

Author names [Jouni Kuha](#) and [Jonathan Jackson](#)
The item count method for sensitive survey questions: modelling criminal behavior
Article (Accepted version)
(Refereed)

Original citation:

Kuha, Jouni and Jackson, Jonathan (2014) *The item count method for sensitive survey questions: modelling criminal behaviour*. [Journal of the Royal Statistical Society: Series C \(Applied Statistics\)](#), 63 (2). pp. 321-341. ISSN 0035-9254 DOI: [10.1111/rssc.12018](#)

© 2013 [Royal Statistical Society](#)

This version available at: <http://eprints.lse.ac.uk/48069/>

Available in LSE Research Online: July 2014

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (<http://eprints.lse.ac.uk>) of the LSE Research Online website.

This document is the author's final accepted version of the journal article. There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

The item count method for sensitive survey questions: Modelling criminal behaviour

Jouni Kuha and Jonathan Jackson

London School of Economics and Political Science, London, United Kingdom

Summary. The item count method is a way of asking sensitive survey questions which protects the anonymity of the respondents by randomization before the interview. It can be used to estimate the probability of sensitive behaviour and to model how it depends on explanatory variables. We analyse item count survey data on the illegal behaviour of buying stolen goods. The analysis of an item count question is best formulated as an instance of modelling incomplete categorical data. We propose an efficient implementation of the estimation which also provides explicit variance estimates for the parameters. We then suggest specifications for the model for the control items, which is an auxiliary but unavoidable part of the analysis of item count data. These considerations and the results of our analysis of criminal behaviour highlight the fact that careful design of the questions is crucial for the success of the item count method.

Keywords: Categorical data analysis; EM algorithm; List experiment; Missing information; Newton-Raphson algorithm; Randomized response

1. Introduction

Asking sensitive questions about behaviour and attitudes is one of the most difficult challenges in survey measurement. In this paper we consider a question on illegal behaviour, but other sensitive areas include sexual activity, use of illicit drugs, and embarrassing or socially undesirable opinions and prejudices. It is easily conceivable that many respondents may not give truthful answers to direct questions on such topics.

Measurement error in answers to sensitive questions may be reduced by some choices in the survey design, such as open-ended questions, asking about behaviour over long reference periods, tolerantly loaded introductions, and self-administration of the sensitive questions (see Tourangeau and Yan 2007 and Groves et al. 2009 for overviews). Another common approach is the *randomized response* method, in which respondents employ a randomizing device to add probabilistic misclassification to their responses and thus conceal their true answers from the interviewer. The original randomized response method was proposed by Warner (1965), and other variants have been developed since (see Chaudhuri and Mukerjee 1988, Lensvelt-Mulders et al. 2005 and Tourangeau and Yan 2007 for overviews).

Another way of protecting the respondents' anonymity is the *item count method* or *list experiment* (Miller (1984); Raghavarao and Federer (1979) proposed a closely related approach), which has become increasingly popular recently (see Blair and Imai (2012) for a list of some applications). Its basic idea can be introduced with the question shown in Table 1, which will be considered in our application. Each respondent is presented with some or all of a list of questions with possible answers of Yes and No. One of these is the *sensitive item* which is the focus of interest; in our case this is item 6, which asks if the respondent has bought stolen goods in the past 12 months. All the other questions are *control items*

which are not of direct interest and not meant to be sensitive. The survey respondents are randomly assigned to either the *control group* whose list includes only the control items, or the *treatment group* who receive both the control items and the sensitive item. In both groups a respondent is asked to report only their total number of Yes answers but not the replies to the individual items.

The intention of the item count method is that respondents in the treatment group should feel able to include a truthful answer to the sensitive item in their response because they would realise that it will be hidden from the interviewer when only the total count is reported. Compared to the classical randomized response method, this has the advantage of avoiding the potentially distracting act of randomization by the respondents themselves during the interview. Potential disadvantages of the item count method are that only the treatment group provides any information about the question of interest, and that the inclusion of the control items complicates the survey design and adds uncertainty to the estimation.

Quantities of interest can be estimated from randomized response data because the randomization mechanism is known. We may be interested in both estimates of the unconditional probability of the sensitive behaviour and regression-type questions about its associations with explanatory variables. For the item-count method, a moment-based (mean difference) estimator has most often been used for the unconditional probability, and straightforward extensions of it for regression modelling (Chaudhuri and Christofides 2007; Tsuchiya et al. 2007; Holbrook and Krosnick 2010; Glynn 2010; Coutts and Jann 2011). These approaches, however, may be inefficient, and the regression methods in particular are not ideally suited for modelling a count response.

The key to more efficient and coherent analysis of any randomized response items is to treat it as a problem of incomplete categorical data. This idea has been applied to the modelling of classical randomized response designs by, for example, Maddala (1983), Scheers and Dayton (1988), Chen (1989), and van den Hout and van der Heijden (2004). For item count data, it was first fully recognised by Imai (2011), whose work represents a major advance in the methodology of the item count technique (see also Blair and Imai (2012); Corstange (2009) also employs models for categorical data, for a related design where in the control group each control item is asked individually).

In this paper we consider the modelling of item counts as categorical data. We expand on the results of Imai (2011) in two main ways. First, we propose a faster implementation of the estimation and an explicit estimate of the variance matrix of the estimated parameters. Second, we propose a more flexible set of choices for the model for the control items, which we argue can have a major impact on conclusions about the model of interest for the sensitive item. These points are developed in Section 3. An application on modelling criminal behaviour is introduced in Section 2 and analysed in Section 5. In Section 4 we discuss the design of item count questions, and in Section 6 offer some conclusions.

2. Using survey data to model predictors of criminal behaviour

Our substantive application concerns criminal behaviour, in particular buying stolen goods. A prominent problem in criminology is understanding what shapes deviant and illegal behaviour. Why people do or do not commit crimes speaks to what motivates people and to how institutions can influence behaviour. In order to explain variation in such behaviour, we need a relatively precise estimate of it. Yet, criminal activity is a classic sensitive is-

Table 1. The item count question on buying stolen goods, as included in the Euro-Justis survey, and the numbers of different responses in the control and treatment groups.

“I am now going to read you a list of five [six] things that people may do or that may happen to them. Please listen to them and then tell me how many of them you have done or have happened to you in the last 12 months. Do not tell me which ones are and are not true for you. Just tell me how many you have done at least once.”

[Items included in both the control and treatment groups:]

1. Attended a religious service, except for a special occasion like a wedding or funeral.
2. Went to a sporting event.
3. Attended an opera.
4. Visited a country outside [your country]?
5. Had personal belongings such as money or a mobile phone stolen from you or from your house.

[Item included in the treatment group only:]

6. Bought something you thought might have been stolen.

	Count							
	0	1	2	3	4	5	6	Total
Control group	269	472	257	133	54	21	–	1206
Treatment group	279	446	281	124	53	20	9	1212

sue, which people tend to want to conceal in survey situations in order to create a good impression.

In the criminological literature, the different motivations and routes of influence can be organised into instrumental and normative motivations. Instrumental motivations are guided by rational choice where people make implicit or explicit calculations of the risks of action against the benefits; consonant modes of crime-control policy focus on deterrence and punishment. Yet, thus far the evidence is mixed. It suggests that most people are not driven by calculations of the risks and benefits of committing crime. Accordingly, criminologists have started to move to trying to understand when and for which crimes deterrence might be an influence.

Normative motivations speak to the idea that reasons of morality explain why most people, most of the time, do not commit crimes. Influenced by socialization, family, friends and so forth, people do not act in ways that they believe are wrong. Social pressure and disapproval plausibly reinforce one’s motivation to act according to guiding moral principles. The important thing then is that people think it is wrong to buy stolen goods, for example, and think it is morally correct to obey the law simply because it is the law (Tyler 2006a, 2006b; Jackson et al. 2012). Criminological research on these issues continues apace, no doubt with many important insights still to come.

An especially interesting area of on-going research draws instrumental and normative motivations into a single model of motivation. For example, Kroneberg et al. (2010) reason that only a certain proportion of individuals make cost-benefit calculations when considering whether to commit a particular act. People who believe that the act is morally wrong do not believe it is an option (see also Wikström et al. 2012). From an analysis of a survey of Germans, on the specific crime of shoplifting, their findings were consistent with the idea that “...respondents with strongly internalized norms disregard instrumental incentives to

shoplift” In other words, if people believe that it is morally wrong to shoplift, then risk-benefit calculations will not predict intentions to commit the crime; but if they think it is not morally wrong to shoplift, then these calculations start to come into play.

We examine similar questions using data from a survey carried out in Italy, Bulgaria and Lithuania in October-November 2010. The surveys were conducted separately by different organizations in the three countries, but coordinated by the principal investigators in the UK. The main purpose of the survey was to measure the legitimacy of the criminal justice systems in these three European countries, as part of a broader project into trust in justice known as Euro-Justis (Hough and Sato, 2011). One of the key outcomes of the Euro-Justis project was the inclusion of a module of questions in round 5 of the European Social Survey in 2010 (European Social Survey, 2011); however, this did not include the item count question considered here. We use the pooled sample of 2549 respondents (1,007 for Bulgaria, 521 for Italy and 1,021 for Lithuania). The surveys are not treated as probability samples from the national populations. The main aim of our analysis is regression modelling of illegal behaviour rather than estimation of population proportions.

