Preventive Veterinary Medicine 147 (2017) 124–131

Contents lists available at ScienceDirect

Preventive Veterinary Medicine

journal homepage: www.elsevier.com/locate/prevetmed

Marginal structural Cox model to estimate the causal effect of clinical mastitis on Québec dairy cow culling risk

Denis Haine[a,*], Hector Delgado[b], Roger Cue[b], Asheber Sewalem[c], Kevin Wade[b], René Lacroix[b,d], Daniel Lefebvre[b,d], Julie Arsenault[a], Émile Bouchard[a], Jocelyn Dubuc[a]

[a] Faculté de médecine vétérinaire, Université de Montréal, 3200 rue Sicotte, St-Hyacinthe, Québec J2S 2M2, Canada
[b] Department of Animal Science, McGill University, 21111 Lakeshore Road, Ste-Anne-de-Bellevue, Québec H9X 3V9, Canada
[c] Guelph Food Research Centre, Agriculture and Agri-Food Canada, 93 Stone Road West, Guelph, Ontario N1G 5C9, Canada
[d] Valacta Inc., 555 boul. des Anciens-Combattants, Ste-Anne-de-Bellevue, Québec H9X 3R4, Canada

ARTICLE INFO

Keywords:
Culling
Dairy cow
Bias
Confounding
Time-varying
Marginalstructural model
Causal inference
Survival

ABSTRACT

Health disorders, such as milk fever, displaced abomasum, or retained placenta, as well as poor reproductive performance, are known risk factors for culling in dairy cows. Clinical mastitis (CM) is one of the most influential culling risk factors. However the culling decision could be based either on the disease status or on the current milk yield, milk production being a significant confounder when modelling dairy cow culling risk. But milk yield (and somatic cell count) are time-varying confounders, which are also affected by prior CM and therefore lie on the causal pathway between the exposure of interest, CM, and the outcome, culling. Including these time-varying confounders could result in biased estimates. A marginal structural model (MSM) is a statistical technique allowing estimation of the causal effect of a time-varying exposure in the presence of time-varying covariates without conditioning on these covariates. The objective of this paper is to estimate the causal effect on culling of CM occurring between calving and 120 days in milk, using MSM to control for such time-varying confounders affected by previous exposure. A retrospective longitudinal study was conducted on data from dairy herds in the Province of Québec, Canada, by extracting health information events from the dairy herd health management software used by most Québec dairy producers and their veterinarians. The data were extracted for all lactations starting between January 1st and December 31st, 2010. A total of 3952 heifers and 8724 cows from 261 herds met the inclusion criteria and were used in the analysis.

The estimated CM causal hazard ratios were 1.96 [1.57–2.44] and 1.47 [1.28–1.69] for heifers and cows, respectively, and as long as causal assumptions hold. Our findings confirm that CM was a risk factor for culling, but with a reduced effect compared to previous studies, which did not properly control for the presence of time-dependent confounders such as milk yield and somatic cell count. Cows experienced a lower risk for CM, with milk production having more influence on culling risk in cows than heifers.

1. Introduction

Health disorders, such as milk fever, displaced abomasum, or retained placenta (Rajala-Schultz and Gröhn, 1999a,b,c; Beaudeau et al., 2000), as well as poor reproductive performance (Schneider et al., 2007; De Vries et al., 2010), are known risk factors for culling in dairy cows. Among these risk factors, one of the most influential is clinical mastitis (CM; Gröhn et al., 1998; Rajala-Schultz and Gröhn, 1999a,b,c; Schneider et al., 2007); with the risk between mastitis and culling being time-dependent (Gröhn et al., 1997, 1998).

However, the culling decision could be based either on the disease status of the cow or on its current milk yield, milk production being a significant confounder when modelling dairy cow culling risk. High producing cows are at greater risk of mastitis (Schukken et al., 1991; Waage et al., 1998; Barnouin et al., 2005; O'Reilly et al., 2006), and a lower milk production compared to herd mates has a significant effect on culling decisions (Beaudeau et al., 1994; Rajala-Schultz and Gröhn, 1999a,b,c; Hadley et al., 2006). Moreover, cows that had an episode of CM are at greater risk for occurrence of other CM episodes later during their lactation (Lam et al., 1997; Zadoks et al., 2001). Similarly, a high somatic cell count (SCC) is a risk factor for mastitis as well as for culling (Caraviello et al., 2005; Sewalem et al., 2006; Steeneveld et al., 2008). The correct estimation of the effect of mastitis on culling requires the inclusion of milk yield (and SCC) in the modelling strategy. However,

* Corresponding author.
E-mail address: denis.haine@umontreal.ca (D. Haine).

http://dx.doi.org/10.1016/j.prevetmed.2017.08.017
Received 23 December 2016; Received in revised form 18 August 2017; Accepted 19 August 2017

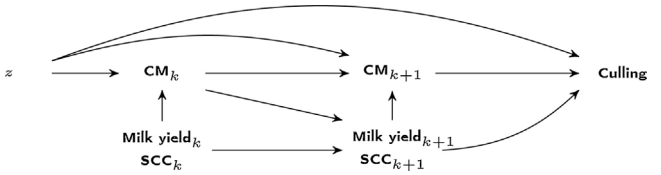

Fig. 1. Directed acyclic graph (DAG) for the effect of clinical mastitis on culling, with time points *k*. *z* is a vector of baseline covariates. CM, clinical mastitis; SCC, somatic cell count.

milk yield and SCC are time-dependent (or time-varying) confounders, which are also affected by prior CM (Rajala-Schultz et al., 1999; Seegers et al., 2003), i.e. intermediate covariates. Therefore these covariates lie on the causal pathway between the exposure of interest, CM, and the outcome, while at the same time being risk factors for culling, as depicted in the directed acyclic graph (DAG) in Fig. 1. Adjusting for variables that are confounders but also affected by prior exposure gives biased estimates of the true or 'causal' total effect. Failing to adjust for milk production and SCC would result in a biased effect estimate, yet adjustment for those variables would also result in biased estimates (Hernán et al., 2004; Cole et al., 2010). This methodological problem has been well described by Robins et al. (2000), Hernán et al. (2004), and Cole and Hernan (2008).

