Estimating the Hidden Number of Scrapie Affected Holdings in Great Britain Using a Simple, Truncated Count Model Allowing for Heterogeneity

Dankmar BÖHNING and Victor Javier DEL RIO VILAS

None of the current surveillance streams monitoring the presence of scrapie in Great Britain provide a comprehensive and unbiased estimate of the prevalence of the disease at the holding level. Previous work to estimate the under-ascertainment adjusted prevalence of scrapie in Great Britain applied multiple-list capture–recapture methods. The enforcement of new control measures on scrapie-affected holdings in 2004 has stopped the overlapping between surveillance sources and, hence, the application of multiple-list capture–recapture models. Alternative methods, still under the capture–recapture methodology, relying on repeated entries in one single list have been suggested in these situations. In this article, we apply one-list capture–recapture approaches to data held on the Scrapie Notifications Database to estimate the undetected population of scrapie-affected holdings with clinical disease in Great Britain for the years 2002, 2003, and 2004. For doing so, we develop a new diagnostic tool for indication of heterogeneity as well as a new understanding of the Zelterman and Chao's lower bound estimators to account for potential unobserved heterogeneity. We demonstrate that the Zelterman estimator can be viewed as a maximum likelihood estimator for a special, locally truncated Poisson likelihood equivalent to a binomial likelihood. This understanding allows the extension of the Zelterman approach by means of logistic regression to include observed heterogeneity in the form of covariates—in case studied here, the holding size and country of origin. Our results confirm the presence of substantial unobserved heterogeneity supporting the application of our two estimators. The total scrapie-affected holding population in Great Britain is around 300 holdings per year. None of the covariates appear to inform the model significantly.

Key Words: Capture–recapture; Chao estimator and bounds; Zelterman estimator.

1. INTRODUCTION

Scrapie is a fatal neuro-degenerative disease that affects small ruminants. It became notifiable in Great Britain in 1993. Clinical cases are reported to the veterinary authorities

Dankmar Böhning is Professor in Statistics, Quantitative Biology and Applied Statistics (QBAS), School of Biological Sciences, Harry Pitt Building, The University of Reading, Earley Gate, Whiteknights Road, Reading, RG6 6FN, UK (E-mail: *d.a.w.bohning@reading.ac.uk*). Victor Javier Del Rio Vilas is Epidemiologist, Veterinary Laboratories Agency, New Haw, Addlestone, Surrey KT15 3NB, UK (E-mail: *v.delriovilas@vla.defra.gsi.gov.uk*).

Journal of Agricultural, Biological, and Environmental Statistics, Volume 13, Number 1, Pages 1–22
DOI: 10.1198/108571108X277904

for confirmatory diagnosis and recorded into the Scrapie Notifications Database (SND). In recent years scrapie has gained importance globally due to potential links with Bovine Spongiform Encephalopathy (BSE) and Variant Creutzfeldt–Jakob disease (vCJD) (Foster et al. 2001; Hunter 2003).

The Department for Environment, Food and Rural Affairs (Defra) in Great Britain aims to reduce the prevalence of scrapie by 40% by 2010 and, eventually, to eradicate it. Although this objective focuses uniquely on scrapie at the individual level, control measures at the holding level are required. To monitor the performance of scrapie eradication measures in Great Britain we need reliable, repeatable, and affordable means of estimating the under-ascertainment adjusted prevalence of scrapie-affected holdings. None of the current surveillance streams in place—the passive surveillance cases reported on SND and active surveillance in the Abattoir (AS) and Fallen Stock (FS) surveys—provide a comprehensive estimate of the prevalence of scrapie-affected holdings. Either they target specific segments of the population, for example relying on the willingness of the farmers to report disease (FS and SND), or they have no immediate application to the holding level due to traceability problems of the individual samples (AS). The under-ascertainment adjusted prevalence has been estimated via anonymous postal surveys (PS) in the past (Hoinville et al. 2000; Sivam et al. 2003). More recently, multiple-list capture–recapture (CRC) methods were applied to estimate the number of scrapie-affected holdings not detected by any of the surveillance streams in place (Del Rio Vilas et al. 2005). Either approach presents limitations. Postal surveys rely on the correct identification of signs of the disease by farmers and are an additional stream that requires extra resources. The traditional CRC approach (see also Bishop et al. 1975, p. 229–256) on the other hand, although using the existing surveillance streams, requires substantial overlapping between them and a favorable regulatory context for a holding to be recorded in more than one source.

With the implementation in 2004 of the Compulsory Scrapie Flock Scheme (CSFS) in Great Britain, following EU Regulations (Anon. 2003), traditional CRC methods are no longer applicable: the confirmation of the index case, from any of the current surveillance sources, leads to the inclusion of the holding in the CSFS. This prevents the overlapping between surveillance sources. Böhning et al. (2004) applied a modification of the traditional CRC methods, based on a single source with repeated entries, to estimate the number of unobserved drug users in Bangkok in 2001. Böhning and colleagues (see also Böhning and Kuhnert 2006; Böhning and Schön 2005; Böhning et al. 2005) identified for each drug user the number of times he/she contacted a treatment institution. This leads to a count distribution $f_1, f_2, \ldots$, where f_j denotes the frequency of drug users with exactly j contacts. The frequency f_0 of those with no contacts is missing in the database and target of the inference. This approach follows a methodology suggested by McKendrick (1926) who estimated the number of cholera-affected households not detected by the officials during an outbreak in India. The extension of this approach to our problem is straightforward, replacing households by animal holdings. In this study we present the application of one-list CRC models to estimate the number of unobserved scrapie-affected holdings in Great Britain in 2002, 2003, and 2004. The article is organized as follows. In Section 2 data sources, case definitions and covariate information are presented. Section 3 discusses the

modeling approaches of Zelterman (1988) and Chao (1987) as well as their extensions in terms of diagnostic tools and covariate modeling. Section 4 presents the results of the analysis of hidden scrapie for the years 2002, 2003, and 2004. An appendix completes the article by providing an approach for variance estimation.

2. MATERIAL

2.1 DATA SOURCES

The SND records details of all premises in Great Britain where a case has been reported since the beginning of 1998. In contrast with more rigorous schemes in place in other EU countries, which involve complete depopulation of the affected holding, Great Britain's moderate approach to scrapie control once an index case was detected allowed the occurrence of more than one case from the same premises (Del Rio Vilas et al. 2006). This generated a count distribution of cases in each holding. We use 2002, 2003, and 2004 SND data in our analyses.

2.2 CASE DEFINITIONS

The process of diagnosing scrapie can be thought of as a cascade of decisions (Kuncheva et al. in press). First, farmers report suspect sheep to veterinary officials (VO). Those animals diagnosed as clinically positive by VOs are then submitted for confirmatory laboratory diagnosis. We are interested in the two groups originating from this process: (1) the total number of reporting holdings and (2) the total number of scrapie-confirmed holdings. The application of our models to the first group will inform the unobserved number of scrapie-like affected holdings in Great Britain. This estimate will permit comparisons with the 2002 PS allowing for similar misclassification in the clinical diagnosis of the condition by the farmers. The second group will inform the number of scrapie-affected holdings with clinical disease confirmed by laboratory tests. As a result of the active surveillance initiated in 2002 in the EU (Anon. 2001), a new form of atypical scrapie that produced discordant responses to the established diagnostic tests was detected throughout Europe (Buschmann et al. 2004a, b). In Great Britain, the first cases of atypical scrapie were detected in 2002 through active surveillance (Everest et al. 2006; Saunders et al. 2006). It was only in 2005 that the first cases of clinical atypical scrapie were reported to the SND (Del Rio Vilas, personal observation). Consequently, the SND data can only inform an analysis of classical scrapie in this study.

