Preventive Veterinary Medicine 176 (2020) 104908

Contents lists available at ScienceDirect

Preventive Veterinary Medicine

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

Assessment of associations between transition diseases and reproductive performance of dairy cows using survival analysis and decision tree algorithms

O. Bogado Pascottini[a,*], M. Probo[b], S.J. LeBlanc[a], G. Opsomer[c], M. Hostens[c]

[a] Population Medicine, Ontario Veterinary College, University of Guelph, ON, N1G 2W1, Canada
[b] Veterinary Teaching Hospital, Department of Veterinary Medicine, University of Milan, via Dell'Università 6, 26900 Lodi, Italy
[c] Department of Reproduction, Obstetrics and Herd Health, Faculty of Veterinary Medicine, Ghent. University, Salisburylaan 133, 9820 Merelbeke, Belgium

ARTICLE INFO

Keywords:
Transition period
Fertility
Machine learning

ABSTRACT

This study aimed to evaluate the associations between transition cow conditions and diseases TD with fertility in Holstein cows, and to compare analytic methods for doing so. Kaplan-Meier, Cox proportional hazard, and decision tree models were used to analyze the associations of TD with the pregnancy risk at 120 and 210 DIM from a 1-year cohort with 1946 calvings from one farm. The association between TD and fertility was evaluated as follows: 1 cows with TD whether complicated with another TD or not TD-all, versus healthy cows, and 2 cows with uncomplicated TD TD-single, versus cows with multiple TD TD+; complicated cases, versus healthy cows. The occurrence of twins, milk fever, retained placenta, metritis, ketosis, displaced abomasum, and clinical mastitis were recorded. Using Kaplan-Meier models, in primiparous cows the 120 DIM pregnancy risk was 62% (95% CI: 57–67 %) for healthy animals. This was not significantly different for TD-single (58%; 95% CI: 51–66 %) but was reduced for TD+ (45%; 95% CI: 33–60 %). Among healthy primiparous cows, 80% (95% CI: 75–84 %) were pregnant by 210 DIM, but pregnancy risk at that time was reduced for primiparous cows with TD-single (72%; 95% CI: 65–79 %) and TD+ (62%; 95% CI: 49–75 %). In healthy multiparous cows, the 120 DIM pregnancy risk was 53% (95% CI: 49–56 %), which was reduced for TD-single (36%; 95% CI: 31–42 %) and TD+ (30%; 95% CI: 24–38 %). The 210 DIM pregnancy risk for healthy multiparous cows was 70% (95% CI: 67–72 %), being higher than the 210 DIM pregnancy risk for multiparous cows with TD-single (47%; 95% CI: 42–53 %) or TD+ (46%; 95% CI: 38–54 %). Cows with TD-all presented similar pregnancy risk estimates as for TD +. Cox proportional hazards regressions provided similar magnitudes of effects as the Kaplan-Meier estimates. Survival analysis and decision tree models identified parity as the most influential variable affecting fertility. Both modeling techniques concurred that TD + had a greater effect than TD-single on the probability of pregnancy at 120 and 210 DIM. Decision trees for individual TD identified that displaced abomasum affected fertility at 120 DIM in primiparous while metritis was the most influential TD at 120 and 210 DIM for multiparous cows. The data were too sparse to assess multiple interactions in multivariable Cox proportional hazard models for individual TD. Machine learning helped to explore interactions between individual TD to study their hierarchical effect on fertility, identifying conditional relationships that merit further investigation.

1. Introduction

Complex pathways of metabolic adaptations occur in peripartum dairy cattle to support milk production. However, inadequate metabolic adaptations may allow negative consequences such as reduced feed intake, excessive insulin resistance, metabolic inflammation, and decreased immune function (Grummer, 1995; Hammon et al., 2006; Bradford et al., 2015). As a result, the vast majority of metabolic and

Abbreviations: TD, transition cow conditions and diseases; TD-all, transition cow conditions and diseases whether complicated with another transition cow condition and disease or not; TD-single, single transition cow conditions and diseases; TD+, multiple (complicated) transition cow conditions amd diseases; DIM, days in milk; RP, retained placenta; MAST, mastitis; METR, metritis; NEB, negative energy balance; KET, ketosis; DA, displaced abomasum; MF, milk fever; TMF, transition management facility; TWIN, twinning; M60, cumulative milk yield to 60 days in milk; Se, sensitivity; Sp, specificity; PPV, positive predictive value; NPV, negative predictive value

* Corresponding author.
E-mail address: osvaldo.bogado@ugent.be (O. Bogado Pascottini).

https://doi.org/10.1016/j.prevetmed.2020.104908
Received 26 September 2019; Received in revised form 10 January 2020; Accepted 27 January 2020

infectious diseases of dairy cattle occur during the transition period (LeBlanc, 2010). Transition cow conditions and diseases (**TD**) affect the profitability of the farm by increasing uneconomical culling (Beaudeau et al., 1993; Chiumia et al., 2013; Probo et al., 2018), lowering milk production (Beaudeau et al., 1995; Hostens et al., 2012), and by decreasing the reproductive performance of the herd (Chapinal et al., 2012; Ribeiro et al., 2013; Dubuc and Denis-Robichaud, 2017). Inflammation that characterizes conditions like retained placenta (**RP**), mastitis (**MAST**), and metritis (**METR**), produces pro-inflammatory mediators that negatively affect uterine function, follicular growth, oocyte quality, and embryo development (Turner et al., 2012; Ribeiro et al., 2016). Aggravation or maladaptation to negative energy balance (**NEB**) increases the risk for ketosis (**KET**) and displaced abomasum (**DA**) (Ospina et al., 2010a), which are in turn associated with reduced reproductive performance (Wathes et al., 2007; Ospina et al., 2010b; Dubuc et al., 2010). Greater NEB during the postpartum period also results in a longer interval from parturition to first ovulation and a decreased probability of pregnancy at first insemination (Opsomer et al., 2000; Patton et al., 2007).

As discussed by Mulligan and Doherty (2008) one of the interesting aspects of TD is the complex linkage among them. Hostens et al. (2012) used a lactation curve modeling approach on milk production data from a transition management facility (**TMF**) to explore the interactive effects of TD on milk production. Subsequently, and using the same data, the effect of TD and their interactions on the risk of culling was modelled using machine learning techniques including decision trees and random forest models (Probo et al., 2018). Their results indicated substantial effects of milk fever (**MF**) and DA as risk factors for culling in the first 120 DIM. Machine-learning methods can accommodate large and complex datasets in the presence of missing values and intricate dependencies among explanatory variables. Some machine learning methods in the field of dairy herd management have been applied to analyze data associated with mastitis (Kamphuis et al., 2010a, b), estrus detection (Firk et al., 2003; Cavero et al., 2008; Sun et al., 2010), reproduction (Caraviello et al., 2006; Shahinfar et al., 2014) and more recently data quality assessment (Hermans et al., 2017). However, machine learning algorithms are not yet widely used in dairy science, although they present a promising method for data mining and development of predictive tools.

Survival analysis has been suggested as the most appropriate method for the analysis of longitudinal data in dairy research (Beaudeau et al., 1995). However, robustness of machine learning modeling might help us to elucidate complex interactions among TD, including multiple degrees of covariation. This can guide us to better understand the complex linkage among TD and to rank their hierarchal effect on the pregnancy risk. Thus, the objective of the present study was to assess the association of uncomplicated TD (**TD-single**), complicated (multiple) TD (**TD+**), and overall TD (**TD-all**; uncomplicated and complicated TD, together) with the short-term (to 120 DIM) and full breeding period (to 210 DIM) pregnancy risk, employing conventional (Kaplan-Meier and Cox proportional hazard models) and machine learning (decision tree) methods.

2. Materials and methods

2.1. General

This retrospective observational study involves the analysis of health, production and reproductive data collected from April 2009 to April 2010. Data were obtained from a dairy located in Mecklenburg-Vorpommern Germany, using Dairy Comp 305 Valley Ag Software, Tulare, CA to record herd data. The herd size was expanding during the data collection and consisted of approximately 2450 lactating Holstein-Friesians. The mean milk production per lactation 305 d was 11,085 kg/cow with an average of 3.6% fat and 3.3% protein (based on monthly measurements by State Control Association for Performance and Quality Testing Mecklenburg-Vorpommern eG during the timeframe of the study). Cows and heifers were housed in a TMF 40 and 42 d before expected parturition, respectively. These barns were composed of 32 sand-bedded free stall pens with a maximum of 32 animals per group, and *ad libitum* access to water. Heifers were transferred to the TMF 40 d before the expected parturition date and cows at dry-off to achieve a dry period of 42 d. Cows and heifers were housed separately within the TMF until moved to the milking barns, typically between 10 and 21 DIM. Separate far-off, close-up, and fresh-cow diets were fed as a TMR once daily, with push-ups every 2 h.