Marginal structural Cox models (MSM) provide the marginal causal relation between a time-varying exposure and a survival outcome (e.g. time to culling), controlling for time-varying confounders without conditioning on those variables (Robins et al., 2000; Hernán et al., 2000; Cole and Hernan, 2008). The regression model relates the exposure history up to time *t* to the counterfactual outcome at time *t*. Propensity scores that estimate the probability of a given level factor (e.g. CM) given measured confounders, are used within a MSM to weight subjects in order to create balanced groups of cows based upon the confounders used in the construction of the scores. An introduction to MSM methodology was described in Martin (2014). The weighting allows the construction, for a risk set at time *t*, of a 'pseudo-population' in which the time-varying confounders no longer predict CM at *t*, i.e. are no longer confounders, and the causal association between CM and culling is the same as in the original population (Hernán et al., 2000). Therefore the estimation of the unconfounded association between the exposure and outcome is now allowed without conditioning on the covariate in the regression model (Robins et al., 2000).

The issue of the direct and indirect effects of milk yield on culling risk due to mastitis was already raised by hn et al. (1997, 1998); hn et al. (1997, 1998). But biases due to time-varying confounders were not identified at that time and have not yet been properly addressed. The objective of this paper is to estimate the causal effect on culling of the time-dependent exposure CM, occurring between calving and 120 days in milk (DIM), by using a marginal structural model (Robins, 1999; Robins et al., 2000) to control for such time-varying confounders affected by previous exposure.

2. Materials and methods

2.1. Dataset

A retrospective longitudinal study was conducted on data from dairy herds in the Province of Québec, Canada, by extracting health information events from *DSA Laitier* (DSAHR Inc., Saint-Hyacinthe, QC, Canada), the dairy herd health management software used by more than half of Québec's producers and their veterinarians. This program uses clearly defined health definitions, ensuring that producers and veterinarians record the same health conditions, using the same definitions. Veterinarian enters health conditions into the herd DSA database, as well as producers for which data are then reviewed by their veterinarian at the herd visit. All information is transferred into the

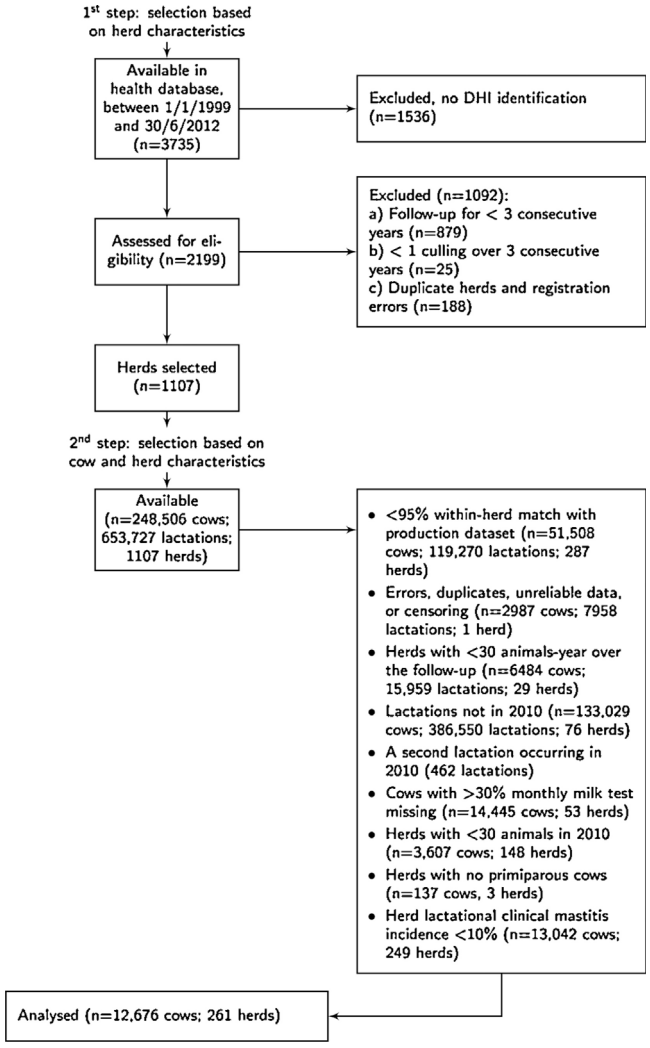

Fig. 2. Flowchart of herds and cows selection.

centralized DSA database by the herd veterinarian, which is then validated. Vets are therefore closely involved in the diagnosis of the disease conditions, as well as their recording and reporting. A purposive sample was created by extracting data for all lactations starting between January 1st, 2001 and December 31st, 2010 (249,536 cows from 3735 herds), keeping herds that had a minimum of three consecutive years of follow-up with *DSA Laitier* and for which at least one culling was recorded over this follow-up (see flowchart in Fig. 2). From this dataset, we extracted data for all lactations starting between January 1st and December 31st, 2010. If a cow had two lactations during the year 2010, one of the two lactations was randomly kept. Production data were obtained from the unique Québec dairy herd improvement (DHI) service provider (Valacta, Sainte-Anne-de-Bellevue, QC, Canada). Health and production data were matched based on herd- and cow-level identification. If not successful, further matching was tried, based on birth date, calving dates, and health and production history. Only herds for which at least 95% of the lactations from the health dataset could be matched with data from the production dataset were kept (42,809 cows from 714 herds). Herds with less than 30 animals, for which more than 30% of the DHI monthly tests were missing, and with a lactational cumulative incidence for CM in 2010 that was less than 10%, were removed. Cows with calving intervals, or an interval between the last calving and the end of data, longer than 580 days were censored at their last calving date. If this censoring was at their first calving date, the observation was dropped. Cows leaving their herd on their calving date were assigned one day of follow-up.