2.3 COVARIATE INFORMATION

Systematic heterogeneity could be informed by the inclusion of the holding size, more specifically numbers of adult sheep, and country of origin for each holding. Previous studies have shown the significant effect of these two variables on the occurrence of scrapie in Great Britain (McLean et al. 1999; Del Rio Vilas et al. 2006). County-Parish-Holding (CPHs) unique identifiers were available for all holdings. Holding size data were retrieved

after matching the SND CPH's with those in the June Agricultural Census for 2004 (Anon. 2004).

3. STATISTICAL METHODS

3.1 THE POISSON MODEL

It is assumed that the surveillance mechanism identifies scrapie-affected holdings with probability $1 - p_0$, so that p_0 is the probability that a scrapie-affected holding remains undetected. Also, we will denote with $p_1, p_2, \ldots p_m$ the probability of identifying a holding with exactly 1, 2, $\ldots m$ cases of scrapie. With these probabilities are associated observed frequencies $f_1, f_2, \ldots f_m$, where f_j corresponds to the frequency of holdings with exactly 1, 2, …, m cases where m is the largest count occurred. Clearly, the frequency f_0 of scrapie-affected holdings is *unobserved* and target of the inference.

We also denote with $N = f_0 + f_1 + \cdots + f_m$ the size of the scrapie-affected holding population and with $n = N - f_0$ that part which is *observed*. We need to distinguish between the true, unobserved number $N = \sum_{i=0}^{m} f_i$ of scrapie-affected holdings and the observed number $n = \sum_{i=1}^{m} f_i$ of scrapie-affected holdings. Also, $s = \sum_{i=1}^{m} i f_i$ is the total number of animals observed with scrapie. N might be partitioned as $N = n + Np_0$ so that the Horvitz–Thompson estimator of N can be found as

$$\hat{N} = n/(1 - p_0)$$

if p_0 is known. If p_0 is unknown, one must assume some model structure for $p_j = p_j(\lambda)$, so that an estimate of $\hat{\lambda}$ can also lead to an estimate of $p_0(\lambda)$, thus by leading to an estimate of $\hat{N} = n/[1 - p_0(\hat{\lambda})]$. In a simple model, one can assume that the counts arise from a Poisson with parameter λ, so that $p_j = \exp(-\lambda)\lambda^j/j!$ and $g(\lambda) = 1 - p_0 = 1 - \exp(-\lambda)$. Estimation could be done by maximizing the likelihood function based upon the zero-truncated Poisson density

$$L(\lambda) = \prod_{i=1}^{m} \left\{ \frac{Po(i, \lambda)}{1 - \exp(-\lambda)} \right\}^{f_i} \tag{3.1}$$

with respect to λ, where m is the largest observed count and $Po(i, \lambda) = \exp(-\lambda)\lambda^i/i!$. The likelihood (3.1) can be easily maximized with the EM algorithm toggling between the *M-Step* $\hat{\lambda} = s/N$ and the *E-Step* $\hat{N} = n/[1-\exp(-\hat{\lambda})]$ so that $\hat{N}$ can be found as a solution from the equation

$$\hat{N} = \frac{n}{1 - \exp(-s/\hat{N})}. \tag{3.2}$$

A closed-form approximation for $\hat{N}$ can be be provided using a modification of a suggestion by Moore (1952), namely $\hat{\lambda} = \frac{s-f_1}{n}$ as an estimate for the Poisson parameter, so that $\hat{N} = n/\{1 - \exp[(s - f_1)/n)]\}$. A variety of estimators have been suggested including the maximum likelihood estimator under Poisson homogeneity as a solution of Equation (3.2) and the Turing estimator defined as $\hat{N} = n/(1 - f_1/s)$ (Good 1953; Darroch and Ratcliff 1980).

An approach that we want to concentrate here on was suggested by Zelterman (1988) and uses the fact that

$$\lambda = (i+1)\frac{Po(i+1,\lambda)}{Po(i,\lambda)}, \tag{3.3}$$

so that the simple estimator $\hat{\lambda}_i = (i+1)\, f_{i+1}/f_i$ can be constructed, typical choices for i are $i = 1$ or $i = 2$. This estimator turns out to be less dependent on the homogeneous Poisson assumption, so that it might still be considered even if a homogenous Poisson model is not valid (Zelterman 1988; Wilson and Collins 1992) and it is likely that for this reason it has found popularity in many applied areas (Hay and Smit 2003; Roberts and Brewer 2006).

3.2 THE ZELTERMAN ESTIMATOR ALLOWING FOR HETEROGENEITY

In more detail, we might think of a Poisson model which truncates all counts except those equal to i or $i+1$, where i is an arbitrary but fixed count larger than 0. Such a truncated Poisson count model will require homogeneity *only* for the count observations i and $i+1$, whereas deviations from the Poisson model for counts larger than $i+1$ are allowed. In particular, $i = 1$ might be chosen so that counts larger than 2 might be allowed to follow a Poisson model with a different parameter. Hence, the Zelterman approach allows some amount of heterogeneity for the count distribution. To be more specific we assume that we have for given and fixed count i a count-specific parameter λ_i associated with $\frac{Po(i,\lambda_i)}{Po(i,\lambda_i)+Po(i+1,\lambda_i)} = \frac{1}{1+\lambda_i/(i+1)}$. An estimate of such a *local* Poisson parameter λ_i can be achieved by considering only neighboring pairs i and $i+1$ for $i = 1, 2, \ldots, m-1$ with all other observations different from i and $i+1$ being truncated. Then, *conditional upon* f_i and f_{i+1}, the likelihood is simply

$$\left(\frac{1}{1+\lambda_i/(i+1)}\right)^{f_i} \times \left(\frac{\lambda_i/(i+1)}{1+\lambda_i/(i+1)}\right)^{f_{i+1}} \tag{3.4}$$

providing the log-likelihood

$$f_{i+1}\log(\lambda) - (f_i + f_{i+1})\log(1+\lambda/(i+1)).$$

Maximizing this log-likelihood in λ delivers $\hat{\lambda}_i = (i+1)\, f_{i+1}/f_i$. Since there is an entire sequence of estimators $2\, f_2/f_1$, $3\, f_3/f_2$, $4\, f_4/f_3$, ... associated with the pairs $(f_1,\ f_2)$, $(f_2,\ f_3)$, ..., respectively , a decision has to be made which of them should be used. Since we want to make a prediction for $i = 0$ we use the pair $(f_1,\ f_2)$ of frequencies of the counts 1 and 2 which are closest to 0 leading to

$$\hat{N}_Z = \frac{n}{1-p_0(\hat{\lambda}_1)} = \frac{n}{1-\exp(-2\, f_2/f_1)}.$$

Also, a variance estimate can be found for the Zelterman estimator $\hat{\lambda} = (i+1)\, f_{i+1}/f_i$ as (see appendix for details)

$$\widehat{\mathrm{var}}\left(\frac{n}{1-\exp(-\hat{\lambda})}\right) = nG(\hat{\lambda})\left[1 + nG(\hat{\lambda})\hat{\lambda}^2\left(\frac{1}{f_i} + \frac{1}{f_{i+1}}\right)\right], \tag{3.5}$$

where $G(\hat{\lambda}) = \frac{\exp(-\hat{\lambda})}{(1-\exp(-\hat{\lambda}))^2}$.

Table 1. Two hypothetical Poisson populations of size 500 each and with means 0.5 and 4, respectively and the associated collapsed population of size 1,000.

	Population I	Population II	Both populations
j	f_j	f_j	f_j
0	305	11	316
1	149	37	186
2	43	75	118
3	3	93	96
4		88	88
5		89	89
6		54	54
7		31	31
8		11	11
9		6	6
10		5	5
N	500	500	1,000

3.2.1 An Illustration of the Zelterman Estimator

Here we would like to illustrate that the Zelterman estimator is less sensitive to the occurrence of heterogeneity than the conventional maximum likelihood estimator. Two populations are considered, each $N = 500$ of size. One is generated from a Poisson with parameter $\lambda = 0.5$ the other from a Poisson with $\lambda = 4$. The data are provided in Table 1.