We consider questions motivated by Kroneberg et al. (2010), using a subset of the concepts they considered. The specific crime we consider is buying stolen goods. An item count question on it was included in the Euro-Justis survey, worded as shown in Table 1 (and translated into the national languages). We consider two explanatory variables of primary substantive interest. The first is the assessment of the morality of the crime, measured by a survey question whose core part was worded as follows: “...please tell me how wrong it is to ... buy something you thought might be stolen”. The response options were “Not wrong at all” (coded in our analysis as 1), “A bit wrong” (2/3) “Wrong” (1/3) and “Seriously wrong” (0). The second variable we focus on is personal financial circumstances, measured by the question “Which of [these descriptions] comes closest to how you feel about your household’s income nowadays?”, with responses “Living comfortably on present income” (0), “Coping on present income” (1/3), “Finding it difficult on present income” (2/3) and “Finding it very difficult on present income” (1). We treat this as a measure of financial need but also as a rough proxy for perceived benefits of financial crime such as buying stolen goods — although acknowledging the obvious limitations of the latter treatment. “Don’t know” responses are coded as missing for both variables. In the survey, the question on morality was asked much earlier than the item count question, which came just before the question on financial need.

Our substantive hypotheses are that lower financial need and a view that buying stolen goods is morally wrong will be associated with lower probability of having committed the crime. A further statistical hypothesis is that there would be a positive interaction between the two explanatory variables. This corresponds to the substantive hypothesis, in the spirit of Kroneberg et al. (2010), that strongly internalized norms lead people to disregard instrumental incentives to commit the crime.

The methodological questions that we consider next are how the item count data should best be analysed to answer the substantive questions, and what aspects of the quality of the data and the survey design affect the chances of success of this analysis.

3. Modelling item count data

3.1. Estimation

Consider an item count survey question which includes J control items and one sensitive item. Suppose we have data on $n = n_0 + n_1$ respondents of whom n_0 have been randomly assigned into the control group where the question included only the control items, and n_1 to the treatment group where the sensitive item was also included. Let t_i , $i = 1, \dots, n$, be a treatment indicator such that $t_i = 1$ for respondents i in the treatment group, and $t_i = 0$ in the control group.

Let $S_i = Z_i + t_i Y_i$ denote a respondent's answer to the item count question, with possible values $0, 1, \dots, J + 1$, and s_i its observed value in the sample. Here Z_i is the total for the control items, and $Y_i = 1$ if the answer to the sensitive item is Yes and $Y_i = 0$ if it is No. For the control group $S_i = Z_i$, but for the treatment group $S_i = Z_i + Y_i$ and Z_i and Y_i are not observed separately (the value of Y_i for the control group is hypothetical and not used). Define $\mathbf{t} = (t_1, \dots, t_n)'$ and vectors \mathbf{s} and \mathbf{Y} similarly, and let $\mathbf{X} = [\mathbf{x}_1 \dots \mathbf{x}_n]'$ where \mathbf{x}_i is a vector of explanatory variables for respondent i , including a constant 1.

We assume that S_i for different respondents are independent given (\mathbf{x}_i, t_i) . The model for S_i is specified through two models, $p_y(y_i|\mathbf{x}_{yi}; \boldsymbol{\beta}) = P(Y_i = y_i|\mathbf{x}_{yi}; \boldsymbol{\beta})$ for the sensitive item and $p_z(z_i|y_i, \mathbf{x}_{zi}; \boldsymbol{\psi}) = P(Z_i = z_i|Y_i = y_i, \mathbf{x}_{zi}; \boldsymbol{\psi})$ for the total of the control items, for $y_i = 0, 1$ and $z_i = 0, 1, \dots, J$. Here \mathbf{x}_{yi} and \mathbf{x}_{zi} are two subsets of the variables in \mathbf{x}_i , which need not be identical, and $\boldsymbol{\beta}$ and $\boldsymbol{\psi}$ are distinct parameter vectors. The substantive interest is on the model for Y_i , for which we use the binary logistic model

$$\pi_{yi} = p_y(1|\mathbf{x}_{yi}; \boldsymbol{\beta}) = \frac{\exp(\mathbf{x}_{yi}'\boldsymbol{\beta})}{1 + \exp(\mathbf{x}_{yi}'\boldsymbol{\beta})}. \quad (1)$$

When $\mathbf{x}_i = 1$, the only unknown parameter in this is $\beta = \log[\pi_y/(1 - \pi_y)]$ where $\pi_y = P(Y_i = 1; \boldsymbol{\beta})$ is the unconditional probability of positive response to the sensitive item. Different specifications for $p_z(z_i|y_i, \mathbf{x}_{zi}; \boldsymbol{\psi})$ are discussed in Section 3.2. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\psi}')'$.

We make the following assumptions:

- (a) The Y_i included in S_i in the treatment group is a truthful answer to the sensitive question, so that $p_y(y_i|\mathbf{x}_{yi}; \boldsymbol{\beta})$ is substantively interesting. In contrast, Z_i does not need to represent the true total for the control items, as long as the assumptions below are satisfied.
- (b) $p_y(y_i|\mathbf{x}_{yi}; \boldsymbol{\beta}) = P(Y_i = y_i|\mathbf{x}_{yi}, t_i; \boldsymbol{\beta})$. This is satisfied by randomization, which makes the joint distribution of (Y_i, \mathbf{x}_{yi}) independent of t_i for any \mathbf{x}_{yi} .
- (c) $p_z(z_i|y_i, \mathbf{x}_{zi}; \boldsymbol{\psi}) = P(Z_i = z_i|Y_i = y_i, \mathbf{x}_{zi}, t_i; \boldsymbol{\psi})$, i.e. that conditional on (Y_i, \mathbf{x}_{zi}) the total reported for the control items is not affected by whether or not the sensitive item was included in the question. This assumption is not related to the randomization, so its plausibility must be considered separately.
- (d) If the n respondents exclude any nonrespondents who did not answer the item count question, the probability of nonresponse is independent of (Y_i, Z_i) given (\mathbf{x}_i, t_i) .

Under these assumptions, the model for the reported total S_i in the observed data is

$$p_s(s_i|\mathbf{x}_i, t_i; \boldsymbol{\theta}) = P(S_i = s_i|\mathbf{x}_i, t_i; \boldsymbol{\theta}) = \pi_{yi} p_z(s_i - t_i|1, \mathbf{x}_{zi}; \boldsymbol{\psi}) + (1 - \pi_{yi}) p_z(s_i|0, \mathbf{x}_{zi}; \boldsymbol{\psi}) \quad (2)$$

where we take $p_z(J + 1|0, \mathbf{x}_{zi}; \boldsymbol{\psi}) = p_z(-1|1, \mathbf{x}_{zi}; \boldsymbol{\psi}) = 0$; these correspond to the two impossible values $(0, J + 1)$ and $(1, 0)$ of (Y_i, s_i) in the treatment group.

When there are no explanatory variables, a simple moment-based estimator of π_y is

$$\tilde{\pi}_y = n_1^{-1} \sum_{i=1}^n t_i s_i - n_0^{-1} \sum_{i=1}^n (1 - t_i) s_i. \quad (3)$$

It follows from (2) that $\tilde{\pi}_y$ is an unbiased estimator of π_y . It is also the maximum likelihood estimator under certain conditions, as discussed in Section 3.2. However, it lacks the flexibility to accommodate less than saturated models for Y_i and Z_i . It also does not provide a generally convenient extension to models with explanatory variables. The most obvious generalisation of (3) in this direction is the linear model $E(S_i) = \mathbf{x}_i' \boldsymbol{\psi} + t_i (\mathbf{x}_{yi}' \boldsymbol{\beta})$ (Holbrook and Krosnick, 2010). This yields a consistent estimator of $\boldsymbol{\beta}$ for a linear probability model for Y_i under certain assumptions about the model for Z_i . However, a linear model for π_{yi} is generally unappealing, and the approach again fails to provide flexibility for modelling Z_i .

A more satisfactory framework for the analysis of item count data is to treat it as the categorical data problem that it clearly is. For convenience of description and computer implementation, we introduce first the device of a set of pseudo-data which consists of two stacked copies of the observed data set. The pseudo-data thus has $m = 2n$ observations where, for each $i = 1, \dots, n$, observations i and $i + n$ have the same values of the observed variables. We then add the pseudo-variables y_i^* and z_i^* with the values $(y_i^*, z_i^*) = (1, s_i - t_i)$ for $i = 1, \dots, n$ and $(y_i^*, z_i^*) = (0, s_i)$ for $i = n + 1, \dots, m$. We denote $p_{yi}^* = p_y(y_i^* | \mathbf{x}_{yi}; \boldsymbol{\beta})$ and $p_{zi}^* = p_z(z_i^* | y_i^*, \mathbf{x}_{zi}; \boldsymbol{\psi})$ for $i = 1, \dots, m$, where by definition p_{yi}^* is π_{yi} for $i = 1, \dots, n$ and $1 - \pi_{yi}$ for $i = n + 1, \dots, m$.