The primary outcome, culling, was defined as a cow being removed from the herd, i.e. due to death, sold to an other herd, or sent for slaughter (Fetrow et al., 2006). A new CM case was considered after a period of 7 days following a preceding CM case for that cow. Using data from *DSA Laitier*, we ascertained the exposure status (CM) for every primiparous and multiparous cow in terms of binary indicators in each monthly interval defined by the DHI monthly test, up to 120 DIM. The following potential baseline confounders were available for analysis: parity (1, 2, 3, 4+), age at first calving, pregnancy status, and occurrence of the following diseases: displaced abomasum, milk fever, and retained placenta. Season (January to July and August to December) was introduced as a time-dependant variable. Québec dairy producers are receiving incentives to produce a more from August to December (i.e. they can go over their quota). For multiparous cows, the following covariates were also retrieved from the previous lactation: occurrence of any CM case, real 305-day milk production, and 305-day SCC geometric mean. These last two covariates were standardized (*z*-score) for each parity strata (primiparous, parities 2, 3, and 4+) within each herd, and then categorized as 3 strata: average (-1 SD $<$ variable $<$ 1 SD; reference), low (≤ -1 SD), high (≥ 1 SD). Missing values were imputed using multivariate imputation by chained equation with the R package MICE (Van Buuren and Groothuis-Oudshoorn, 2011), using default settings. The time-varying confounders monthly milk production and SCC (on a log-scale) were also standardized by test, parity, and herd, and the same categories as above were created.

A CM variable was defined based on udder observations made by either the dairy producer or the farm veterinarian in which abnormalities of the udder and/or secretion were readily observable. Severity can vary but went from changes in milk, such as flakes, clots, and watery appearance, to acute mastitis, with a sudden onset, redness, swelling, hardness, pain, grossly abnormal milk, and reduced milk yield.

2.2. Data analysis

Age at first calving was missing for 8.6% of the cows (parity 2 and over). The missing values were imputed using multivariate imputation by chained equation as described above and assuming age at first calving was missing at random. Logistic regression between missingness of age at first calving and other variables revealed a single significant association, with parity, i.e. older cows were more likely to have missing information for their age at first calving.

Two time-varying confounders (monthly SCC and milk production) were confirmed acting as confounders and mediators, i.e. that they were longitudinally associated with later CM case, were predicted by CM, and were associated with culling independently of CM. Then we conducted two separate sets of analyses for primiparous and multiparous cows. First, time-varying confounders were not included and only CM was considered time-varying at the monthly intervals defined by the DHI monthly tests. Both the exposure and the covariates were assumed to be constant during these intervals. The crude and adjusted risks of culling associated with CM were estimated using extended Cox models for time-dependent variables. As the aim is to define a marginal, population average effect, non-clustered time-to-event Cox model was chosen over stratified or random effect model (Glidden and Vittinghoff, 2004; Munda and Legrand, 2014).

Second, time-varying confounders were added to the framework of the MSMs based on the method described by Hernán et al. (2000). Marginal structural models are constructed as a four-step modelling strategy. First, the propensity scores for exposure and censoring were estimated by separate logistic regression models with herd as a random effect (He, 2014). Second, the inverse probability of the treatment/exposure weights (IPTWs) were used to create a weighted sample in which the exposure is unconfounded by the covariates (Cole and Hernan, 2008). The IPTWs are the inverse of the probability of being exposed at each monthly test, i.e. the propensity score. Large weights lead to large standard error and noisier estimates of causal effects, as large weights indicate near violations of the positivity assumption (Cole and Hernan, 2008; Lee et al., 2011). Therefore weights were stabilized in order to reduce their variability and the standard errors of the estimated hazard ratios, by multiplying the IPTW by the marginal probability of being actually exposed. With the same objective, weights were also trimmed at the 1st and 99th percentiles (Lee et al., 2011; Austin, 2014). The exposure and censoring weights were then multiplied to get the overall weights in each one-month interval. Third, balance diagnostics of the weight were conducted by checking their range and distribution (Cole and Hernan, 2008). Finally, extended Cox models for time-dependent variables were fit using these weights, with a robust variance estimator, to estimate the average effect of CM over the follow-up period. Baseline covariates were included in these models, since the stabilization of the weights create a pseudo-population where there might still be residual confounding (Cole and Hernan, 2008). Model fit and functional form of the MSMs were assessed based on martingale residuals. Presence of outliers and influential observations were checked with deviance residuals and delta-beta values, respectively.

Note that “risk” in this paper should be understood as “hazard”.

All statistical analyses were performed with R version 3.4.1 (R Core Team, 2015). An example of R code is given in the appendix.

3. Results

Table 1 presents the characteristics of the 3952 primiparous cows and 8724 multiparous cows (N = 12,676; 261 herds) that met the inclusion criteria. A total of 25.2% animals were culled during a mean follow-up time of 287 days (20.2%—307 days and 27.5%—278 days for heifers and cows, respectively). The herd size ranged from 30 to 205 cows (median: 52). More than 95% of the cows were Holsteins. Between calving and 120 DIM, 15.3% of the primiparous and 17.0% of the multiparous cows had at least one episode of clinical mastitis. The median time between the first and second CM episode was 21 days (interquartile range: 14–45) for primiparous and 29 days (14–58) for multiparous cows. The intervals between the second and third cases were 22 days (12–33) and 21 days (14–31).

