Methods in Ecology and Evolution

DR AMY HINSLEY (Orcid ID : 0000-0002-5590-7617)
DR FREYA ANNA VICTORIA ST JOHN (Orcid ID : 0000-0002-5707-310X)

MISS ANA NUNO (Orcid ID : 0000-0003-4680-2378)

Article type : Review

Editor : Aaron Ellison

Asking sensitive questions using the Unmatched Count Technique: Applications and guidelines for conservation

Amy Hinsley[1*], Aidan Keane[2], Freya A. V. St. John[3], Harriet Ibbett[1], Ana Nuno[4]

[1] Department of Zoology, University of Oxford, Oxford, OX1 3PS, UK

[2] School of GeoSciences, University of Edinburgh, EH9 3FF, Edinburgh, UK

[3] School of Natural Sciences, Bangor University, Bangor, LL57 2UW, UK

[4] Centre for Ecology and Conservation, College of Life and Environmental Sciences, University of Exeter, Penryn, Cornwall TR10 9FE, UK

* corresponding author: amy.hinsley@zoo.ox.ac.uk

Running reader: **Guidelines for using UCT in Conservation**

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/2041-210X.13137

ABSTRACT

1. Researchers and practitioners are increasingly using methods from the social sciences to address complex conservation challenges. This brings benefits but also the responsibility to understand the suitability and limitations of these methods in different contexts. After years of use in other disciplines, the Unmatched Count Technique (UCT) has recently been adopted by conservation scientists to investigate illegal and socially undesirable human behaviours. Here we provide guidance for practitioners and researchers on how to apply UCT effectively, and outline situations where it will be most and least appropriate.

2. We reviewed 101 publications in refereed journals that used UCT to draw conclusions on its use to date and provide recommendations on when and how to use the method effectively in conservation. In particular, we explored: type of studies undertaken (e.g. disciplines; behaviour being studied; rationale for using UCT); survey administration (e.g. sample size, pilot studies, administration mode); UCT outcomes (e.g. type of analyses, estimates, comparison with other methods); and type of recommendations.

3. We show that UCT has been used across multiple disciplines and contexts, with 10 studies that focus on conservation and natural resource use. The UCT has been used to investigate topics falling into five categories: socially undesirable behaviours, socially undesirable views, illegal or non-compliant behaviours, socially desirable behaviours; and personal topics (e.g. being HIV positive). It has been used in 51 countries and is suitable to several situations, but limitations do exist, and the method does not always improve reporting of sensitive topics.

4. We provide best-practice guidance to researchers and practitioners considering using UCT. We highlight that alternate methods should be considered if sample sizes are likely to be small, the behaviour in question is likely to be extremely rare, or if the behaviour is not particularly sensitive. UCT can be a useful tool for estimating the extent of non-compliance within a

conservation context, but as with all scientific investigation, careful study design, robust sampling and consistent implementation are required in order for it to be effective.

Sumário:

1. Investigadores e outros conservacionistas têm vindo cada vez mais a usar métodos das ciências sociais para abordar desafios complexos de conservação. Isso traz benefícios, mas também a responsabilidade de entender a adequabilidade e as limitações desses métodos em diferentes contextos. Após anos de uso noutras disciplinas, a técnica de contagem de itens (TCI) foi adotada recentemente por cientistas da conservação para investigar comportamentos humanos ilegais e socialmente indesejáveis. Neste artigo, fornecemos orientações para investigadores e outros profissionais sobre como aplicar a TCI de maneira eficaz e descrevemos situações em que esta técnica poderá ser mais e menos apropriada.

2. 101 artigos que usaram a TCI em publicações arbitradas foram revistos de modo a tirar conclusões sobre o seu uso e fornecer recomendações sobre quando e como usar este método efetivamente em conservação. Em particular, explorámos: tipo de estudos realizados (por exemplo: disciplinas, comportamento em estudos razão para o uso da TCI); administração da pesquisa (ex.: tamanho da amostra, estudos piloto, modo de administração); resultados da TCI (ex.: tipo de análises, estimativas, comparação com outros métodos) e tipo de recomendações.

3. Mostramos que a TCI tem sido usada em várias disciplinas e contextos, incluindo 10 estudos sobre a conservação e uso de recursos naturais. A TCI tem sido usada para investigar tópicos que se enquadram em cinco categorias: comportamentos socialmente indesejáveis; opiniões socialmente indesejáveis; comportamentos ilegais ou incumpridores; comportamentos socialmente desejáveis; e tópicos pessoais (por exemplo, ser HIV positivo). Esta técnica já foi usada em 51 países e é adequada para várias situações, mas existem limitações, e o método nem sempre melhora o relato de tópicos sensíveis.

4. Fornecemos orientação sobre as melhores práticas para investigadores e outros profissionais que estejam a considerar o uso da TIC. Métodos alternativos devem ser considerados se os tamanhos amostrais forem relativamente pequenos, o comportamento em questão extremamente raro, ou se o comportamento não for particularmente sensível. A TCI pode ser uma ferramenta útil para estimar a extensão do incumprimento de regras num contexto de conservação, mas, como em todas as investigações científicas, é necessário planear o estudo cuidadosamente, efetuar uma amostragem robusta e uma implementação consistente para que seja eficaz.

Keywords: conservation regulations; conservation social sciences; indirect questioning; monitoring and evaluation; rule-breaking; sensitive questions; social desirability bias; specialized questioning techniques; unmatched count technique

INTRODUCTION

Researchers and practitioners working on complex conservation challenges are increasingly encouraged to adopt social science methods to better understand human behaviour (St. John, Keane & Milner-Gulland, 2013; Bennett et al., 2016). The creation of an interdisciplinary toolbox brings benefits, but also a responsibility to understand the requirements, feasibility and potential limitations of methods new to our field. Both qualitative and quantitative social science methods have proven valuable for investigating human dimensions of conservation. However, traditional methods face limitations when researching socially sensitive topics (Tourangeau & Yan, 2007). When asked directly, participants may conceal their true attitudes, beliefs or behaviours if the behaviour in question is illegal (Warner, 1965), or may temper their answers to appear more socially acceptable (social desirability bias), especially if data collection is observed by third-parties. For example, teenagers interviewed with their parents present

were more likely to deflate their smoking activity, compared to those surveyed privately (Tourangeau & Yan, 2007). Perceived invasions of privacy, distrust of the interviewer and their research intentions, or fear of reprisal may also lead respondents to refuse to answer questions altogether (refusal bias) (Tourangeau & Yan, 2007). Within conservation, these issues are particularly pertinent when researchers work in remote, rural communities, where levels of literacy may be low, power-relations prevalent, and distrust of outsiders, foreigners, and authorities high (e.g. Razafimanahaka et al., 2012). In such situations, researchers may not be perceived as neutral, due to associations with non-governmental organisations or government agencies, which may exacerbate topic-sensitivity.

In recognition of the biases associated with direct questioning about sensitive or stigmatising topics, social scientists developed specialised questioning approaches, known as indirect questioning techniques, with the aim of providing respondents with greater levels of privacy and anonymity (Chaudhuri & Christofides, 2013). The most well-known is the Randomised Response Technique (RRT), proposed by Warner (1965) but many other techniques now exist (Nuno & St. John, 2015). One such method is the Unmatched Count Technique (UCT), which was developed to investigate topics such as racism (Kuklinski, Cobb & Gilens, 1997). Since an initial application in 2013 (Nuno et al., 2013), UCT has increasingly been used to understand sensitive conservation topics (e.g. illegal wildlife trade: Hinsley, Nuno, Ridout, St. John, & Roberts, 2017) with five studies deploying UCT in 2017 alone (see Appendix I).