2.2. Transition management

Within 24 h after parturition, 500 mL of propylene glycol (Bernd-Dieter, Dusseldorf, Germany) diluted in 50 L of warm water was administered once by oral tube to all animals. At the same time, multiparous cows also received 500 mL of a 38% calcium borogluconate solution intravenously (Calcilift Forte, Albrecht GmbH, Aulendorf, Germany) once. In the TMF, all diagnoses and treatments were executed and documented according to specific protocols by specialised farm personnel that were always supervised by the farm veterinarian. Briefly, after milking, animals were routinely head-locked (at least one hour while at the TMF, after the morning milking) and examined daily for rectal temperature (cut-point for pyrexia < 39.5 °C), ketonuria (Ketostix, Bayer AG, Leverkusen, Germany), and feed intake (visually evaluated), until cows were moved to the milking barns when producing milk as expected (steady increase) and free from signs of disease (between 10–21 DIM). The following TD were recorded in the present study: twinning (**TWIN**), MF, RP, METR, KET, DA, and MAST.

The diagnosis of (clinical) MF was based on signs (within 48 h of calving) of cold ears and extremities, muscular tremors, ataxia, decreased ruminal motility, and/or inability to rise. The MF treatment consisted of 500 mL of calcium borogluconate solution (38%) intravenously for up to three consecutive days (in case of recumbency), and 6–8 s.c. injections of calcium borogluconate into widely separated sites (50 mL per site). Diagnosis of MF in cows that did not respond to calcium treatment was confirmed by assessing the total calcium serum concentration (cut point < 1.5 mmol/L). Cows were considered to have RP if the foetal membranes remained visible for more than 24 h after parturition (daily inspected in head-locked cows). Retained placenta was not medically treated, however, RP cows were carefully monitored for pyrexia during subsequent days. The diagnosis of METR was based on detection of fetid red-brown uterine discharge, with temperature > 39.5 °C or systemic illness (toxemia, dullness, anorexia, etc.) within 21 DIM. For this, a gloved hand was used to check the vaginal discharge in cows presenting pyrexia (or other systemic illness, as stated above). Cows diagnosed with METR were treated (i.m.) with Ceftiofur hydrochloride (Excenel RTU, Pfizer Animal Health GmbH, Berlin, Germany) for three consecutive days (2.2 mg/kg of BW), and a single intrauterine infusion of oxytetracycline at a dose of 15 mg/kg of BW (Oxy-Sleecol 200 LA, Albrecht GmbH, Aulendorf, Germany).

Ketone bodies acetoacetate were measured twice daily as described above by farm personnel using urinary test strips. Ketosis was defined as reduced feed intake detected by farm personnel followed by midstream urinary ketone bodies exceeding 500 μmol/L. Cows with KET were treated differently according to the ketone body concentration: between 1500 and 4000 μmol/L treatment consisted of 500 mL propylene glycol orally once daily for 3 consecutive days. Animals with urinary ketones exceeding 4000 μmol/L were treated intravenously with 500 mL of a 40% glucose solution (B. Braun, Melsungen, Germany), 40 mL of a vitamin B complex (Vitamin B-Komplex, Serumwerk Bernburg, Bernburg, Germany), and an i.m. injection of dexamethasone at a dose of 0.08 mg/kg of BW (Rapidexon, Albrecht GmbH). The DA diagnosis was done by the detection of tympanic resonance ("ping") during percussion of the left flank of cows that did not eat or appeared to be dull or lethargic during the daily inspection while in the headlocks. Initially, DA was treated by the "roll-and-toggle" method. Recurrent cases were treated with surgical abomasopexy.

Although TWIN is not a disease, it is a risk factor for other transition diseases (Sawa et al., 2015). Thus, TWIN was considered as a TD in this study. Clinical mastitis (visibly abnormal milk or swelling of the udder) diagnosed before 30 DIM was also considered in the analysis. Mastitis is not a transition disease but was included as a TD due to its strong association with impaired metabolic health (Jánosi et al., 2003; Nyman et al., 2008; Moyes et al., 2009). Cows with MAST were treated with an intra-mammary infusion of 88.8 mg of cefquinome sulfate per infected quarter (Cobactan LC, Intervet Deutschland GmbH) for three consecutive days. In case of relapse (or severe clinical cases, i.e. systemically ill cow), cefquinome was additionally administered intramuscularly at a dose of 1 mg/kg of BW (Cobactan 2.5%, Intervet Deutschland GmbH) in combination with a single intravenous injection of flunixin meglumine at a dose of 2.2 mg/kg of BW (Finadyne, Intervet Deutschland GmbH).

2.3. Reproductive management and milk production data

All cows were enrolled weekly on a reproductive management program for the first insemination. Briefly, cows were given 2 cloprostenol injections ($PGF_{2\alpha}$, Veyx forte, Veyx-Pharma, Schwarzenborn, Germany), one at each of 35 and 45 DIM, then tail-chalked and observed twice daily for heat and inseminated if detected in estrus. Cows that failed to be detected in estrus were submitted to an Ovsynch56 protocol as described by Dewey et al. (2010). The schedule started at 72 DIM using gonadorelin (Gonavet, Veyx-Pharma, Schwarzenborn, Germany), cloprostenol 7 d later, gonadorelin 56 h after cloprostenol, and timed AI 16 h later. All injections, estrus detection, and inseminations were done by trained personnel following breeding protocols which remained consistent over the entire study period. Pregnancy diagnosis and confirmation were performed weekly using ultrasound (BCF Easi Scan, Gormanston, Ireland) starting at 32 and 60 d after AI, respectively. Cows found open at 32 d were resynchronized using Ovsynch56 as for the first synchronization protocol. However, if standing heat was detected between treatments, cows were inseminated (i.e., cows were continuously eligible for AI).

Monthly milk weights assessed by State Control Association for Performance and Quality Testing Mecklenburg-Vorpommern eG were exported from Dairy Comp 305 (Microsoft Corp., Redmond, WA) and fitted with a MilkBot model using a proprietary maximum likelihood algorithm (DairySight LLC, Argyle, NY). Details regarding milk production and the MilkBot function parameters are described by Hostens et al. (2012).

2.4. Statistical analyses

Data were exported from Dairy Comp 305 to a Microsoft Excel file (Microsoft Corp., Redmond, WA). All statistical analyses were performed using R-core (version 3.6.1; R Core Team, Vienna, Austria). The cow was considered as the unit of interest. The significance and tendency levels were set at $P < 0.05$ and $P < 0.1$, respectively. Time to pregnancy (and pregnancy risk) was evaluated at 2 economically important cut-points: 1) 120 DIM, economically optimal time for pregnancy (i.e. higher milk sales, calf sales, lower replacement costs and lower relative reproductive costs;Cabrera, 2014), and 2) 210 DIM, economical endpoint (often uneconomical to continue breeding after this time point;Cabrera, 2014). Left censored cows (culled or dead before the end of the voluntary waiting period) were not excluded from the survival and decision models since this may be informative (associated with TD).

The association between TD and the 120 and 210 DIM pregnancy risk was evaluated as follows: 1) cows with TD-all (whether complicated with another TD or not) versus healthy cows (two groups), and 2) cows with a TD-single (uncomplicated cases), versus TD+ (complicated cases), versus healthy cows (three groups). The association between each individual TD and pregnancy risk was evaluated similarly: 1) cows with an individual TD (e.g., RP) whether complicated with another TD (e.g., RP and METR) or not, versus healthy cows (two groups), and 2) cows with an individual TD-single (e.g., RP, uncomplicated cases), versus individual TD+ (e.g. RP and METR, complicated cases), versus healthy cows (three groups). Within each cow presenting TD+, no further distinction was made between diseases (e.g., relapsed cases of the same TD or two different TD within 30 DIM). Parity was classified as primiparous and multiparous. Season of calving was included as winter, spring, summer, and fall. Cumulative milk yield to 60 DIM (**M60**) was included as a binomial variable, with cows producing milk above the median were categorised as "high" and below the median as "low" yielding cows, stratified by parity (primiparous and multiparous). Full lactation or 305 d milk yield was not included in the models since it may introduce bias in the present context: cows with TD and/or that are not pregnant may be selectively culled before completing a 305 d milk production record.

2.4.1. Survival analysis

Kaplan-Meier (package survival, function survfit; Therneau, 2015) and Cox proportional hazard models (package survival, function coxph; Therneau, 2015) were used to assess the association of TD groups (i) healthy vs. TD-all or ii) healthy vs. TD-single vs. TD + or iii) healthy vs. each of TWIN, MF, RP, METR, KET, DA, or MAST) with pregnancy risk to 120 and 210 DIM, with parity, season of calving, and M60 as covariates. Since the interaction between parity and TD-all, TD-single or TD + was significant and had a substantial effect on the pregnancy risk (120 and 210 DIM; $P < 0.01$), all survival models were stratified to account for differences by parity. The log-log survival function was plotted against the time to pregnancy to confirm that curves were parallel to test the assumption for the Kaplan-Meier models. The Schoenfeld residuals were used to check the proportional hazards assumptions (package survival, function cox.zph; Therneau, 2015). To estimate the differences between groups, the chi-square test for each model (package survival, function survdiff; Therneau, 2015) was conducted to compare the log-likelihood ratios. Multivariable Cox proportional hazard models were built to evaluate the confounding effect that season or M60 may have on TD-all, TD-single, TD+, or individual TD. First, univariable models were constructed and variables with $P < 0.2$ were offered to the multivariable models. Next, variables and first-order interactions of covariates with TD with $P < 0.05$ were retained in the final model via manual stepwise backward elimination. A multivariable Cox proportional hazard regression analysis including all individual TD in a single model was not possible to fit given the high number of categorical variables and interactions because the data were too sparse for that large number of covariate patterns. For all converged models, Goodness of Fit test by examining Schoenfeld residuals were supported by a non-significant relationship between residuals and time ($P < 0.05$).