We are interested in estimating the size $N = 1{,}000$ of the merged populations where membership to the two populations could be described by a simple dummy covariate. Using the covariate stratification and homogeneous MLE estimation we find $\hat{N}_I + \hat{N}_{II} = 505.65 + 498.06 = 1003.71$ where $\hat{N}_I$ and $\hat{N}_{II}$ are the conventional maximum likelihood estimates under homogeneity. The estimate appears quite reasonable. Now, we ignore the stratification and estimate N under homogenous Poisson: $\hat{N} = 719.74$ and the bias occurs as expected. Let us do the same procedure for the Zelterman estimate. We find $\hat{N}_Z = 942.31$ for the stratified case and $\hat{N}_Z = 951.54$ for the unstratified case, the latter being considerable less biased despite strong heterogeneity.

3.2.2 Other Estimators

Modifications of the Turing estimator $\hat{N} = n/(1 - f_1/s)$ were suggested by Chao and Lee (1992) to adjust for heterogeneity in the capture probabilities leading to an additional term in the Turing estimator determined by the coefficient of variation of the capture probabilities. Two estimators for the coefficient of variation were suggested by Chao and Lee (1992) (see Equations (2.12 and (2.23)) and the associated modifications of the Turing estimator are conventionally called ACE and ACE1 (the latter of the two is suggested for highly heterogeneous situations). All these estimators (except the Zelterman estimator) can be easily computed by means of the software tool SPADE (Chao and Shen 2006) includ-

ing the lower bound estimators suggested by Chao (1987) which we will discuss in more detail in the following section. These estimators will also serve as comparative estimators to the Zelterman approach and will be applied for estimating the number of hidden scrapie holdings as well.

3.3 DIAGNOSTIC DEVICE FOR HETEROGENEITY AND CHAO'S LOWER BOUND ESTIMATOR

Chao (1984, 1987) suggested the estimator $\hat{N}_C = n + f_1^2/(2f_2)$ which is based on a mixed Poisson model with $p_i = \int_t e^{-t}t^i/i!\lambda(t)dt$ for $i = 0, 1, 2, \ldots$ and arbitrary density $\lambda(t)$ modelling the heterogeneity in the parameter. Applying the inequality of Cauchy-Schwartz, which states that $[E(XY)]^2 \leq E(X^2)E(Y^2)$ for arbitrary random variables X and Y, to any mixed Poisson probability leads to

$$\left(\int_t e^{-t}t^i\lambda(t)dt\right)^2 \leq \int_t e^{-t}t^{i-1}\lambda(t)dt \int_i e^{-t}t^{i+1}\lambda(t)dt, \tag{3.6}$$

or

$$(i!p_i)^2 \leq (i-1)!p_{i-1}(i+1)!p_{i+1}, \tag{3.7}$$

or

$$ip_i/p_{i-1} \leq (i+1)p_{i+1}/p_i, \tag{3.8}$$

from where the Chao bound is derived with $i = 1$ as $p_0 \geq p_1^2/(2p_2)$, and Chao's estimator follows by replacing p_i by f_i for $i = 1, 2$. The Chao estimator provides a lower bound for the population size which

- is independent of the form of the heterogeneity distribution $\lambda(t)$, and
- avoids estimating the heterogeneity distribution $\lambda(t)$ nonparametrically.

We will use the lower bound estimate of Chao mainly as a tool for double-checking the plausibility of Zelterman's estimate. We expect that $\hat{N}_C \leq \hat{N}_Z$ will be fulfilled in all cases (see also the discussion for this point). A variance formula for $\hat{N}_C$ can also be derived with the approach presented in the appendix:

$$\widehat{\text{var}}(\hat{N}_C) = \frac{1}{2}\frac{f_1^2}{f_2}\left(1 - \frac{f_1^2}{2f_2n + f_1^2}\right) + \frac{f_1^3}{f_2^2}\left(1 + \frac{1}{4}\frac{f_1}{f_2}(1 - f_2/n)\right). \tag{3.9}$$

Chao (1984, 1987) also suggested a bias-corrected version defined as $n + f_1(f_1-1)/(2f_2+2)$ which is always smaller than $n + f_1^2/(2f_2)$.

In addition, it can be seen that the ratios ip_i/p_{i-1} (3.8) are increasing with increasing i if there is heterogeneity. Now, ip_i/p_{i-1} can be simply estimated by if_i/f_{i-1}, so that we take a monotone increasing pattern in these ratios as an *indication of presence of heterogeneity* in which case we suggest using the Zelterman estimator in connection with Chao's estimator as a fail-safe method.

Table 2. Ten holdings with number of confirmed scrapie cases, the location of the holding and its size

ID	Confirmed cases y_i	δ_i	Country	Size of holding
1	1	0	Wales	34
2	2	1	England	310
3	1	0	England	31
4	2	1	England	232
5	2	1	England	17
6	1	0	England	49
7	1	0	Scotland	591
8	3	–	Scotland	19
9	1	0	England	249
10	1	0	England	29

3.4 EXTENDING THE MODEL FOR COVARIATES

It is one of the favorable aspects of the Zelterman estimator (in comparison to other estimators adjusting for heterogeneity) that it can easily be extended to include covariate information. We have seen that $\hat{\lambda}_i$ is also a maximum likelihood estimator with respect to the likelihood (3.4) which occurs if all counts are truncated unless $Y = i$ or $Y = i + 1$. Let us concentrate on the case $i = 1$ which is most important one for practical purposes. Note that this likelihood can also be written as

$$\left(\frac{1}{1+\lambda/2}\right)^{f_1} \times \left(\frac{\lambda/2}{1+\lambda/2}\right)^{f_2} = (1-p)^{f_1} p^{f_2} \tag{3.10}$$

which is proportional to a *binomial likelihood* with event parameter $p = \frac{\lambda/2}{1+\lambda/2}$, the probability for $Y = 2$. Note that λ is connected uniquely to p via $\lambda = 2\frac{p}{1-p}$ and $p = \lambda/(2+\lambda)$. The likelihood is also valid for *case data* in *ungrouped* form. Table 2 provides an illustration of this situation. There are 10 holdings listed with the number y_i of confirmed scrapie cases. Whenever $y_i = 2$, then the binary indicator $\delta_i = 1$, and if $y_i = 1$, then $\delta_i = 0$. If $y_i > 2$ the case is ignored. See Table 2 for detailed illustration. In addition, two important additional covariates, the location of the holding and its size, are provided.

Suppose covariates are available in the form of a vector $\mathbf{x}_i$ for the ith unit in the sample. In a generalized linear model connecting the binary outcome probability p_i with the *linear predictor* $\eta_i = \beta^T \mathbf{x}_i$ with a log-link we have that

$$p_i = \frac{e^{\eta_i}}{1+e^{\eta_i}}.$$

On the other hand, p_i and the local Poisson parameter λ_i are connected via

$$p_i = \frac{\lambda_i/2}{1+\lambda_i/2},$$

so that λ_i and the linear predictor η_i are simply connected via $\lambda_i = 2e^{\eta_i}$. Maximum likelihood estimates $\hat{\beta}$ of β are readily available by means of *logistic regression* so that the

Table 3. Distribution of cases within holding for holdings with reported/confirmed scrapie