Originally the 'block total response' method (Raghavarao & Federer, 1979) and also commonly known as the item count technique and list experiment (Glynn, 2013), UCT involves randomly assigning individuals into two groups: control and treatment. The control group receive a list of non-sensitive statements or 'items' whilst the treatment group receive the same list of innocuous items, along with a sensitive item. Individuals in both groups are asked to indicate

how many, but not which items apply to them (Fig.1). Prevalence is estimated by calculating the difference in means between the two groups (Droitcour et al., 1991): p = mean (treatment group) – mean (control group), where p is the proportion of participants engaged in sensitive behaviour.

Please read the following statements and tell us how many are true for you. You do not need to tell us which statements are true for you, just the total number

Control Group
I have never bought orchids at an orchid show
I am a member of a Facebook orchid group
I have a species [orchid] collection
I have been a member of an orchid society for more than a year

Treatment group
I have never bought orchids at an orchid show
I am a member of a Facebook orchid group
I have personally sent or carried an orchid across an international border without obtaining the required CITES paperwork
I have a species [orchid] collection
I have been a member of an orchid society for more than a year

Figure 1: UCT lists used online to estimate the prevalence of orchid-related CITES infractions (Hinsley et al., 2017). Respondents were randomly assigned to either the control or treatment group and answered using a drop-down list.

Respondents never reveal having the sensitive characteristic as long as they report a value lower than the total number of items on the treatment list. However, secrecy is removed if someone reports possessing all characteristics in the treatment list, meaning that they may under-report their true answer to avoid admitting directly to the sensitive item ('ceiling effects'). Conversely, if someone only possesses the sensitive characteristic, they may over-

report the number of non-sensitive characteristics they possess, to conceal their answer ('floor effects') (Zigerell, 2011). Ways to minimize these issues are discussed later in this paper.

Interest in using UCT has grown in conservation (see Appendix I) as researchers trying to understand the prevalence of illegal or otherwise sensitive behaviours looked beyond their field for appropriate methods. In other disciplines, UCT studies were producing higher prevalence estimates of sensitive behaviours than both direct questioning (Tsuchiya, Hirai & Ono, 2007) and other indirect methods, such as RRT (Coutts & Jann, 2011). Furthermore, UCT was promoted as easy to administer; participants simply state how many items on a pre-prepared list apply to them (Dalton, Wimbush & Daily, 1994). Indeed, the first conservation study to use UCT demonstrated that it can be adapted for use in areas of limited literacy, where respondents reported being comfortable with UCT as a questioning technique (Nuno, Bunnefeld, Naiman, & Milner-Gulland, 2013). However, UCT has limitations. It is unsuitable for very rare behaviours due to its lower precision, and requires large sample sizes (Ulrich, Schröter, Striegel, & Simon, 2012). For example, a UCT survey of >1000 households surrounding the Serengeti, Tanzania, returned an estimate of hunting prevalence with a ±5% standard error (Nuno et al., 2013) whilst an online questionnaire of 814 orchid growers estimated a smuggling prevalence with a ±6% standard error (Hinsley et al., 2017). Depending on specific research questions and research-user needs (e.g. monitoring behaviour prevalence over time), this uncertainty might undermine the use of this information to guide decisions, meaning trade-offs between accuracy and precision must be carefully considered.

Stimulated by a need to increase method efficiency, variations of UCT have emerged, including the double-list UCT (Glynn, 2013) and single sample count (Petroczi et al., 2011) (Fig. 2). In the double-list UCT, participants act simultaneously as control and treatment groups by answering two lists, one of which always has the sensitive statement. Again, respondents are randomly

allocated into two groups but on this occasion, half receive Control A and Control B + sensitive item, whilst the remainder view Control B and Control A + sensitive item (See Appendix II). Estimates of the sensitive characteristic are averaged across the two groups to derive its prevalence (Droitcour et al., 1991; Glynn, 2013):

$p = \frac{1+\ 2}{2}$, with

p_x= mean (treatment group $_x$)-mean (control group $_x$), where p is the proportion of participants engaged in sensitive behaviour.

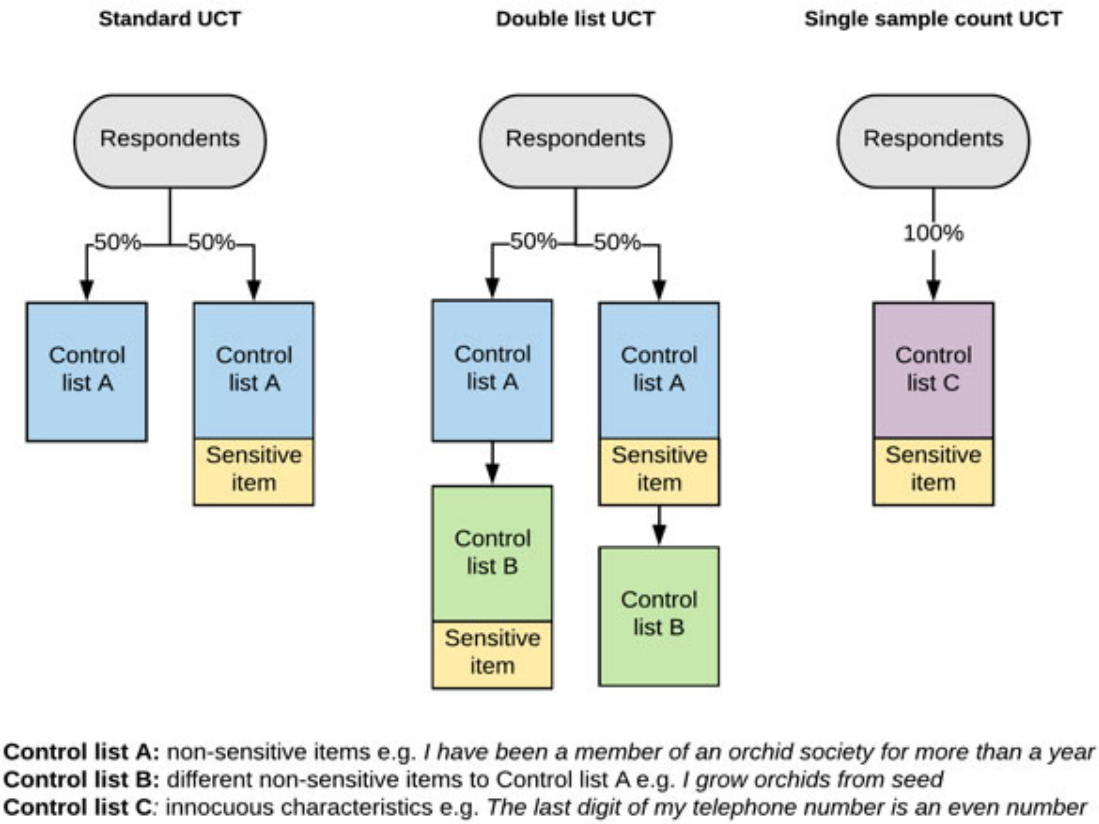

Figure 2: Structure of the standard UCT experiment and modified versions: the double-list and single sample count UCT.