2.4.2. Decision tree algorithms

Conditional inference classification trees (ctree function, package partykit; Hothorn et al., 2015) were built to assess the associations of TD-all, TD-single or TD + as well as parity (primiparous vs. multiparous), season of calving (winter vs. spring vs. summer vs. fall), and M60 (high vs. low) with the pregnancy risk at 120 and 210 DIM. Moreover, the effects of individual TD were also fitted in decision tree models as TWIN, MF, RP, METR, KET, DA, MAST, parity (primiparous vs. multiparous), season of calving (winter vs. spring vs. summer vs.

fall), and M60 (high vs. low). Data were randomly subdivided in a training set (a subset to train the model; 75%) and a test set (a subset to test the training model; 25%). The training set was used to fit the model and the test set was used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyperparameters. The mincriterion was set at 0.95. Sampling was done with replacement. The number of variables randomly sampled as candidates at each split (mtry) was set based on the mtry with the lowest out-of-bag error. The number of trees to be grown (ntree) was based on the stability of the error rate of each model. The minimum size of the end nodes (nodesize) was set on one, the default value for classification trees (Liaw and Wiener, 2002). To assess the performance of the decision trees as described above, test set predictions of the pregnancy outcome by 120 and 210 DIM (at the individual sample level) were computed. The prediction results were interpreted as continuous variables ranging from 0 to 1. A dummy variable based on the prediction outcome was created using a threshold level of 0.5 (prediction 0.5 = 50% chance to become pregnant). All the values > 0.5 were considered as positive predictors for pregnancy, and all variables ≤ 0.5 were considered as negative predictors for pregnancy. The dichotomized outcome was compared to the respective actual pregnancy status using the function confusionMatrix of the package caret (Kuhn, 2011) to determine accuracy, sensitivity (**Se**), specificity (**Sp**), positive predictive value (**PPV**), and negative predictive value (**NPV**). The degrees of accuracy, Se, Sp, PPV, and NPV values were considered as follows: less than 0.2 as low, 0.2 to 0.4 as fair, 0.4 to 0.6 as moderate, 0.6 to 0.8 as substantial, and 0.8 as high (Dohoo et al., 2009).

3. Results

3.1. Descriptive statistics

A total of 1946 cows were followed over 1 year, including 542 28% primiparous and 1404 72%) multiparous cows. All details of culling rates and milk production parameters were reported in previous publications using the same dataset (Hostens et al., 2012; Probo et al., 2018). The incidences of individual TD, TD-all, TD-single, and TD + by parity and corresponding 120 and 210 DIM pregnancy risks are shown in Tables 1 and 2. Fig. 1 shows the proportion of combinations of TD+, stratified by parity.

3.2. Survival analysis

Fig. 2 illustrates the effects of TD-all, TD-single, and TD + on time to pregnancy. In the univariable Kaplan-Meier analysis, parity influenced both the 120 DIM pregnancy risk (primiparous 59% (95% CI: 55–63 %), multiparous 46% (95% CI: 44–49 %); $P < 0.001$) and the 210 DIM pregnancy risk (primiparous 76% (95% CI: 72–79 %), multiparous 61% (95% CI): 59–64 %; $P < 0.001$). The interaction between parity and TD-all, TD-single or TD+ had a significant effect on the pregnancy risk (120 and 210 DIM; $P < 0.01$), therefore, all survival models were stratified to account for differences by parity. Summer was the single season with a detrimental effect on 120 DIM pregnancy risk ($P = 0.05$) for primiparous animals while for multiparous cows, season did not have an effect ($P = 0.5$). Similarly, summer tended to have a detrimental effect on the 210 DIM pregnancy risk for primiparous cows ($P = 0.06$) but did not have an effect for multiparous cows ($P = 0.7$). The M60 did not have an effect on the 120 nor 210 DIM pregnancy risk of primiparous cows ($P < 0.2$). Multiparous cows with low M60 tended to have a reduced pregnancy risk at 120 DIM (M60 low 49% (95% CI: 45–53 %) vs. M60 high 56% (95% CI: 52–60 %); $P < 0.06$) and at 210 DIM the pregnancy risk was lower for low M60 in multiparous cows (M60 low 64% (95% CI: 61–68 %) vs. M60 high 74% (95% CI: 70–77 %); $P < 0.03$). The Kaplan-Meier estimates indicate that in primiparous cows, the 120 DIM pregnancy risk was 62% (95% CI: 57–67 %) for healthy animals, and 55% (95% CI: 48–62 %; $P = 0.04$) for cows with TD-all. The 120 DIM pregnancy risk for primiparous cows with TD-single or TD+ were 58% (95% CI: 51–66 %; $P = 0.21$) and 45% (95% CI: 33–60 %; $P = 0.02$), respectively. The 210 DIM pregnancy risk for healthy primiparous cows was 80% (95% CI: 75–84 %), while it was 70% (95% CI: 64–76 %; $P = 0.006$) for primiparous cows with TD-all. Similarly, the 210 DIM pregnancy risk for primiparous cows with TD-single or TD+ were 72% (95% CI: 65–79 %; $P = 0.06$) and 62% (95% CI: 49–75 %; $P = 0.005$), respectively. The Kaplan-Meier estimates indicated that in multiparous cows, the 120 DIM pregnancy risk was 53% (95% CI: 49–56 %) for healthy animals, and 34% (95% CI: 30–39 %; $P < 0.001$) for cows with TD-all. The 120 DIM pregnancy risk for multiparous cows with TD-single or TD+ were 36% (95% CI: 31–42 %; $P < 0.001$) and 30% (95% CI: 24–38 %; $P < 0.001$), respectively. The 210 DIM pregnancy risk for healthy multiparous cows was 70% (95% CI: 67–72 %) while it was 47% (95% CI: 43–52 %; $P < 0.001$) for multiparous cows with TD-all. Similarly, the 210 DIM pregnancy risk for multiparous cows with TD-single or TD + were 47% (95% CI: 42–53 %; $P < 0.001$) and 46% (95% CI: 38–54 %; $P < 0.001$), respectively.

Supplementary Table S1 presents the results of the multivariable Cox proportional hazard ratios for pregnancy for TD-all in primiparous and multiparous cows at 120 and 210 DIM. Supplementary Tables S2 and S3 show the results of the multivariable Cox proportional hazard ratios for pregnancy for TD-single and TD + in primiparous and multiparous, respectively. Data in Table 3 show the overall and individual

Table 1
Descriptive statistics showing the incidence of individual transition cow conditions and diseases (TD) and the pregnancy by 120 and 210 DIM for primiparous dairy cows (n = 542) from 1 herd over 1 year. The 120 and 210 DIM pregnancy for healthy primiparous cows was 62 and 82%, respectively.

Variable[1]	Lactational incidence risk[2]	120 DIM pregnancy risk[2]	210 DIM pregnancy risk[2]	TD-all incidence[3]	120 DIM pregnancy risk[3]	210 DIM pregnancy risk[3]
TWIN	0.4% (n = 2)	50% (n = 1)	50% (n = 1)	0.9% (n = 5)	40% (n = 2)	40% (n = 2)
TWIN+	0.5% (n = 3)	33.3% (n = 1)	33.3% (n = 1)			
RP	1.3% (n = 7)	57.1% (n = 4)	71.4% (n = 5)	3.5% (n = 19)	52.6% (n = 10)	63.1% (n = 12)
RP+	2.2% (n = 12)	50% (n = 6)	58.3% (n = 7)			
METR	20.7% (n = 112)	61.6% (n = 69)	78.5% (n = 88)	28.2% (n = 153)	58.8% (n = 90)	75.1% (n = 115)
METR+	7.5% (n = 41)	51.2% (n = 21)	65.8% (n = 27)			
KET	3.9% (n = 21)	71.4% (n = 15)	71.4% (n = 15)	8.8% (n = 48)	41.6% (n = 20)	70.8% (n = 34)
KET+	4.9% (n = 27)	55.5% (n = 15)	70.3% (n = 19)			
DA	0.4% (n = 2)	0% (n = 0)	50% (n = 1)	2.2% (n = 12)	16.6% (n = 2)	50% (n = 6)
DA+	1.8% (n = 10)	20% (n = 2)	50% (n = 5)			
MAST	2.8% (n = 15)	26.6% (n = 4)	40% (n = 6)	4.6% (n = 25)	32% (n = 8)	48% (n = 12)
MAST+	1.8% (n = 10)	40% (n = 4)	60% (n = 6)			

TWIN = twinning; RP = retained placenta; METR = metritis; KET = ketosis; DA = displaced abomasum; MAST = mastitis. + = multiple TD.
[1]From univariable analyses.
[2]Uncomplicated (TD-single) and complicated (TD+) cases, separately.
[3]Uncomplicated and complicated TD cases, together (TD-all).