The observed-data log-likelihood is $\ell(\boldsymbol{\theta}; \mathbf{s}) = \sum_{i=1}^n \log p_s(s_i | \mathbf{x}_i, t_i; \boldsymbol{\theta})$ or, in terms of the pseudo-data, $\ell(\boldsymbol{\theta}; \mathbf{s}) = \sum_{i=1}^n \log[\pi_{yi} p_{zi}^* + (1 - \pi_{yi}) p_{z,i+n}^*]$. One convenient way to maximize it is with the EM algorithm (Dempster et al., 1977). This is based on viewing the values of Y_i as missing data. If they and thus also all $Z_i = s_i - t_i Y_i$ were observed, the *complete-data log-likelihood* would be

$$\begin{aligned} \ell(\boldsymbol{\theta}; \mathbf{Y}, \mathbf{s}) &= \sum_{i=1}^n \log p(Y_i, s_i | \mathbf{x}_i, t_i; \boldsymbol{\theta}) \\ &= \sum_{i=1}^n \log p_y(Y_i | \mathbf{x}_{yi}, t_i; \boldsymbol{\beta}) + \sum_{i=1}^n \log p_z(s_i - t_i Y_i | Y_i, \mathbf{x}_{zi}, t_i; \boldsymbol{\psi}) = \ell(\boldsymbol{\beta}; \mathbf{Y}) + \ell(\boldsymbol{\psi}; \mathbf{Y}, \mathbf{s}). \end{aligned} \quad (4)$$

To express this in terms of the pseudo-data, let $Y_i = Y_{i+n}$ there be equal to the true Y_i for each $i = 1, \dots, n$. Then

$$\ell(\boldsymbol{\theta}; \mathbf{Y}, \mathbf{s}) = \sum_{i=1}^m I(Y_i = y_i^*) \log p_{yi}^* + \sum_{i=1}^m I(Y_i = y_i^*) \log p_{zi}^* \quad (5)$$

where $I(Y_i = y_i^*) = 1$ if $Y_i = y_i^*$ and 0 otherwise. This means that if Y_i were observed, we would know which of the two paired rows of the pseudo-data corresponded to the true value of Y_i . As we do not know this, the EM algorithm gives weight to both possibilities. The algorithm proceeds by alternating between two steps:

E-step: Calculate the conditional expected value of the complete-data log likelihood (5) given the observed data and current estimate $\boldsymbol{\theta}_{(t)}$ of $\boldsymbol{\theta}$, as

$$\begin{aligned} Q(\boldsymbol{\theta} | \boldsymbol{\theta}_{(t)}) &= Q(\boldsymbol{\beta} | \boldsymbol{\theta}_{(t)}) + Q(\boldsymbol{\psi} | \boldsymbol{\theta}_{(t)}) = E[\ell(\boldsymbol{\theta}; \mathbf{Y}, \mathbf{s}) | \mathbf{s}, \mathbf{X}, \mathbf{t}; \boldsymbol{\theta}_{(t)}] \\ &= \sum_{i=1}^m w_{i(t)}^* \log p_{yi}^* + \sum_{i=1}^m w_{i(t)}^* \log p_{zi}^* \end{aligned} \quad (6)$$

where

$$\begin{aligned} w_{i(t)}^* &= P(Y_i = y_i^* | s_i, \mathbf{x}_i, t_i; \boldsymbol{\theta}_{(t)}) \\ &= \frac{p_{yi(t)}^* p_{zi(t)}^*}{p_{yi(t)}^* p_{zi(t)}^* + p_{yi'(t)}^* p_{zi'(t)}^*} = \frac{p_{yi(t)}^* p_{zi(t)}^*}{p_{yi(t)}^* p_{zi(t)}^* + (1 - p_{yi(t)}^*) p_{zi'(t)}^*}, \end{aligned} \quad (7)$$

$i' = i + n$ for $i = 1, \dots, n$ and $i' = i - n$ for $i = n + 1, \dots, m$. The subscripts (t) indicate that all the probabilities in $w_{i(t)}^*$ are calculated with the parameter values $\boldsymbol{\theta}_{(t)}$.

M-step: Maximise $Q(\boldsymbol{\theta} | \boldsymbol{\theta}_{(t)})$ to obtain updated estimates $\boldsymbol{\theta}_{(t+1)} = (\boldsymbol{\beta}'_{(t+1)}, \boldsymbol{\psi}'_{(t+1)})'$. The two terms $Q(\boldsymbol{\psi} | \boldsymbol{\theta}_{(t)})$ and $Q(\boldsymbol{\beta} | \boldsymbol{\theta}_{(t)})$ can be maximised separately. These are weighted log likelihoods for the complete-data models for Y_i and Z_i , with weights $w_{i(t)}^*$. The updated estimates can thus be obtained by fitting the models to the pseudo-data using standard software, as long as these allow the fractional frequency weights $w_{i(t)}^*$.

Maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is obtained by iterating the algorithm to convergence, starting with some initial values $\boldsymbol{\theta}_{(0)}$. This is the estimation method proposed by Imai (2011) and implemented in the R software (R Core Team, 2012) with the package `list` (Blair and Imai, 2010).

Two well-known disadvantages of the EM algorithm are that it can be slow to converge and that it does not automatically provide an estimate of the variance matrix of $\hat{\boldsymbol{\theta}}$. Here the speed difference to our Newton-Raphson (N-R) algorithm described below is enhanced by the fact that the EM implementation requires two iterative procedures at each M-step where the N-R estimation involves only a single noniterative step. In our application the difference is mostly in convenience than real practical significance: a typical model might take less than 1 second and 15 iterations with N-R and 12 seconds and 150 (M-step) iterations with EM. For the variance matrix, Imai (2011) used numerical differentiation of $\ell(\boldsymbol{\theta}; \mathbf{s})$, as implemented in the `optim` function in R, to approximate the observed-data information matrix. It is in general preferable to replace such numerical methods with explicit expressions where possible.

To remedy both of these shortcomings, we make use of an elegant but apparently relatively little-used result which is implicit in some earlier literature on the EM algorithm but which was first stated explicitly by Oakes (1999). This shows that the function $Q(\boldsymbol{\theta} | \boldsymbol{\theta}_{(t)})$ derived at the E-step of the EM algorithm can also be used to calculate both the observed-data score function, as

$$u(\boldsymbol{\theta}_{(t)}) = \left. \frac{\partial \ell(\boldsymbol{\theta}; \mathbf{s})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{(t)}} = \left. \frac{\partial Q(\boldsymbol{\theta} | \boldsymbol{\theta}_{(t)})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{(t)}}, \quad (8)$$

and the observed-data information matrix as

$$I(\boldsymbol{\theta}_{(t)}) = - \left. \frac{\partial^2 \ell(\boldsymbol{\theta}; \mathbf{s})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{(t)}} = - \left[\frac{\partial^2 Q(\boldsymbol{\theta} | \boldsymbol{\theta}_{(t)})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} + \frac{\partial^2 Q(\boldsymbol{\theta} | \boldsymbol{\theta}_{(t)})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'_{(t)}} \right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_{(t)}}, \quad (9)$$

both of which hold for any value of $\boldsymbol{\theta}_{(t)}$. In these expressions it is crucial that $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_{(t)}$ are treated as distinct quantities.

The results (8) and (9) allow us, first, to speed up to convergence of the estimation substantially by replacing the M-step of EM with a Newton-Raphson update step

$$\boldsymbol{\theta}_{(t+1)} = \boldsymbol{\theta}_{(t)} + I(\boldsymbol{\theta}_{(t)})^{-1} u(\boldsymbol{\theta}_{(t)}).$$

Second, the estimated variance matrix of the estimates $\hat{\boldsymbol{\theta}}$ is given by (9) as $\widehat{\text{var}}(\hat{\boldsymbol{\theta}}) = I(\hat{\boldsymbol{\theta}})^{-1}$. Because the N-R algorithm may also diverge, to achieve convergence it is important to use sensibly chosen starting values and/or shorten the update step for some iterations if needed.

For the model for item count data, $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\psi}')'$ and $Q(\boldsymbol{\theta}|\boldsymbol{\theta}_{(t)})$ is given by (6). Here p_{yi}^* and p_{zi}^* depend only on $\boldsymbol{\theta}$ and $w_{i(t)}^*$, defined by (7), only on $\boldsymbol{\theta}_{(t)}$. Denote $\mathbf{X}_y = [\mathbf{x}_{y1} \dots \mathbf{x}_{yn}]'$ and $\boldsymbol{\pi}_y = (\pi_{y1}, \dots, \pi_{yn})'$, evaluated at $\boldsymbol{\beta}$. Define $\boldsymbol{\pi}_{y(t)}^* = (\pi_{y1(t)}^*, \dots, \pi_{yn(t)}^*)'$ where $\pi_{yi(t)}^* = P(Y_i = 1 | s_i, \mathbf{x}_i, t_i; \boldsymbol{\theta}_{(t)})$; note that $\pi_{yi(t)}^*$ are equal to $w_{i(t)}^*$ for $i = 1, \dots, n$. Let \mathbf{D}_1 be the matrix with rows $\partial \log p_{zi}^* / \partial \boldsymbol{\psi}'$ for $i = 1, \dots, n$, and \mathbf{D}_0 the matrix with rows $\partial \log p_{zi}^* / \partial \boldsymbol{\psi}'$ for $i = n+1, \dots, m$, and define $\mathbf{D}_{1(t)}$ and $\mathbf{D}_{0(t)}$ analogously with rows $\partial \log p_{zi(t)}^* / \partial \boldsymbol{\psi}'$ where $p_{zi(t)}^*$ is like p_{zi}^* but evaluated with $\boldsymbol{\theta}_{(t)}$. Let $\text{diag}(a_i)$ in general denote a diagonal matrix with the elements of a vector (a_1, \dots, a_n) on the diagonal, and $\mathbf{1}$ a vector of ones of appropriate length. The score function (8) is then

$$\begin{aligned} \partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}_t) / \partial \boldsymbol{\beta} &= \mathbf{X}_y' (\boldsymbol{\pi}_{y(t)}^* - \boldsymbol{\pi}_y) \\ \partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}_t) / \partial \boldsymbol{\psi} &= (\mathbf{D}_1 - \mathbf{D}_0)' \boldsymbol{\pi}_{y(t)}^* + \mathbf{D}_0' \mathbf{1}. \end{aligned} \quad (10)$$

For the first term of (9),

$$\begin{aligned} \frac{\partial^2 Q(\boldsymbol{\theta}|\boldsymbol{\theta}_{(t)})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} &= -\mathbf{X}_y' \text{diag}[\pi_{yi}(1 - \pi_{yi})] \mathbf{X}_y \\ \frac{\partial^2 Q(\boldsymbol{\theta}|\boldsymbol{\theta}_{(t)})}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'} &= \sum_{i=1}^m w_{i(t)}^* \frac{\partial^2 \log p_{zi}^*}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'} \end{aligned} \quad (11)$$

and $\partial^2 Q(\boldsymbol{\theta}|\boldsymbol{\theta}_{(t)}) / \partial \boldsymbol{\beta} \partial \boldsymbol{\psi}' = [\partial^2 Q(\boldsymbol{\theta}|\boldsymbol{\theta}_{(t)}) / \partial \boldsymbol{\beta} \partial \boldsymbol{\psi}']'$ are zero matrices. The second term of (9) is

$$\frac{\partial^2 Q(\boldsymbol{\theta}|\boldsymbol{\theta}_{(t)})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'_{(t)}} = \begin{bmatrix} \mathbf{X}_y' \\ (\mathbf{D}_1 - \mathbf{D}_0)' \end{bmatrix} \text{diag}[\pi_{yi(t)}^* (1 - \pi_{yi(t)}^*)] \begin{bmatrix} \mathbf{X}_y & (\mathbf{D}_{1(t)} - \mathbf{D}_{0(t)}) \end{bmatrix} \quad (12)$$

which is symmetric when \mathbf{D}_1 and \mathbf{D}_0 are evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_{(t)}$ to become $\mathbf{D}_{1(t)}$ and $\mathbf{D}_{0(t)}$.