The single sample count (Petróczi et al., 2011), utilises existing data on population prevalence of innocuous characteristics (e.g. birth-month or final digit of telephone number). This approach avoids the need to estimate the prevalence of non-sensitive items meaning all participants are asked the sensitive question (See Appendix II). The prevalence estimate from single sample count data is then calculated as:

$$p = (\lambda / n) - b$$

where p is the estimated population distribution of the sensitive item, λ is the observed number of 'yes' answers, n is the sample size, and b is the population mean value of responses for the baseline non-sensitive questions. However, there may be challenges in locating data on non-sensitive characteristics that complement sensitive topics of conservation interest.

Again, drawing inspiration from UCT, Trappmann, Krumpal, Kirchner and Jann (2014) proposed the item sum technique which enables researchers to estimate quantities of sensitive activities (e.g. number of hours engaged in undeclared work). Respondents randomly assigned to the control group respond to a question such as 'How many hours did you spend in village meetings during the last three months?' whilst the treatment group answers the control and a sensitive question (e.g. 'How many hours did you spend hunting in the last three months') simultaneously. Ideally, but not essentially, activities should be measured on the same scales (e.g. hours: Trappmann et al., 2014). The quantity of the sensitive act is estimated by computing the mean difference of answers between control and treatment groups.

Given the growing interest in UCT amongst conservation researchers and practitioners (Appendix I), it is essential to acknowledge and understand the uncertainty about if and when UCT provides better estimates. This is especially important given the economic and temporal expense involved in conducting UCT surveys, particularly given demanding samples sizes. We

therefore conducted a systematic review of all empirical applications of UCT in peer-reviewed literature, which aimed to:

1) Define the scope of subjects to which UCT has been applied;
2) Identify trends in design, implementation, and analysis methods; and
3) Identify the reported success of the method, and the challenges encountered.

We use this information, and our own experience, to provide best practice guidelines to conservationists, highlighting when to use UCT, potential pitfalls, and robust study design tips. Ultimately, this will provide a better understanding of the challenges and opportunities for employing UCT, allowing more critical and robust applications of this method for improving data collection about sensitive behaviours, and informing conservation decisions.

METHODS

We searched Web of Science and Scopus using the keywords "unmatched count technique", "item count technique", "list experiment", "list response technique", "list method", "list randomization" or "block total response method". At the search date (14th April 2018), this set of criteria identified 661 papers. All titles and abstracts were read by AN, who assessed if each paper potentially used a form of UCT; 121 studies met these criteria. In order to ensure inclusion of recent conservation publications, an additional search in Google Scholar was undertaken; another five papers were identified and included because they specifically dealt with UCT in conservation. A total of 126 articles were thus transferred to the following step.

These 126 papers were randomly assigned for review by AH or AN. For each paper, we recorded: discipline/field; behaviour(s) being studied; rationale for using UCT; location; spatial

scale; survey administration (e.g. sample size, pilot study, administration mode and material); and survey outcomes and conclusion (e.g. type of analyses, estimates, comparison with other methods) (see Appendix III for full details). Of 126 papers, 14 (11.1%) were theoretical only and/or only used data reported in other studies, 6 (4.8%) used non-directly comparable methods (e.g. item sum technique), 2 (1.6%) applied UCT with no sensitive statement to test baseline list design, 2 (1.6%) could not be accessed, even after contacting authors, and 1 (0.8%) was not written in English; thus, these 25 papers (19.8%) were excluded from further analysis. The remaining 101 papers were then analyzed and summarized (see Appendix IV for all reviewed papers). To investigate potential effects on binary variables (e.g effects on likelihood of checking design assumptions or not), generalized linear models with binomial error distribution and a logit link were fitted. In order to account for the quantitative nature of the information but without making assumptions about the distance between ordered categories, an ordered logistic regression was used to assess trends in use of increasingly complex statistical approaches over time.

In the final 101 papers, authors often used multiple UCT lists to explore, for example, different behaviours or different modes of survey administration. To explore potential effectiveness of UCT, we focused on the 73 studies comparing UCT to direct questions, which had a total of 229 separate UCT lists. Using an ordered logistic regression, UCT outcomes measured in terms of "success" (i.e. increase in social desirability bias, no significant difference and reduction in bias when compared to direct questioning) were analysed in function of key study characteristics (administration mode, number of control statements, match in topic of statements, design assumptions checked, and pilot conducted) to explore potential predictors of UCT effectiveness.

RESULTS

Types of studies

We found that UCT has been applied to a wide range of topics, including abortion, anti-immigration, plagiarism, and voting. In studies related to natural resource use, all topics related to non-compliance with conservation regulations (e.g. illegal fishing, hunting). The 101 reviewed studies justified the use of UCT due to their topic(s) of interest being in one or more of the following categories: socially undesirable behaviours (e.g. promiscuity) (n=38 papers); non-compliant/illegal behaviours (e.g. smuggling) (n=26); socially undesirable views (e.g. racism) (n=21); socially desirable behaviours (e.g. recycling) (n=13); and personal topics linked to possessing a socially stigmatised characteristic (e.g. being HIV positive) (n=7).

The most frequent study fields were political science (n=41), sociology (n=12) and health (n=10). Ten studies specifically related to biodiversity conservation or natural resource management. Other fields included: statistics (n=9), development studies (n=7), psychology and organization studies (n=4 each), and migration studies and veterinary science (n=2 each). Studies were conducted in 51 countries, with the highest number of studies in the USA (n=41), and few or no studies in Asian and African countries (Fig.3). The majority of studies were at the national level (n=55), followed by regional (n=28), local (n=10) and international (i.e. multiple countries; n=7). No studies used UCT to study changes in behaviour prevalence over time.

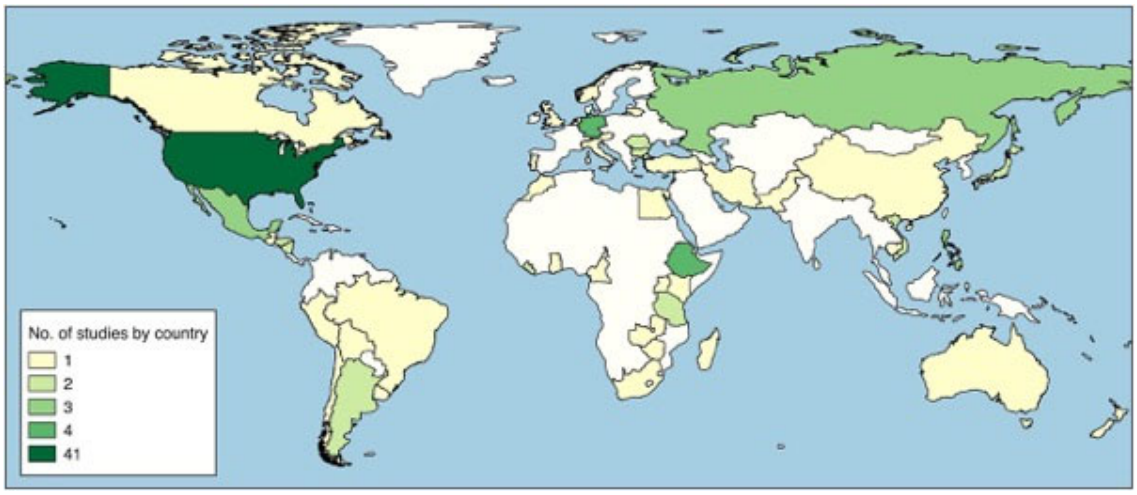

Figure 3. World map showing the locations of published UCT studies (n=101).

