Agence canadienne d'inspection des aliments | Canadian Food Inspection Agency

Capture-recapture methods to estimate commercial turkey farm population in Canada

Farouk El Allaki[1a], Jette Christensen[1b], and André Vallières[1a]
[1]Canadian Food Inspection Agency, [a]St-Hyacinthe, [b]Charlottetown, Canada

BACKGROUND

Capture-recapture analysis (CR) are methods to estimate the population size from a series of population samples. They were first used by Pierre Simon Laplace to estimate the population of France in 1786; later used to monitor the abundance of wildlife and increasingly applied in public health[1]. In surveillance, the aim of CR methods is to estimate the number of individuals with the characteristics of interest that are not detected by any of the surveillance sources in place[2].

The basic idea behind these CR methods is that animals are captured, marked and returned to the population. The estimate of population size is made based on the relative numbers of marked and unmarked animals in subsequent samples.

OBJECTIVE

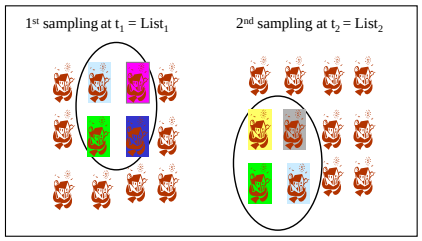
The objectives of this study were to:

- estimate the turkey farm population size in three areas of Canada: British Columbia (BC), Ontario (ON) and the rest of Canada using the 2008-2011 Canadian Notifiable Avian influenza Surveillance System (CanNAISS) data.
- compare the turkey farm population size estimates using two different capture-recapture methods: (i) Log-linear modelling and (ii) Bailey's binomial model estimation with population data from 2007 (N_{2007}).

MATERIALS AND METHODS

We used closed population CR methods to estimate the Turkey farm population size by area. A ''list'' referred to all farms sampled in CanNAISS during a given year (2008, 2009, 2010, 2011). The Bailey's farm population size estimates (N_B) and their 95% confidence intervals were calculated in R using the 'mrclosed' function in 'FSA' package[3,4]. The Log-linear model farm population size estimates (N_{LLM}) were calculated in SAS[5].

Figure 1: Two-list capture-recapture analysis - Bailey's estimator for sampling with replacement[3]



$N_B = [n_1(n_2+1)]/(m_2+1)$, where n_1 = number of sampled and marked farms in the first sample ($list_1$) that were returned to the population; n_2 = number of farms in the second sample ($list_2$); m_2 = number of recapture in the second sample ($list_2$)

Log-linear model: 4-list capture-recapture analysis

$$\text{Log}\,E(m_{ijkl}) = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{4(l)} + u_{12(ij)} + u_{13(ik)} + u_{14(il)} + u_{23(jk)} + u_{24(jl)} + u_{34(kl)} + u_{123(ijk)} + u_{124(ijl)} + u_{134(ikl)} + u_{234(jkl)} + u_{1234(ijkl)}$$

m_{ijkl} = Expected number of farms for the cell $ijkl$

u_1, u_2, u_3, u_4 = Four "main effects"–log odds against appearing on each list u_1, u_2, u_3, and u_4 (e.g. $u_{1(i=0)} = 0$, $u_{1(i=1)} = u_1$)

u_{12}, u_{123}, u_{1234} = Examples of interaction terms

Goodness of fit of the model[6] : Deviance (G^2)

Model selection[6]: Akaike Information Criterion (AIC)

AIC= G^2 - 2(df)

Best log-linear model was the one with the lowest AIC

Table 1: Turkey farm population size for three areas in Canada — Bailey's estimates & 2007 population data

Area	Compared lists (S)	n_1	n_2	m_2	Population size estimate N_B (95% CI)	N_{2007}
BC	S_{2010},S_{2011}	30	31	22	42 (36-55)	60
	S_{2009},S_{2010}	37	41	28	54 (46-68)	60
	S_{2008},S_{2009}	20	50	18	54 (39-78)	60
ON	S_{2010},S_{2011}	100	151	59	253 (211-311)	299
	S_{2009},S_{2010}	122	105	44	216 (186-259)	299
	S_{2008},S_{2009}	152	127	80	240 (214-278)	299
Rest of Canada	S_{2010},S_{2011}	144	236	108	313 (275-362)	332
	S_{2009},S_{2010}	114	162	64	286 (240-348)	332
	S_{2008},S_{2009}	150	115	68	252 (221-297)	332

Figure 2: Absolute difference between farm population size estimates and 2007 population data