The crude, unadjusted hazard ratio (HR) suggests an increased culling risk from CM (HR = 1.84 [1.48–2.29] and HR = 2.07 [1.82–2.35] for primiparous and multiparous cows, respectively). The same model with baseline covariates and time-varying pregnancy and season variables gave the same result (HR = 2.00 [1.61–2.50] and HR = 1.67 [1.47–1.90] for primiparous and multiparous cows, respectively). Table 2 shows the results of MSMs using the monthly interval approach. The models showed satisfying model fits and functional forms, while few outlier observations were present but with minimal influence. The stabilized IPTWs used in the marginal structural models have a mean of 0.99 for primiparous and 1.00 for multiparous cows (standard deviation of 0.04 and 0.08 for primiparous and multiparous, respectively). They ranged from 0.64 to 1.10 for primiparous and from 0.53 to 1.48 for multiparous cows, respectively. The estimated average CM causal HRs over time were 1.96 [1.57–2.44] and 1.47 [1.28–1.69] for primiparous and multiparous cows, respectively. Other effects on culling were still present after controlling for the time-varying confounders. Non-pregnancy was a major culling risk factor for both primiparous and multiparous cows. Multiparous cows showed additional risk factors from increasing parity, milk fever, and season. Their past history also provided information on their culling risk. Having already experienced a CM during the previous lactation increased their culling risk by 12%. Multiparous cows for which milk production was below the average of their herdmates, in respect to their parity, had an increased culling risk. On the other hand, above average milk production was a protective factor for culling. The SCC over their previous lactation was also a risk factor in reference to the herd/parity average value, for above average cows as well as below average ones.

Table 1
Characteristics of primiparous and multiparous cows by culling status (n = 12,676; 3952 primiparous, 8724 multiparous).

	Primiparous		Multiparous	
	Censored N = 3155	Culled N = 797	Censored N = 6327	Culled N = 2397
Parity				
2			2486 (39%)	614 (26%)
3			1675 (26%)	561 (23%)
4+			2166 (34%)	1222 (51%)
Age at first calving (months)				
> 24	477 (15%)	112 (14%)	907 (15%)	333 (16%)
24–26	1394 (44%)	332 (42%)	2523 (43%)	864 (41%)
26–28	760 (24%)	219 (27%)	1369 (23%)	527 (25%)
> 28	524 (17%)	134 (17%)	1059 (18%)	389 (18%)
Pregnancy				
Not pregnant	315 (10%)	721 (90%)	1252 (20%)	2174 (91%)
Before 90 DIM	1087 (34%)	33 (4%)	1762 (28%)	89 (4%)
90–120 DIM	513 (16%)	18 (2%)	861 (14%)	43 (2%)
> 120 DIM	1240 (39%)	25 (3%)	2452 (39%)	91 (4%)
Mean 305-day milk production,[a] kg (SD)	8561 (± 1450)	6985 (± 1945)	10,277 (± 1832)	9795 (± 1855)
Calved between August and December	1117 (35%)	279 (35%)	2259 (36%)	909 (38%)
Milk fever	1 (0%)	2 (0%)	278 (4%)	231 (10%)
Displaced abomasum	118 (4%)	28 (4%)	230 (4%)	118 (5%)
Dystocia	352 (11%)	133 (17%)	447 (7%)	218 (9%)
Retained placenta	156 (5%)	44 (6%)	507 (8%)	230 (10%)
Metritis	104 (3%)	12 (2%)	278 (4%)	79 (3%)
Clinical mastitis in previous lactation			1166 (18%)	587 (24%)
Clinical mastitis during follow-up				
None	2710 (86%)	637 (80%)	5311 (84%)	1926 (80%)
One case	401 (13%)	138 (17%)	875 (14%)	392 (16%)
Two cases	33 (1%)	19 (2%)	109 (2%)	53 (2%)
Three cases	11 (0%)	3 (0%)	32 (1%)	26 (1%)
First mastitis case, mean DIM (SD)	23 (± 33)	13 (± 22)	33 (± 36)	26 (± 33)
Second mastitis case, mean DIM (SD)	67 (± 33)	36 (± 24)	61 (± 32)	55 (± 32)
Third mastitis case, mean DIM (SD)	75 (± 29)	41 (± 21)	82 (± 24)	75 (± 23)

[a] Based on real production; DIM = days in milk; SCC = somatic cell count; SD = standard deviation.

4. Discussion

Compared with estimates from the standard adjusted model, estimates from the MSMs were about the same for primiparous and 12% lower for multiparous cows. Adjustment for milk yield and SCC was more pronounced in multiparous than primiparous cows, suggesting that the culling decision based on these factors was less stringent for heifers than for cows. Hazard ratios reported previously in the literature for models including milk production and SCC or both as covariates ranged from 1.6 to greater than 2. Some of these results were from random effect models (Caraviello et al., 2005; Schneider et al., 2007), i.e. population-average estimates which give larger estimates. Other studies by Gröhn et al. (1997) and Rajala-Schultz and Gröhn (1999) modelled milk production as a time-dependent covariate. However none of these studies addressed the bias introduced by including a time-dependent confounder in the model.

The effect of the fixed-time confounding variables were on par with what is found in the literature (Beaudeau et al., 1994; Gröhn et al., 1998; Rajala-Schultz and Gröhn, 1999). Lower producing cows have already been reported as being more at risk for culling (Beaudeau et al., 1995; Gröhn et al., 1998; Schneider et al., 2007). The effect of the cow's SCC characteristics from previous lactation was also related to previous studies. Somatic cell count is an important risk marker for CM (Beaudeau et al., 1998; Suriyasathaporn et al., 2000; Green et al., 2004). It has also been shown that high SCC herds have a higher culling rate and that culling occurs earlier in lactation (Caraviello et al., 2005). Moreover, cows having a high SCC for their lactation were probably chronically infected, which could increase the bulk tank SCC (Madouasse et al., 2010). Their removal from the herd is therefore a good strategy for the dairy producer to manage his/her bulk tank SCC regulatory limit. In this study, cows with lower previous lactation SCC compared to herd mates have a higher culling risk as well. Additionally to these health conditions, pregnancy remains however a determinant factor in the culling decision-making (Rajala-Schultz and Gröhn, 1999; Schneider et al., 2007).