UCT implementation among surveyed studies

Most studies were administered face-to-face (n=42) or online (n=31); other modes of administration included: phone (n=11) and self-administered questionnaires on paper or digital devices (n=8), whilst six studies deployed multiple modes. Three did not report how surveys were administered. Most studies used a list of statements read by/to participants; only 6.9% (n=7) used pictures, and one study provided a counting device (stones) to help participants count the number of items that applied to them. Three studies explicitly mentioned using a training question before the actual UCT. Nine studies used the double-list design. Baseline lists included between two (n=1) and seven (n=2) control items, with the majority of studies using three (n=40) or four (n=46). Ideally, baseline items should match the subject of the sensitive item so that they do not stand out (Glynn, 2013), which was true in 68.3% (n=69) of the studies.

Twenty-nine (28.7%) studies explicitly mentioned piloting, but this information was often brief and/or placed in appendices, so it is possible that more studies pre-tested in some way without mentioning this in the manuscript. Pilot studies were used to: identify appropriate non-sensitive items for control lists; obtain baseline prevalence rates for non-sensitive items to avoid using

those which were either too rare or too common; estimate correlation between items to reduce variability; and refine wording and order of questions.

The number of survey participants receiving UCT questions ranged from 50 to 24,020 people (median=1000, lower quartile=562, upper quartile=1605). Surprisingly, the sample size was not reported for two studies. We expected that face-to-face studies would have limited sample sizes due to the amount of resources they require, but we found face-to-face studies with sample sizes ranging from 50 to 13,686 respondents. Larger sample sizes were often achieved by embedding UCT questions within existing nationwide surveys (e.g. Çarkoğlu & Aytaç, 2015) or using professional survey companies (e.g. Kiewiet De Jonge, 2015).

Outcomes and conclusions from surveyed studies

After data collection, 30.7% (n=31) of studies checked if at least some UCT design assumptions, such as randomization or design effects, were met; with more recent publications more likely to report checking assumptions ($z=3.4$, $p<0.001$). Most studies (n=87) provided prevalence estimates of the sensitive item, but only around half of these clearly reported a full estimate with standard errors (n=44), or confidence intervals (n=10). Some (n=28) provided no information about estimate uncertainty, with four of these only reporting results graphically. Difference-in-means was the most commonly applied approach for calculating prevalence (n=52 studies), followed by linear regression, generalized linear modelling or generalized linear mixed modelling (n=32). Sixteen studies used more advanced multivariate analyses, such as non-linear least squares and Bayesian versions of estimation procedures, while one study did not estimate any behaviour prevalence rates. More complex analytical procedures were increasingly more likely over time ($t=1365$, $p<0.002$).

Overall, 73 studies compared UCT to direct questioning, with ten studies comparing directly to at least one other indirect questioning approach, such as RRT (Fig. 4). Across these 73 studies, 229 separate UCT question lists were asked, ranging from 1 list (n=61) to 16 lists (n=1) per study. When compared to estimates from direct questioning, online, phone and self-administrated surveys were less likely to result in successful (i.e. reduce bias vs no significant effect or increase bias) UCT applications when compared to face-to-face questions (Table 1; Fig. 4).

Table 1. Parameter estimates obtained from ordered logistic regression fitted to "UCT success" (i.e. increase/no significant difference/decrease in social desirability bias compared to direct questioning) for the 223 UCT lists that reported all variables. A negative estimate shows less success in reducing bias. Reference levels: face-to-face survey; no match between statement topics; design assumptions not checked; no pilot.

Parameters	**Estimate (S.E.)**	**t-value**	**Confidence intervals**	
			2.5%	**97.5%**
Mode of administration: Multiple	-19.991 (0.001)	-0.001	-19.991	19.991
Online	-1.513 (0.460)	-3.289	-2.415	-0.612
Phone	-1.856 (0.684)	-2.714	-3.197	-0.516
Self-administrated questionnaire	-1.134 (0.453)	-2.506	-2.021	-0.247
Number of control statements	-0.166 (0.150)	-1.103	-0.460	0.129
Match between topic of statements: medium	0.788 (0.449)	1.754	-0.093	1.669

high	-0.523 (0.377)	-1.389	-1.261	0.215
Design assumptions: checked	0.084 (0.374)	0.247	-0.649	0.816
Pilot study: yes	0.474 (0.398)	1.191	-0.306	1.254

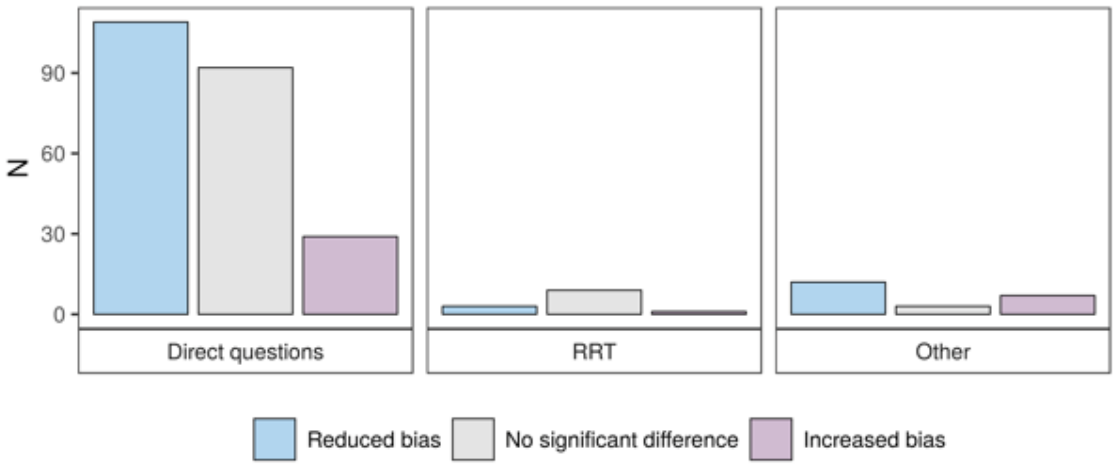

Figure 4: Outcome of all UCT lists with comparisons to direct questions (n=229), the Randomised Response Technique (RRT) (n=13), and other methods (n=22). A UCT that significantly increased reporting of a socially undesirable, illegal, or personal question, or decreased reporting of a socially desirable behaviour was defined as having reduced social desirability bias, with the opposite result defined as an increase in bias.

Recommendations for implementing UCT

Some surveyed studies (n=27) provided specific recommendations for future implementation of UCT (Table 2).

