INTERPRETIVE SUMMARY

DHI data and insemination outcome, Hudson

Early lactation milk constituent concentrations are commonly used as a proxy for energy balance in dairy herds; this study aimed to evaluate associations between these and insemination outcome during early lactation (whilst accounting for other routinely recorded factors). A number of milk constituent predictors demonstrated statistically significant associations with the outcome, but accounted for a very small proportion of the observed variation in herd-year conception risk. Around 40% of this variation was accounted for by a herd-level random effect, suggesting there are unmeasured or unmeasurable factors at herd level which are highly influential in determining conception risk.

Associations between routinely collected dairy herd improvement data and insemination outcome in UK dairy herds

C.D. Hudson*, M.J. Green*

***School of Veterinary Medicine and Science, University of Nottingham, Sutton Bonington, LE12 5RD, United Kingdom**

Corresponding author: Christopher David Hudson; School of Veterinary Medicine and Science, University of Nottingham, Sutton Bonington, LE12 5RD, United Kingdom

chris.hudson@nottingham.ac.uk

ABSTRACT

Milk constituent concentrations in samples taken during early lactation are often used to generate proxy measures for energy balance in dairy herds. This study aimed to explore associations between these and other measures routinely recorded by dairy herd improvement schemes and insemination outcome, with an emphasis on the likely predictiveness of such measures for conception risk (the proportion of inseminations that are successful) at herd level. Data from 312 United Kingdom (UK) dairy herds were restructured so that each unit of data represented an insemination at less than 100 DIM. Milk constituent concentrations from first and second test day (corrected for the effects of season and DIM at sampling) were used as potential predictors of insemination outcome in a logistic regression model. Other predictors included representations of milk yield and other information routinely collected by DHIAs; random effects were used to account for clustering at cow and herd level. The final model included a large number of predictors, with a number of interaction and non-linear terms. The relative effect sizes of the measures of early lactation milk constituent concentrations were small. The full model predicted just under 64% of observed variation in herd-year conception risk (i.e. the proportion of inseminations that were successful in each herd in each calendar year): however, around 40% was accounted for by the herd-level random effect. The predictors based on early lactation milk constituent concentrations accounted for less than 0.5% of observed variation, representations of milk yield (both overall level of yield and early lactation curve shape) for around 7%, with the remaining 15% accounted for by DIM at insemination, parity, inter-service interval, year and month. These results suggest that early lactation milk constituent information is unlikely to predict herd conception risk to a useful extent. The large proportion of observed variation explained by the herd-level random effect suggests that there are unmeasured (in this study) or unmeasurable factors which are consistent within herd and are highly influential in determining herd conception risk.

KEYWORDS

Fertility, conception risk, dairy cow, DHI data

INTRODUCTION

It is widely recognised that improving reproductive performance has great scope to improve the profitability of an individual dairy unit, and has a role in securing the economic and environmental sustainability of the industry as a whole (Archer et al., 2015). Broadly, a herd's overall level of reproductive performance is determined by two factors. The first is the proportion of eligible cows coming into estrus and being inseminated per unit time (commonly termed submission rate [**SR**], and dependent on cow cyclicity, and in herds using artificial insemination also on estrus detection or cycle manipulation). The second is the proportion of inseminations which lead to a pregnancy (variously known as pregnancy rate, conception rate and conception risk [**CR**]). In the United Kingdom (UK), as in many other dairying nations, the medium term trend has been a decline in overall reproductive performance (Norman et al., 2009; Hudson et al., 2010; Morton, 2011). Efforts to mitigate or reverse this decline have mostly focused on improving SR, as this tends to much more amenable to manipulation, and for most herds there is substantial scope for improvement. There is some evidence that these strategies have been at least partially successful (Hanks and Kossaibati, 2016), and as SRs increase the relative importance of CR becomes greater.

A wide range of factors associated with the outcome of an insemination have been described; one of the most widely explored and likely most important is early lactation energy balance. A period of negative energy balance (**NEB**, defined as daily energy output in excess of energy intake) is considered to be normal in modern early lactation dairy cows (Jorritsma et al., 2003), but the severity and duration of NEB has been shown to affect CR during early lactation via a number of physiological pathways (Villa-Godoy et al., 1988; Butler, 2001; Wathes et al., 2007; Leroy et al., 2008a). Monitoring early lactation energy balance is therefore likely to be important in maximising a herd's CR, as it will allow problems to be identified early, and strategies to improve EB to be put in place. A number of monitoring approaches exist, with body condition scoring and evaluation of blood metabolites (e.g. beta-hydroxybutryate and non-esterified fatty acid) concentrations amongst the most popular. However, these require time and incur cost, so alternative monitoring approaches are attractive.