These expressions apply when the model for Y_i is the logistic model (1). The specific forms of (10)–(12) depend on the choice of the model for $p_z(z_i | y_i, \mathbf{x}_{zi}; \boldsymbol{\psi})$. It can be seen that only the first and second derivatives of the logarithms of these probabilities are needed to complete the calculations. Explicit expressions for the four models that we will consider are given in Appendix A. Finally, the observed-data log-likelihood at the maximum likelihood estimates can be calculated from the pseudo-data as $\ell(\hat{\boldsymbol{\theta}}; \mathbf{s}) = \sum_{i=1}^n \log(\pi_{yi} p_{zi}^* + (1 - \pi_{yi}) p_{z,i+n}^*)$, with the probabilities evaluated at $\hat{\boldsymbol{\theta}}$.

3.2. Specification of a model for the control items

The formulation of the problem in the previous section makes it clear that any analysis of item count data involves a model $p_z(z_i | y_i, \mathbf{x}_{zi}; \boldsymbol{\psi})$ for the totals Z_i of the control items, whether or not this is explicit in the formulas of estimators. This model is a distinctive element of the method which does not arise in classical forms of randomized response. It is a nuisance element which is of no substantive interest in itself. Nevertheless, it still needs to be specified appropriately, lest errors there distort estimates of the model of interest for Y_i . In this section we compare possible choices for the model for the control items.

There are two parts to the specification of the model for Z_i : the choice of the distribution itself, and how it may depend on the explanatory variables Y_i and \mathbf{x}_{zi} . For the distribution, Imai (2011) and the computer implementation by Blair and Imai (2010) considered the binomial and the beta-binomial. For the explanatory variables they used \mathbf{x}_{yi} or $(\mathbf{x}'_{yi}, Y_i \mathbf{x}'_{yi})'$, i.e. the same variables as in the model for Y_i , or these interacted fully with Y_i . Here we suggest some generalisations for both of these elements of the model. Details of the specific models we use are given in Appendix A.

Let Z_{ij} , $j = 1, \dots, J$, denote respondent i 's unobserved answer to control item j , with values 0 for No and 1 for Yes, so that $Z_i = \sum_j Z_{ij}$. Suppose that each Z_{ij} follows a Bernoulli distribution with probability p_{ij} , and that different items Z_{ij} may be dependent, with covariances $\text{cov}(Z_{ij}, Z_{ik})$. Then Z_i has mean $E(Z_i) = J\bar{p}_i$ and variance

$$\text{var}(Z_i) = J\bar{p}_i(1 - \bar{p}_i) - \sum_{j=1}^J (p_{ij} - \bar{p}_i)^2 + \sum_{j \neq k} \text{cov}(Z_{ij}, Z_{ik}). \quad (13)$$

where $\bar{p}_i = J^{-1} \sum_j p_{ij}$. The mean is equal to that of a binomial distribution with index J and probability \bar{p}_i . The first term of (13) is the variance of this binomial distribution, while the last two terms represent over- or underdispersion relative to this variance. The second term, which is due to heterogeneity of the probabilities p_{ij} , is always negative and thus contributes underdispersion. The third term can be positive or negative, depending on the pattern of dependencies among the Z_{ij} .

If there is neither heterogeneity nor dependence, the last two terms of (13) are both zero and the distribution of Z_i is binomial. If there is no dependence, the last term is zero and we get the Poissonian binomial distribution in the sense of Johnson et al. (1992, S. 3.12.2), which is always underdispersed relative to the binomial. If there is no heterogeneity, the second term is zero. If we then further assume that the covariances $\text{cov}(Z_{ij}, Z_{ik})$ are all equal, we get $\text{var}(Z_i) = J\bar{p}_i(1 - \bar{p}_i)[1 + (J - 1)\rho]$ where ρ is the common ‘‘intra-class correlation’’ between Z_{ij} for respondent i . The beta-binomial distribution has a variance of this form. Its standard motivation as a mixture distribution implies that ρ is nonnegative, but more generally the distribution is also well-defined for some negative values of ρ (Prentice, 1986).

It is, however, undesirable to consider only such special cases of the distribution of Z_i . First, even a general version of the beta-binomial, say, cannot accommodate a distribution that is strongly underdispersed relative to the binomial. Such a distribution for the control items would in fact be ideal from a design point of view (see discussion in Section 4), so we should prefer a distribution which can represent such items if we do manage to create them. Second, in the item count context it may not be enough to consider only the mean and variance of Z_i . Expression (2) of the probabilities for the observed data shows that these involve all the probabilities of individual values of Z_i . An adequate model for all of them is thus needed to correctly disentangle them from the model for Y_i .

We would suggest that by default the distribution of Z_i should be specified with maximum flexibility, as a multinomial distribution with index 1 and probabilities $p_z(z|y_i, \mathbf{x}_{zi}; \boldsymbol{\psi})$ for $z = 0, \dots, J$. In the examples below we compare the multinomial with the binomial and beta-binomial models. For dependence of the multinomial probabilities on explanatory variables we consider two possibilities, the multinomial logistic model which ignores the ordering of the values of Z_i , and the ordinal logistic (proportional odds) model which takes the ordering into account. The ordinal model is in principle appealing because it is flexible in the response distribution but relatively parsimonious in the parametrisation of the effects of the explanatory variables.

Consider now choices for the explanatory variables for the model for Z_i . We denote these by \mathbf{v}_i , which include at least \mathbf{x}_{zi} but possibly also Y_i and even products (interactions) between Y_i and some or all of \mathbf{x}_{zi} . Note that \mathbf{x}_{zi} does not need to include the same variables as \mathbf{x}_{yi} , and there may be some gain in efficiency if it does not. Throughout, models with nested choices of \mathbf{v}_i may be compared using likelihood ratio tests.

The most consequential aspect of the model for Z_i is the extent to which it depends on Y_i . In particular, the estimates will be most efficient when it does not, i.e. when Z_i and Y_i are conditionally independent given the explanatory variables. If this is the case, (2) shows that $p_s(s_i|\mathbf{x}_i, t_i = 0; \boldsymbol{\theta}) = p_z(s_i|\mathbf{x}_{zi}; \boldsymbol{\psi})$. The parameters $\boldsymbol{\psi}$ could then even be estimated directly by fitting the model for Z_i for S_i for the control group only, and the data in the treatment group will contribute information mostly about the model for Y_i . If the conditional independence does not hold, a smaller amount of information is available on both models, and both are mixed up in both groups.

Further insight into this loss of information is provided by the form of the information matrix $I(\boldsymbol{\theta})$ in (9). Its second term can be seen as the “missing information” due to the fact that Y_i are not observed. All of its terms depend on the quantities $\pi_{yi(t)}^*(1 - \pi_{yi(t)}^*)$, which are the predictive variances of the unknown Y_i given (s_i, \mathbf{x}_i, t_i) and $\boldsymbol{\theta}_{(t)}$. For $\boldsymbol{\psi}$, and through the cross-derivative terms of $I(\boldsymbol{\theta})$ also for $\boldsymbol{\beta}$, the missing information also involves contributions from $\mathbf{D}_1 - \mathbf{D}_0$, which are of the form $\partial \log(p_{zi}^*/p_{z,i+n}^*)/\partial \boldsymbol{\psi}$. These describe how different the gradients of $\log p_{zi}$ are at the two possible values 0 and 1 of Y_i . The magnitude of these differences, and thus the amount of missing information, depend on the specification of the model for Z_i . The key feature of the case where Z_i is conditionally independent of Y_i is that then $\log(p_{zi}^*/p_{z,i+n}^*) = 0$ for all observations in the control group, so they do not add anything to this component of the missing information.

A model of special interest is one with $\mathbf{x}_{yi} = 1$ and multinomially distributed Z_i with $\mathbf{x}_{zi} = (1, Y_i)'$. This has no other explanatory variables and has $2J + 1$ parameters to model the $(J + 1) + J = 2J + 1$ free probabilities $p_s(s|t)$ in the table of randomization group t by observed item count s (c.f. Table 1). The model is thus saturated, and the maximum likelihood estimators of $p_s(s|t)$ are the observed sample proportions $\hat{p}_s(s|t)$. Solving the expressions of these probabilities in (2) for π_y and $p_z(z|y) = P(Z = j|Y = y)$, we get as estimate of π_y the mean difference $\tilde{\pi}_y$ given by (3), and for $p_z(z|y)$

$$\tilde{p}_z(j|0) = [\hat{p}_s(j|1) - \tilde{\pi}_y \tilde{p}_z(j-1|1)]/(1 - \tilde{\pi}_y) \quad \text{and} \quad (14)$$

$$\tilde{p}_z(j|1) = [\hat{p}_s(j|0) - (1 - \tilde{\pi}_y) \tilde{p}_z(j|0)]/\tilde{\pi}_y \quad (15)$$

for $j = 0, \dots, J$, starting with $\tilde{p}_z(-1|1) = 0$ (see also Glynn 2010, who gives corresponding expressions for the probabilities of Y given Z). These are equal to the maximum likelihood estimates of π_y and $p_z(z|y)$ obtained as in Section 3.1, if $\tilde{p}_z(j|y)$ are all non-negative (the case where they are not is discussed in the next section). This equivalence demonstrates that any gain in efficiency obtained by the maximum likelihood estimators over the mean difference (3) is not due to the formulation of the problem as a model for categorical data, but is only realised if we are able to assume that Z_i and Y_i are independent.

