Which farmers to ask?
Methods for identifying timely and reliable respondents for syndromic surveillance of farmers' observations

CN PFEIFFER[1]*, SM FIRESTONE[2], AJD CAMPBELL[1], JWA LARSEN[1], MA STEVENSON[2]
[1]Mackinnon Project, The University of Melbourne, Werribee, Australia; [2]Asia-Pacific Centre for Animal Health, The University of Melbourne,Parkville, Australia

*caitlin.pfeiffer@unimelb.edu.au

Abstract
Syndromic surveillance of farmers' disease observations can improve animal health data capture from extensive farming systems. It is rarely feasible to recruit a random sample of farmers to provide observational reports. Therefore, a targeted sample of farmers aligned with surveillance objectives should be invited to participate. Farmers who will report reliably are desirable participants; to target these farmers, characteristics associated with reliable participation must be known at recruitment. We describe survival analysis and classification and regression tree analysis of a pilot farmer syndromic surveillance system in Victoria, Australia, to identify farm level factors associated with reliable participation. Sheep and beef farms with high stocking rates and multi-species production were less reliable than smaller, simpler sheep or beef farms. Once farm characteristics associated with reliable participation are known, they can be incorporated into surveillance system design in accordance with the objectives of the system.

Keywords: *syndromic surveillance, evaluation, animal health, survival analysis, regression tree*

Introduction
Maintaining animal health surveillance satisfactory to regulatory and industry requirements is a global challenge. In countries such as Australia, extensive pasture-based livestock production and diminishing veterinary contact with farmers can limit the effectiveness of traditional passive surveillance (1). Syndromic surveillance can be used to capture additional animal health data from farms through regular structured reporting of farmers' disease observations (2). However, the timeliness of responses and overall response rate must be acceptable for such a system to form a sustainable part of surveillance strategy.

While there have been several examples of syndromic surveillance used to collect farmers' observations, evaluation of such programs is infrequently reported. Evaluation provides specific information for internal system improvement, and findings may be generalised to inform the development of future surveillance programs. Appropriate methods to evaluate the timeliness of farmer syndrome reporting have not been previously reported.

In this paper, we present an evaluation of timeliness and response behaviour in a syndromic surveillance system collecting farmers' disease observations in Victoria, Australia. The objective was to assess whether easily determined farm characteristics were associated with timeliness of reporting and overall response rate.

Methods
The Local Area Network for Disease Information (LANDI) collects monthly disease reports from sheep and beef farms in Victoria, Australia. All disease observations on participant farms each month are described in a single report, using 10 simplified syndromic categories (2). An online survey interface (using REDCap software (3) hosted at the University of Melbourne) is used to collect disease reports, with invitations to provide reports sent to participants via email with an embedded web link each month. Where a report is not provided in response to the initial invitation, up to four reminder emails are sent at approximately 10 day intervals.

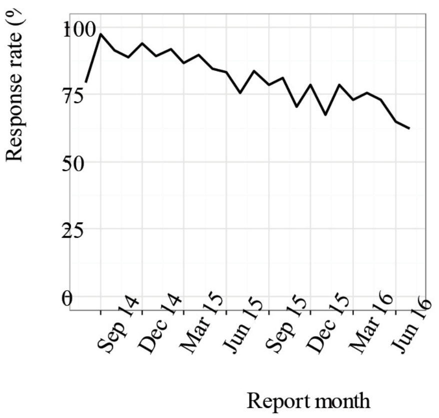
Reports over the first two years of LANDI operation from August 2014 to July 2016, from 39 sheep and/or beef farms in Victoria, Australia were evaluated. Data included the date of each email, total number of emails per report, whether the salutation of the email was personalised ('Dear John') or generic ('Dear participant'), whether a report was provided, and farm characteristic data including the species kept, number of sheep, number of cattle, total grazing area, and approximate winter stocking rate (measured as the number of dry sheep equivalents per hectare, DSE/ha). Survival analysis was used to quantify the timeliness of responses, using time from email invitation to report completion as time-to-event data and non-response as the "survival" outcome. In months where no report was received, the report was considered censored at 90 days after the initial invitation for the report; this occurred in 171/872 cases. Difference between survival curves was tested using both the log-rank and Peto-Peto modification of the Gehan-Wilcoxon tests, with non-significance designated at $p>0.05$. A Cox proportional hazards model was then fitted to the data using participant identifier as a cluster term to account for repeated outcomes over time and derive robust sandwich variance estimators. Classification and regression tree (CART) analysis investigated the major farm characteristics influencing non-response. Analyses were performed with R statistical software v3.1.1 using the *survival* and *rpart* packages.