Milk constituent concentrations (from routinely collected dairy herd improvement/milk recording samples) in early lactation have been used as proxy measures of energy balance (Coulon and Rémond, 1991; de Vries and Veerkamp, 2000), with the ratio of butterfat to protein concentration (fat:protein ratio [**FPR**]) being the most widely reported. Although FPR has been associated with reduced fertility (Heuer et al., 1999; Loeffler et al., 1999; Podpečan et al., 2008) and increased risk of early lactation disease (Geishauser et al., 1998; Heuer et al., 1999), there is little evidence to suggest that it is a useful direct predictor of energy balance (Duffield et al., 1997). A study using a large dataset derived from UK herds (Madouasse et al., 2010) found that a model including several measures of early lactation milk constituent concentration and yield was predictive of calving to conception interval, but that FPR at either first or second test day of lactation had no significant association with this outcome when the other predictor variables were included.

The objective of this study was to explore associations between the information routinely recorded as part of dairy herd improvement or milk recording schemes and the outcome of early lactation inseminations, with an emphasis on the potential predictiveness of such measures for conception risk at herd level.

MATERIALS AND METHODS

Data Collection and Restructuring

A convenience sample of routinely recorded management data was collected from 312 dairy herds across England and Wales; data collection and initial audit is previously described in Hudson et al. (2012). Data from lactations beginning between 1999 and 2008 were used, and lactations were only included where a milk recording test day occurred within both intervals 5-35 days in milk (**DIM**) and 36-65 DIM. These intervals were considered to represent typical first and second test days of a lactation. Where there was more than one recording event within an interval, the nearest recording to the centre of the interval was selected. At each test day, daily milk yield and concentrations of protein, butterfat and lactose were recorded. In order to account for the possibility that milk constituent concentrations and yields would be affected by DIM at time of sampling (such that a daily yield of 45 litres at a first test day at 10 DIM was treated differently to the same yield at a first test day at 30 DIM) and day of the year, milk recording data were standardised using the approach described in Madouasse et al. (2010). Briefly, continuous-outcome linear regression models were constructed using the complete dataset, taking each milk recording variable in turn as the outcome. Polynomial terms (order <=6) representing DIM and trigonometric functions representing day of the year at time of sampling were used as predictor variables, and models built using forward selection. These models were then used to estimate an expected value for each variable at each milk recording event in the dataset (given the DIM and day of year of that recording). The expected values for each lactation were then used to standardise the observed values by subtracting expected from observed and dividing by standard deviation, such that the standardised values for each variable had mean 0 and standard deviation 1 across the dataset. In order to illustrate the changes in these variables associated with DIM and stage of year at time of sampling, these models were used to generate predictions across the range of DIM (5-65) and day of year (1-365) and the predictions represented as 3-dimensional surface plots.

The dataset was then restructured such that each unit of data represented an insemination at up to 100 DIM, with a total of 134,520 inseminations from 77,803 cows in 312 herds. Potential explanatory variables at lactation level (including those related to first and second milk recording test days) were replicated and aligned with each insemination from the same lactation. Additional insemination level variables were also included; a binary variable was used to represent the outcome of the insemination. Inseminations were eligible for inclusion if they were from parities which ended in either a subsequent calving date or an exit date, and where herd calving and exit records were available for at least one year after insemination date. Herd-years were excluded if they failed to meet data quality criteria and/or recorded less than 50 inseminations. Inseminations were classified as successful either if they were followed by a calving event 266-296 days later (McGuirk et al., 1998, 1999) or if the animal exited the herd following a positive pregnancy diagnosis. Pregnancy diagnosis results were not used where a subsequent calving event was recorded. Where multiple inseminations occurred 266 to 296 days before a calving event, the insemination giving a gestation period closest to 282 days was classified as successful, and the others as unsuccessful. Where a positive pregnancy diagnosis was followed by an exit from the herd, the last recorded insemination was considered to be successful unless the pregnancy diagnosis event specifically recorded pregnancy to an earlier insemination. The potential predictor variables used in model building are shown in Table 1. Data restructuring and standardisation was carried out using R version 3.0.0 (R Core Development Team, 2010).

Inter-service interval (number of days since previous insemination) was included as potential explanatory variable as it was considered plausible that it could influence CR (for example, CR may be lower where the interval from the previous insemination is very short, potentially representing insemination outside of an estrus event). In order to explore the potential importance of the choice of inter-service interval categories (reflecting the possibility that 18 to 24 days does not represent an appropriate choice of "normal" range, (Remnant et al., 2015; Blavy et al., 2016), model building was repeated using categories based on a 19 to 26 day "normal". This comparison of category choices (18 to 24 day versus 19 to 26 day normal) was repeated using a dataset comprising inseminations up to 200 DIM (cf 100 DIM for the main analysis described).