Marginal structural models are not subject to collider-stratification bias (Greenland, 2003), since the confounding effect of time-dependent confounders that are affected by prior mastitis status is controlled by weighting instead of conditioning. But MSMs have several key assumptions that must be satisfied: exchangeability, consistency, positivity, and correct model specification (Cole and Hernan, 2008). The exchangeability assumption, or no unmeasured confounding, also includes that there should be no informative censoring due to unmeasured covariates. Unmeasured confounding is a major source of bias in observational studies. Sensitivity analyses to evaluate unmeasured confounders were developed for linear-, Poisson-, and logistic-MSMs (Brumback et al., 2004). In a survival analysis with Cox regression, Klungsøyr et al. (2009) developed a sensitivity analysis for a point exposure design (constant exposure or single assessment). However no sensitivity analyses are readily available for repeated exposures in a Cox survival model. The exchangeability assumption cannot be verified empirically and we assumed therefore that the measured covariates included in the analysis, which include major, known, confounders, were sufficient to control for confounding bias. Consistency, i.e. that a cow's potential outcome under her observed mastitis history is precisely her observed outcome (Robins et al., 2000), is also difficult to verify. However, the health data registry used in this study provides clearly defined health definitions reviewed by the herd veterinarian, putting confidence into the consistency assumption. Positivity requires that at every level of the confounders, cows in the population have a nonzero probability of experiencing every level of exposure, which implies that the average causal effect of mastitis can be estimated in each subset of the population defined by the confounders. Clinical mastitis cases occur most often early in lactation (Barkema et al., 1998; Sargeant et al.,

Table 2
Estimates of association between clinical mastitis and culling using marginal structural Cox models.

Adjusted for	Primiparous		Multiparous	
	HR[a]	95% CI	HR[a]	95% CI
Clinical mastitis	1.96	1.57, 2.44	1.47	1.28, 1.69
Clinical mastitis in previous lactation			1.12	1.01, 1.24
Parity				
3 vs 2			1.15	1.02, 1.29
4+ vs 2			1.50	1.35, 1.66
Milk production during previous lactation				
Below average (vs avg)			1.17	1.05, 1.30
Above average (vs avg)			0.83	0.73, 0.94
SCC ranking for previous lactation				
Below average (vs avg)			1.36	1.00, 1.84
Above average (vs avg)			1.37	1.23, 1.54
Pregnancy	0.02	0.01, 0.02	0.02	0.02, 0.03
Dystocia	1.16	0.95, 1.42	1.12	0.96, 1.31
Displaced abomasum	0.74	0.51, 1.09	1.01	0.84, 1.23
Milk fever			1.49	1.23, 1.80
Retained placenta	1.00	0.74, 1.36	1.07	0.93, 1.23
Age at first calving				
24–26 months (vs < 24)	0.97	0.79, 1.20	0.93	0.82, 1.05
26–28 months (vs < 24)	1.02	0.82, 1.28	0.96	0.83, 1.10
> 28 months (vs < 24)	0.92	0.72, 1.17	0.86	0.74, 1.00
Season (August–December vs January–July)	1.04	0.94, 1.15	0.76	0.71, 0.81
Concordance	0.68		0.72	
Likelihood ratio test (df, *p*-value)	1028.04	(9, < 0.001)	3294.56	(17, < 0.001)

CI, confidence interval; DIM, days in milk; HR, hazard ratio; SCC, somatic cell count; df, degrees of freedom.

[a] Estimated from a marginal structural Cox model, adjusted for baseline covariates.

1998; Olde Riekerink et al., 2008). By focusing on the lactation period between calving and 120 DIM, we excluded periods of zero exposure probability from the data set, meeting the positivity requirement. The positivity assumption also applies to the presence of clusters in the analysis, here the various herds. All herds selected for analysis had at least one case of CM, and were selected for their good reporting of health events, therefore meeting the positivity assumption. Correct model specification implies that appropriate functional forms are used in the logistic models used to determine the weights, and in the final weighted model. While positivity and correct model specification are mainly working assumptions, they can be further confirmed by the absence of extreme weights and by having a mean weight close to 1 (Cole and Hernan, 2008; Howe et al., 2011), which were validated here.

Finally, the analysis is based on the assumption of random measurement error. Using a retrospective, observational data set following a user-defined event recording scheme, both the exposure and the outcome are at risk of measurement errors that are correlated with each other, potentially leading to differential misclassification of culling by CM status. We tried to overcome this issue by selecting herds with a comprehensive record of health events, by extracting herds with at least 10% CM lactational incidence. The herds were deemed representative of Québec dairy herds using a monthly DHI service for individual cow milk recording, and a computerized data management system for reproduction and health management. But we still lack precision on CM as no information on its severity was available, and dairy producers might be more likely to report the most severe cases. We have, however, good confidence on the reporting of culling, veterinary-treated and veterinary-supervised conditions, as well as pregnancy status. Likewise, we restricted the analysis to herds with satisfactory matches between health and production record data sets, to improve our confidence in the milk production information.

To authors' knowledge, this study is the first application of MSM in veterinary medicine, even if the method was described previously in this journal (Martin, 2008, 2014). While epidemiologists are encourage to rely less on cross-sectional studies and move to cohort studies (Martin, 2008), the association measures from observational studies cannot generally be interpreted as effect measures because the exposed and unexposed subjects are not exchangeable (Hernan and Robins, 2006; Martin, 2014). With IPTW, a re-weighted pseudo-sample is created, mimicking the random assignment of a randomized experiment. The assignment probability can even be allowed to vary between exposure levels when using stabilized IPTWs. Therefore weighting, stabilized or not, leads to causal measures of association as in a randomized trial, as long as all the required assumptions hold (Daniel et al., 2013). The implementation of MSM can be easily realized in standard statistical software, using the weighted version of logistic or Cox regression models. However, the presence of extreme weights can make the MSM unstable and/or inefficient even if this issue can be attenuated by using stabilized weights. Other approaches to estimate causal effect in the presence of time-dependent covariates are also available, like the g-computation formula (Robins, 1986) or the g-estimation of structural nested models (Robins et al., 1992). Their implementation is not as straightforward, sometimes requiring heavy computations or are simply not available in standard software. They all however lack extensive research and applications in hierarchical settings. Guidelines on computing IPTWs for multilevel structures are available (He, 2014), but population-average estimation of multilevel marginal structural survival model is still not easily accomplished and requires further developments.