Table 2
Descriptive statistics showing the incidence of transition cow conditions and diseases TD and the pregnancy by 120 and 210 DIM for multiparous dairy cows n = 1,404 from 1 herd over 1 year. The 120 and 210 DIM pregnancy for healthy multiparous cows was 53 and 70%, respectively.

Variable[1]	Lactational incidence risk[2]	120 DIM pregnancy risk[2]	210 DIM pregnancy risk[2]	TD-all incidence[3]	120 DIM pregnancy risk[3]	210 DIM pregnancy risk[3]
TWIN	2.2% (n = 31)	35.4% (n = 11)	48.3% (n = 15)	5.8% (n = 81)	35.8% (n = 29)	50.6% (n = 41)
TWIN+	3.6% (n = 50)	36% (n = 18)	52% (n = 26)			
MF	5.5% (n = 78)	44.8% (n = 35)	48.7% (n = 38)	9.1% (n = 128)	38.2% (n = 49)	44.5% (n = 57)
MF+	3.6% (n = 50)	28% (n = 14)	38% (n = 19)			
RP	0.9% (n = 13)	15.3% (n = 2)	46.1% (n = 6)	3.9% (n = 55)	29.1% (n = 16)	45.4% (n = 25)
RP+	3.0% (n = 42)	33.3% (n = 14)	45.2% (n = 19)			
METR	4.1% (n = 58)	39.6% (n = 23)	55.1% (n = 32)	9.5% (n = 134)	34.3% (n = 46)	51.4% (n = 69)
METR+	5.4% (n = 76)	30.2% (n = 23)	48.6% (n = 37)			
KET	4.8% (n = 67)	34.3% (n = 23)	46.2% (n = 31)	11.2% (n = 158)	27.2% (n = 43)	41.1% (n = 65)
KET+	6.4% (n = 91)	21.9% (n = 20)	37.3% (n = 34)			
DA	1.0% (n = 14)	21.4% (n = 3)	35.7% (n = 5)	3.6% (n = 51)	23.5% (n = 12)	41.1% (n = 21)
DA+	2.6% (n = 37)	24.3% (n = 9)	43.2% (n = 16)			
MAST	4.0% (n = 56)	35.7% (n = 20)	44.6% (n = 25)	5.8% (n = 82)	34.1% (n = 28)	45.1% (n = 37)
MAST+	1.8% (n = 26)	30.7% (n = 8)	46.1% (n = 12)			

TWIN = twinning; RP = retained placenta; METR = metritis; KET = ketosis; DA = displaced abomasum; MAST = mastitis. + = multiple TD.
[1]From univariable analyses.
[2]Uncomplicated (TD-single) and complicated (TD+) cases, separately.
[3]Uncomplicated and complicated TD cases, together (TD-all).

Cox hazard proportional ratio of each TD (TWIN, MF, RP, METR, KET, DA, and MAST) non-complicated and complicated cases (together), on the pregnancy risk at both 120 and 210 DIM (stratified by parity). Tables 4 and 5 illustrate the overall and individual Cox proportional hazard ratios of each TD (TWIN, MF, RP, METR, KET, DA, and MAST) either uncomplicated or complicated with another TD, on the pregnancy risk at 120 and 210 DIM (stratified by parity).

3.3. Decision tree algorithms

The conditional inference decision trees show a hierarchical order of the association between TD-all, TD-single or TD+, season, parity, and M60 with the time to pregnancy to 120 and 210 DIM. For both 120 and 210 DIM decision trees for TD-all (Supplementary Fig. S4), parity was the most influential variable affecting fertility (lower pregnancy risk in multiparous cows). The effect of season (higher fertility in fall, lower in summer) had a significant influence on reproductive performance in primiparous cows. In the conditional inference decision trees for TD-single and TD+ (Supplementary Fig. S6), TD + was identified as the variable with the greatest effect on fertility in multiparous cows at both 120 and 210 DIM. Complicated TD was also harmful for primiparous cows in the summer (210 DIM). For individual TD, the conditional inference decision trees as presented in Fig. 3 indicated that parity was the most influential variable affecting fertility (lower pregnancy risk in multiparous cows) at 120 and 210 DIM. Displaced abomasum affected fertility at 120 DIM in primiparous cows followed by the seasonal effect of summer at both 120 and 210 DIM. For multiparous cows, METR was the most influential TD in the short and long term. Table 6 indicates the accuracy, Se, Sp, PPV, and NPV between the decision tree modeling predictions and the actual pregnancy status by 120 and 210 DIM for individual TD. Tables S5 and S7 summarize the performance values as described above for the decision tree of the TD-all, TD-single, and TD + models.

4. Discussion

This study explored the short- and long term effects of individual and multiple TD on the pregnancy risk via Kaplan-Meier, Cox hazard proportional, and decision tree modeling. Survival analysis and decision tree models indicated parity as the most influential variable affecting fertility (above TD-all, TD-single or TD+) with a higher pregnancy risk in primiparous cows at both 120 and 210 DIM. Both modeling techniques concurred that TD + had a greater effect than TD-single on pregnancy at the short- and long term for primiparous and multiparous cows, and season had an important effect on the pregnancy risk of primiparous cows (lower fertility in summer and higher fertility in fall). For individual TD, multivariable models could not be fit (no model convergence) given the high number of categorical variables and their interactions because the data were too sparse to adequately fill all covariate patterns. Decision trees for individual TD identified that DA affected fertility at 120 DIM in primiparous while METR was the most influential TD at 120 and 210 DIM for multiparous cows.

During the evaluation of interactions for TD-all, TD-single and TD +, parity was identified to have a major effect on the pregnancy risk at 120 and 210 DIM. Instead of including the interaction of parity with every TD in the Cox proportional hazard we stratified the models by parity (primiparous and multiparous). For the decision trees, parity was the component of all the ‘root nodes’. These findings indicate that parity was the most influential variable associated with fertility, surpassing any TD in these data. The probability of pregnancy was higher for healthy or TD-all primiparous cows than healthy multiparous cows. Independently of TD, studies performed in Europe similarly found better fertility in primiparous cows (Inchaisri et al., 2010; Pascottini et al., 2017). Interestingly, the negative effect of TD on fertility was always more pronounced in multiparous than primiparous cows. This may be associated with a concomitant, higher metabolic demand in multiparous than primiparous cows due to greater milk yield (Wathes et al., 2007; Walsh et al., 2011). Briefly, elevated metabolic stress due to increased milk production is often translated to a postpartum metabolic profile characterized by an altered lipid and energy metabolism status and reduced neutrophil function as reviewed by Ingvartsen and Moyes (2013). This may have played a role in the greater incidence of TD and a more significant (negative) effect of TD on fertility in multiparous cows.

The survival analysis indicated that when a TD was complicated with another TD (TD+), the effect on fertility was always more pronounced. For the decision trees, only TD + appeared to have an effect on fertility, especially in multiparous cows. Ribeiro et al. (2006) demonstrated the carryover consequences of inflammatory diseases in the postpartum period on the reproductive performance by reducing fertilization and development, and by increasing early embryonic death. They demonstrated that cows with TD had an impaired conceptus