Table 2. Summary of published recommendations for improvement of UCT implementation

Recommendation	**Example**	**No. papers recommending**
Investigate the use of UCT in different contexts	Behaviour over time; in different cultures; when the sensitive behaviour is highly prevalent	7
Investigate the cognitive process involved in answering UCT, and how this influences answers	Understand bias in responses created by people understanding the method differently e.g. how education level affects comprehension of UCT	6
Improve statistical analysis	Develop and improve multivariate analyses for UCT analysis	5
Use UCT alongside other specialised methods & compare	More comparisons to existing methods; develop new methods to test UCT against	4
Improve the design	More efficient design of baseline lists	3
Use UCT alongside direct questions	As validation for direct answers; to test accuracy of UCT	2

DISCUSSION

When to use, and when not to use UCT

Whilst UCT has developed into a useful research tool, it is clearly not suitable in all situations. Furthermore, grey-literature and informal sources (e.g. blogs) reveal varied opinions regarding the utility of UCT to return robust prevalence estimates (Gelman, 2014), suggesting many authors leave negative results unpublished. As we show in Figure 4, UCT does not always perform better than other methods. Key factors such as the sensitivity of the question being asked, the likely prevalence of the behaviour within the study population, and the study design itself need to be considered before the method is used.

These factors should be carefully considered, as recent conservation studies, similar to our findings in the wider literature, have reported mixed success with UCT. For example, whilst positive estimates were derived by asking people questions directly, Nuno et al. (2018) reported a negative, although overlapping with zero and thus inconclusive, UCT estimate of illegal turtle egg consumption amongst 560 participants in the Cayman Islands. Similarly, whilst 8.6% of 616 Cambodian respondents reported hunting birds when asked directly, Ibbett et al. (2017) again reported inconclusive results, with no significant difference between UCT control and treatment estimates.

Steps to designing an effective UCT

As with all research methodologies, the quality of UCT data depends on the quality of the survey instrument, and the care with which data are collected. Prior to designing a survey that incorporates UCT, a clear research question should be established, relevant literature should be reviewed, and the study context considered. Ultimately, careful and thorough study design

underpins this process (St. John, Keane, Jones & Milner-Gulland, 2014). Our findings suggest that, while UCT is increasingly applied in a wide range of contexts and locations, some questions and concerns remain in terms of reliability, and it is clear that the method is not applicable in all situations. Our findings of ongoing design and statistical developments, a more critical consideration of UCT success, and multiple comparisons between UCT and other methods are, however, encouraging and acknowledge both challenges and opportunities for this method. Here, we offer practical recommendations for applying UCT in conservation, drawn from our review findings, recommendations published in the methodological literature, and our personal experiences in applying the method.

Step 1. Sensitive item definition

One of the first choices that must be made is exactly how the sensitive item is to be defined. Sensitive items in conservation are typically an action (e.g. hunting wild animals), but the level of specificity (e.g. "hunting" or "setting wire snares"; whether the target is a specific species or a higher taxonomic group) and timescale (e.g. "in the last month" or "in the last year") may vary according to study needs. It is also important to consider the unit of analysis required to answer your research question (e.g. are you interested in studying the behaviour at the individual or household level). Multiple UCT questions can be used within a single survey to ask about different behaviours (e.g. four behaviours relating to illegal orchid trade: Hinsley et al. 2017), however including too many questions may induce respondent fatigue.

Step 2. Selection of non-sensitive items

The next step is to choose a set of non-sensitive, control list items. Although conceptually straightforward, selecting control items raises a series of subtle challenges which are not always fully appreciated. Ideally, a long list of potential items should be piloted to select the final list

(e.g. see Hinsley et al., 2017). Deciding on the number of non-sensitive items is important; typically, applications of UCT in the literature use three to four control items. While shorter lists place less cognitive burden on respondents and provide greater statistical precision, they must be carefully designed to avoid ceiling and floor effects, in order to protect individual respondents (Zigerell, 2011). Longer lists naturally mitigate these risks, but at the cost of lower statistical power. Another way to minimise these risks is to avoid using items that are likely to be very high or very low prevalence in your sample. However, increasing the overall level of variation in responses can also reduce the statistical power of the technique (Glynn, 2013). Consequently, the best current advice is to choose a set of control items that have at least one pair of negatively-associated items i.e. a respondent saying yes to one item will be very likely to say no the other (Glynn, 2013). For example, Hinsley et al. 2017 used a pilot study to measure association between candidate control statements to ensure that some included in the final design were negatively correlated (e.g. "I have never been to an orchid show" and "I have won awards for my orchids"). If that is not possible, a mixture of high and low prevalence items should be used (Tsuchiya et al., 2007). Prevalence and association between different potential items can be determined using pilot data. Ideally, the control items should be reasonably familiar to the respondent and sufficiently similar in nature and specificity to the sensitive item so that it does not stand out (Droitcour et al., 1991; Kuklinski et al., 1997). For example, Nuno et al. (2013) estimated hunting prevalence by presenting it amongst other livelihood strategies. However, it is also important to avoid control items that might themselves be subject to biased, for example by being sensitive in some unanticipated way, or socially-desirable and therefore prone to positive exaggeration.

Step 3. Pre-testing and piloting

Piloting and pre-testing is an essential part of a UCT project, but many empirical papers did not mention this stage, or give details of pilot studies that would enhance replicability of the

experiment. The need to balance respondent secrecy, statistical precision, and ease-of-comprehension means that researchers must have an excellent understanding of the system they wish to study. The early stages of a UCT design should draw heavily on qualitative understanding, existing literature and local expertise, and initial design ideas and "long lists" of potential control items should be carefully pre-tested, ideally with respondents who are representative of the target group. These pre-tests might take the form of in-depth qualitative interviews and/or focus group discussions, with the primary focus on assessing cultural acceptability, determining the most appropriate mode of administration (e.g. avoiding written descriptions if illiteracy is likely; providing physical tokens such as beans or stones to assist counting if required), and checking whether the chosen approach is understandable and captures the data intended. Photographs or illustrations are sometimes incorporated into UCT cards, but care must be taken that respondents are used to are locally appropriate (e.g. that icons or illustrations are interpreted correctly or that respondents are used to interpreting photographs and can recognise the items pictured) (Keane, Ramarolahy, Jones & Milner-Gulland, 2011). Once a good candidate design has been chosen, it should also be formally piloted to determine whether further refinements are required and to ensure that respondents are comfortable with the method and confident that their privacy is being protected.

Step 4: Choosing between standard UCT design and its variants and extensions

In order to overcome some UCT limitations, several variants of the basic method exist. For example, in some settings we may be interested in more than one sensitive item (e.g. we want to obtain estimates of hunting prevalence separately for two threatened species). In this case, rather than dividing the sample into two groups ("control" and "treatment"), the sample can be divided into three or more groups ("control", "treatment 1" and "treatment 2"), provided there is a large sample size.

A common challenge for the application of UCT is the need for large samples. The double-list UCT can reduce the overall cost of obtaining a prevalence estimate of the sensitive behaviour with the desired precision (Glynn 2013), but is harder to design (e.g. it can be challenging to find enough reasonable control items), may be harder to explain to research assistants, and increases the risk of respondent fatigue. Another modification designed to increase the precision of estimates is to ask respondents who do not receive the sensitive item separately about each of the control items (Corstange, 2009). Other authors have argued that after a UCT, respondents should be asked about the sensitive item directly whenever this is ethical and feasible (Blair & Imai, 2012). However, comparison between UCT and direct questions in conservation have sometimes shown no significant differences, a finding that reflects our review of the wider literature (Fig. 4). This has been reported to be because the behaviours were not sensitive enough to create bias (Thomas, Gavin and Milfont, 2015) or because the survey was carried out online, where participants felt sufficiently protected (Hinsley et al, 2017).