We conducted a limited simulation study to learn more about the possible impact of the specification of the model for Z_i . Four situations were considered, with 1000 data sets simulated in each. There were no explanatory variables. In each case the sample size was 2400 with 1200 observations in each of the treatment and control groups, Y_i was drawn from a Bernoulli distribution with probability $\pi_y = 0.1$, and there were 5 control items. The four cases differ in the model for Z_i , and represent different combinations of its distribution and

Table 2. Results of a simulation study of estimators of the probability of a sensitive item obtained from item count data. The table shows means and root mean squared errors of estimators over 1000 simulated data sets (multiplied by 1000). See the text for the details of the simulation specifications.

True model for Z :	Estimated model for Z :						
	Binomial		Beta-Binomial		Multinomial logistic		Ordinal logistic
	Indep.*	Dep.	Indep.	Dep.	Indep.	Dep.	Dep.
Mean (true value is 100):							
Binomial, independent	102	101	102	101	102	102	97
Binomial, dependent	131	103	118	104	113	100	98
Multinomial, independent	289	437	112	89	100	110	92
Multinomial, dependent	308	437	172	79	154	112	89
Root Mean Squared Error:							
Binomial, independent	38	40	49	40	40	40	44
Binomial, dependent	45	37	40	39	40	43	44
Multinomial, independent	191	340	33	55	28	53	46
Multinomial, dependent	209	338	79	68	61	52	50

* "Indep.": Z and Y are modelled as independent; "Dep.": Z and Y are modelled as dependent.

whether this depends on Y_i . In cases 1 and 2, Z_i follow a binomial distribution, in case 1 with probability 0.25 for all respondents, and in case 2 with probability 0.25 when $Y_i = 0$ and 0.355 when $Y_i = 1$. In cases 3 and 4, Z_i are drawn from a multinomial distribution, in case 3 with probabilities (0.2, 0.2, 0.2, 0.2, 0.1, 0.1) for $Z = 0, \dots, 5$ for every respondent; in case 4 these probabilities apply when $Y_i = 0$ and the probabilities (0.1, 0.15, 0.15, 0.2, 0.2, 0.2) when $Y_i = 1$. The multinomial probabilities are chosen so that they are not well-represented by a binomial distribution.

Table 2 shows the results of the simulation for estimates of π_y . We consider maximum likelihood estimators under the four models listed in Appendix A, each both assuming Z_i and Y_i to be independent and allowing them to be dependent, for a total of seven estimators (the multinomial logistic and ordinal logistic model are equivalent under independence). There were a handful of simulations where some estimators converged to a very small value (less than -10 on the logit scale). These may represent false convergence of the algorithm, so simulations where this happened for any estimator are excluded; there were no more than 19 instances of this in any of the four cases.

The simulation means in the upper part of Table 2 show that all of the estimators are approximately unbiased when the model for Z_i is correctly specified or overparametrised. When this model is incorrectly specified, however, the estimator of π_y is biased, in many cases dramatically so. This happens whether the misspecification concerns the distribution of Z_i or the association between Z_i and Y_i . It is worth noting that at least in these cases the one additional parameter of the beta-binomial model reduces its bias substantially relative to the binomial model when the true model is multinomial.

The main message of the lower part of Table 2 concerns the loss of efficiency when we have to assume dependence between Z_i and Y_i . This can be seen by comparing the results under independence models for correctly specified distributions. It can be seen that this loss of efficiency is small when the true distribution is binomial but much larger when it is multinomial. In the latter case the use of an ordinal model for the dependence improves the efficiency slightly but still leaves it far lower than that of the independence model.

4. Considerations on the design of item count questions

Careful design of the survey items is clearly a prerequisite for the success of the item count methodology. Here we consider briefly some elements of design (for more extensive discussions, see for example Glynn (2010), Blair and Imai (2012), and references cited therein). We focus on the technical questions of articulating the assumptions that the items should satisfy, how these affect the model specification, and how they may be checked in the analysis. This is informative on, but does not directly answer, the central practical challenge of design, which is how we should choose the items in order to have a good chance of satisfying the assumptions. This question is not amenable to simple technical analysis or solutions, because the success of the exercise is ultimately dependent on how the survey respondents react to the questions. For understanding and predicting these reactions, the designer of an item count question will benefit from a good knowledge of the general theories and empirical evidence on the psychology of survey response (Tourangeau et al., 2000).

The item count technique or any other randomized response method is also likely to involve psychological peculiarities of its own. These relate to what might be termed the “weirdness factor” of the method, which is created when the interviewer appears to the respondent to deviate from the implicit social contract of what a survey interview should involve. With a classical randomized response question this happens when the respondent is suddenly asked to do something like spin a dial to choose at random which question they should answer. In an item count question, the weirdness arises from being presented with a list of apparently disparate items with no indication of why the total of them might be of interest to the interviewer. The strangeness of this may be lesser than that of the dial-spinning but it can still be nonnegligible. It can thus not be taken for granted that the respondents will react to an item count list exactly as intended, even if all the individual items are ostensibly simple and easy to answer.

Validity of item count measurement requires that the assumptions (a)–(d) stated in the beginning of Section 3.1 are satisfied. Of them, (b) is satisfied by the randomization unless it is undermined by failure of assumption (d), i.e. differential nonresponse. The other assumptions cannot be guaranteed through design.

Assumption (a) of no lying is the motivation of the item count technique in the first place, in that it is designed to reduce reasons for lying by guaranteeing anonymity. This protection will fail completely in one situation, the “ceiling effect” of a respondent in the treatment group whose truthful answer to all of the items would be affirmative, in which case a truthful total of $J + 1$ would logically reveal the answer to the sensitive item. Direct evidence of the prevalence of this problem is given by the proportion of counts of J in the control group. In design, the aim should then be to select control items for which few respondents would give only affirmative answers. One way to achieve this is to use items which are individually rare. Another one, which also reduces the chances of the floor effect discussed below, is to choose a control set where some pairs of items are negatively correlated (Glynn, 2010).

Often discussed alongside the ceiling effect is the “floor effect” of a respondent in the treatment group whose truthful answers would be negative to all the control items but affirmative to the sensitive item. The argument is that such a person might judge that a truthful count of 1 would be known to correspond to the sensitive item. This, however, follows logically only if the interviewer can reasonably conclude from their observation of that respondent that his or her answer to all of the control items must be No, a situation which should not be allowed to arise with a sensible set of control items. In other cases,

concern with the floor effect involves a less compelling argument which requires reference to a population of other respondents, i.e. a judgement that the control items are such that most people's answer to all of them is likely to be negative. This does not necessarily follow even if all of the items are individually rare.

A violation of the nonresponse assumption (d) can also lead to violation of assumption (c). Apart from that, (c) is essentially a requirement of compliance, that the respondents actually respond to the question as stated and report a sum of the items rather than somehow react to the list as a whole (in which case it could matter whether the sensitive item was on it or not). Assuring this at the design stage is a considerable challenge. In the analysis, there is one obvious if partial way of examining the validity of the assumption. This is to use the logical conditions that $P(S \leq s|T = 0) \geq P(S \leq s|T = 1)$ for all $s = 0, \dots, J - 1$ and $P(S \leq s - 1|T = 0) \leq P(S \leq s|T = 1)$ for all $s = 1, \dots, J$ (Blair and Imai, 2012). If these do not hold for all sample proportions, some of the moment estimates (14)-(15) of the probabilities of Z will also be negative. This may occur because of sampling variation even when assumption (c) is satisfied, so a significance test for the conditions is needed; such a test is proposed by Blair and Imai (2012). A test result that does not detect a significant violation is not sufficient evidence that the assumption is satisfied, but a significant test result does provide strong evidence that it is not. Furthermore, apart from sensitivity analysis of the possible biases there is nothing that can really be done in the analysis to adjust for such a violation. The conclusion from this part of the analysis may thus be the disheartening one that an item count question is irretrievably flawed.

Another element of validity is that of the model specification, especially of the model for the control items as we argued in Section 3.2. Even when correct, this model also affects the *efficiency* of the analysis, i.e. the extent of missing information that reduces the precision of the estimates of the parameters of interest. This depends on the complexity of the model for the control items, most of all on whether or not they are conditionally independent of the sensitive item. This conditional independence should be one aim of the design of an item count question. It should not be impossible to achieve when the control items are unrelated in content to the sensitive item, but it cannot be guaranteed in advance.

These considerations suggest that the ideal set of control items would be one for which every respondent would report the same count z with $0 < z < J$, achieved in such a way that the z would not be the *same* items for everyone (this to avoid a version of the floor effect). Such items would both satisfy all of the assumptions for validity and maximize the efficiency of the estimates. They are of course unachievable in practice but worth keeping in mind as a general aim.

How then do the Euro-Justis item count questions measure up against these criteria? When the questions were designed, there was a conscious attempt to avoid very common items and some aim to include ones which would be weakly or even negatively correlated with each other. To try to reduce the weirdness factor, the list consists of relatively general and not strikingly peculiar inquiries. Furthermore, item 5, which asks if the respondent has been a victim of crime, was included to try make the item count list seem a little less out of place in a survey that was otherwise mostly about crime and criminal justice. This, however, may have been disadvantageous in a different way, by introducing an association between the sensitive item and the total of the control items. It is clear from the analysis below that there is indeed such an association, and it may well be due to the fact that victimisation and criminal behaviour tend to be correlated.