5. Conclusion

Our findings confirm that CM is a risk factor for culling, but with reduced effect compared to previous studies, which did not properly control for time-dependent confounders. Heifers and cows also experienced the same risk for CM, milk production having less influence on the culling decision in heifers than cows. However, after controlling for the potential confounders, the culling risk was still large and other factors might influence the culling decision process. Unmeasured confounders require further evaluation through sensitivity analyses to be developed in the framework of time-varying exposure.

Conflict of interest

None.

Acknowledgements

The authors thank Agriculture and Agri-Food Canada, Fonds de recherche du Québec – Nature et technologies (FQRNT), Novalait, and Ministère de l'Agriculture, des Pêcheries et de l'Alimentation du Québec (MAPAQ) for supporting this study (grant LT-163341).

Appendix A. R code

```
## Required packages
library(lme4)

## Compute weights
# formulas
denom_form <- formula(mastitis ~ shift.mast + milk + scc + covariate_list +
                          log(tstop) + (1|herd/id))
num_form <- formula(mastitis ~ shift.mast + covariate_list +
                        log(tstop) + (1|herd/id))
denomcen_form <- formula(cens ~ shift.mast + milk + scc + covariate_list +
                             log(tstop) + (1|herd/id))
numcen_form <- formula(cens ~ shift.mast + covariate_list +
                           log(tstop) + (1|herd/id))
# generate propensity scores
msmps1 <- glmer(denom_form,
                data = dat,
                family = binomial)
dat$msm_ps1 <- fitted(msmps1)

msmps0 <- glmer(num_form,
                data = dat,
                family = binomial)
dat$msm_ps0 <- fitted(msmps0)

# stabilize weight by dividing propensity score
dat$msm_iptws[dat$mastitis == 1] <- dat$msm_ps0[dat$mastitis == 1] /
    dat$msm_ps1[dat$mastitis == 1]
dat$msm_iptws[dat$mastitis == 0] <- (1 - dat$msm_ps0)[dat$mastitis == 0] /
    (1 - dat$msm_ps1)[dat$mastitis == 0]

# trimming weights
lower_wts <- quantile(dat$msm_iptws, 0.01)
upper_wts <- quantile(dat$msm_iptws, 0.99)
dat$iptws_trim <- dat$msm_iptws
dat$iptws_trim <- with(dat, ifelse(iptws_trim < lower_wts, lower_wts, iptws_trim))
dat$iptws_trim <- with(dat, ifelse(iptws_trim > upper_wts, upper_wts, iptws_trim))

# generate censoring weights
pcen1 <- glmer(denomcen_form,
```

```
                data = dat,
                family = binomial)
dat$pcen1 <- fitted(pcen1)

pcen0 <- glmer(numcen_form,
                data = dat,
                family = binomial)
dat$pcen0 <- fitted(pcen0)

dat$msm_ipcws[dat$cens == 1] <- dat$pcen0[dat$cens == 1] / dat$pcen1[dat$cens == 1]
dat$msm_ipcws[dat$cens == 0] <- (1 - dat$pcen0)[dat$cens == 0] /
    (1 - dat$pcen1)[dat$cens == 0]
dat$msm_ipcws <- with(dat, ifelse(is.na(msm_ipcws), 1, msm_ipcws))

# trimming weights
lower_wts <- quantile(dat$msm_ipcws, 0.01)
upper_wts <- quantile(dat$msm_ipcws, 0.99)
dat$ipcws_trim <- dat$msm_ipcws
dat$ipcws_trim <- with(dat, ifelse(ipcws_trim < lower_wts, lower_wts, ipcws_trim))
dat$ipcws_trim <- with(dat, ifelse(ipcws_trim > upper_wts, upper_wts, ipcws_trim))

# cumulative weights
c_iptws <- unsplit(sapply(split(dat$iptws_trim, dat$id), cumprod), dat$id)
c_ipcws <- unsplit(sapply(split(dat$ipcws_trim, dat$id), cumprod), dat$id)

dat <- cbind(dat, c_iptws, c_ipcws)
dat$cum_wts <- dat$c_iptws * dat$c_ipcws

## Stabilized treatment and censoring weights
msm <- coxph(Surv(tstart, tstop, status) ~ mastitis + covariate_list +
                cluster(herd),
             data = dat,
             weights = cum_wts)
```

References

Austin, P.C., 2014. The use of propensity score methods with survival or time-to-event outcomes: reporting measures of effect similar to those used in randomized experiments. Stat. Med. 33 (7), 1242–1258.

Barkema, H.W., Schukken, Y.H., Lam, T.J.G.M., Beiboer, M.L., Wilmink, H., Benedictus, G., Brand, A., 1998. Incidence of clinical mastitis in dairy herds grouped in three categories by bulk milk somatic cell counts. J. Dairy Sci. 81 (2), 411–419. URL http://jds.fass.org/cgi/content/abstract/81/2/411.

Barnouin, J., Bord, S., Bazin, S., Chassagne, M., 2005. Dairy management practices associated with incidence rate of clinical mastitis in low somatic cell score herds in France. J. Dairy Sci. 88 (10), 3700–3709. http://dx.doi.org/10.3168/jds.S0022-0302(05)73056-3.

Beaudeau, F., Ducrocq, V., Fourichon, C., Seegers, H., 1995. Effect of disease on length of productive life of French Holstein dairy cows assessed by survival analysis. J. Dairy Sci. 78 (1), 103–117. URL http://www.sciencedirect.com/science/article/pii/S0022030295766218.

Beaudeau, F., Frankena, K., Fourichon, C., Seegers, H., Faye, B., Noordhuizen, J.P.T.M., 1994. Associations between health disorders of French dairy cows and early and late culling within the lactation. Prev. Vet. Med. 19 (3–4), 213–231. URL http://www.sciencedirect.com/science/article/pii/0167587794900906.