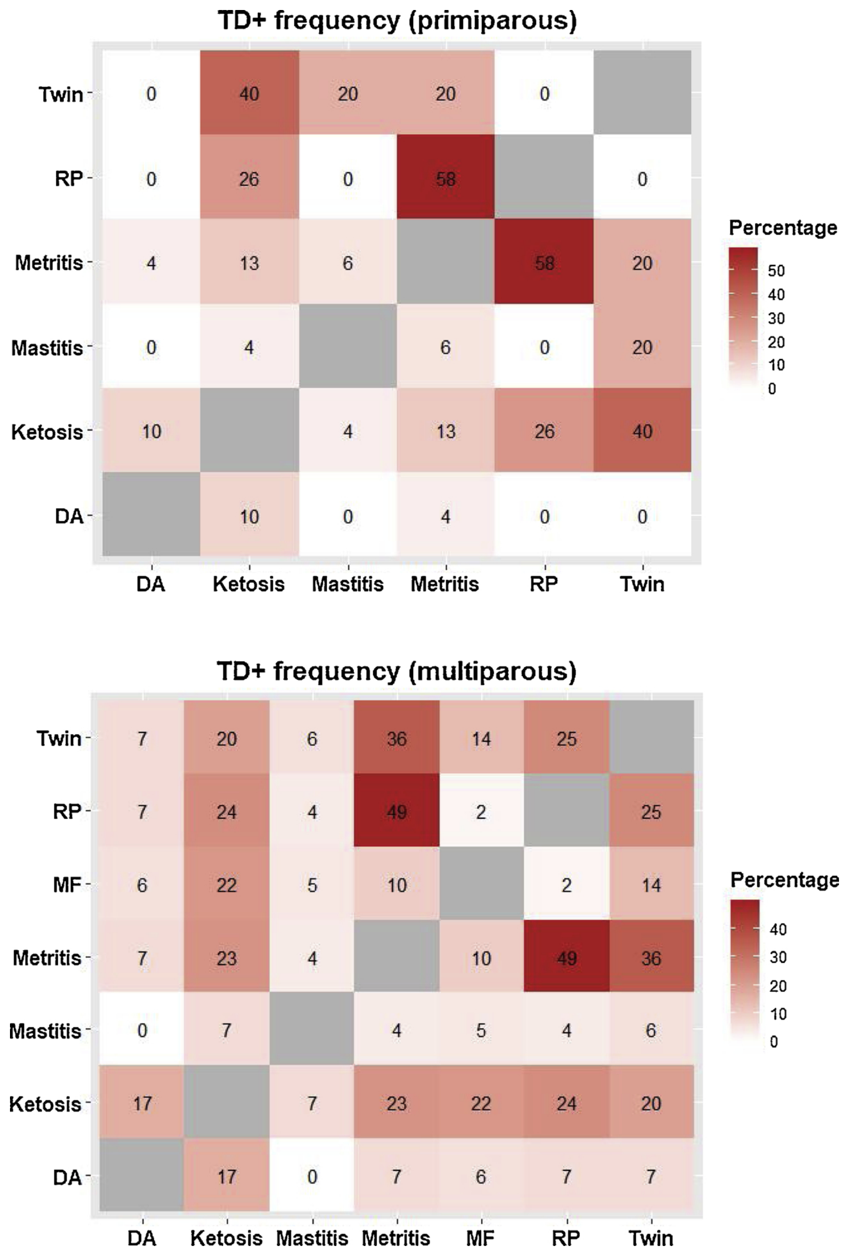

Fig. 1. Heatmaps showing the proportion of combinations of transition cow conditions and diseases TD diagnosed during the transition period from 1 heard over 1 year. The heatmap in the top is showing the proportion of combinations of TD + in primiparous cows n = 542 and the heatmap in the bottom the proportion of combinations of TD + in multiparous cows n = 1404.

elongation and reduced secretion of interferon-τ. These changes were concurrent with an altered gene expression profile of the conceptus that was also associated with the increased pregnancy loss. In the present study we found an inverse proportional effect of greater inflammatory status (TD+) with lower pregnancy risk. Future studies should further evaluate the pathophysiology behind the intensity of the inflammatory condition and its proportional deleterious effect on the pre-implantation embryo physiology.

The seasonal effect for primiparous cows (lower fertility in summer, higher fertility in fall) was congruent for survival analysis and the decision trees. Generally, for most of the variables analysed in this study we found agreement between traditional modeling techniques and machine learning, with the exception of M60. Our multivariable Cox proportional hazard models interestingly showed a link between the presence of TD-all, TD-single or TD + accompanied with low M60 with lower fertility in the short- and long term, in multiparous cows. Using the same dataset, Hostens et al. (2012) similarly demonstrated that the lactation curve profile was affected in TD-all, TD-single or TD + in comparison to healthy cows. Decision trees did not identify M60 as an influential variable associated with fertility.

Decision tree models indicated that METR was the most influential TD affecting fertility for multiparous cows in the short- and long term. Univariable Cox hazard proportional models revealed that the effect of METR (complicated or uncomplicated cases) on pregnancy at 120 DIM was not significant in primiparous cows, but METR + had a significant effect on pregnancy at 210 DIM in primiparous cows. Displaced abomasum was often complicated with another TD (DA+), specially KET (Fig. 1). Subsequently, DA + had a profound effect on the short- and

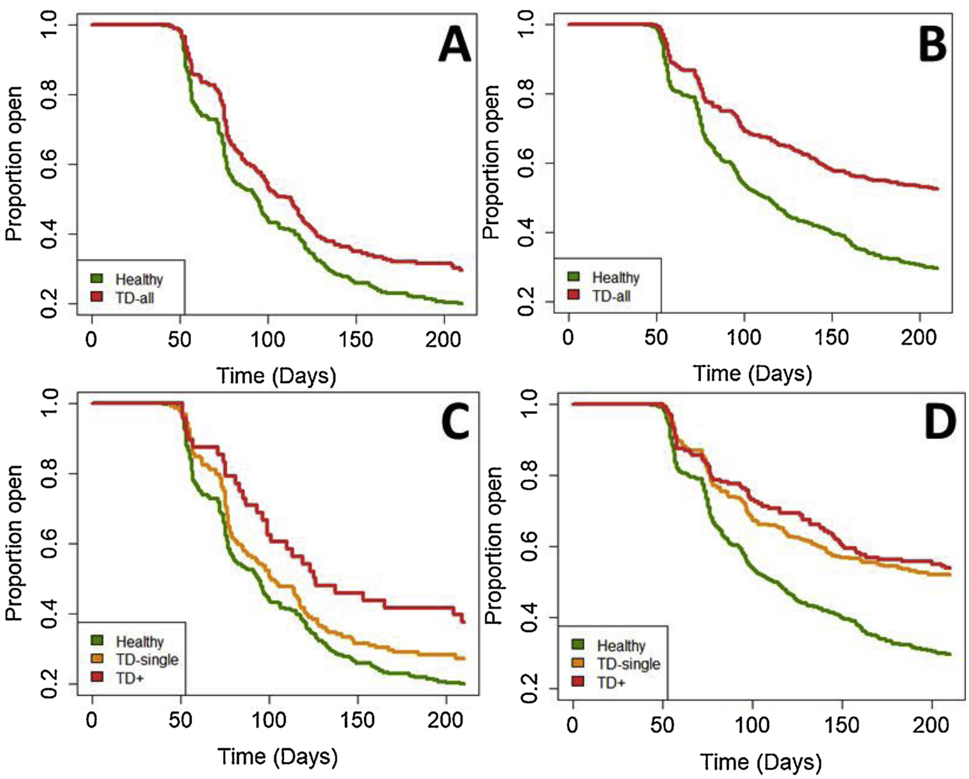

Fig. 2. Kaplan-Meier plots of time to pregnancy to 210 DIM. Top row (A and B), cows with transition cow conditions and diseases (TD) whether complicated with another TD or not (TD-all; red line), versus healthy cows (green line). Bottom row (C and D), cows with uncomplicated TD (TD-single; orange line), versus complicated cases of TD (TD+; red line), versus healthy cows (green). The left-hand column represents primiparous (A and C) while the right-hand column multiparous cows (B and D). These analyses considered only the effect of TD (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

long term fertility in both primiparous and multiparous cows in the univariable Cox hazard proportional models. Interestingly, DA was the most influential TD at 120 DIM for primiparous cows by the decision tree. Only two out of twelve primiparous cows with DA became pregnant by 120 DIM. This may be associated with the high culling rate in DA cows as described by Probo et al. (2018) using the same dataset. The "roll-and-toggle" method to treat DA might not be ideal since recurrent DA cases may cause additional distress. Ideally, cows diagnosed with DA should be immediately treated with abomasopexy. Consequently, the effect of DA on fertility should be carefully interpreted. Cox proportional hazard models revealed an association of KET with the 120 and 210 DIM pregnancy risk in multiparous cows (uncomplicated and complicated cases). Ketosis is well known to negatively affect the oocyte's microenvironment, which has significant carryover effects on embryo quality and viability (Aardema et al., 2011; Van Hoeck et al., 2011). However, in the Cox proportional hazard models for

Table 3

Univariable Cox hazard proportional models showing the effects of transition cow conditions and diseases (TD) on time to pregnancy to 120 and 210 DIM stratified by parity. Each TD includes uncomplicated and complicated cases (together; TD-all). Models considered only the effect of TD.

Variable[1]	Primiparous (n = 542)			Multiparous (n = 1,404)		
	Hazard ratio[2]	95% CI	*P*-value	Hazard ratio[2]	95% CI	*P*-value
Healthy	Referent			Referent		
TD 120 DIM	0.8	0.6-0.9	0.04	0.6	0.5-0.7	< 0.001
TD 210 DIM	0.7	0.6-0.9	0.006	0.5	0.4-0.6	< 0.001
TWIN 120 DIM[3]	0.5	0.1-2.0	0.34	0.6	0.4-0.9	0.005
TWIN 210 DIM[3]	0.3	0.1-1.3	0.13	0.5	0.4-0.8	< 0.001
MF 120 DIM	–	–	–	0.6	0.4-0.7	< 0.001
MF 210 DIM	–	–	–	0.5	0.4-0.6	< 0.001
RP 120 DIM[3]	0.8	0.4-1.4	0.39	0.5	0.3-0.8	0.002
RP 210 DIM[3]	0.6	0.3-1.2	0.18	0.5	0.3-0.7	< 0.001
METR 120 DIM[3]	0.8	0.7-1.1	0.22	0.6	0.5-0.8	< 0.001
METR 210 DIM[3]	0.8	0.6-1.0	0.11	0.6	0.4-0.7	< 0.001
KET 120 DIM[3]	0.9	0.6-1.3	0.53	0.4	0.3-0.6	< 0.001
KET 210 DIM[3]	0.7	0.5-1.0	0.14	0.4	0.3-0.5	< 0.001
DA 120 DIM	0.2	0.1-0.7	0.01	0.4	0.2-0.6	< 0.001
DA 210 DIM	0.3	0.1-0.8	0.01	0.4	0.2-0.6	< 0.001
MAST 120 DIM	0.4	0.2-0.8	0.01	0.6	0.4-0.8	0.005
MAST 210 DIM	0.4	0.2-0.7	0.002	0.5	0.3-0.7	< 0.001

[1]TWIN = twinning; MF = milk fever; RP = retained placenta; METR = metritis; KET = ketosis; DA = displaced abomasum; MAST = mastitis.
[2]Hazard ratio (HR) = Relative risk (constantly proportional) of pregnancy per day, compared to unaffected cows (the implied hazard ratio for the referent is 1, so HR < 1 indicates lesser pregnancy risk per day, i.e., longer time to pregnancy).
[3]Retrospective calculation of statistical power (post-hoc analysis) shows < 80% of power in primiparous cows.