Cases count within holding	1	2	3	4	5	6	7	8+	n
Frequency reported 2002	95	28	19	8	7	2	4	14	177
Frequency confirmed 2002	74	23	15	6	8	3	3	12	144
Frequency reported 2003	82	30	16	1	4	2	4	18	157
Frequency confirmed 2003	66	29	12	2	3	2	3	17	134
Frequency reported 2004	142	37	19	6	5	9	0	9	227
Frequency confirmed 2004	83	29	14	6	5	6	0	8	151

generalized Zelterman estimate is found as

$$\hat{N}_Z = \sum_{i=1}^{n} \frac{1}{1-\exp(-\hat{\lambda}_i)} = \sum_{i=1}^{n} \frac{1}{1-\exp(-2e^{\hat{\eta}_i})}, \tag{3.11}$$

where $\hat{\eta}_i = \hat{\beta}^T \mathbf{x}_i$. In the Equation (3.12) we provide a variance estimate formula for (3.11) and give details in an appendix (see also formula (A.3) in the appendix):

$$\widehat{\text{var}}(\hat{N}_Z) = \sum_{i=1}^{n} \frac{1-w_i}{w_i^2} + \left(\sum_i \nabla w_i(\hat{\beta})^T\right) \text{cov}(\hat{\beta}) \left(\sum_i \nabla w_i(\hat{\beta})\right), \tag{3.12}$$

where

$$\nabla w_i(\hat{\beta}) = \frac{(1-w_i)v_i}{w_i^2}\mathbf{x}_i,$$

and $v_i = -2e^{\hat{\eta}_i}$, $w_i = 1-\exp(v_i)$.

4. RESULTS

4.1 DESCRIPTIVE STATISTICS AND DIAGNOSTICS

Table 3 shows the frequency distributions of cases within holding, separated for holdings with reported and confirmed scrapie, for the years 2002, 2003, and 2004. For example, in 2002 there are 95 holdings with one reported case of scrapie and 74 with one confirmed case of scrapie. Also, there were 28 holdings in 2002 that had two reported cases of scrapie and 23 with two confirmed cases of scrapie. Note that about 70% of holdings have one or two cases, so that the majority of holdings is represented by f_1 and f_2.

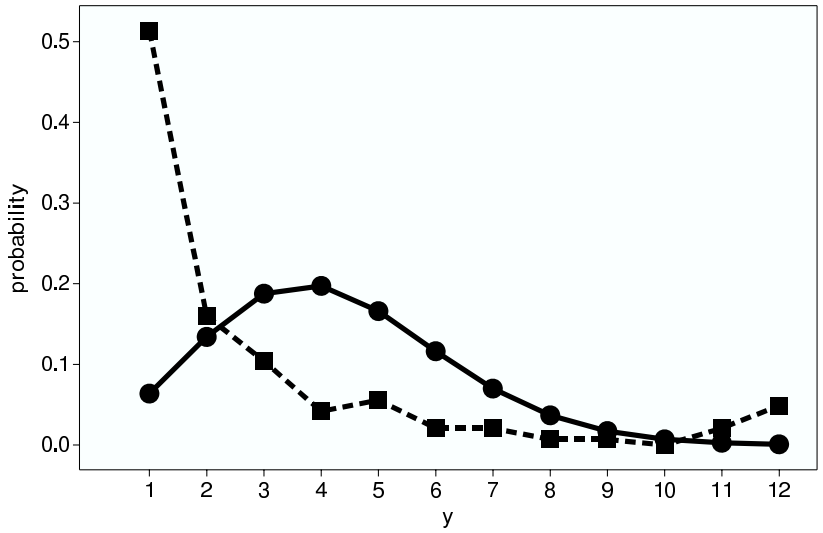

Figure 1. Graph of observed relative frequencies (squares with dashed line) and fitted probabilities using a simple Poisson model (bullets with solid line) versus y for the year 2002.

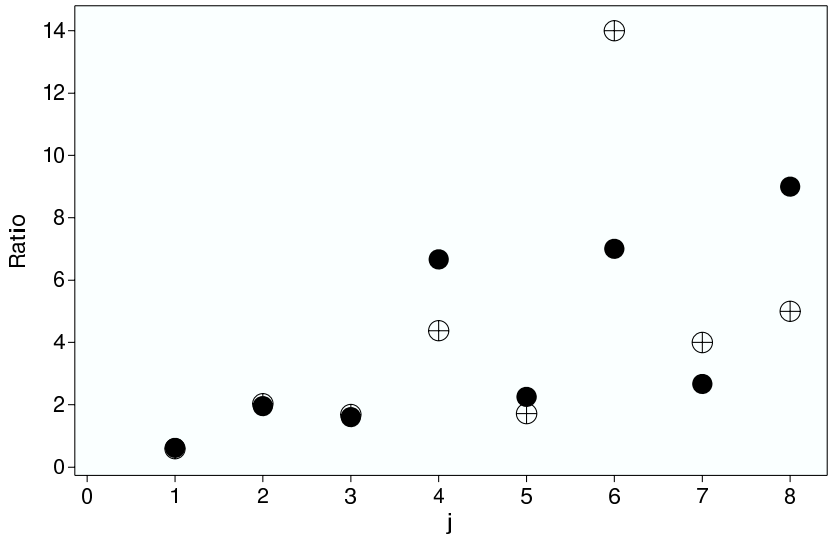

Figure 2. Graph of $(i+1)\, f_{i+1}/f_i$ versus i for the year 2002; black bullets are confirmed cases, circles with a plus are reported cases

It is clear from Figure 1 that a simple Poisson model of the form $\Pr(Y = y) = e^{-\lambda}\lambda^y/y!$ does not fit the observed frequency distribution by any acceptable degree (note that for creating the fitted Poisson distribution λ was replaced by its maximum likelihood estimate). If we consider the ratios $(i+1)\,f_{i+1}/f_i$ for the data of 2002 it is evident from Figure 2 that the graph of the function $i \to (i+1)\,f_{i+1}/f_i$ has a distinctive pattern. Although there is no strict monotonicity of these ratios, a clear monotone increasing pattern is apparent, in particular, for the distribution of holdings with confirmed cases. This is similar for the other years 2003 and 2004. We take this phenomenon as indicative of the presence of heterogeneity in the Poisson model. This indicates that using a simple Poisson model is most inappropriate and alternatives incorporating the heterogeneity of the Poisson parameter will be needed.

4.2 ESTIMATES FOR HIDDEN SCRAPIE ADJUSTING FOR UNOBSERVED HETEROGENEITY

Given the results from the previous section we will develop estimates for hidden scrapie using the robust approach developed in Section 3.2. Table 4 provides the results for the years 2002, 2003, 2004, separated for holdings with reported and confirmed scrapie cases. The Zelterman estimate N_Z of the size of the population of holdings with scrapie is given in the first rows in the three subtables of Table 4 with associated 95% confidence interval.

The lower bound estimator of Chao follows in rows 4 of the subtable with associated 95% confidence interval. The lower bound estimator is provided here as an additional tool for supporting the trust in the robust estimator supplied by the Zelterman approach. Note that in *all* cases the lower bound estimator is indeed smaller than the Zelterman estimator. Finally, in the last rows of the subtables the proportion of observed holdings with scrapie within all (observed and unobserved) holdings with scrapie is provided including a 95% confidence intervals. This statistic measures the extent to which the surveillance stream is able to capture the occurrence of the disease, in other words, it measures the *completeness of identification*. For the reported cases, these measures lie in the range 0.40–50, whereas the range for the confirmed cases is roughly five units higher, namely 0.45–0.55. The situation is very similar for 2002 and 2004. Also, for 2003 the observed sizes are similar to 2002 and 2004; however, due to a different distributional pattern, the estimate for the hidden size of scrapie affected holdings is slightly higher in 2002 and 2004. There are *no* significant differences in the population size estimate *between* the three years.

In addition, Table 4 also provides results using available alternative estimates of population size by means of the program SPADE (Chao and Shen 2006). The homogeneous model estimate is the Turing estimate and the homogenous MLE is the solution of (3.2)—both lower than the others—as expected in the situation of heterogeneity. This is followed by Chao's lower bound estimate and its bias-corrected version and the ACE and ACE1 estimators. Typically, the Zelterman estimator is close to the latter two estimators which also adjust for heterogeneity.