The survey organisations in the three countries of the Euro-Justis survey each produced a field report where the interviewers summarised their own experiences and common reac-

tions by the respondents. It is encouraging to note that none of the three reports mentioned any concerns about the item count question. This contrasts with comments, which came consistently from all the countries, that many respondents had reacted negatively to direct questions elsewhere in the survey on other sensitive topics such as family income and attitudes toward crime.

Only 5.6% of the respondents in the control group and 4.7% in the treatment group refused to answer the item count question, so nonresponse is unlikely to be a major source of bias. The potential for ceiling effects is also minimal, as only 21 of 1206 respondents in the control group gave the maximum value of 5. Less reassuring is the finding that two of the cumulative proportions in the observed sample are inconsistent. These are the ones for counts of 0 and 2, where the cumulative probability is smaller in the control than in the treatment group (0.223 vs. 0.230 and 0.828 vs. 0.830). These differences are not statistically significantly negative, so they can be due to sampling variation. Nevertheless, they give some reason to worry that assumption (c) of consistency of responses may be violated.

5. Item count estimates of criminal behaviour

Table 3 shows estimates for models without explanatory variables for the item count question in the Euro-Justis survey. The quantity of main interest is the probability π_y of the sensitive item Y of having bought stolen goods in the past year. The table also includes estimates of the probabilities of different counts Z for the five control items. Here the focus is on how the estimates of π_y are affected by different choices for the model for the control items. We consider the moment-based estimators given by (3), (14) and (15), and maximum likelihood estimators with each of the four models for Z listed in Appendix A, each of the latter both with and without the assumption that Z and Y are independent. For assessment and comparison of model fits, the table includes the AIC statistic $AIC = -2\ell(\hat{\theta}; \mathbf{s}) + 2q$ where q is the number of estimated parameters in a model, and p -value for χ^2 test of goodness of fit which compares the fitted counts for reported totals S from each model to the observed counts that are shown in Table 1.

As discussed in Section 4, two of the moment-based estimates of the probabilities of Z are negative; these probabilities have boundary estimates of 0 in the saturated multinomial logistic model where Y and Z are dependent. For each model for Z , the hypothesis of independence between Z and Y is clearly rejected, and there is a substantial difference between the estimated probabilities of Z conditional on the two values of Y . The probabilities given $Y = 0$ are generally similar to the ones obtained when Z and Y are assumed independent, while the probabilities given $Y = 1$ are much less stable.

For these data, the estimated proportion of people who have bought stolen goods varies widely depending on the model for the control items. It is 12–14% when the clearly inappropriate assumption of independence between Z and Y is made, but 2–9% without it. Different dependence models also produce rather different results. The binomial and beta-binomial give the higher estimates of 6–9%. The goodness of fit of these two models is inadequate according to the χ^2 test even when Z and Y are dependent. Only the multinomial models where Z depends on Y yield a good fit, both with a multinomial logistic and an ordinal logistic model for the dependence. Estimates of π_y are then 4.1% with the multinomial and 1.7% with the ordinal model. In this example the estimate of the probability of substantive interest is thus very sensitive to assumptions about the model for the control items, the nuisance element of item count analysis.

Table 3. Probabilities of buying stolen goods ($\hat{\pi}_y$) and of different counts for the five control items, estimated from the item count question in the Euro-Justis survey. The table shows estimates under different model assumptions for the control items.

Estimator (model for Z)	Model for the control counts Z							$\hat{\pi}_y$	$p(\chi^2)^*$	AIC
	Conditional on Y ?	0	1	2	3	4	5			
Moment-based	$Y = 0$.24	.39	.22	.11	.04	.01	.026	—	—
	$Y = 1$	-.27	.62	-.10	.21	.25	.29			
Multinomial logistic	$Y = 0$.24	.39	.22	.11	.04	.01	.041	.91 [†]	7280
	$Y = 1$.00	.50	.00	.16	.17	.18			
	Independent	.24	.39	.21	.10	.04	.02	.123	< .001	7291
Ordinal logistic	$Y = 0$.23	.39	.23	.11	.04	.01	.017	.67	7278
	$Y = 1$.00	.02	.05	.14	.36	.43			
	Independent	.24	.39	.21	.10	.04	.02	.123	< .001	7291
Beta- binomial	$Y = 0$.25	.38	.25	.10	.02	.00	.064	.005	7290
	$Y = 1$.00	.04	.15	.30	.34	.16			
	Independent	.27	.33	.24	.12	.04	.01	.144	< .001	7353
Binomial logistic	$Y = 0$.25	.40	.26	.08	.01	.00	.087	.003	7291
	$Y = 1$.01	.05	.18	.34	.31	.16			
	Independent	.21	.38	.28	.11	.02	.00	.149	< .001	7509

* p -value for a χ^2 test of goodness of fit compared to the observed counts shown in Table 1.

† With 2 degrees of freedom, allowing post hoc for the two estimated probabilities of 0.

Table 4. Regression models for the item count question in the Euro-Justis survey. The table shows estimated coefficients and (in parentheses) their standard errors for binary logistic models for having bought stolen goods in the past 12 months, and for ordinal logistic models for the total count of the control items.

	Model 1		Model 2		Model 3		Model 4	
Model for sensitive item Y (buying stolen goods):								
Constant	-2.97		-3.11		-3.82		-4.72	
Age	-0.01	(0.01)	-0.01	(0.01)				
Morality	2.23	(0.95)	1.72	(1.48)	0.71	(0.64)	2.87	(1.90)
Need	1.05	(1.04)	1.60	(1.21)	2.22	(0.86)	3.41	(1.34)
Morality*Need							-2.58	(2.16)
Model for the total Z of the control items:								
Age	-0.02	(0.00)	-0.02	(0.00)	-0.02	(0.00)	-0.02	(0.00)
Morality	-0.17	(0.16)	-0.28	(0.32)				
Need	-1.38	(0.16)	-1.60	(0.28)	-1.71	(0.21)	-1.72	(0.20)
Y			0.99	(0.40)	1.11	(0.42)	0.99	(0.38)
Log-likelihood	-3226.5		-3222.5		-3223.8		-3223.1	

The constant terms for every model for Z are approximately (-3.1, -1.2, 0.1, 1.2, 2.4).

In Table 4 we turn to regression modelling of the item count question. The first analyses have given us some reason to approach this with caution, as the apparent association between the sensitive and control items will reduce the available information, and as there is some indication of noncompliance by respondents. However, it is still of interest to see what the item count question can reveal about associations between explanatory variables and self-reported criminal behaviour.

We consider two substantively interesting explanatory variables, the respondents' judgement of the moral acceptability of buying stolen goods and their self-reported economic circumstances. The motivation and definitions of these variables were given in Section 2. The respondent's age in years is also included as a control variable. For the control items we use an ordinal logistic model. This provides flexibility for the choice of the distribution by treating it as multinomial, but is more parsimonious than a multinomial logistic model in how the effects of the explanatory variables are specified.

Estimates for four models are shown in Table 4. Models 1 and 2 include all three explanatory variables in the models for both Y and Z . They differ in that in Model 1 the outcomes Y and Z are conditionally independent given the predictors, while in Model 2 Z depends also on the main effect of Y as an additional explanatory variable (the model with interactions between Y and the other predictors was not significantly different from this). The difference between these models is statistically significant, so the previous conclusion that Y is significantly associated with Z holds even after controlling for the three explanatory variables. Comparison of the estimated coefficients and their standard errors between these two models shows clearly that conclusions about the model of interest may be strongly affected by whether or not an association between Z and Y is included, and that uncertainties are substantially increased if it needs to be included.

In Model 3 we remove two explanatory variables from Model 2, age from the model for Y and moral judgement from the model for Z . Neither is significant in Model 2, and for the morality variable there is also no substantive motivation for considering it as a predictor for the control items. In this model, the associations involving Z are strongly significant and substantially sensible. They indicate that older people and people who are struggling on their present income tend to have engaged in fewer of the activities on the control list. The effect of Y is that people who have bought stolen goods tend to report higher totals for the control items. As discussed in Section 4, a possible substantive explanation of this involves the control item on having been a victim of crime.

The model of interest in Model 3 includes moral judgement and financial need as explanatory variables. These are taken to represent aspects of normative and instrumental motivations of criminal behaviour respectively. Their estimated effects are in expected directions: respondents who have a higher need are more likely to have bought stolen goods, as are people who do not regard such action as morally wrong. The coefficient of moral judgement is not significant, but that of financial need is. For at least one explanatory variable the item count has thus provided enough information for us to be able detect a positive association between it and the sensitive item, separate from its (negative) association with the control items.

Finally, in Model 4 we examine the last substantive hypothesis discussed in Section 2, that of an interaction between morality and need. The point estimate of this interaction is negative. This would be an intriguing conclusion in that it would suggest that moral judgement makes a difference only when need is low, and need only makes a difference when an act is judged immoral in general — which would be the exact opposite of the hypothesis proposed by Kroneberg et al. (2010). However, the interaction is clearly not

significant so no firm conclusions should be drawn. It is apparent that estimating such an interaction represents a bridge too far in what the information from these item count data can reliably support.

6. Conclusions

The item count method is a valuable and increasingly commonly used addition to the methodology of asking sensitive survey questions. It has some definite advantages over both direct questioning and other randomized response methods. Statistical analysis of item count data can be implemented elegantly and efficiently with methods for categorical data analysis for incomplete data. We illustrated this in our application, where substantively plausible models for illegal behaviour were obtained.