Table 4
Univariable Cox hazard proportional models showing the effects of transition cow conditions and diseases (TD) on time to pregnancy to 120 DIM stratified by parity. Each TD includes uncomplicated (TD-single) and complicates cases (TD +), separately. Models considered only the effect of TD.

Variable[1]	Primiparous (n = 542)			Multiparous (n = 1404)		
	Hazard ratio[2]	95% CI	*P*-value	Hazard ratio[2]	95% CI	*P*-value
Healthy	Referent			Referent		
TD	0.8	0.7-1.0	0.21	0.6	0.5-0.7	< 0.001
TD +	0.6	0.4-0.9	0.02	0.5	0.3-0.7	< 0.001
TWIN[3]	0.7	0.1-5.2	0.76	0.6	0.3-1.0	0.07
TWIN + [3]	0.4	0.1-2.7	0.34	0.6	0.4-0.9	0.02
MF	–	–	–	0.7	0.5-0.9	0.04
MF +	–	–	–	0.4	0.2-0.7	< 0.001
RP[3]	0.8	0.3-2.0	0.61	0.2	0.1-0.9	0.03
RP + [3]	0.7	0.3-1.7	0.49	0.5	0.3-0.9	0.02
METR[3]	0.9	0.7-1.2	0.58	0.7	0.4-1.0	0.05
METR + [3]	0.7	0.4-1.1	0.10	0.5	0.3-0.7	< 0.001
KET[3]	1.1	0.6-1.8	0.69	0.5	0.4-0.8	< 0.001
KET + [3]	0.7	0.4-1.2	0.24	0.3	0.2-0.5	< 0.001
DA	–	–	–	0.3	0.1-0.9	0.04
DA +	0.1	0.1-0.8	0.03	0.4	0.2-0.7	0.002
MAST	0.3	0.1-0.9	0.02	0.6	0.4-0.9	0.02
MAST + [3]	0.5	0.2-1.3	0.16	0.5	0.3-1	0.08

[1]TWIN = twinning; MF = milk fever; RP = retained placenta; METR = metritis; KET = ketosis; DA = displaced abomasum; MAST = mastitis.
[2]Hazard ratio (HR) = Relative risk (constantly proportional) of pregnancy per day, compared to unaffected cows (the implied hazard ratio for the referent is 1, so HR < 1 indicates lesser pregnancy risk per day, i.e., longer time to pregnancy).
[3]Retrospective calculation of statistical power (post-hoc analysis) shows < 80% of power in primiparous cows.

Table 5
Univariable Cox hazard proportional models showing the effects of transition cow conditions and diseases (TD) on time to pregnancy to 210 DIM stratified by parity. Each TD includes uncomplicated and complicates cases, separately. Models considered only the effect of TD.

Variable[1]	Primiparous (n = 542)			Multiparous (n = 1404)		
	Hazard ratio[2]	95% CI	*P*-value	Hazard ratio[2]	95% CI	*P*-value
Healthy	Referent			Referent		
TD	0.8	0.6-1.0	0.05	0.5	0.4-0.6	< 0.001
TD +	0.5	0.4-0.8	0.005	0.4	0.3-0.6	< 0.001
TWIN[3]	0.5	0.1-3.5	0.49	0.5	0.3-0.9	0.02
TWIN + [3]	0.2	0.1-1.8	0.18	0.6	0.4-0.9	0.01
MF	–	–	–	0.5	0.4-0.8	0.001
MF +	–	–	–	0.4	0.2-0.6	< 0.001
RP[3]	0.7	0.3-1.8	0.56	0.4	0.2-1.0	0.05
RP + [3]	0.6	0.3-1.3	0.21	0.5	0.3-0.8	0.004
METR[3]	0.9	0.7-1.1	0.48	0.6	0.4-0.9	0.02
METR +	0.6	0.4-0.9	0.03	0.5	0.3-0.7	< 0.001
KET[3]	0.8	0.5-1.4	0.58	0.5	0.3-0.7	< 0.001
KET + [3]	0.7	0.4-1.1	0.14	0.3	0.2-0.5	< 0.001
DA[3]	0.3	0.1-2.5	0.3	0.3	0.1-0.8	0.02
DA +	0.3	0.1-0.8	0.02	0.4	0.2-0.7	0.001
MAST	0.3	0.1-0.7	0.007	0.5	0.3-0.7	0.001
MAST + [3]	0.5	0.2-1.2	0.13	0.5	0.3-0.9	0.04

[1]TWIN = twinning; MF = milk fever; RP = retained placenta; METR = metritis; KET = ketosis; DA = displaced abomasum; MAST = mastitis.
[2]Hazard ratio (HR) = Relative risk (constantly proportional) of pregnancy per day, compared to unaffected cows (the implied hazard ratio for the referent is 1, so HR < 1 indicates lesser pregnancy risk per day, i.e., longer time to pregnancy).
[3]Retrospective calculation of statistical power (post-hoc analysis) shows < 80% of power in primiparous cows.

primiparous cows, KET and KET + were not significantly associated with the 120 and 210 DIM pregnancy risk.

Mastitis was associated with lower fertility in both uncomplicated and complicated cases in multiparous cows (univariable Cox hazard proportional models; 120 and 210 DIM pregnancy risk). The negative effect of MAST on the insemination outcome is well documented (Dahl et al., 2017). Moreover, in most cases, cows with chronic MAST are deliberately chosen not to be bred, or they leave the farm (sold) before 210 DIM (Santos et al., 2004). In primiparous cows, the lack of detection of an effect of MAST in the univariable Cox hazard proportional models is probably due to the low incidence of cases in this group of animals or by the efficient diagnosis and prompt treatment for MAST adopted in this dairy. Similar to other studies, MF affected around 10% of multiparous cows (Fourichon et al., 1999 and DeGaris and Lean, 2008) and its effect on fertility was mainly associated with the elevated number of left-censored cows (Dohoo and Martin, 1984 and Milian-Suazo et al., 1988). As indicated by Probo et al. (2018) and using the same dataset, the decision tree indicated MF as the most influential variable associated with the 120 DIM culling risk (40% of culling in the first 120 DIM). Milk fever was strongly associated with other TD, such that more than 50% of MF cases suffered from other TD and a markedly reduced pregnancy both at 120 and 210 DIM. Calcium metabolism has numerous effects on metabolism through disruptions in the homeostatic and homeorhetic metabolism that contribute to the development of secondary TD.

For the univariable Cox hazard proportional models, the adverse effect of TWIN on fertility was only visible in multiparous cows and the effect was similar in both TWIN and TWIN + . The lack of effect of TWIN in primiparous cows may be associated with its low incidence in this group of animals. A risk factor associated with TWIN, RP was only significantly associated with fertility in multiparous cows (univariable Cox hazard proportional models; short- and long term effect). The lack of statistical power (< 80%) due to the low overall incidence of the disease and the low proportion of complicated cases may have influenced the associations detected. Similarly, TWIN and RP did not appear in the decision tree plots, indicative of the secondary importance (probably associated with their inherent low incidence) of these TD on fertility as compared to the others TD evaluated in this study.

Predictive values for models evaluating 210 DIM estimates of pregnancy were greater than those for 120 DIM. In general, pregnancy predictions had a moderate accuracy with a low Se, but a high Sp. The NPV and PPV were always on a fair-to-moderate scale. Thus, our decision tree algorithms were accurate to predict that those cows presenting TD in the transition period had high chances to remain open at 210 DIM. However, presence (or absence) of TD in the transition period was not a good predictor for pregnancy. While disease is generally accepted as an important event in a cow's life, it should be unsurprising that the predictive accuracy of disease for pregnancy was low, because many other factors influence reproduction including semen quality, insemination technique, heat detection efficiency and accuracy, delayed ovulation, uterine health, nutrition, heat stress, and genetics.