Table 4. Estimates of number of holdings with reported/confirmed scrapie for 2002, 2003, and 2004

	reported 2002		confirmed 2002	
estimator/model	$\hat{N}$	(95% C.I.)	$\hat{N}$	(95% C.I.)
Zelterman	397	(267,528)	311	(200,422)
Homogenous Model	241	(219,276)	191	(173,220)
Homogeneous (MLE)	187	(182,196)	151	(147,159)
Chao lower bound	338	(269,458)	263	(208,366)
Chao-bc*	331	(266,444)	256	(205,352)
ACE	328	(272,416)	250	(208,321)
ACE1	410	(310,587)	302	(230,435)
Completeness $\frac{n}{\hat{N}_Z}$	0.45	(0.29,0.57)	0.46	(0.28,0.60)
	reported 2003		confirmed 2003	
estimator/model	$\hat{N}$	(95% C.I.)	$\hat{N}$	(95% C.I.)
Zelterman	303	(211,394)	229	(162,296)
Homogenous Model	210	(190,242)	174	(158,202)
Homogeneous (MLE)	164	(160,172)	139	(136,147)
Chao lower bound	269	(220,357)	209	(174,275)
Chao-bc*	264	(217,347)	206	(172,268)
ACE	307	(251,397)	247	(202,322)
ACE1	409	(298,606)	318	(233,476)
Completeness $\frac{n}{\hat{N}_Z}$	0.52	(0.34,0.65)	0.58	(0.39,0.72)
	reported 2004		confirmed 2004	
estimator/model	$\hat{N}$	(95% C.I.)	$\hat{N}$	(95% C.I.)
Zelterman	559	(395,723)	300	(200,422)
Homogenous Model	348	(311,402)	209	(188,243)
Homogeneous (MLE)	269	(255,289)	169	(162,183)
Chao Lower Bound	500	(398,662)	270	(218,363)
Chao-bc*	490	(393,646)	264	(215,353)
ACE	510	(418,646)	280	(230,361)
ACE1	691	(511,986)	344	(206,394)
Completeness $\frac{n}{\hat{N}_Z}$	0.41	(0.28,0.51)	0.50	(0.33,0.63)

* Bias-correction of Chao's lower bound estimator (see Section 3.3).

Table 5. Estimates of number of holdings with confirmed scrapie adjusted for the covariates "log-size of holding" and "country of origin (England is reference)" using a logistic regression

	Logistic regression					Size estimation		
Source	Covariates	Coef.	SE(Coef.)	z	$P > \|z\|$	n	$\hat{N}_Z$	95%CI
2002	log-size	0.0254	0.2289	0.11	0.912	125	308	167–448
	Wales	−0.2537	0.8603	−0.29	0.768			
	Scotland	0.2334	0.6205	0.38	0.707			
	intercept	−1.4621	1.0114	−1.45	0.148			
2003	log-size	0.0379	0.1877	0.20	0.840	122	226	142–310
	Wales	−0.1209	0.6669	−0.18	0.856			
	Scotland	−0.8755	0.7004	−1.25	0.211			
	intercept	−0.8445	0.8040	−1.05	0.294			
2004	log-size	0.2006	0.2268	0.88	0.376	135	265	167–364
	Wales	−0.2909	0.4858	−0.60	0.549			
	Scotland	−1.0413	0.7678	−1.36	0.175			
	intercept	−1.6279	1.0773	−1.51	0.131			

4.3 ESTIMATES FOR HIDDEN SCRAPIE ADJUSTING FOR OBSERVED COVARIATES AND UNOBSERVED HETEROGENEITY

We now include the covariate *log-size* of holding and country of origin with England (as reference) and *Wales* and *Scotland* as categories into the analysis. In the Zelterman approach, only holdings with one *or* two scrapie cases are considered. We restrict our analysis to confirmed scrapie. The results of the logistic regression are shown in Table 5. None of the covariates appears to have a significant effect for any of the three years. Consequently, the generalized Zelterman estimator (adjusting for the covariates) does not differ considerably from the unadjusted version provided in Table 4. However, as a negative effect, the variance (computed with the multivariate δ-method provided in (A.2) and (A.3) in the appendix) of the generalized Zelterman estimator increases slightly leading to enlarged confidence intervals. Since for 2003 the analysis gives an estimate slightly different from the other two years, we also looked at bootstrap confidence intervals which is found for 2003 to be 139–300, quite similar to the the δ-method-achieved confidence interval of 142–310. Note that the size of the holding is missing for about 10–15% of the holdings leading to a reduced population size estimator. This fact as well as the nonsignificance of the covariates led to prefer the unadjusted population size estimators in the final analysis.

5. DISCUSSION

5.1 SCRAPIE-RELATED ISSUES

Our results show that the number of scrapie-affected and scrapie-like affected holdings in Great Britain appears to decline in 2003. In contrast, 2004 shows an increase in both

groups. This is a reflection of the greater number of holdings observed in that year (Table 1). Note that the estimate of hidden scrapie-affected holdings will increase with number of observed affected holdings if the ratio of f_2/f_1 stays constant which is roughly the case for the years 2002 to 2004. This increase is likely to be a consequence of the introduction of the CSFS in July 2004. The generous compensation payments to owners of scrapie-affected holdings may have driven the reporting of scrapie. This would be similar to previous findings by Kuchler and Hamm (2000) who reported the additional supply of suspect scrapie cases in the U.S. with payments for affected animals above the market price. Furthermore, it is obvious from Table 1 that in 2004 a larger proportion of holdings reported one or two cases compared to the two previous years. Both Chao and Zelterman estimators react sensitively to any increased weight in the left side of the distribution producing consequently larger estimates for f_0 as it can be seen in the year 2004.

The extension of our models to the count distribution of reported holdings allowed comparisons with the PS assuming that the clinical misclassification is similar between the two approaches. Sivam et al. (2003) estimated that more than 38% of the farmers that thought they had scrapie reported it to the veterinary authorities in 2002. Our estimates of the completeness of identification, a measure of the sensitivity of the surveillance source to detect the event under study, for the same year, was 45%. Both are in the same range suggesting the adequacy of our models by comparison with the only and most used "gold standard" measure of completeness of scrapie reporting to date.

In contrast with the multiple-list CRC approach, this one-list CRC is less demanding in terms of data requirements, mainly because it relies on one list only. In the veterinary surveillance context there is no abundance of diseases monitored by more than one source with enough overlapping to guarantee stable multiple-list CRC methodologies. The models suggested here appear as a straightforward alternative. Nevertheless, the application of these models requires repeated entries from the same holding. This was possible in the context of Great Britain's approach to scrapie controls up to 2004. Del Rio Vilas et al. (2005) stated that the multiple-list CRC was not applicable on other EU countries where the confirmation of the index case involved the depopulation of the holding avoiding any future overlapping with any other surveillance source. The one-list model would not be applicable in these circumstances as there would not be any repeated entries from the same holding. This limitation appears to render ineffective these models for their application from 2004 onwards, once the CSFS was enforced. Another limitation concerns the scope of our results; they are limited by the nature of the source. We can only inform the number of scrapie-affected holdings with clinical disease. Hagenaars et al. (2006) reported that the majority of cases do not develop clinical disease. Consequently, we are still missing, potentially, a large number of scrapie-affected holdings where the disease does not manifest itself clinically. Future research is needed to extend these models to the current situation, under the CSFS, where a count distribution of cases may arise after the TSE testing performed upon each holding entering the scheme. The application of our models to the CSFS would, ideally, return a more comprehensive estimate of the scrapie-affected population, one that would account for clinical disease as well as infection. The study of the CSFS data is in progress.

