Randomized response: a survey technique for eliminating evasive answer bias. J. Am. Stat. Assoc. 60, 63–69.

White, P.C.L., Jennings, N.V., Renwick, A.R., Barker, N.H.L., 2005. REVIEW: Questionnaires in ecology: a review of past use and recommendations for best practice. J. Appl. Ecol. 42, 421–430.

Xie, H., Kulik, L., Tanin, E., 2011. Privacy-aware collection of aggregate spatial data. Data Knowl. Eng. 70, 576–595.

Yu, J.-W., Tian, G.-L., Tang, M.-L., 2008. Two new models for survey sampling with sensitive characteristic: design and analysis. Metrika 67, 251–263.

Zigerell, L.J., 2011. You wouldn't like me when I'm angry: list experiment misreporting. Soc. Sci. Q. 92, 552–562.

Zink, T., Levin, L., Wollan, P., Putnam, F., 2006. Mothers' comfort with screening questions about sensitive issues, including domestic violence. J. Am. Board Fam. Med. 19, 358–367.

Table 1. Summary of methods reported in this study and a non-exhaustive list of studies in which these techniques were used

Technique	Previously used in conservation or natural resource management?	Methods comparison studies completed	Evidence that method can be adapted for use in illiterate community?	Possible data outputs
Randomised response technique (RRT; Warner et al. 1965)	Yes (Solomon et al., 2007; Blank et al. 2009; St. John et al., 2010, 2012)	RRT with direct questions (Solomon et al. 2007); RRT with UCT (Coutts and Jann, 2011); RRT with nominative (St. John et al., 2010)	Yes	Proportion of sample population engaging in sensitive behaviour + link to explanatory variables associated with behaviour
Nominative technique (Miller, 1985)	Yes (St. John et al., 2010)	Nominative with RRT and direct questions (St. John et al., 2010)	Yes	Proportion of sample population engaging in sensitive behaviour
Unmatched-count technique (UCT; Droitcour et al., 1991)	Yes (Nuno et al., 2013b)	UCT with direct questions (Tsuchiya et al., 2007); UCT with RRT (Coutts and Jann, 2011)	Yes	Proportion of sample population engaging in sensitive behaviour + link to explanatory variables associated with behaviour
Grouped answer method (Droitcour and Larson, 2002)	No	None	Yes	Proportion of sample population engaging in sensitive behaviour
Crosswise model (CM; Yu et al. 2008), Triangular model (TM; Yu et al. 2008), Diagonal model (DM; Groenitz 2014) Hidden sensitivity model (HSM; Tian et al., 2007)	No	CM with direct questions (Jann et al. 2012)	Maybe	Proportion of sample population engaging in sensitive behaviour + link to explanatory variables associated with behaviour
Surveys with negative questions (Esponda and Guerrero, 2009)	No	None	Maybe	Proportion of sample population engaging in sensitive behaviour
Bean method (BM; Lau et al. 2011)	No	BM with direct questions (Lau et al. 2011)	Yes	Proportion of sample population engaging in sensitive behaviour

FIGURE LEGENDS

Figure 1. An example instruction card for the forced response randomized response technique. Respondents are provided with an opaque beaker, two dice and a set of question cards each displaying the instructions. The dice are rolled and the instructions followed. Depending upon how the survey is administered, respondents provide their answers either by saying ‘yes’ or ‘no’ out load to an interviewer, or by personally recording their answer. The respondent never reveals the result of the dice role. Killing a leopard is used here (and in Figures 2 and 3) as an example of an activity of conservation concern that may be illegal in some study systems.

Figure 2. An example instruction card for the additive randomized response technique. Respondents are provided with a cloth sack containing numbered balls with a known distribution. Respondents select one ball from the sack and add the number shown on the ball to their numeric response to the question. The respondent never reveals the number displayed on the ball they select. Respondent may call their answers out load to an interviewer or record them personally.

Figure 3. Example questions for the nominative technique. This method could be administered through a face-to-face interview or self-administered using pen-and-paper, or computer. *Randomised selection requires respondents to write down the initials of each friend and number them from 1 to the end of the list; predefined instructions (e.g. if the number of close friends reported in question 1 is 5, ask about friend number 2 on the list) in order to identify which friend they should think about when answering the sensitive question(s).

Figure 4. An example of baseline and treatment unmatched-count technique (UCT) lists viewed by survey respondents randomly allocated to either baseline or treatment groups. Respondents are required to report the total number of items that apply to them without identifying any individual item. “Hunting” is used here (and all figures thereafter) as an example of an activity of conservation concern that may be conducted illegally in some study systems and/or under certain conditions.

Figure 5. An example of cards used for the grouped answer method. Depending upon the treatment group they are assigned to, respondents are required to report which group on Card 1 or 2 they belong to without identifying which items apply to them.

Figure 6. An example of a question card to be used in studies applying either the crosswise model or the triangular model. Respondents are asked to provide a joint answer to both questions following different rules according to specific technique.

Figure 7. An example of a question card to be used in studies applying the hidden sensitivity model. Respondents are asked to answer A, B, C or D according to the card instructions; people that have done any of the sensitive activities are required to answer irrespectively of their actual birthday, protecting their answers.

Figure 8. An example of a question card to be used in studies applying the diagonal model. After being read or shown two questions (one sensitive and the other non-sensitive), respondents should report the number (1, 2, 3 or 4) in the table that provides the required answer depending on both questions simultaneously.