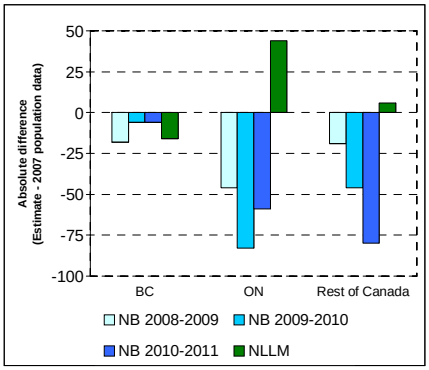


Figure 3: Absolute difference (estimate-2007 population data), sample size per analysis, and population size (2007 data) by area and analysis

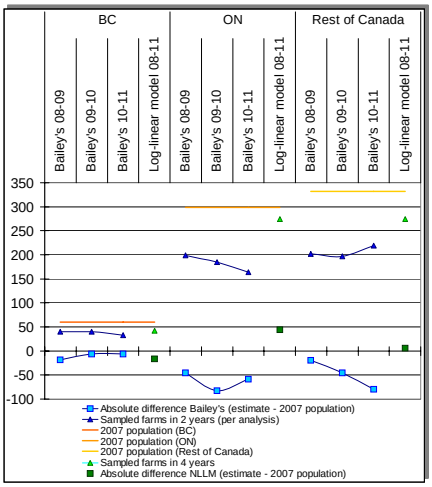


Table 2: Turkey farm population size Log-linear model estimation

Areas	Model*	Deviance	df	p value	AIC	Population size estimate (95% CI**)	N_{2007}
BC	S_1,S_2,S_3,S_4, S_3S_4	15,5562	9	0,0767	-2,444	44 (22-91)	60
ON	S_1,S_2,S_3,S_4, S_1S_4,$S_1S_2S_4$, $S_1S_3S_4$	1,0807	3	0,7817	-4,919	343 (215-572)	299
Rest of Canada	S_1,S_2,S_3,S_4, S_1S_2, S_1S_3, S_1S_4, S_3S_4, $S_1S_2S_3$	9,4367	4	0,0511	1,4367	338 (219-530)	332

*Only best models were presented **Profile Likelihood Method

RESULTS

Bailey's estimates:

Bailey's point estimates were lower than the 2007 population data in the 3 areas (Table 1 and Figure 2).

Four out of nine Bailey's confidence intervals did not include the 2007 farm population size (Table 1).

The bailey's point estimates were not stable from one analysis to another. For example, in two analysis (BC 2008-2009 and BC 2009-2010 the same number of farms were sampled (40) but the absolute differences were -18 and -6, respectively (Figure 3). This may be explained by dependencies between sampling lists.

The Bailey's estimates for BC were close to the 2007 population data. BC differed from the other two areas because the population size was about 60 farms compared to about 300 farms in the other areas (Figure 3) even though a smaller proportion of farms were sampled in BC (70%) compared to Ontario (92%) and the rest of Canada (83%).

Log-linear model estimates:

The Log-linear modelling approach yielded estimates closer to the 2007 population data than the Bailey's analysis for turkey farm populations in Ontario and the rest of Canada when the farm population size was about 300 farms and a larger proportion of farms were sampled (Figure 3).

All estimated confidence intervals contained the reference population size.

The dependencies between two sampling lists were assessed and accounted for using the log-linear modelling (Table 2).

CONCLUSION

- Both Bailey's analysis and the log-linear models indicated dependencies between annual sampling lists from CanNAISS.
- Bailey's analysis performed well in the small population in BC (60 farms).
- The log-linear models to estimate the turkey farm population size in Canada performed better than Bailey's analysis in the larger (300 farm) populations where a large proportion of farms were sampled (>80%).
- This study demonstrated that capture-recapture methods may be applied to estimate farm population size based on ongoing surveillance data.

REFERENCES

[1] Pollock, Nichols, Brownie, Hines. 1990. Statistical inference for capture-recapture experiments. Wildl. Monogr. 107. 97pp.
[2] Del Vilas, Pfeiffer. 2010, The evaluation of bias in scrapie surveillance: a review. Vet J 185, 259-264.
[3] Bailey. 1951, On estimating the size of mobile populations from recapture data. Biometrika 38, 293-306.
[4] R Copyright (C) 2011 The R Foundation for statistical Computing).
[5] SAS version 9.1 (Cary, NC, U.S.A.)
[6] Hook, Regal. 1997, Validity of methods for model selection, weighting for model uncertainty, and small sample adjustment in capture-recapture estimation. Am J Epi. 145, 1138-1144.