Another seemingly obvious benefit of the one-list CRC in comparison with multiple-list models is the lower dependence of the former on the need of perfect matches. However, one-list models still rely on the consistent identification of individuals to guarantee the count distributions. Böhning et al. (2004) reported the potential occurrence of this problem if there were multiple centers recording the events and different identifiers were used to record the individuals. If individuals, holdings in our case, were recorded differently when they report a case, from one occasion to the next one, we could miss events in our count distribution. This is not the case in our study where holdings are unequivocally recorded with their unique County-Parish-Holding identifier. Matching problems may still occur. Confirmation of holdings is based ultimately on the fair detection of an animal in a holding. Due to the lack of proper individual records, traceability of a particular animal can be difficult. Whereas there is no room for any mismatches between repeated entries in the case of reporting farmers, this might occur for the confirmed set of holdings, specifically among those reporting purchased animals, for which the tracking of their natal holding, where confirmation regularly takes place, is not always successful.

5.2 STATISTICAL METHODOLOGY

The focus of this study was two-fold: first, a diagnostic device was developed that could be used to support evidence of heterogeneity. Second, a robust estimator previously suggested by Zelterman (1988) was taken and extended for covariates. Though there is evidence that the Zelterman approach is beneficial in comparison to a simple homogeneous Poisson model, it was neither intended to claim that the Zelterman approach is superior to the many existing alternatives, nor was such a comparative approach intended. Our focus was the development and application of a reasonable technique to estimate the burden of hidden scrapie. Nevertheless, it might be helpful and illustrative to see how the Zelterman estimator behaves in simulated data. $N = 100$ counts Y_i were sampled from a two-component mixture of two Poisson distributions: $Y_i \sim p\text{Po}(\lambda) + (1-p)\text{Po}(\mu)$ with $p = 0.5$ and $\lambda = 0.5$, for $i = 1, \ldots, N$. Note that the true population size was $N = 100$. μ was chosen to be from one of the five populations determined by 0.5 (homogeneity), 1, 2, 3, 4 (largest heterogeneity). Counts $Y_i = 0$ were truncated and the reduced sample of size $n = N - f_0$ was used to determine the estimators. Besides the Zelterman estimator itself the following estimators were considered: Chao's lower bound, Turing and ACE. An unknown reviewer suggested a modification of the Zelterman estimator:

$$\frac{f_1 + f_2}{1 - \exp(-2f_2/f_1)} + n - (f_1 + f_2).$$

The motivation behind this estimator is that the first term is a Horvitz–Thompson estimate based on counts 1 and 2, whereas the second terms corresponds to counts larger than 2. We have also included this modification (denoted as Zelterman-R) into the study. The results in terms of bias and variance are provided in Table 6.

Column 4 in Table 6 contains the expected value of the sample size n expected under the model used in this simulation: $E(n) = (1-p_0)N$ with $p_0 = p\exp(-\lambda)+(1-p)\exp(-\mu)$. Under homogeneity, all estimators behave similar with the Turing estimator having the

Table 6. A comparison of selected estimates by means of a simulation study; $Y \sim p\text{Po}(\lambda) + (1-p)\text{Po}(\mu)$ with $p = 0.5$ and $\lambda = 0.5$, true $N = 100$, replication size is 1,000; $\Delta = \exp\left(-\frac{2p_2}{p_1}\right) - p_0$ and $p_i = p\exp(-\lambda)\lambda^i/i! + (1-p)\exp(-\mu)\mu^i/i!$

estimator/model	mean	variance	$E(n) = N(1-p_0)$	Δ
$\mu = 0.5$			39.35	0.
Zelterman	114.390	2978.38		
Zelterman-R	111.770	2832.59		
Chao	113.940	2789.21		
Turing	110.510	1953.11		
ACE	116.570	2361.39		
$\mu = 1$			51.28	0.026
Zelterman	101.770	708.13		
Zelterman-R	97.067	587.80		
Chao	100.530	569.98		
Turing	96.043	272.12		
ACE	101.030	427.36		
$\mu = 2$			62.91	0.072
Zelterman	93.902	319.41		
Zelterman-R	85.528	181.37		
Chao	89.557	177.88		
Turing	82.636	62.74		
ACE	87.328	109.85		
$\mu = 3$			67.18	0.062
Zelterman	95.677	372.13		
Zelterman-R	82.999	132.69		
Chao	86.326	131.82		
Turing	77.343	31.83		
ACE	82.031	61.24		
$\mu = 4$			68.76	0.006
Zelterman	105.910	646.00		
Zelterman-R	84.777	135.32		
Chao	87.352	133.53		
Turing	75.298	23.92		
ACE	79.832	45.13		

smallest variance. With increasing heterogeneity the Zelterman remains having a fairly small bias [a property noted by Collins and Wilson (1992)], whereas all others increase their bias with Chao's lower bound estimator being best among the remaining four. To understand the behavior of the bias of the Zelterman estimator consider the difference Δ of

$$\exp\left(-\frac{2p_2}{p_1}\right) = \exp\left(-\frac{p\exp(-\lambda)\lambda^2 + (1-p)\exp(-\mu)\mu^2}{p\exp(-\lambda)\lambda + (1-p)\exp(-\mu)\mu}\right),$$

for which the estimate $\exp(-2f_2/f_1)$ is used in the Zelterman estimation, to $p\exp(-\lambda) + (1-p)\exp(-\mu)$ which should have been used in the Horvitz–Thompson estimation. This difference (included as the last column in Table 6) will determine the bias of the Zelterman estimator. Bias will be small is p is close to zero (or one) or if μ is close to λ. It is interesting to see what happens if μ becomes large while λ remains fixed, indicating a contamination of large size (large count). Then $\exp(-\frac{2p_2}{p_1}) \to \exp(-\lambda)$ while $p_0 \to p\exp(-\lambda)$. Since $\exp(-\lambda) \geq p\exp(-\lambda)$ this shows that the Zelterman estimate will overestimate the population size by an amount which depends on the fraction p of contamination. Hence, for large μ the Zelterman estimator will provide an upper bound for the population size where this upper bound is the larger the smaller the proportion p is (the higher the amount of contamination). We note further that Zelterman's estimator has often the largest variance among all estimators considered. It might be an interesting question to improve the precision of the Zelterman estimator while keeping its relative low bias.

We modeled the observed heterogeneity by means of the inclusion of holding size and country of origin as covariates in our models. Logistic regression analysis shows that none of them had a significant effect on our response variable (holding with two cases versus holding with one case). This was also found when a zero-truncated regression of the count of cases per holding on the covariates size of holding and country of origin was applied. Note that the logistic regression analysis uses only the reduced dataset of those holdings with exactly one or two cases of scrapie. However, when estimating the hidden number of holdings with scrapie the entire dataset with known values for the covariates of interest is used.

Another related question is if the log-size of holding should be treated as an *offset* in the modeling. Recall that an offset is a covariate with known coefficient. Suppose that the number of cases of scrapie Y_i would be proportional to the holding size m_i: $E(Y_i) = \lambda_i = m_i\mu_i$. Since

$$p_i = \frac{\lambda_i/2}{1+\lambda_i/2} = \frac{e^{\log m_i/2+\eta_i}}{1+e^{\log m_i/2+\eta_i}} = \frac{m_i e^{\eta_i}/2}{1+m_i e^{\eta_i}/2},$$

using $\mu_i = e^{\eta_i}$, it follows that *proportionality can easily be handled* by using $\log m_i/2$ as an offset in the logistic regression model. If we take a look at the scatterplot of log-number of cases per holding versus the holding size we would expect a straight line relationship—under ideal circumstances under proportionality. Figure 3 (for 2002) shows a different picture. In fact, if we include the LOWESS-smoother in the diagram it shows only a mild trend. Hence, we had included the size of the holding as free covariate and it became not significant in the analysis for the reasons visible in Figure 3.