Model construction

A multilevel logistic regression model was built to explore associations between the outcome (establishment of pregnancy following insemination) and the potential predictors. A 3-level structure was used to account for the hierarchically clustered structure of the data (with inseminations nested within cow, which were nested within herds). A 4-level structure (with lactations as the additional level) was rejected owing to the large number of cows contributing inseminations from a single lactation. The model took the form:

$$\text{Preg}_{ijkl} \sim \text{Bernoulli}(\text{mean}=\mu_{ijkl})$$

$$\ln\left(\frac{\mu_{ijkl}}{1-\mu_{ijkl}}\right)=\alpha+\boldsymbol{\beta}_1\mathbf{X}_{ijkl}+\boldsymbol{\beta}_2\mathbf{X}_{jkl}+u_{kl}+v_l \tag{1}$$

$$v_l \sim \text{normal distribution } (0, \sigma_v^2) \tag{2}$$

$$u_{kl} \sim \text{normal distribution } (0, \sigma_u^2) \tag{3}$$

where i represents a given insemination from lactation j of cow k in herd l; μ_{ijkl} the fitted probability of Preg_{ijkl} (the outcome insemination i leading to a pregnancy); α the regression intercept; $\boldsymbol{\beta}_1$ the vector of coefficients corresponding to the vector of insemination-level predictors $\mathbf{X}_{ijkl}$; $\boldsymbol{\beta}_2$ the vector of coefficients corresponding to the vector of lactation level predictors $\mathbf{X}_{jkl}$; u_{kl} the random effect to reflect variation between individual cows and v_l the random effect representing variation between herds, with σ_u^2 and σ_v^2 the variances of the normal distributions of the random effects terms representing cow and herd respectively.

Model building was by forward selection, with terms retained in the model if the magnitude of the estimated coefficient was greater than double the standard error of the estimate. Univariable associations between the proportion of successful inseminations and each predictor variable in turn were evaluated, and where this suggested a non-linear pattern a polynomial representation of that predictor variable was tested in the model. Categorical variables where several categories had similar parameter estimates were recoded by combining categories for model parsimony. All possible first order interactions were tested in the model, and retained where they met the criteria described above, or altered the estimate for at least one other parameter by at least 10%. For terms relating to early lactation milk records, interactions with the natural logarithm of DIM at insemination were tested, to allow for the possibility that these have decreasing strength of association with inseminations further into lactation. All rejected predictor variables were re-tested in the final model and retained if they met the criteria above. Model building was carried out in MLwiN version 2.29 (Rasbash et al., 2010), with iterative generalized least squares used for exploratory model building and Markov chain Monte Carlo (MCMC) with diffuse prior distributions used over 20,000 iterations for final parameter estimation (Browne, 2009). Visual assessment of MCMC chain behavior was carried out to ensure satisfactory convergence had occurred.

In order to evaluate model fit, full posterior predictions were generated for each insemination using the full MCMC chain for each parameter. The dataset was subset in a variety of ways, including subsets based on variables included in the model (e.g. subset by parity) and subsets based on other variables (e.g. subset by month of calving). Model fit was considered acceptable where the observed CR across a subset of inseminations fell within the 95% coverage interval of the predicted posterior distribution for that subset. MCMC chains were exported to R version 3.0.0 for generation and analysis of model predictions.

Illustration of results using posterior predictions

Posterior predictions were also used for out-of-sample predictions, to demonstrate how the probability of a successful insemination would be expected to vary if one predictor was varied over a given range while the others were held at their population means. For each milk recording variable, the range chosen was -2 to 2: as these variables were standardised, this represented 2 standard deviations either side of the population mean. Line plots were used to represent each relationship. The same approach was used to illustrate interactions between predictor variables.

In order to evaluate the proportion of variation in a herd's CR explained by each element of the model, the data were subset into herd-years (such that each subset contained all the inseminations for one herd in one calendar year; herd-years containing less than 50 inseminations were excluded). Different elements of the model (e.g. full fixed and random effects, fixed effects only, fixed effects for restricted groups of predictors) were used to generate a predicted CR for each herd-year, which was compared to the observed CR in that herd year, with the overall relationship presented using scatterplots and Pearson correlation coefficients. Comparison of r^2 values for correlations between each set of predicted herd-year CRs and the corresponding observed values allowed estimation of the proportion of variation in a herd's CR explained by the different model components.