The method also has its disadvantages and peculiar methodological challenges. Most of these stem from the distinctive feature of an item count question, which is the list of the control items. Even though these items are of no direct substantive interest themselves, careful attention must be paid to them so as not to compromise information about the sensitive item of interest. We have argued that at the analysis stage sufficiently flexible model specification for the total of the control items is crucial, in particular that it should usually be treated as multinomially distributed.

Most of the effort and ingenuity in the design of an item count question should also be devoted to the control items. For validity, responses to them should not be affected by the presence of the sensitive item on the list or to give respondents reasons to lie about it, and for efficiency the control items should ideally be independent of the sensitive item. At the design stage it is not easy to be confident that these conditions will be satisfied. At the analysis stage, failures of them cannot always be detected and even when they can, are typically not correctable.

All of this makes for a substantial challenge for designers of surveys on sensitive topics, and one which will no doubt generate much future research. One practical recommendation that it suggests is that we should aim to build up a body of knowledge about specific item count questions, so that control items which have been found to work well in the past could be used again, even in item count surveys of different sensitive topics.

Acknowledgements: The Euro-Justis project (full title: *Scientific Indicators of Confidence in Justice: Tools for Policy Assessment*) was funded by the 7th Framework Programme of the European Commission under grant agreement 217311. We are grateful to Daniel Oberski for introducing the item count method to us, and to Ben Lauderdale for helpful comments.

A. Details of models for the control items

The model expressions in Section 3 involve probabilities of the form $p_{zi} = P(Z_i = z_i | Y_i = y_i, \mathbf{x}_{zi}; \psi)$ for observations i , for different values of z_i and y_i . We obtain them by specifying a model for $\pi_{zij} = P(Z_i = j | Y_i = y_i, \mathbf{x}_{zi}; \psi)$, for all $j = 0, 1, \dots, J$. Defining δ_{ij} for $j = 0, \dots, J$ such that $\delta_{ij} = 1$ if $z_i = j$ and $\delta_{ij} = 0$, we have $p_{zi} = \sum_{j=0}^J \delta_{ij} \pi_{zij}$. We denote by \mathbf{v}_i the vector of functions of y_i and \mathbf{x}_{zi} , possibly including interactions between them, which enters into the linear predictor of the model. Let \mathbf{V}_1 and \mathbf{V}_0 denote the matrices

whose rows are \mathbf{v}_i for the first and last n observations of the pseudo-data respectively, where $y_i = 1$ and $y_i = 0$ respectively, and let $\mathbf{V} = [\mathbf{V}'_1 \mathbf{V}'_0]'$.

We also need to define the quantities that depend on the model for Z_i and that appear in the estimation procedure described in Section 3.1. These include the matrix in (11) and \mathbf{D}_1 , \mathbf{D}_0 , $\mathbf{D}_{1(t)}$ and $\mathbf{D}_{0(t)}$ in (10) and (12). For the latter, we give the formulas for \mathbf{D}_1 , where $z_i = z_i^* = s_i - t_i$ and $y_i = y_i^* = 1$ for the first n rows of the pseudo-data, \mathbf{v}_i are the rows of \mathbf{V}_1 and all the probabilities are calculated at parameter value $\boldsymbol{\psi}$. The forms of the other three matrices are similar, except that for \mathbf{D}_0 and $\mathbf{D}_{0(t)}$ the data are $z_i = z_i^* = s_i$, $y_i = y_i^* = 0$ and \mathbf{V}_0 for the last n rows of the pseudo-data, and that for $\mathbf{D}_{1(t)}$ and $\mathbf{D}_{0(t)}$ the parameters are set at $\boldsymbol{\psi}_{(t)}$.

Below we define these quantities for four different models.

A.1. Binomial logistic model

Here Z_i follows a binomial distribution with index J and probability $\gamma_i = [1 + \exp(-\mathbf{v}'_i \boldsymbol{\psi})]^{-1}$. Then $\pi_{zij} = \binom{J}{j} \gamma_i^j (1 - \gamma_i)^{(J-j)}$. Let γ_i^* be γ_i for the observations $i = 1, \dots, m$ of the pseudo-data, evaluated at $\boldsymbol{\psi}$. Then $\mathbf{D}_1 = \text{diag}[z_i^* - J\gamma_i^*] \mathbf{V}_1$ for $i = 1, \dots, n$, and (11) is $\partial^2 Q(\boldsymbol{\theta} | \boldsymbol{\theta}_{(t)}) / \partial \boldsymbol{\psi} \partial \boldsymbol{\psi}' = -\mathbf{V}' \text{diag}[J w_{i(t)}^* \gamma_i^* (1 - \gamma_i^*)] \mathbf{V}$ for $i = 1, \dots, m$.

A.2. Beta-binomial logistic model

We specify the beta-binomial model as in Prentice (1986). Let $\gamma_i = [1 + \exp(-\mathbf{v}'_i \boldsymbol{\psi}_\gamma)]^{-1}$ as for the binomial logistic model, and let ρ denote the intraclass correlation. For simplicity we consider here only models where ρ is a constant, but this can easily be generalised. Then $\boldsymbol{\psi} = (\boldsymbol{\psi}'_\gamma, \rho)'$ and

$$\pi_{zij} = \binom{J}{j} \frac{\prod_{l=0}^{j-1} (\gamma_i + \tau l) \prod_{l=0}^{J-j-1} (1 - \gamma_i + \tau l)}{\prod_{l=0}^{J-1} (1 + \tau l)}$$

where $\tau = \rho(1 - \rho)^{-1}$. Here any product with upper limit of -1 is taken to be 1, and any such sum is taken to be 0. Let γ_i^* be γ_i for the observations $i = 1, \dots, m$ of the pseudo-data, evaluated at $\boldsymbol{\psi}$. Then $\mathbf{D}_1 = [\text{diag}(a_i) \mathbf{V}_1 \text{diag}(b_i) (1 - \rho)^{-2} \mathbf{1}]$ where $a_i = \left[\sum_{l=0}^{z_i^*-1} (\gamma_i^* + \tau l)^{-1} - \sum_{l=0}^{J-z_i^*-1} (1 - \gamma_i^* + \tau l)^{-1} \right] \gamma_i^* (1 - \gamma_i^*)$ and $b_i = \sum_{l=0}^{z_i^*-1} l(\gamma_i^* + \tau l)^{-1} + \sum_{l=0}^{J-z_i^*-1} l(1 - \gamma_i^* + \tau l)^{-1} - \sum_{l=0}^{J-1} l(1 + \tau l)^{-1}$ for $i = 1, \dots, n$. The elements of $\partial^2 Q(\boldsymbol{\theta} | \boldsymbol{\theta}_{(t)}) / \partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'$ are $\partial^2 Q(\boldsymbol{\theta} | \boldsymbol{\theta}_{(t)}) / \partial \boldsymbol{\psi}_\gamma \partial \boldsymbol{\psi}_\gamma' = \mathbf{V}' \text{diag}(w_{i(t)}^* c_i) \mathbf{V}$, $\partial^2 Q(\boldsymbol{\theta} | \boldsymbol{\theta}_{(t)}) / \partial \boldsymbol{\psi}_\gamma \partial \rho = \mathbf{V}' \text{diag}(w_{i(t)}^* (1 - \rho)^{-2} d_i) \mathbf{1}$ and $\partial^2 Q(\boldsymbol{\theta} | \boldsymbol{\theta}_{(t)}) / \partial \rho^2 = \mathbf{1}' \text{diag}(w_{i(t)}^* (1 - \rho)^{-3} e_i) \mathbf{1}$ for $i = 1, \dots, m$, where

$$\begin{aligned} c_i &= - \left[\sum_{l=0}^{z_i^*-1} (\gamma_i^* + \tau l)^{-2} + \sum_{l=0}^{J-z_i^*-1} (1 - \gamma_i^* + \tau l)^{-2} \right] [\gamma_i^* (1 - \gamma_i^*)]^2 + a_i (1 - 2\gamma_i^*), \\ d_i &= \left[- \sum_{l=0}^{z_i^*-1} l(\gamma_i^* + \tau l)^{-2} + \sum_{l=0}^{J-z_i^*-1} l(1 - \gamma_i^* + \tau l)^{-2} \right] \gamma_i^* (1 - \gamma_i^*) \quad \text{and} \\ e_i &= \left[- \sum_{l=0}^{z_i^*-1} l^2 (\gamma_i^* + \tau l)^{-2} - \sum_{l=0}^{J-z_i^*-1} l^2 (1 - \gamma_i^* + \tau l)^{-2} + \sum_{l=0}^{J-1} l^2 (1 + \tau l)^{-2} \right] (1 - \rho)^{-1} + 2b_i. \end{aligned}$$

A.3. Multinomial logistic model

Here Z_i follows a multinomial distribution with index 1 and probabilities $\pi_{zij} = \exp(\mathbf{v}'_i \boldsymbol{\psi}_j) [1 + \sum_{l=1}^J \exp(\mathbf{v}'_l \boldsymbol{\psi}_l)]^{-1}$ for $j = 1, \dots, J$ and $\pi_{zi0} = 1 - \sum_{j=1}^J \pi_{zij}$, so that $\boldsymbol{\psi} = (\boldsymbol{\psi}'_1, \dots, \boldsymbol{\psi}'_J)'$. Let π_{zij}^* be π_{zij} for the observations $i = 1, \dots, m$ of the pseudo-data, evaluated at $\boldsymbol{\psi}$, and define, for $i = 1, \dots, m$, $\delta_{ij}^* = 1$ if $z_i^* = 1$ and $\delta_{ij}^* = 0$ otherwise. Then

$$\mathbf{D}_1 = [\text{diag}(\delta_{i1}^* - \pi_{zi1}^*) \mathbf{V}_1 \dots \text{diag}(\delta_{iJ}^* - \pi_{ziJ}^*) \mathbf{V}_1]$$

for $i = 1, \dots, m$, and (11) is

$$\begin{aligned} \frac{\partial^2 Q(\boldsymbol{\theta} | \boldsymbol{\theta}_{(t)})}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'} = & - \begin{bmatrix} \mathbf{V}' \text{diag}(w_{i(t)}^* \pi_{zi1}^*) \mathbf{V} & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{V}' \text{diag}(w_{i(t)}^* \pi_{ziJ}^*) \mathbf{V} \end{bmatrix} \\ & + \begin{bmatrix} \mathbf{V}' \text{diag}(w_{i(t)}^* \pi_{zi1}^* \pi_{zi1}^*) \mathbf{V} & \dots & \mathbf{V}' \text{diag}(w_{i(t)}^* \pi_{zi1}^* \pi_{ziJ}^*) \mathbf{V} \\ \vdots & \ddots & \vdots \\ \mathbf{V}' \text{diag}(w_{i(t)}^* \pi_{ziJ}^* \pi_{zi1}^*) \mathbf{V} & \dots & \mathbf{V}' \text{diag}(w_{i(t)}^* \pi_{ziJ}^* \pi_{ziJ}^*) \mathbf{V} \end{bmatrix} \end{aligned}$$

for $i = 1, \dots, m$.