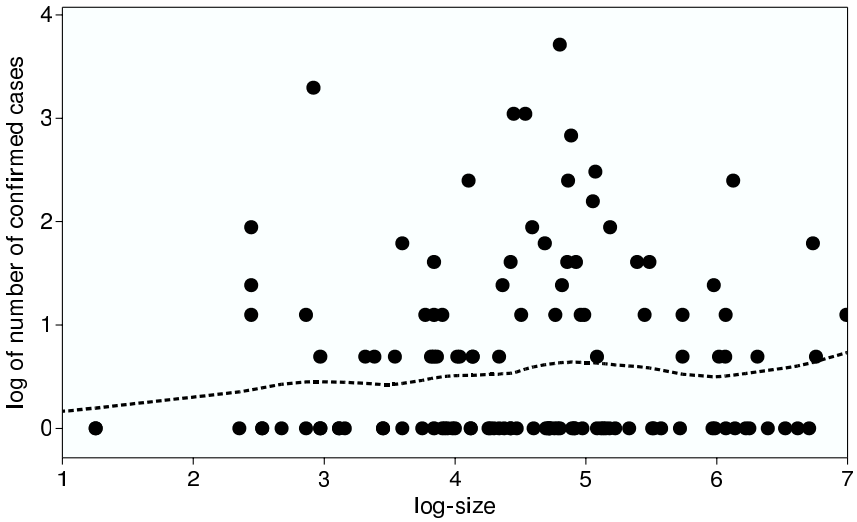

Figure 3. Scatterplot of `log-number of confirmed cases` versus `log-size of holding` in 2002; dotted line corresponds to LOWESS-smoother.

5.3 GENERAL

The present study and results apply to classical scrapie only. Holdings with atypical scrapie cases, in the current regulatory situation, are not placed under any restriction. Hence, atypical cases from the same holding could be detected by more than one source. This would allow the application of multiple-CRC models too. For the application of one-list models we would require the occurrence of more than one case from the same holding. Although there have been holdings with more than one case of atypical scrapie (Del Rio Vilas, personal observation) these are the exception. The lack of enough data on atypical scrapie does not support the application of any of the two CRC approaches as yet. A further complication would be the occurrence of the two diseases in the same holding. This requires cautious case definition and it is the subject of further research.

The low within-holding incidence of scrapie (Gubbins et al. 2005; Del Rio Vilas et al. 2006) supports the application of Chao and Zelterman estimators: most of the holdings appear distributed in the left side of the count distribution. Although this low probability of detecting more cases from the same holding might amount to a problem in multiple-list approaches, stopping any overlapping between sources by depletion of cases and, hence, becoming mutually exclusive (Chang et al. 1999), this is not the case in the one-list models. The severity factor (Chang et al. 1999) however, by which holdings with more cases are more likely to be detected is also applicable here. This again supports the application of methods which allow the incorporation of unobserved heterogeneity as modeled by the Zelterman and Chao estimators. Diseases with higher within-holding occurrence however, would not fit well under the Chao and Zelterman approaches.

A. APPENDIX

A.1 VARIANCE FORMULA FOR THE GENERALIZED ZELTERMAN ESTIMATOR

Note that the uppper summation index n in (3.11) is a random quantity itself. We use the techniques of *conditioning* to develop a variance estimator of (3.11). We have that (see Ross 1985, p. 125)

$$\text{var}(N_Z) = \text{var}_n[E(N_Z|n)] + E_n[\text{var}(N_Z|n)], \tag{A.1}$$

where moments inside the brackets are computed *conditional* upon n and moments outside the bracket refer to the marginal distribution of n. Consider $E(N_Z|n)$ and its estimate

$$\widehat{E(N_Z|n)} = \sum_{i=1}^{n} \frac{1}{w_i} = \sum_{i=1}^{N} \frac{\Delta_i}{w_i},$$

where $w_i = 1 - \exp(-2e^{\hat{\eta}_i})$ and Δ_i is an indicator which is 1 (holding is sampled) with probability w_i and 0 (holding is not sampled) with probability $1 - w_i$. Consequently,

$$\text{var}_n\left(\sum_{i=1}^{N} \frac{\Delta_i}{w_i}\right) = \sum_{i=1}^{N} \text{var}_n(\frac{\Delta_i}{w_i}) = \sum_{i=1}^{N} w_i(1-w_i)/w_i^2 = \sum_{i=1}^{N}(1-w_i)/w_i,$$

for which an unbiased estimator can be provided as

$$\widehat{\text{var}}_n\left(\sum_{i=1}^{N} \frac{\Delta_i}{w_i}\right) = \sum_{i=1}^{N} \Delta_i(1-w_i)/w_i^2 = \sum_{i=1}^{n}(1-w_i)/w_i^2. \tag{A.2}$$

We move on to consider the second term, $E_n[\text{var}(N_Z|n)]$ involved in (A.1). We write $\text{var}(N_Z|n) = \text{var}\left(\sum_{i=1}^{N} \frac{\Delta_i}{w_i}|\Delta_1, \cdots, \Delta_N\right)$, so that

$$\text{var}(N_Z|n) = \text{var}\left(\sum_{i=1}^{n} \frac{1}{w_i}\right).$$

Recall that $w_i = 1 - \exp(v_i)$ and $v_i = -2e^{\hat{\eta}_i}$, so that

$$w_i = w_i(\hat{\beta}) = 1 - \exp(v_i) = 1 - \exp(-2e^{\hat{\eta}_i}) = 1 - \exp(-2e^{\hat{\beta}^T \mathbf{x}_i}).$$

Consequently, $w_i(\hat{\beta})$ and $w_j(\hat{\beta})$ will be *not* independent for $i \neq j$, since both depend on a common $\hat{\beta}$. An application of the multivarate δ-method as done similarly in van der Heijden et al. (2003) or Huggins (1989) provides

$$\left(\sum_i \nabla w_i(\hat{\beta})^T\right) \text{cov}(\hat{\beta}) \left(\sum_i \nabla w_i(\hat{\beta})\right), \tag{A.3}$$

where

$$\nabla w_i(\hat{\beta}) = \frac{(1-w_i)v_i}{w_i^2}\mathbf{x}_i.$$

Note that (A.3) leads to (3.5) in the case that there are no covariates. Whereas we find that (A.3) provides a reasonable approximation, it is sometimes desirable to have an alternative way of computing the variance available. We therefore consider the nonparametric bootstrap as an alternative. Note that we consider the bootstrap only for approximating $\text{var}\left(\sum_{i=1}^{n} \frac{1}{w_i}\right)$ which is a variance conditional upon the n observed units. The application of the nonparamatric bootstrap proceeds as follows:

1. Draw n units $(\delta_1^*, \mathbf{x}_1^*), \ldots, (\delta_n^*, \mathbf{x}_n^*)$ with replacement from the original data $(\delta_1, \mathbf{x}_1)$, $\ldots, (\delta_n, \mathbf{x}_n)$.
2. Compute N_Z^* for the resample.
3. Repeat steps 1 and 2 B times leading to $N_{Z,1}^*, \ldots, N_{Z,B}^*$
4. Compute $\widehat{\text{var}}^*(\hat{N}_Z|\Delta_1, \ldots, \Delta_N) = \frac{1}{B}\sum_{b=1}^{B}(N_{Z,b}^* - \overline{N_Z^*})^2$.

Finally, we use

$$\sum_{i=1}^{n}(1-w_i)/w_i^2 + \widehat{\text{var}}^*(\hat{N}_Z|\Delta_1, \ldots, \Delta_N)$$

as an estimate for the variance of $\hat{N}_Z$.

ACKNOWLEDGMENTS

This work was funded by Great Britain's Department for Environment, Food and Rural Affairs (Defra), research project SE0243.

[Received TKKK. Revised TKK.]