Including a direct question also allows researchers to test some core assumptions of the method, such as the lack of design effects (i.e. to verify that responses to the non-sensitive items are not affected by the presence or absence of the sensitive item), the honesty of responses (i.e. to assess whether respondents include the sensitive item in their count if it applies to them), and the ignorability of treatment assignment (i.e. that the allocation of respondents to treatment or control groups is truly random) (Imai, 2011). Procedures for testing UCT assumptions are implemented in specialised software designed for UCT analysis, such as the list package (Blair & Imai, 2010) in R.

Step 5: Deciding what else to ask

Although UCT is inherently a quantitative, large-sample technique, it is almost always a good idea to use it together with other forms of data collection. Simple follow-up questions can help researchers understand how sensitive the topic was for participants, how much they trusted the protection offered, and how easy they found the technique to follow (e.g. Thomas et al. 2015). It can also provide opportunity to collect additional information that can improve understanding of the context in which the sensitive behaviour takes place. For example, multivariate analyses that incorporate UCT data and socio-demographic or socio-psychological information can improve our understanding of why people engage in sensitive behaviours of conservation concern (e.g. Hinsley et al. 2017).

Step 6: Implementing UCT in the field

The success of even the best-designed UCT study depends on how the method is implemented in the field. In some conservation contexts, choosing a survey mode may be straightforward, for example where respondents are located in areas where online or telephone surveys are not possible. While our findings suggest that online, phone or self-administered UCTs are less successful than face-to-face UCTs at reducing social desirability bias compared to a direct question, there are exceptions (e.g. a successful online UCT about drug use: Coutts & Jann, 2011), and UCTs that are not face-to-face should still be used where pilot results suggest that direct questions are not ethical or appropriate.

In conservation, several other aspects of implementation deserve particular attention, the most important of which is the ethics of studying sensitive or illegal behaviours. Techniques such as UCT can provide an additional level of anonymity and protection for individual respondents, but it is still important that free prior informed consent is sought from all participants. It should

also be noted that UCT does not assure perfect protection, as individual anonymity may not automatically translate into anonymity for a group (St. John et al., 2016).

For an unfamiliar technique such as UCT, careful thought should therefore be given to how the research will be introduced and explained. An introductory script should be planned and included in the pre-testing and piloting to ensure that it uses appropriate language and can be easily understood. Planning should also include practicing agreed verbal explanations, in case respondents have poor eyesight or do not understand the words or images shown. A practice UCT question on an innocuous topic can also be useful. Although this was rarely implemented in our reviewed studies, it can help participants to understand what they need to do, while also acting as a warm-up exercise, which can help build rapport between participant and researcher.

Other simple considerations can also help to improve the quality and usefulness of UCT data. Although UCT is usually intended to be answered by an individual (whether for themselves or on behalf of a household), in conservation settings it may be difficult in practice to create a situation where the intended respondent is alone when giving their answers. The presence of other people (e.g. other household or community members) may influence the participants' responses and their willingness to answer honestly so it is important to have a clear plan in place for how to deal with this. Careful preparation can also help to make it easier for research assistants to follow experimental design and to record the data accurately, while also ensuring that the technique is presented in a simple, clear way to participants (e.g. randomisation to treatment or control groups should be carried out in advance).

What to use instead

Considering the time and resources needed to follow these steps and design and implement a good quality UCT, it is clear that several factors must be considered in order to assess the suitability of the method in any given study (Fig. 5).

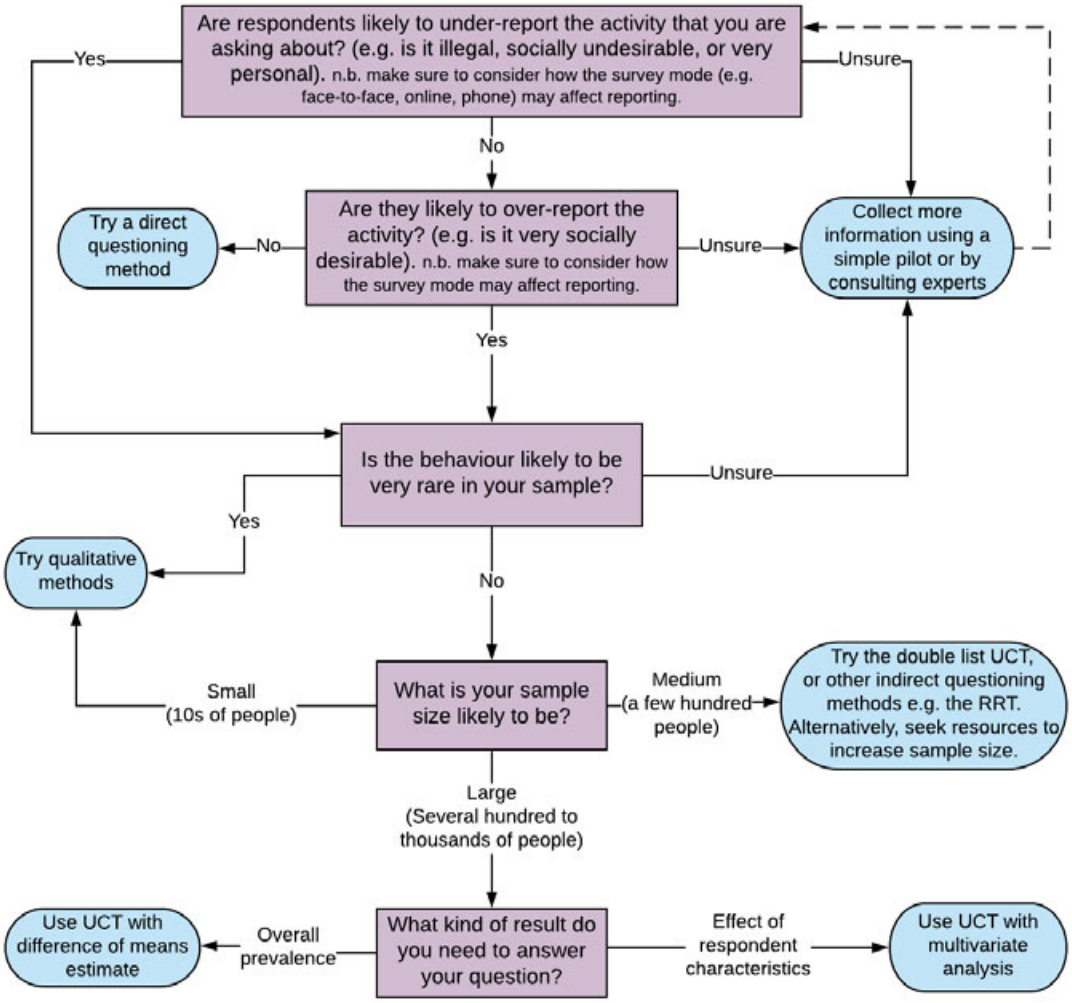

Figure 5: Decision tree to assess when UCT is suitable to use, and when other methods may be more appropriate.