Cross-validation

To evaluate potential predictiveness of the model on new data, cross validation was performed. This involved subsetting the data randomly (stratified for insemination outcome and herd) into a training dataset containing 80% of the inseminations, and a testing dataset containing the remainder. The training dataset was then used to estimate model parameters (using the same random effects structure as described above, but using least-squares estimation rather than MCMC for computational reasons), which were then used to derive predictions for the inseminations in the test dataset. This process was repeated 10 times (with a different stratified random split of data into training and testing sets each time), resulting in a set of predictions twice as big as the original dataset. These predictions were then summarized as mean CR in each herd-year (excluding those with less than 50 inseminations), which were compared to the observed CR in each herd-year as described for the main model.

RESULTS

A total of 190,324 inseminations at up to 100 DIM were retrieved from herd-years meeting the data quality criteria. Of these, 12,655 were excluded as no outcome was determined by the rules described in the Method section; a further 43,149 were excluded due to other lactation-level data quality issues (most commonly missing test day milk recording information). A dataset containing 134,520 inseminations was therefore used for the final analysis, of which 53,909 (40%) were determined to have led to pregnancy (by subsequent calving in approximately 96% of successful inseminations; in approximately 4% the outcome was determined by a positive pregnancy diagnosis prior to culling).

Regression planes illustrating the relationship between early lactation milk recording parameters and DIM/day of year at sampling are shown in Figure 1. Interactive versions of these plots are available online (https://plot.ly/~cdhudsonx/73/, https://plot.ly/~cdhudsonx/76/, https://plot.ly/~cdhudsonx/79/, https://plot.ly/~cdhudsonx/67/, https://plot.ly/~cdhudsonx/82/, https://plot.ly/~cdhudsonx/85/). The concentrations of butterfat, protein and lactose all fell markedly while milk yield increased over the first 30 DIM. FPR also increased through early lactation, showing a peak which was earlier and more pronounced than the nadir demonstrated by butterfat or protein concentration. Daily milk yields showed a seasonal trend, increasing through winter to a peak in spring followed by a decline through the summer months. Milk constituent concentrations tended to show a converse trend (decreasing as daily yield increased), with butterfat percentage also decreasing sharply in spring.

Parameter estimates for the logistic regression model are shown in Table 2. Of the potential predictors (see Table 1) based on milk constituent concentrations (standardized for DIM and day of year at sampling), butterfat, protein and lactose percentages at first test day and protein and lactose percentages at second test day were significantly associated with the probability of pregnancy to an insemination (CR). Of these, lactose concentration (at both test days) and protein concentration at second test day had significant interaction terms with DIM at insemination (broadly such that the effect of each was greater on inseminations earlier in lactation). These relationships are illustrated using model predictions in Figure 2. It is worth noting that these graphs illustrate the direct components of each relationship only, as they show the relationship between the outcome and one predictor variable after accounting for the effects of all the other predictor variables in the model. For example, if the same factors influence the concentration of lactose at both first and second test day in the same direction, the observed relationship between lactose at first test day and the outcome would be expected to appear stronger if lactose at second test day was not accounted for.

Associations between milk yield related predictors and the outcome are illustrated using predictions in Figure 3. There was a different relationship between 305-day lactation yield and CR for first lactation animals and mature cows; although in each category predicted CR increased to a peak around 6,000 litres then declined with increasing yield beyond this. For first lactation animals, the size of the relationship was larger and the peak CR occurred at a slightly lower yield. Test day yield at both first and second test day was also significantly associated with CR, and there was a significant interaction between the two test day yield terms.

For a given level of lactation yield, predicted CR was generally higher where yields at first and second test day were higher (i.e. where yield increased more quickly after calving). This relationship was relatively simple where corrected first test day yield was greater than zero (i.e. yield at first test day was greater than predicted for a test day at that DIM and day of year), but below zero became more complex. This relationship is illustrated using a predicted regression plane in Figure 4, and an interactive version is available at https://goo.gl/F35QR9. The associations between CR and the yield-based predictors were generally much larger than those with the constituent-based predictors.

Inseminations during the summer months and in later years were associated with lower CR, and CR increased with DIM although the gradient of this increase became smaller at around 70 DIM. Predicted CR was lowest for parity 1 and highest for parity 2; parities 3 and 4 were similar to 2. In order to explore this association further, predictors were removed from the model sequentially and parameters re-estimated. The decrease in predicted CR in parities above 1 was observed where the terms relating to 305-day adjusted lactation milk yield were retained in the model; where these terms were removed the association changed such that increasing parity was associated with decreasing CR.