Beaudeau, F., Seegers, H., Ducrocq, V., Fourichon, C., Bareille, N., 2000. Effect of health disorders on culling in dairy cows: a review and a critical discussion. Ann. Zootech. 49, 293–312. URL http://prodinra.inra.fr/record/4291.

Beaudeau, F., Seegers, H., Fourichon, C., Hortet, P., 1998. Association between milk somatic cell counts up to 400,000 cells/ml and clinical mastitis in French Holstein cows. Vet. Rec. 143 (25), 685–687.

Brumback, B.A., Hernán, M.A., Haneuse, S.J.P.A., Robins, J.M., 2004. Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures. Stat. Med. 23 (5), 749–767.

Caraviello, D.Z., Weigel, K.A., Shook, G.E., Ruegg, P.L., 2005. Assessment of the impact of somatic cell count on functional longevity in Holstein and Jersey cattle using survival analysis methodology. J. Dairy Sci. 88 (2), 804–811. URL http://www.sciencedirect.com/science/article/pii/S0022030205727454.

Cole, S.R., Hernan, M.A., 2008. Constructing inverse probability weights for marginal structural models. Am. J. Epidemiol. 168 (6), 656–664.

Cole, S.R., Platt, R.W., Schisterman, E.F., Chu, H., Westreich, D., Richardson, D., Poole, C., 2010. Illustrating bias due to conditioning on a collider. Int. J. Epidemiol. 39 (2), 417–420. http://dx.doi.org/10.1093/ije/dyp334.

Daniel, R., Cousens, S., De Stavola, B., Kenward, M.G., Sterne, J.A.C., 2013. Methods for dealing with time-dependent confounding. Stat. Med. 32 (9), 1584–1618. http://dx.doi.org/10.1002/sim.5686.

De Vries, A., Olson, J.D., Pinedo, P.J., 2010. Reproductive risk factors for culling and productive life in large dairy herds in the eastern United States between 2001 and 2006. J. Dairy Sci. 93 (2), 613–623.

Fetrow, J., Nordlund, K.V., Norman, H.D., 2006. Invited review: culling: nomenclature, definitions, and recommendations. J. Dairy Sci. 89 (6), 1896–1905.

Glidden, D.V., Vittinghoff, E., 2004. Modelling clustered survival data from multicentre clinical trials. Stat. Med. 23 (3), 369–388.

Green, M.J., Burton, P.R., Green, L.E., Schukken, Y.H., Bradley, A.J., Peeler, E.J., Medley, G.F., 2004. The use of Markov Chain Monte Carlo for analysis of correlated binary data: patterns of somatic cells in milk and the risk of clinical mastitis in dairy cows. Prev. Vet. Med. 64 (2–4), 157–174.

Greenland, S., 2003. Quantifying biases in causal models: classical confounding vs collider-stratification bias. Epidemiology 14 (3), 300–306.

Gröhn, Y.T., Ducrocq, V., Hertl, J.A., 1997. Modeling the effect of a disease on culling: an illustration of the use of time-dependent covariates for survival analysis. J. Dairy Sci.