Where UCT is not appropriate, there are several other options. Six other types of indirect questioning technique have been identified, all of which ensure respondent privacy to encourage honest reporting (Nuno & St. John, 2015). The widely used randomised response technique (RRT) (Warner, 1965) uses a randomising device (e.g. dice or a spinner) to force some respondents to answer 'yes' or 'no' to the sensitive questions. By considering the probability of forced yes responses, the population prevalence of the sensitive act can be calculated. Further, a specialised form of logistic regression can be applied to investigate predictors of rule-breaking (Heck, 2018; Chang, Cruyff & Giam, 2018). Different forms of RRT have been applied in various conservation contexts to estimate the proportion of people engaged in sensitive acts (Solomon, Jacobson, Wald & Gavin, 2007; Nuno & St. John 2015) and the frequency of such acts (St. John et al., 2018). However, RRT takes time to explain to participants, can be cognitively burdensome, and some do not like being 'forced' to give an answer that implies they are a rule-breaker, when their true answer is otherwise. This latter effect may explain why RRT estimates of hunting by men living near Kerinci Seblat National Park, Indonesia were negative for three of four study species (St. John et al., 2018). In our reviewed studies, most UCT and RRT results were not significantly different, so consideration of the various limitations and trade-offs should be used to choose the best method.

Where the aim of a study is simply to gauge the group-level prevalence of a sensitive behaviour and multivariate analysis of individual-level data is not required, the simplicity of the bean method (Lau, Yeung, Mui, Tsui & Gu, 2011) offers considerable potential. Participants are presented with one large and one small jar containing a known number of different coloured beans (or counters). The quantity of beans in each jar should be sufficient that the movement of beans between jars cannot be detected visually. In private, respondents are instructed to move a specified type of bean from the small jar to the large jar if their answer is yes; a different type of pre-specified bean should be moved if their answer is no. Once multiple people have

participated, changes in bean composition are used to estimate the prevalence of the sensitive acts (Lau et al., 2011).

Other indirect methods are described in Nuno & St. John (2015). However, whilst in various publications we have advocated for the use of specialised methods when studying sensitive topics (Nuno & St. John, 2015; St. John et al., 2016; Nuno et al., 2018) there is nothing to be gained from deploying such techniques when topics are not sufficiently sensitive to bias responses to direct questions. Illegality does not always translate to unwillingness to divulge information. For example, whilst respondents close to protected areas were less likely to admit to eating sifaka (legally protected) when asked directly compared to via RRT, participants far from the protected area complex were equally willing to report their involvement in this act directly as they were via RRT (Razafimanahaka et al., 2012). Such signals may only become visible once a study has been completed. However, initial enquiries amongst local experts coupled with informal interviews with individual's familiar with typical survey respondents can provide valuable insight to study design.

Where there is reason to believe that indirect questioning techniques are not required or are deemed inappropriate, methods including hypothetical scenarios, vignettes and interviews may provide useful insights. For example, Travers, Clements and Milner-Gulland (2016) used hypothetical scenario questioning to explore how socio-economic changes and increased law enforcement effort might affect households' decisions to illegally clear land in Cambodia. Results revealed that doubling enforcement effort would have little impact on rule-breaking due to weakly perceived rule of law and high corruption levels (Travers et al., 2016). Vignettes, short descriptions of hypothetical situations, can be used to elicit participants' behavioural responses in situations similar to the one described, without individuals having to reveal their actual behaviour (Eifler, 2007). However, it is also important to note that discrepancies often

exist between what people say they would do and what people actually do (Eifler, 2007). Where access to known rule-breakers is possible, interviews can help develop an in-depth picture of factors underpinning deviant acts. For example, Gore et al. (2013) used interviews to explore local conceptions of corruption and drivers of rule-breaking behaviour in eastern Madagascar. Similarly, semi-structured interviews with self-declared poachers in Tanzania revealed information on hunting methods, motivations and seasonality (Knapp, Peace & Bechtel, 2017).

Conclusions

The UCT is simple to administer and the data are easy to derive prevalence estimates from, meaning that it often seems like a 'silver bullet' that will enable researchers to rapidly collect data on conservation rule-breaking. It is true that with careful design, UCT can provide useful results in situations where direct questioning is difficult, or help to validate the answers from other methods. However, it is the responsibility of researchers to understand the limitations of the methods they are using, and the contexts in which they are most suitable. With better understanding of how best to use them, methods such as UCT have real potential to allow researchers and practitioners to produce reliable findings that can be used to underpin conservation decision-making.

Acknowledgements

AH acknowledges the support of the Oxford Martin Programme on the Illegal Wildlife Trade. AK and HI were funded by the UK National Environment Research Council (Grant ref. NE/N001370/1). AN acknowledges the support of the Darwin Initiative.

Authors' contributions

AH and AN developed the data collection protocol, analysed data and wrote the methods and results. FSJ and HI wrote the introduction and part of the discussion. AK wrote most of the discussion. All authors contributed to the figures and edited the text.

Data accessibility

All data used in the analyses are freely available in the University of Oxford research archive at this link: https://ora.ox.ac.uk/objects/uuid:556a8a97-2d3d-4bf2-8fc1-359ce9786986. Data were gathered from 101 English language publications that empirically tested the UCT method. For each paper, information on 17 variables was collected, including the context (e.g. discipline, behaviour studied, rationale for using UCT, location), details of survey administration (e.g. whether a pilot study was conducted, whether design assumptions were checked), type of analysis, and comparisons to other methods (e.g. direct questions). Full data collection protocol and citations for all 101 reviewed publications are included in Appendix III and IV.

References

1. Bennett, N.J., Roth, R., Klain, S.C., Chan, K.M.A., Clark, D.A., Cullman., G.,…Verissimo, D. 2016. Mainstreaming the social sciences in conservation. *Conservation Biology*, 31(1),56–66.doi:10.1111/cobi.12788
2. Blair, G. & Imai, K., 2010. list: Statistical Methods for the Item Count Technique and List Experiment. Available at: https://CRAN.R-project.org/package=list.
3. Blair, G. & Imai, K., 2012. Statistical Analysis of List Experiments. Political analysis: an annual publication of the Methodology Section of the American Political Science Association, 20(1),47–77.
4. Çarkoğlu, A. & Aytaç, S.E., 2015. Who gets targeted for vote-buying? Evidence from an augmented list experiment in Turkey. European Political Science Review, 7(4),547-566.
5. Chang, C.H., Cruyff, M.J.L.F. & Giam, X., 2018. Examining conservation compliance with randomized response technique analyses. *Conservation Biology*.
6. Chaudhuri, A. & Christofides, T., 2013. *Indirect Questioning in Sample Surveys* 1st ed., Springer-Verlag Berlin Heidelberg.
7. Corstange, D., 2009. Sensitive Questions, Truthful Answers? Modeling the List Experiment with LISTIT. Political analysis: an annual publication of the Methodology Section of the American Political Science Association, 17(1),45–63.
8. Coutts, E. & Jann, B., 2011. Sensitive questions in online surveys: Experimental results for the randomized response technique (RRT) and the unmatched count technique (UCT). *Sociological Methods and Research*, 40(1),169–193.
9. Dalton, D.R., Wimbush, J.C. & Daily, C.M., 1994. Using the Unmatched Count Technique (UCT) To Estimate Base Rates for Sensitive Behavior. Personnel Psychology, 47(4),817–829.
10. Kiewiet De Jonge, C.P., 2015. Who lies about electoral gifts? Experimental evidence from Latin America. Public Opinion Quarterly, 79(3),710-739.
11. 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 S. Biemer et al. ed. Measurement Errors in Surveys. New York. Wiley.185–210.