Results

The 39 participating farms included 32 with sheep (median flock size 2908, range 202 to 17490 sheep), and 23 with cattle (median herd size 141, range 30 to 890 cattle). The farms varied in size with a median of 360 hectares (range 32 to 2900ha) and had a median estimated winter stocking rate of 10 DSE/ha (range 3 to 22 DSE/ha). These farms were recruited into the LANDI project by convenience but are considered typical of the range of sheep and beef farms in Victoria, Australia.

The overall response rate was 80%, with a mean monthly response rate of 81% (monthly response pattern in Figure 1). A negative trend can be appreciated, driven by an intermittent then consistent non-response from nine farmers over the duration of the study.

Figure 1. Monthly online disease report response rate for 39 sheep and beef farmers in Victoria, Australia.

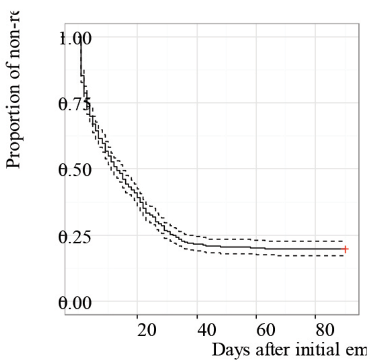


Survival analysis quantified the influence of farm-level factors on the time to reporting after the initial email request for a report. Responses were received steadily over the first three weeks after the first email, with a median of 13 days (95% CI 11-15 days) between initial email and report completion.

Figure 2 shows this in inverse as the proportion of non-responders remaining over time, with few reports received beyond 40 days after the initial invitation; this reflects the cessation of reminders for that month's report. The first email request yielded a response in 57% (398/701) of all cases where a report was received; in the remaining cases, up to four reminder emails were required to obtain a report. Emails with personalised salutations were not significantly more likely to elicit a response (Fisher's exact and Chi-squared tests p >0.05). Timeliness of responses did not vary significantly between months.

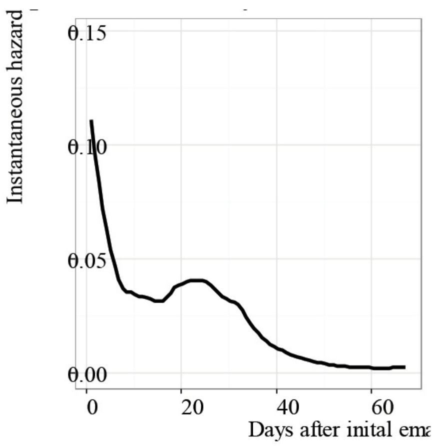
The association of disease status with non-response was not determined, because disease status could not be determined when non-response occurred. However, 69% (487/701) of responses had disease to report, with no significant difference in the Kaplan-Meier estimated curves for disease positive and negative reports. Farms keeping both sheep and cattle were significantly slower to respond and more likely to have cases of non-response, compared to farms with only one species (p<0.001). These farms were overrepresented as seven of the group of nine farmers who became consistent non-responders. Farms with a grazing area above the population median (>360ha) took significantly longer (p<0.001) to provide reports, with a median of 21 days (95% CI 18 to 22) compared to a median of seven days (95% CI 5 to 9) for farms smaller than 360ha. Timeliness of responses was also significantly slower for those farms with stocking rates in the top quartile of the population, with no difference between the remaining three quartiles.

Figure 2. Kaplan-Meier estimated non-response over time for monthly online disease reports submitted by sheep and beef farmers in Victoria, Australia.



The instantaneous hazard function (Figure 3) shows the probability of response was highest in the days immediately after the first email, with a second peak at around 22 days due to the second reminder email, which was sent approximately 20 days after the initial email.

Figure 3. Smoothed instantaneous hazard function for response to monthly online disease report requests.