A.4. Ordinal logistic model

Redefine \mathbf{v}_i by omitting its first element which corresponds to the constant term of the model, and let $\mathbf{v}_{ij} = (\mathbf{d}'_j, -\mathbf{v}'_i)'$ for $j = 0, \dots, J-1$, where \mathbf{d}_i is a J -vector of 0s, except for a 1 as its $(j+1)$ th element, and let $\mathbf{v}_{i,-1}$ and \mathbf{v}_{iJ} be vectors of 0s. Let $\boldsymbol{\psi} = (\psi_{00}, \dots, \psi_{0,J-1}, \boldsymbol{\psi}'_v)'$. Here Z_i follows a multinomial distribution with index 1 and probabilities $\pi_{zij} = \alpha_{ij} - \alpha_{i,j-1}$ for $j = 0, \dots, J$, where $\alpha_{i,-1} = 0$, $\alpha_{iJ} = 1$ and $\alpha_{ij} = \{1 + \exp(-\mathbf{v}'_{ij} \boldsymbol{\psi})\}^{-1} = \{1 + \exp(-[\psi_{0j} - \mathbf{v}'_j \boldsymbol{\psi}_v])\}^{-1}$ for $j = 0, \dots, J-1$. Let $\mathbf{V}_{(j)}$ and $\mathbf{V}_{(j-1)}$ be matrices whose rows $i = 1, \dots, m$ are \mathbf{v}_{ij} and $\mathbf{v}_{i,j-1}$ respectively when $z_i^* = j$, and let $\mathbf{V}_{1(j)}$ and $\mathbf{V}_{1(j-1)}$ be the first n rows of these matrices respectively. Let $\alpha_{iz} = \alpha_{ij}$ and $\alpha_{i,z-1} = \alpha_{i,j-1}$ when $z_i^* = j$. Then $\mathbf{D}_1 = \text{diag}(\alpha_{iz}(1 - \alpha_{iz})/p_{zi}^*) \mathbf{V}_{1(j)} - \text{diag}(\alpha_{i,z-1}(1 - \alpha_{i,z-1})/p_{zi}^*) \mathbf{V}_{1(j-1)}$ and (11) is given by $\mathbf{V}'_{(j)} \text{diag}(a_i) \mathbf{V}_{(j)} + \mathbf{V}'_{(j)} \text{diag}(b_i) \mathbf{V}_{(j-1)} + \mathbf{V}'_{(j-1)} \text{diag}(b_i) \mathbf{V}_{(j)} - \mathbf{V}'_{(j-1)} \text{diag}(c_i) \mathbf{V}_{(j-1)}$ where $a_i = w_{i(t)}^* \alpha_{iz}(1 - \alpha_{iz})[(1 - 2\alpha_{iz})p_{zi}^* - \alpha_{iz}(1 - \alpha_{iz})]/(p_{zi}^*)^2$, $b_i = w_{i(t)}^* \alpha_{iz}(1 - \alpha_{iz})\alpha_{i,z-1}(1 - \alpha_{i,z-1})/(p_{zi}^*)^2$, and $c_i = w_{i(t)}^* \alpha_{i,z-1}(1 - \alpha_{i,z-1})[(1 - 2\alpha_{i,z-1})p_{zi}^* + \alpha_{i,z-1}(1 - \alpha_{i,z-1})]/(p_{zi}^*)^2$ for $i = 1, \dots, m$.

References

- Blair, G. and K. Imai (2010). list: Statistical methods for the item count technique and list experiment. Available at The Comprehensive R Archive Network (CRAN) at <http://CRAN.R-project.org/package=list>.
- Blair, G. and K. Imai (2012). Statistical analysis of list experiments. *Political Analysis* 20, 47–77.
- Chaudhuri, A. and T. C. Christofides (2007). Item count technique in estimating the proportion of people with a sensitive feature. *Journal of Statistical Planning and Inference* 137, 589–593.

- Chaudhuri, A. and R. Mukerjee (1988). *Randomized Response: Theory and Techniques*. New York: Marcel Dekker.
- Chen, T. T. (1989). A review of methods for misclassified categorical data in epidemiology. *Statistics in Medicine* 8, 1095–1106.
- Corstange, D. (2009). Sensitive questions, truthful answers? Modeling the list experiment with LISTIT. *Political Analysis* 17, 45–63.
- Coutts, E. and B. Jann (2011). Sensitive questions in online surveys: Experimental results for the randomized response technique (RRT) and the unmatched count technique (UCT). *Sociological Methods & Research* 40, 169–193.
- Dempster, A. P., N. W. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm(with discussion). *Journal of the Royal Statistical Society B* 39, 1–38.
- European Social Survey (2011). Trust in justice: Topline findings from the European Social Survey. *ESS Topline Results Series, Issue 1*. By J. Jackson, M. Hough, B. Bradford, T. M. Pooler, K. Hohl and J. Kuha.
- Glynn, A. N. (2010). What can we learn with statistical truth serum? Design and analysis of the list experiment. Technical report, Department of Government, Harvard University.
- Groves, R. M., F. J. Fowler, Jr, M. P. Couper, J. M. Lepkowski, E. Singer, and R. Tourangeau (2009). *Survey Methodology* (Second ed.). Hoboken, NJ: Wiley.
- Holbrook, A. L. and J. A. Krosnick (2010). Social desirability bias in voter turnout reports: Tests using the item count technique. *Public Opinion Quarterly* 74, 37–67.
- Hough, M. and M. Sato (Eds.) (2011). *Trust in Justice: Why It Is Important for Criminal Policy, and How It Can Be Measured: Final Report of the Euro-Justis Project*. Helsinki: European Institute for Crime Prevention and Control, affiliated with the United nations (HEUNI).
- Imai, K. (2011). Multivariate regression analysis for the item count technique. *Journal of the American Statistical Association* 106, 407–416.
- Jackson, J., B. Bradford, M. Hough, A. Myhill, P. Quinton, and T. R. Tyler (2012). Why do people comply with the law? Legitimacy and the influence of legal institutions. *British Journal of Criminology*. Forthcoming.
- Johnson, N. L., S. Kotz, and A. W. Kemp (1992). *Univariate Discrete Distributions* (Second ed.). New York: Wiley.
- Kroneberg, C., I. Heintze, and G. Mehlkop (2010). The interplay of moral norms and instrumental incentives in crime causation. *Criminology* 48, 259–294.
- Lensvelt-Mulders, G. J. L. M., J. J. Hox, and P. G. M. van der Heijden (2005). How to improve the efficiency of randomised response designs. *Quality & Quantity* 39, 253–265.
- Maddala, G. S. (1983). *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

- Miller, J. D. (1984). *A New Survey Technique for Studying Deviant Behavior*. Phd thesis, The George Washington University.
- Oakes, D. (1999). Direct calculation of the information matrix via the EM algorithm. *Journal of the Royal Statistical Society B* 61, 479–482.
- Prentice, R. L. (1986). Binary regression using an extended beta-binomial distribution, with discussion of correlation induced by covariate measurement errors. *Journal of the American Statistical Association* 81, 321–327.
- R Core Team (2012). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <http://www.R-project.org>.
- Raghavarao, D. and W. T. Federer (1979). Block total response as an alternative to the randomized response method in surveys. *Journal of the Royal Statistical Society B* 41, 40–45.
- Scheers, N. J. and C. M. Dayton (1988). Covariate randomized response models. *Journal of the American Statistical Association* 83, 969–974.
- Tourangeau, R., L. J. Rips, and K. Rasinski (2000). *The Psychology of Survey Response*. New York: Cambridge University Press.
- Tourangeau, R. and T. Yan (2007). Sensitive questions in surveys. *Psychological Bulletin* 133, 859–883.
- Tsuchiya, T., Y. Hirai, and S. Ono (2007). A study of the properties of the item count technique. *Public Opinion Quarterly* 71, 253–272.
- Tyler, T. R. (2006a). Psychological perspectives on legitimacy and legitimation. *Annual Review of Psychology* 57, 375–400.
- Tyler, T. R. (2006b). *Why People Obey the Law*. Princeton: Princeton University Press.
- van den Hout, A. and P. G. M. van der Heijden (2004). The analysis of multivariate misclassified data with special attention to randomized response data. *Sociological Methods & Research* 32, 384–410.
- Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association* 60, 63–69.
- Wikström, P.-O. H., D. Oberwittler, K. Treiber, and B. Hardie (2012). *Breaking Rules: The Social and Situational Dynamics of Young People's Urban Crime*. Oxford: Oxford University Press.