12. Eifler, S., 2007. Evaluating the Validity of Self-Reported Deviant Behavior Using Vignette Analyses. *Quality & Quantity*, 41(2),303–318.
13. Gelman, A., 2014. Thinking of doing a list experiment? Here's a list of reasons why you should think again. *Statistical Modeling, Causal Inference, and Social Science Blog*. http://andrewgelman.com/2014/04/23/thinking-list-experiment-heres-list-reasons-think/
14. Glynn, A.N., 2013. What can we learn with statistical truth serum? *Public Opinion Quarterly*, 77(S1),159–172.
15. Gore, M.L., Ratsimbazafy, J. & Lute, M.L., 2013. Rethinking Corruption in Conservation Crime: Insights from Madagascar. Conservation Letters, 6(6), 430–438.
16. Heck, D., 2018. RRreg manual. https://cran.r-project.org/web/packages/RRreg/vignettes/RRreg.html
17. Hinsley, A. Nuno, A., Ridout, M., St. John, F.A.V. & Roberts, D. 2017. Estimating the Extent of CITES Noncompliance among Traders and End-Consumers; Lessons from the Global Orchid Trade. Conservation Letters, 10(5),602–609.doi:10.111/conl.12316
18. Ibbett, H., Lay, C., Phlai, P., Song, D., Hong, C., Mahood, S.P., & Milner-Gulland, E. J. 2017. Conserving a globally threatened species in a semi-natural, agrarian landscape. Oryx,1-11.
19. Imai, K., 2011. Multivariate Regression Analysis for the Item Count Technique. Journal of the American Statistical Association, 106(494),407–416.
20. Keane, A., Ramarolahy, A.A., Jones, J.P., & Milner-Gulland, E.J. 2011. Evidence for the effects of environmental engagement and education on knowledge of wildlife laws in Madagascar. Conservation Letters, 4(1),55-63.
21. Knapp, E.J., Peace, N. & Bechtel, L., 2017. Poachers and Poverty: Assessing Objective and

Subjective Measures of Poverty among Illegal Hunters Outside Ruaha National Park, Tanzania. *Conservation*, 15(1),24–32.

22. Kuklinski, J.H., Cobb, M.D. & Gilens, M., 1997. Racial attitudes in the 'New South'. *The Journal of Politics*, 59(2),323–349.
23. Lau, J. T., Yeung, N. C., Mui, L. W., Tsui, H. Y., & Gu, J. 2011. A simple new method to triangulate self-reported risk behavior data—the bean method. Sexually transmitted diseases, 38(9),788-792.
24. Nuno, A. & St. John, F.A.V., 2015. How to ask sensitive questions in conservation: A review of specialized questioning techniques. *Biological Conservation*, 189,5–15.
25. Nuno, A., Bunnefeld, N., Naiman., L.C. & Milner-Gulland, E.J. 2013. A Novel Approach to Assessing the Prevalence and Drivers of Illegal Bushmeat Hunting in the Serengeti. *Conservation Biology*, 27(6),1355–1365.doi:10.1111/cobi.12124
26. Nuno, A., Blumenthal, J.M., Austin, T.J., Bothwell, J., Ebanks-Petrie, G., Godley, B.J. and Broderick, A.C., 2018. Understanding implications of consumer behavior for wildlife farming and sustainable wildlife trade. Conservation Biology, 32(2),390-400.
27. Petroczi, A., Nepusz, T., Cross, P., Taft, H., Shah., S., Deshmukh, N…Naughton, D.P. 2011. New non-randomised model to assess the prevalence of discrimination behaviours: A pilot study on mephedrone. *Substance abuse treatment, prevention, and policy*, 6(20),1–18.doi: 10.1186/1747-597X-6-20
28. Raghavarao, D. & Federer, W.T., 1979. Block total response as an alternative to the randomized response method in surveys. *Journal of the Royal Statistical Society: Series B (Methodological)*,41(1),40–45.
29. Razafimanahaka, J.H., Jenkins, R.K., Andriafidison, D., Randrianandrianina, F., Rakotomboavonjy, V., Keane, A.M., & Jones, J.P. (2012). Novel approach for quantifying illegal bushmeat consumption reveals high consumption of protected species in Madagascar. Oryx,46(4),584-592.
30. Solomon, J., Jacobson, S.K., Wald, K.D., & Gavin, M. (2007). Estimating illegal resource use

at a Ugandan park with the randomized response technique. Human Dimensions of Wildlife,12(2),75-88.

31. St. John, F.A.V., Keane, A.M. & Milner-Gulland, E.J., 2013. Effective conservation depends upon understanding human behaviour. Key Topics in Conservation Biology 2,344–361.
32. St. John, F.A.V., Keane, A.M., Jones, J.P., & Milner-Gulland, E.J. 2014. Robust study design is as important on the social as it is on the ecological side of applied ecological research. Journal of Applied Ecology, 51(6),1479-1485.
33. St. John, F.A.V., Brockington, D., Bunnefeld, N., Duffy, R., Homewood, K., Jones, J. P., ... & Razafimanahaka, J. 2016. Research ethics: Assuring anonymity at the individual level may not be sufficient to protect research participants from harm. Biological Conservation, 196,208–209.
34. St. John, F.A.V., Linkie, M., Martyr, D.J., Milliyanawati, B., McKay, J.E., Mangunjaya, F.M...Struebig, M.J. 2018. Intention to kill: Tolerance and illegal persecution of Sumatran tigers and sympatric species. Conservation Letters, e12451.
35. Thomas, A.S., Gavin, M.C. & Milfont, T.L., 2015. Estimating non-compliance among recreational fishers: Insights into factors affecting the usefulness of the Randomized Response and Item Count Techniques. Biological Conservation,189,pp.24-32.
36. Tourangeau, R. & Yan, T., 2007. Sensitive Questions in Surveys. Psychological Bulletin, 133(5),859–88
37. Travers, H., Clements, T. & Milner-Gulland, E.J., 2016. Predicting responses to conservation interventions through scenarios: A Cambodian case study. Biological Conservation, 204,403–410.
38. Trappman, M., Krumpal, I., Kirchner, A. & Jann, B. 2013. Item Sum: A new technique for asking quantitative sensitive questions. Journal of Survey Statistics and Methodology. 2(1),58-77
39. Tsuchiya, T., Hirai, Y. & Ono, S., 2007. A study of the properties of the item count technique. Public Opinion Quarterly, 71(2),253–272.

40. Ulrich, R., Schröter, H., Striegel, H., & Simon, P. (2012). Asking sensitive questions: a statistical power analysis of randomized response models. Psychological Methods, 17(4),623–641.
41. Warner, S.L., 1965. Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. Journal of the American Statistical Association, 60(309),63–69.
42. Zigerell, L.J., 2011. You Wouldn't Like Me When I'm Angry: List Experiment Misreporting*. Social Science Quarterly, 92(2),552–562.