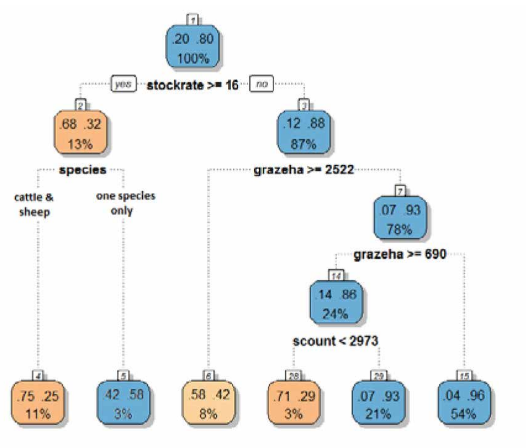
The Cox proportional hazards model (Table 1), after accounting for clustering of outcomes at the farm level, showed that the interaction term for farms keeping both species at high stocking rates was significant and of moderate magnitude, indicating these farms were poorer respondents. Keeping larger numbers of cattle, compared with small numbers or no cattle was associated with improved response times. Farms with larger grazing areas were less timely in providing their reports. Both these latter terms had hazard ratios of ≈1 and likely have a negligible effect.

Table 1. Regression coefficients and their (robust) standard errors from a Cox proportional hazards model for non-response in monthly online disease reporting by farmers in Victoria, Australia. The hazard ratio represents the relative probability that a report will be provided.

	Coef.	Robust SE	p-value	Hazard ratio	95% CI
Sheep only	Ref.			1.00	
Cattle only	0.3458	0.5788	0.550	1.41	0.45 to 4.39
Both sheep and cattle	0.7819	0.4992	0.117	2.18	0.82 to 5.81
Stocking rate	-0.0097	0.0482	0.841	0.99	0.90 to 1.08
Cattle numbers	0.0025	0.0006	<0.001*	1.01	1.01 to 1.01
Grazing area	-0.0006	0.0001	<0.001*	0.99	0.99 to 0.99
Sheep x stocking rate	Ref.			1.00	
Cattle x stocking rate	-0.0917	0.0572	0.109	0.91	0.81 to 1.02
Both sheep and cattle x stocking rate	-0.1508	0.0546	0.006*	0.86	0.77 to 0.95

* indicates significance of p-value.

Figure 4. Classification tree for farms participating in monthly online disease reporting. Numbers in node boxes indicate: proportion of non-responses, proportion of responses, percentage of total population. Light grey = majority non-response, dark grey = majority response.



These effects were confirmed by CART analysis (Figure 4) using response or non-response each month as an outcome. The strongest predictor of response was the stocking rate of the farm, with farms of ≥16DSE/ha less likely to provide a response. Within those high stocking rate farms, those with both sheep and cattle were less likely to respond than those with sheep or cattle only. In the farms with <16DSE/ha stocking rates, farms with effective grazing area ≥2522ha were less likely to respond. Within smaller farms, those with effective grazing area <690ha were the most likely to respond.

Discussion

Survival analysis and CART analysis provide insights into the timeliness and response behaviour of these farmers providing monthly online syndromic disease reports. Writing a personalised email salutation was no more effective than a generic greeting, and is non-essential for systems where personalised greetings cannot be generated automatically. Disease status did not influence the timeliness of reports, suggesting that the extra effort to provide disease occurrence data when disease had occurred did not affect timely reporting. While the 13 day median response time suggests this system is not suited to exotic disease detection, participation in disease reporting may increase farmers' disease awareness and the probability they will report unusual disease by other methods.

Farms with high stocking rates, especially those keeping multiple species, were poorer responders. In Australia, these farms are typically pushing the productivity of their land and manage their farms rigorously. These farmers may have less time and inclination for extra duties such as disease reporting. However, these farms may also be the most valuable for surveillance, due to an increased risk of disease. A balance between those most likely to respond (perhaps useful for endemic disease monitoring), and those who are at highest risk for disease outbreaks (such as emerging diseases) is therefore desirable. The objectives of the surveillance system must inform targeted recruitment.

Univariate and two multivariate methods have been demonstrated in this analysis to determine farm-level explanatory variables to target recruiting for future similar surveillance systems in comparable farming contexts. Effective syndromic surveillance of farmers' observations is challenging to implement, but can play an important role in wider surveillance strategies when design takes advantage of all available evidence.

Acknowledgements

The contribution of the 39 participating farms is gratefully acknowledged. The LANDI project is funded by the Victorian Sheep and Goat Compensation Fund and the Victorian Cattle Compensation Fund, administered by Agriculture Victoria.

References

1. **Frawley PT.** *Report to Commonwealth of Australia*, 2003
2. **Pfeiffer CP *et al.*** *Proc Aust Sheep Vet,* 2016
3. **Harris PA *et al.*** *J Biomed Inform*, 2009