REFERENCES

Anon. (2001), Regulation (EC) No 999/2001 laying down rules for the prevention, control and eradication of certain transmissible spongiform encephalopathies.

Anon. (2003), Commission Regulation (EC) No 1915/2003 amending Annexes VII, VIII and IX to Regulation (EC) No 999/2001 as regards the trade and import of ovine and caprine animals and the measures following the confirmation of transmissible spongiform encephalopathies in bovine, ovine and caprine animals.

Anon (2004), The June Agricultural Census. Available online at: *www.defra.gov.uk/esg/-work_htm/-publications/cs/-farmstats_web/-datamap_links/search_menu.asp*.

Bishop, Y.M.M., Fienberg, S.E., and Holland, P.W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, MA: MIT Press.

Böhning, D., Dietz, E., Kuhnert, R., and Schön, D. (2005), "Mixture Models for Capture–Recapture Count Data," *Statistical Methods & Applications. Journal of the Italian Statistical Society*, 14, 29–43.

Böhning, D., and Kuhnert, R. (2006), "The Equivalence of Truncated Count Mixture Distributions and Mixture of Truncated Count Distributions," *Biometrics*, 62, 1207–1215.

Böhning, D., and Schön D. (2005), "Nonparametric Maximum Likelihood Estimation of the Population Size Based upon the Counting Distribution," *Journal of the Royal Statistical Society*, Series C, Applied Statistics, 54, 721–737.

Böhning, D., Suppawattanabodee, B., Kusolvisitkul, W., and Viwatwongkasem, C. (2004), "Estimating the Number of Drug Users in Bangkok 2001: A Capture–Recapture Approach Using Repeated Entries in one List," *European Journal of Epidemiology*, 19, 1075–1083.

Buschmann, A., Biacabe, A.G., Ziegler, U., Bencsik, A., Madec, J.Y., Erhardt, G., Luhken, G., Baron, T., and Groschup, M.H. (2004a), "Atypical Scrapie Cases in Germany and France are Identified by Discrepant Reaction Patterns in BSE Rapid Tests," *Journal of Virological Methods*, 117, 27–36.

Buschmann, A., Luhken, G., Schultz, J., Erhardt, G., and Groschup, M.H. (2004b), "Neuronal Accumulation of Abnormal Prion Protein in Sheep Carrying a Scrapie-Resistant Genotype (PrPARR/ARR)," *Journal of General Virology*, 85, 2727–2733.

Chang, Y-F., La Porte, R.E., Aaron, D.J., and Songer, T.J. (1999), "The Importance of Source Selection and Pilot Study in the Capture–Recapture Application," *Journal of Clinical Epidemiology*, 52, 927–928.

Chao, A. (1984), "Nonparametric Estimation of the Number of the Classes in a Population," *Scandinavian Journal of Statistics*, 11, 265–270.

——— (1987), "Estimating the Population Size for Capture–Recapture Data with Unequal Capture Probabilities," *Biometrics*, 43, 783–791.

Chao, A., and Lee, S.-M. (1992), "Estimating the Number of Classes via Sample Coverage," *Journal of the American Statistical Association*, 87, 210–217.

Chao, A., and Shen, T.-J. (2006), User's Guide for Program SPADE. Available online at: *http://chao.stat.nthu.edu.tw/softwareCE.html*.

Darroch, J.N., and Ratcliff, D. (1980), "A Note on Capture–Recapture Estimation," *Biometrics*, 36, 149–153.

Del Rio Vilas, V.J., Guitian, J., Pfeiffer, D.U., and Wilesmith, J.W. (2006), "Analysis of Data from the Passive Surveillance of Scrapie in Great Britain Between 1993 and 2002," *Veterinary Record*, 159, 799–804.

Del Rio Vilas, V.J., Sayers, R., Sivam, K., Pfeiffer, D.U., Guitian, J., and Wilesmith, J.W. (2005), "A Case Study of Capture–Recapture Methodology using Scrapie Surveillance Data in Great Britain," *Preventive Veterinary Medicine*, 67, 303–317.

Everest, S.J, Thorne, L., Barnicle, D.A., Edwards, J.C., Elliot, H., Jackman, R., and Hope, J. (2006), "Atypical Prion Protein in Sheep Brain Collected during the British Scrapie-Surveillance Programme," *Journal of General Virology*, 87, 471–477.

Foster, J.D., Parhham, D., Chong, A., Goldman, W., and Hunter, N. (2001), "Clinical Signs, Histopathology and Genetics of Experimental Transmission of BSE and Natural Scrapie to Sheep and Goats," *Veterinary Record*, 148, 165–171.

Gubbins, S. (2005), "A Modelling Framework for the Spread of Scrapie Between Sheep Flocks in Great Britain," *Preventive Veterinary Medicine*, 67, 143–156.

Hagenaars, T.J., Donnelly, C.A., and Ferguson, N.M. (2006), "Epidemiological Analysis of Data for Scrapie in Great Britain," *Epidemiology and Infection*, 134, 359–367.

Hay, G., and Smit, F. (2003), "Estimating the Number of Drug Injectors from Needle Exchange Data," *Addiction Research and Theory*, 11, 235–243.

Hoinville, L.J., Hoek, A., Gravenor, M.B., and Mclean, A.R. (2000), "Descriptive Epidemiology of Scrapie in Great Britain: Results of a Postal Survey," *Veterinary Record*, 146, 455–461.

Huggins, R.M. (1989), "On the Statistical Analysis of Capture Experiments," *Biometrika*, 76, 133–140.

Hunter, N. (2003), "Scrapie and Experimental BSE in Sheep," *British Medical Bulletin*, 66, 171–183.

Kuncheva, L., Del Rio Vilas, V.J., and Rodriguez, J.J. (2006), "Diagnosing Scrapie in Sheep: A Classification Experiment," *Computers in Biology and Medicine*, (in press).

McKendrick, A.G. (1926), "Application of Mathematics to Medical Problems," *Proceedings of the Edinburgh Mathematical Society* 44, 98–130.

Mclean, A.R., Hoek, A.R., Hoinville, L.J., and Gravenor, M.B. (1999), "Scrapie Transmission in Britain: A Recipe for a Mathematical Model," *Proceedings of the Royal Society of London*, Ser. B, 266, 2531–2538.

Moore, P.G. (1952), "The Estimation of the Poisson Parameter from a Truncated Distribution," *Biometrika*, 39, 247–251.

Roberts, J.M., and Brewer, D.D. (2006), "Estimating the Prevalence of Male Clients of Prostitute Women in Vancouver with a Simple Capture–Recapture Method," *Journal of the Royal Statistical Society*, Ser. A, 169, 745–756.

Ross, S.M. (1985), *Introduction to Probability Models* (3rd ed.), Orlando: Academic Press.

Saunders, G.C., Cawthraw, S., Mountjoy, S.J., Hope, J., and Windl, O. (2006), "PrP Genotypes of Atypical Scrapie Cases in Great Britain," *Journal of General Virology*, 87, 3141–3149.

Sivam, K., Baylis, M., Gravenor, M.B., Gubbins, S., and Wilesmith, J.W. (2003), "Occurrence of Scrapie in GB: Results of a Postal Survey in 2002," *Veterinary Record*, 153, 782–783.

van der Heijden, P.G.M., Bustami, R., Cruy, M., Engbersen, G., and van Houwelingen, H. C. (2003), "Point and Interval Estimation of the Population Size Using the Truncated Poisson Regression Model," *Statistical Modelling—An International Journal*, 3, 305–322.

Wilson, R.M., and Collins, M.F. (1992), "Capture–Recapture Estimation with Samples of Size one Using Frequency Data," *Biometrika*, 79, 543–553.

Zelterman, D. (1988), "Robust Estimation in Truncated Discrete Distributions with Applications to Capture–Recapture Experiments," *Journal of Statistical Planning and Inference*, 18, 225–237.