An interesting approach of the present study is that the dairy was equipped with a TMF, where special efforts were made to accurately diagnose all possible cases of TD. However, lack of association of some individual TD, especially for primiparous cows, should be interpreted with caution due to their low incidence and the limited number of animals. Nevertheless, an added value offered by machine-learning modeling is that the low incidence of TD plays a role in the hierarchical order of importance of TD associated with fertility. So, we were able to hierarchically order TD and covariates, accounting for the prevalence of each disease. This should be helpful in identifying and prioritizing conditions that are influential for fertility at the herd level, rather than just having a large effect in the small proportion of affected cows. Still,

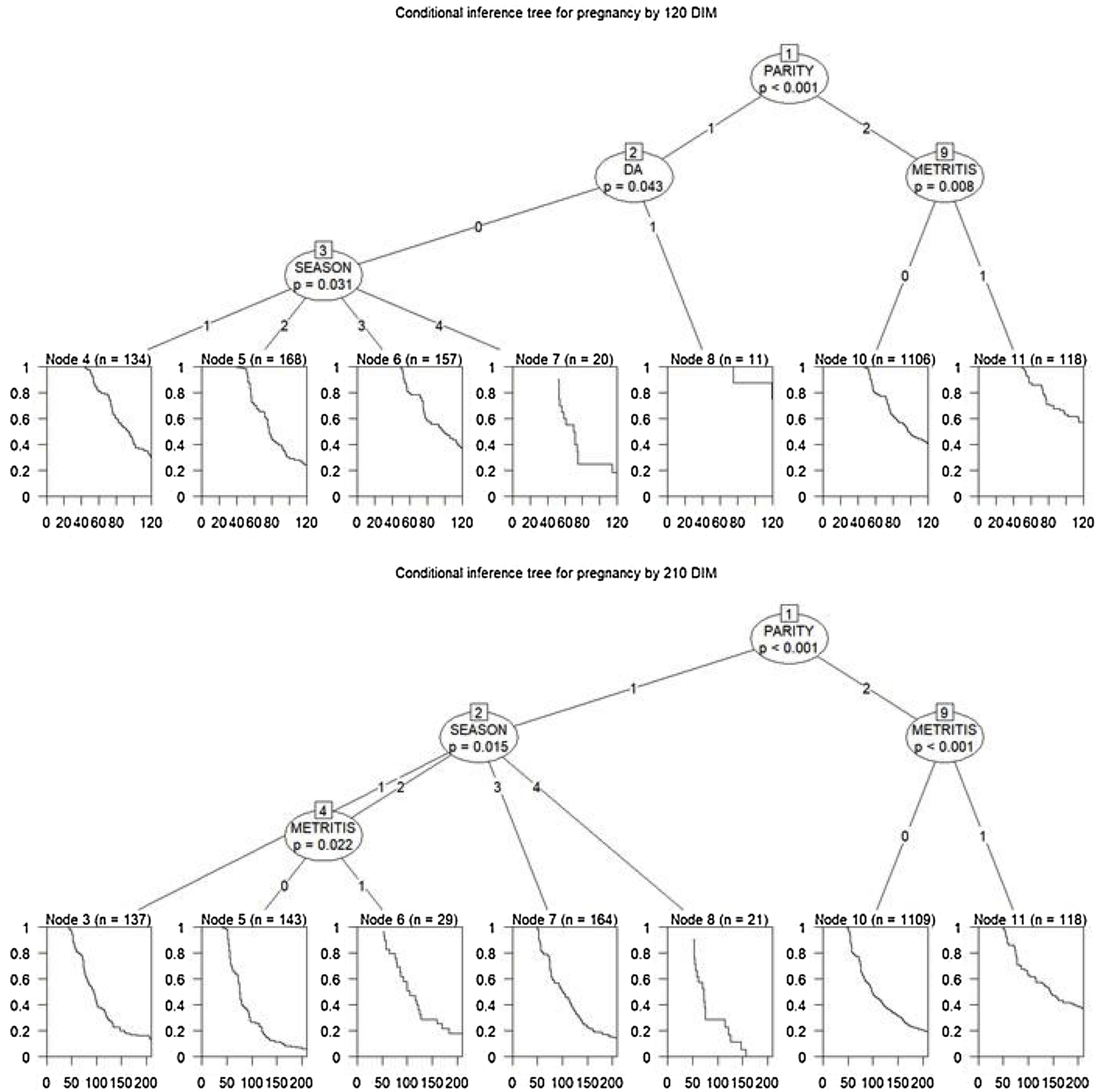

Fig. 3. Conditional inference trees showing the hierarchical effects of parity (1 = primiparous and 2 = multiparous), season (1 = winter, 2 = spring, 3 = summer, and 4 = fall), and individual transition cow conditions and diseases (TD) on the pregnancy by 120 and 210 DIM. For both 120 and 210 DIM, parity was the most influential variable affecting fertility (lower pregnancy risk in multiparous cows). Displaced abomasum (DA) affected fertility at 120 DIM in primiparous followed by the seasonal effect of summer (120 and 210 DIM). For multiparous cows, metritis was the most influential TD at the short and long term (120 and 210 DIM). The number in the node size corresponds to the number of animals that were randomly selected to fit the decision tree algorithm.

Table 6
Performance of decision tree algorithms. The decision tree outcomes of individual transition cow conditions and diseases were used to compute predictions on the pregnancy by 120 and 210 DIM. Predictions < 0.5 were categorised as pregnant and at ≤ 0.5 as non-pregnant and compared to the actual pregnancy status.

Variable (%)	Decision tree 120 DIM	Decision tree 210 DIM
Accuracy	60	74
Sensitivity	13	15
Specificity	95	95
PPV[1]	68	53
NPV[2]	59	76

[1]Positive predictive value.
[2]Negative predictive value.

it would be noteworthy to explore TD using much larger data sets to tease out the most frequent combinations of TD + that have the greatest impact on fertility.

5. Conclusion

Survival analysis and decision outcomes were similar for most of the variables evaluated in this study. The effect of parity on the 120 and 210 DIM pregnancy risk was more pronounced than any TD. Multiple TD had an additive negative effect on reproductive performance. Metritis in multiparous cows was the most important individual TD associated with reduced fertility at 120 and 210 DIM. For primiparous cows, DA was the most influential individual TD at the short-term. Machine learning methods are a valid approach to evaluate the hierarchical effect of complex data, especially in situations where it is not possible to fit multivariable survival models due to the high number of categorical variables and interactions with missing values. Machine learning helped to explore interactions between individual TD to study their hierarchical effect on fertility, identifying conditional relationships that merit further investigation.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.prevetmed.2020.104908.