80 (8), 1755–1766. URL http://www.sciencedirect.com/science/article/pii/S0022030297761095.
Gröhn, Y.T., Eicker, S.W., Ducrocq, V., Hertl, J.A., 1998. Effect of diseases on the culling of Holstein dairy cows in New York State. J. Dairy Sci. 81 (4), 966–997. URL http://www.sciencedirect.com/science/article/pii/S0022030298756577.
Hadley, G.L., Wolf, C.A., Harsh, S.B., 2006. Dairy cattle culling patterns explanations and implications. J. Dairy Sci. 89 (6), 2286–2296. URL http://www.sciencedirect.com/science/article/pii/S0022030206723001.
He, J., 2014. Causal modeling under complex dependency in clustered and longitudinal observations (Ph.D. thesis). University of Pennsylvania. URL http://repository.upenn.edu/edissertations/1311.
Hernán, M., Brumback, Á., Robins, B.J.M., 2000. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. Epidemiology 11 (5), 561–570.
Hernán, M.A., Hernández-Díaz, S., Robins, J.M., 2004. A structural approach to selection bias. Epidemiology 15 (5), 615–625.
Hernan, M.A., Robins, J.M., 2006. Estimating causal effects from epidemiological data. J. Epidemiol. Community Health 60 (7), 578–586. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2652882/.
Howe, C.J., Cole, S.R., Chmiel, J.S., Muñoz, A., 2011. Limitation of inverse probability-of-censoring weights in estimating survival in the presence of strong selection bias. Am. J. Epidemiol. 173 (5), 569–577.
Klungsøyr, O., Sexton, J., Sandanger, I., Nygård, J.F., 2009. Sensitivity analysis for unmeasured confounding in a marginal structural Cox proportional hazards model. Lifetime Data Anal. 15 (2), 278–294.
Lam, T.J.G.M., Schukken, Y.H., van Vliet, J.H., Grommers, F.J., Tielen, M.J.M., Brand, A., 1997. Effect of natural infection with minor pathogens on susceptibility to natural infection with major pathogens in the bovine mammary gland. Am. J. Vet. Res. 58 (1), 17–22.
Lee, B.K., Lessler, J., Stuart, E.A., 2011. Weight trimming and propensity score weighting. PLoS One 6 (3).
Madouasse, A., Huxley, J.N., Browne, W.J., Bradley, A.J., Green, M.J., 2010. Somatic cell count dynamics in a large sample of dairy herds in England and Wales. Prev. Vet. Med. 96 (1–2), 56–64. http://dx.doi.org/10.1016/j.prevetmed.2010.05.005.
Martin, W., 2008. Linking causal concepts, study design, analysis and inference in support of one epidemiology for population health. Prev. Vet. Med. 86 (3–4), 270–288.
Martin, W., 2014. Making valid causal inferences from observational data. Prev. Vet. Med. 113 (3), 281–297. URL http://linkinghub.elsevier.com/retrieve/pii/S0167587713002845.
Munda, M., Legrand, C., 2014. Adjusting for centre heterogeneity in multicentre clinical trials with a time-to-event outcome. Pharm. Stat. 13 (2), 145–152.
Olde Riekerink, R.G.M., Barkema, H.W., Kelton, D.F., Scholl, D.T., 2008. Incidence rate of clinical mastitis on Canadian dairy farms. J. Dairy Sci. 91 (4), 1366–1377. URL http://www.sciencedirect.com/science/article/pii/S0022030208712645.
O'Reilly, K.M., Green, M.J., Peeler, E.J., Fitzpatrick, J.L., Green, L.E., 2006. Investigation of risk factors for clinical mastitis in British dairy herds with bulk milk somatic cell counts less than 150,000 cells/ml. Vet. Rec. 158, 649–653.
R Core Team, 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org.
Rajala-Schultz, P.J., Gröhn, Y.T., 1999a. Culling of dairy cows. Part I: Effects of diseases on culling in Finnish Ayrshire cows. Prev. Vet. Med. 41 (2–3), 195–208. URL http://www.sciencedirect.com/science/article/pii/S016758779900046X.
Rajala-Schultz, P.J., Gröhn, Y.T., 1999b. Culling of dairy cows. Part II: Effects of diseases and reproductive performance on culling in Finnish Ayrshire cows. Prev. Vet. Med. 41 (4), 279–294. URL http://www.sciencedirect.com/science/article/pii/S0167587799000458.
Rajala-Schultz, P.J., Gröhn, Y.T., 1999c. Culling of dairy cows. Part III: Effects of diseases, pregnancy status and milk yield on culling in Finnish Ayrshire cows. Prev. Vet. Med. 41 (4), 295–309. URL http://www.sciencedirect.com/science/article/pii/S0167587799000471.
Rajala-Schultz, P.J., Gröhn, Y.T., McCulloch, C.E., Guard, C.L., 1999. Effects of clinical mastitis on milk yield in dairy cows. J. Dairy Sci. 82 (6), 1213–1220. URL http://www.sciencedirect.com/science/article/pii/S0022030299753440.
Robins, J., 1986. A new approach to causal inference in mortality studies with a sustained exposure period – application to control of the healthy worker survivor effect. Math. Model. 7, 1393–1512.
Robins, J.M., 1999. Association, causation, and marginal structural models. Synthese 121 (1), 151–179.
Robins, J.M., Blevins, D., Ritter, G., Wulfsohn, M., 1992. G-estimation of the effect of prophylaxis therapy for *Pneumocystis carinii* pneumonia on the survival of AIDS patients. Epidemiology 319–336. URL http://www.jstor.org/stable/3702734.
Robins, J.M., Hernan, M.A., Brumback, B., 2000. Marginal structural models and causal inference in epidemiology. Epidemiology 11 (5), 550–560.
Sargeant, J.M., Scott, H.M., Leslie, K.E., Ireland, M.J., Bashiri, A., 1998. Clinical mastitis in dairy cattle in Ontario: frequency of occurrence and bacteriological isolates. Can. Vet. J. 39 (1), 33–38. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1539829/.
Schneider, M.d.P., Strandberg, E., Emanuelson, U., Grandinson, K., Roth, A., 2007. The effect of veterinary-treated clinical mastitis and pregnancy status on culling in Swedish dairy cows. Prev. Vet. Med. 80 (2–3), 179–192. URL http://www.sciencedirect.com/science/article/pii/S0167587707000487.
Schukken, Y.H., Grommers, F.J., van de Geer, D., Erb, H.N., Brand, A., 1991. Risk factors for clinical mastitis in herds with a low bulk milk somatic cell count. 2. Risk factors for *Escherichia coli* and *Staphylococcus aureus*. J. Dairy Sci. 74 (3), 826–832. URL http://linkinghub.elsevier.com/retrieve/pii/S0022030291782313.
Seegers, H., Fourichon, C., Beaudeau, F., 2003. Production effects related to mastitis and mastitis economics in dairy cattle herds. Vet. Res. 34 (5), 475–491. http://dx.doi.org/10.1051/vetres:2003027.
Sewalem, A., Miglior, F., Kistemaker, G.J., Van Doormaal, B.J., 2006. Analysis of the relationship between somatic cell score and functional longevity in Canadian dairy cattle. J. Dairy Sci. 89 (9), 3609–3614. URL http://www.sciencedirect.com/science/article/pii/S0022030206724006.
Steeneveld, W., Hogeveen, H., Barkema, H.W., van den Broek, J., Huirne, R.B.M., 2008. The influence of cow factors on the incidence of clinical mastitis in dairy cows. J. Dairy Sci. 91 (4), 1391–1402. URL http://www.sciencedirect.com/science/article/pii/S0022030208712669.
Suriyasathaporn, W., Schukken, Y.H., Nielen, M., Brand, A., 2000. Low somatic cell count: a risk factor for subsequent clinical mastitis in a dairy herd. J. Dairy Sci. 83 (6), 1248–1255. URL http://www.journalofdairyscience.org/article/S0022030200749915/fulltext.
Van Buuren, S., Groothuis-Oudshoorn, K., 2011. MICE: multivariate imputation by chained equations in R. J. Stat. Softw. 45 (3), 1–67.
Waage, S., Sviland, S., Ødegaard, S.A., 1998. Identification of risk factors for clinical mastitis in dairy heifers. J. Dairy Sci. 81 (5), 1275–1284. URL http://www.journalofdairyscience.org/article/S0022030298756899/fulltext.
Zadoks, R.N., Allore, H.G., Barkema, H.W., Sampimon, O.C., Wellenberg, G.J., Gröhn, Y.T., Schukken, Y.H., 2001. Cow- and quarter-level risk factors for *Streptococcus uberis* and *Staphylococcus aureus* mastitis. J. Dairy Sci. 84 (12), 2649–2663. URL http://www.ncbi.nlm.nih.gov/pubmed/11814021.