References

Aardema, H., Vos, P.L.A.M., Lolicato, F., Roelen, B.A.J., Knijn, H.M., Vaandrager, A.B., Helms, J.B., Gadella, B.M., 2011. Oleic acid prevents detrimental effects of saturated fatty acids on bovine oocyte developmental competence. Biol. Reprod. 85, 62–69.
Beaudeau, F., Henken, A., Fourichon, C., Frankena, K., Seegers, H., 1993. Associations between health disorders and culling of dairy cows: a review. Livest. Prod. Sci. 35, 213–236.
Beaudeau, F., Ducrocq, V., Fourichon, C., Seegers, H., 1995. Effect of disease on length of productive life of French Holstein cows assessed by survival analysis. J. Dairy Sci. 78, 103–117.
Bradford, B.J., Yuan, K., Farney, J.K., Mamedova, L.K., Carpenter, A.J., 2015. Invited review: inflammation during the transition to lactation: new adventures with an old flame. J. Dairy Sci. 98, 6631–6650.
Cabrera, V.E., 2014. Economics of fertility in high-yielding dairy cows on confined TMR systems. Animal 8, 211–221.
Caraviello, D.Z., Weigel, K.A., Craven, M., Gianola, D., Cook, N.B., Nordlund, K.V., Fricke, P.M., Wiltbank, M.C., 2006. Analysis of reproductive performance of lactating cows on large dairy farms using machine learning algorithms. J. Dairy Sci. 89, 4703–4722.
Cavero, D., Tölle, D., Henze, K.H., Buxadé, C., Krieter, J., 2008. Mastitis detection in dairy cows by application of neural networks. Livest. Sci. 114, 280–286.
Chapinal, N., Carson, M.E., LeBlanc, S.J., Leslie, K.E., Godden, S., Capel, M., Santos, J.E., Overton, M.W., Duffield, T.F., 2012. The association of serum metabolites in the transition period with milk production and early-lactation reproductive performance. J. Dairy Sci. 95, 1301–1309.
Chiumia, D., Chagunda, M.G.G., Macrae, A.I., Roberts, D.J., 2013. Predisposing factors for involuntary culling in Holstein-Friesian dairy cows. J. Dairy Res. 80, 45–50.
Dahl, M.O., Maunsell, F.P., De Vries, A., Galvão, K.N., Risco, C.A., Hernandez, J.A., 2017. Evidence that mastitis can cause pregnancy loss in dairy cows: a systematic review of observational studies. J. Dairy Sci. 100, 8322–8329.
DeGaris, P.J., Lean, I.J., 2008. Milk fever in dairy cows: a review of pathophysiology and control principles. Vet. J. 176, 58–69.
Dewey, S.T., Mendonca, L.G., Lopes Jr, G., Rivera, F.A., Guagnini, F., Chebel, R.C., Bilby, T.R., 2010. Resynchronization strategies to improve fertility in lactating dairy cows utilizing a presynchronization injection of GnRH or supplemental progesterone: I. Pregnancy rates and ovarian responses. J. Dairy Sci. 93, 4086–4095.
Dohoo, T.R., Martin, S.W., 1984. Disease, production and culling in Holstein-Friesian cows. IV. Effects of disease on production. Prev. Vet. Med. 2, 775.
Dohoo, I., Martin, W., Stryhn, H., 2009. Veterinary Epidemiological Research. Charlottetown, Prince Edward Island, Canada.
Dubuc, J., Denis-Robichaud, J., 2017. A dairy herd-level study of postpartum diseases and their association with reproductive performance and culling. J. Dairy Sci. 100, 3068–3078.
Dubuc, J., Duffield, T.F., Leslie, K.E., Walton, J.S., LeBlanc, S.J., 2010. Risk factors for postpartum uterine diseases in dairy cows. J. Dairy Sci. 93, 5764–5771.
Firk, R., Stamer, E., Junge, W., Krieter, J., 2003. Improving oestrus detection by combination of activity measurements with information about previous oestrus cases. Livest. Prod. Sci. 82, 97–103.
Fourichon, C., Seegers, H., Bareille, N., Beaudeau, F., 1999. Effects of disease on milk production in the dairy cow: a review. Prev. Vet. Med. 41, 1–35.
Grummer, R.R., 1995. Impact of changes in organic nutrient metabolism on feeding the transition dairy cow. J. Anim. Sci. 73, 2820–2833.
Hammon, D.S., Evjen, I.M., Dhiman, T.R., Goff, J.P., Walters, J.L., 2006. Neutrophil function and energy status in Holstein cows with uterine health disorders. Vet. Immunol. Immunopathol. 113, 21–29.
Hermans, K., Waegeman, W., Opsomer, G., Van Ranst, B., De Koster, J., Van Eetvelde, M., Hostens, M., 2017. Novel approaches to assess the quality of fertility data stored in dairy herd management software. J. Dairy Sci. 100, 4078–4089.
Hostens, M., Ehrlich, J., Van Ranst, B., Opsomer, G., 2012. On-farm evaluation of the effect of metabolic diseases on the shape of the lactation curve in dairy cows through the MilkBot lactation model. J. Dairy Sci. 95, 2988–3007.
Hothorn, T., Hornik, K., Zeileis, A., 2015. ctree: Conditional Inference Trees. https://CRAN.R-project.org/package=partykit.
Inchaisri, C., Hogeveen, H., Vos, P., Van der Weijden, G., Jorritsma, R., 2010. Effect of milk yield characteristics, breed, and parity on success of the first insemination in Dutch dairy cows. J. Dairy Sci. 93, 5179–5187.
Ingvartsen, K.L., Moyes, K., 2013. Nutrition, immune function and health of dairy cattle. Animal. 7, 112–122.
Jánosi, S., Kulcsár, M., Kóródi, P., Kátai, L., Reiczigel, J., Dieleman, S.J., Nikolic, J.A., Sályi, G., Ribiczey-Szabó, P., Huszenicza, G., 2003. Energy imbalance related predisposition to mastitis in group-fed high-producing postpartum dairy cows. Acta Vet. Hung. 51, 409–424.
Kamphuis, C., Mollenhorst, H., Feelders, A., Pietersma, D., Hogeveen, H., 2010a. Decision-tree induction to detect clinical mastitis with automatic milking. Comput. Electron. Agric. 70, 60–68.
Kamphuis, C., Mollenhorst, H., Heesterbeek, J.A.P., Hogeveen, H., 2010b. Detection of clinical mastitis with sensor data from automatic milking systems is improved by using decision tree induction. J. Dairy Sci. 93, 3616–3627.
Kuhn, M., 2011. The Caret Package. http://CRAN.R-project.org/web/packages/caret/vignettes/caretTrain. .
LeBlanc, S., 2010. Monitoring metabolic health of dairy cattle in the transition period. J. Reprod. Dev. 56, S29–S35.
Milian-Suazo, F., Erb, H.N., Smith, R.D., 1988. Descriptive epidemiology of culling in dairy cows from 34 herds in New York State. Prev. Vet. Med. 6, 243–251.
Moyes, K.M., Larsen, T., Friggens, N.C., Drackley, J.K., Ingvartsen, K.L., 2009. Identification of potential markers in blood for the development of subclinical and clinical mastitis in dairy cattle at parturition and during early lactation. J. Dairy Sci. 92, 5419–5428.
Mulligan, F.J., Doherty, M.L., 2008. Production diseases of the transition cow. Vet. J. 176, 3–9.
Nyman, A.K., Emauelson, U., Holtenius, K., Ingvartsen, K.L., Larsen, T., Persson Waller, K., 2008. Metabolites and immune variables associated with somatic cell counts of primiparous dairy cows. J. Dairy Sci. 91, 2996–3009.
Opsomer, G., Grohn, Y.T., Hertl, J., Coryn, M., Deluyker, H., de Kruif, A., 2000. Risk factors for post partum ovarian dysfunction in high producing dairy cows in Belgium: a field study. Theriogenology 53, 841–857.
Ospina, P.A., Nydam, D.V., Skokol, T., Overton, T.R., 2010a. Evaluation of non-esterified fatty acids and β-hydroxybutyrate in transition dairy cattle in the northeastern United States: critical thresholds for prediction of clinical diseases. J. Dairy Sci. 93, 546–554.
Ospina, P.A., Nydam, D.V., Skokol, T., Overton, T.R., 2010b. Association between the proportion of sampled transition cows with increased non-esterified fatty acids and β hydroxybutyrate and disease incidence, pregnancy, and milk production at the herd level. J. Dairy Sci. 93, 3595–3601.
Pascottini, O.B., Hostens, M., Sys, P., Vercauteren, P., Opsomer, G., 2017. Cytological endometritis at artificial insemination in dairy cows: prevalence and effect on pregnancy outcome. J. Dairy Sci. 100, 588–597.
Patton, J., Kenny, D.A., McNamara, S., Mee, J.F., O'Mara, F.P., Diskin, M.G., Murphy, J.J., 2007. Relationships among milk production, energy balance, plasma analytes, and reproduction in Holstein-Friesian cows. J. Dairy Sci. 90, 649–658.
Probo, M., Pascottini, O.B., LeBlanc, S.J., Opsomer, G., Hostens, M., 2018. Association between metabolic diseases and the culling risk of high-yielding dairy cows in a transition management facility using survival and decision tree analysis. J. Dairy Sci. 101, 9419–9429.
Ribeiro, E.S., Lima, F.S., Greco, L.F., Bisinotto, R.S., Monteiro, A.P.A., Favoreto, M., Ayres, H., Marsola, R.S., Martinez, N., Thatcher, W.W., Santos, J.E.P., 2013. Prevalence of periparturient diseases and impacts on fertility of seasonally calving grazing dairy cows supplemented with concentrates. J. Dairy Sci. 96, 5682–5697.
Ribeiro, E.S., Gomes, G., Greco, L.F., Cerri, R.L.A., Vieira-Neto, A., Monteiro Jr, P.L.J., Lima, F.S., Bisinotto, R.S., Thatcher, W.W., Santos, J.E.P., 2016. Carryover impact of postpartum inflammatory diseases on developmental biology and fertility in lactating dairy cows. J. Dairy Sci. 99, 2201–2220.
Santos, J.E.P., Thatcher, W.W., Chebel, R.C., Cerri, R.L.A., Calvão, K.N., 2004. The effect of embryonic death rates in cattle on the efficacy of estrus synchronization programs. Anim. Reprod. Sci. 83, 513–535.
Sawa, A., Bogucki, M., Glowska, M., 2015. Effect of single and multiple pregnancies on performance of primiparous and multiparous cows. Arch. Anim. Breed. 58, 43–48.
Shahinfar, S., Page, D., Guenther, J., Cabrera, V., Fricke, P., Weigel, K., 2014. Prediction of insemination outcomes in Holstein dairy cattle using alternative machine learning algorithms. J. Dairy Sci. 97, 731–742.
Sun, Z., Samarasinghe, S., Jago, J., 2010. Detection of mastitis and its stage of progression by automatic milking systems using artificial neural networks J. Dairy Res. 77, 168–175.
Therneau, T.M., 2015. A Package for Survival Analysis in R. https://CRAN.R-project.org/package=survival.
Turner, M.L., Healey, G.D., Sheldon, I.M., 2012. Immunity and inflammation in the uterus. Reprod. Domest. Anim. 47, 402–409.
Van Hoeck, V., Sturmey, R.G., Bermejo-Alvarez, P., Rizos, D., Gutierrez-Adan, A., Leese, H.J., Bols, P.E., Leroy, J.L., 2011. Elevated non-esterified fatty acid concentrations during bovine oocyte maturation compromise early embryo physiology. PLoS One 6, e23183.
Walsh, S.W., Williams, E., Evans, A., 2011. A review of the causes of poor fertility in high milk producing dairy cows. Anim. Reprod. Sci. 123, 127–138.
Wathes, D.C., Fenwick, M., Cheng, Z., Bourne, N., Llewellyn, S., Morris, D.G., Kenny, D., Murphy, J., Fitzpatrick, R., 2007. Influence of negative energy balance on cyclicity and fertility in the high producing dairy cow. Theriogenology. 68, S232–S241.