Inseminations at an interval of 18 to 24 days after a previous insemination had a very slightly higher CR than first inseminations; other categories of inter-service interval (**ISI**) were associated with lower CR. Odds of a successful insemination were lowest for ISI less than 18 days, at around 50% lower than for a first insemination. Adoption of ISI categories based on a 19 to 26 day "normal" interval made very little difference to model fit (as measured by deviance information criterion, (Spiegelhalter et al., 2002)); although when analysis was repeated with a dataset containing inseminations up to 200 DIM, use of the alternative ISI categories did improve model fit.

Figure 5 illustrates the use of model predictions to partition observed variation between herd-year subsets (i.e. the subset of inseminations from each herd in each year) of the data. Predictions based on the full model, including herd- and cow-level random effects, accounted for around 64% of the variation in observed herd-year CR; removing the cow-level random effect made negligible difference to this, while removing the herd-level random effect reduced the r^2 value to around 22%. A fixed-effect model without any milk constituent predictors accounted for a very similar proportion of variation (just below 22%), and removal of milk constituent and yield predictors reduced this to 15%, representing the proportion of observed variation in herd-year CR explained by days in milk, parity, inter-service interval, year and month of insemination.

These changes in r^2 value were used to partition variance in herd-year CR across the fixed and random effects in the model; this is shown in Figure 5 (f). When predictions generated using 10-fold cross validation were used, the fixed-effect model explained around 19% of the observed variation in herd-year CR. This is similar to the value derived using the full dataset both to estimate model parameters and to generate predictions (22%), suggesting that the model would be similarly predictive if applied to new data from the same population.

The 95% coverage interval of model posterior predictions for a wide selection of different subsets of the data included the observed result for each subset, confirming that the model fitted the data well. Visual assessment of MCMC chain behavior revealed a small number of chains amongst the milk constituent concertation predictors where convergence had not clearly been achieved. Parameters were re-estimated using a larger number of iterations (100,000, compared to 20,000 initially): this resulted in very similar parameter estimates, although in some cases chains had again not clearly converged. Recoding the problematic variables from continuous values into five categories each and removal of their interaction terms with DIM resulted in a model with good chain behavior. Again, parameter estimates were very similar to the original model, and deviance information criterion was higher. This suggested that the model could be reparamaterized to improve MCMC chain behavior, and that this resulted in a model that gave very similar information but had a poorer fit to the data. The model using continuous milk constituent predictors was therefore reported.

DISCUSSION

The main objective of this study was to investigate the relationship between routinely recorded dairy herd management information and insemination outcomes at up to 100 DIM. One aspect of interest was the association between CR and early lactation milk constituent concentrations, as it is highly plausible that CR is the element of the reproductive process most influenced by energy balance, and a number of milk constituent based indicators are commonly used as proxy measure for herd-level energy status. As in previous work using a similar approach in data from UK herds (Madouasse et al., 2010), a large number of statistically significant associations were revealed (Table 2). Many of these relationships were not simple to interpret from model parameters, as there were a number of interaction terms, both with early lactation variables and between these and stage of lactation. Graphical presentation of these results (Figure 2) using model predictions provides a more intuitive interpretation. Broadly, these findings agree with earlier work (Madouasse et al., 2010), with increased protein concentration at either of the first two test days and decreased butterfat concentration at the first test both generally associated with an increased probability of pregnancy. The association between early lactation lactose concentration and the outcome was more variable with DIM at insemination; this was especially marked for lactose at first test day.

For most of the relationships, the predicted CR varied relatively little over the range illustrated (2 SD below to 2 SD above population mean). For example, an insemination at 50 DIM which was average in every respect would be expected to have a CR of just over 30% if lactose concentration at first test day was – 2 (i.e. 2 SD below population mean); this would increase to just over 40% for a concentration of +2 (2 SD above the population mean). This range is likely to represent almost the full range of observed lactose concentrations (as 95% of values would be expected to lie within 2 SD of the mean). So although this is one of the larger associations between milk constituent concentration and CR, lactose concentration would have to alter from close to the lowest observed level to close to the highest observed level in order to produce a meaningful change in CR.

Taking all the milk constituent variables together, it appears that they collectively account for an extremely small proportion of observed variation in herd-year CR (Figure 5). There are a number of possible explanations for this: for example, that these parameters are not reliable predictors of early lactation energy balance in UK herds, or that early lactation energy balance has little impact on CR in early lactation. The latter seems highly unlikely, as there as a considerable body of evidence demonstrating a strong link between energy balance and CR (Butler, 2003; Roche, 2006; Leroy et al., 2008b; Ospina et al., 2010). This study would therefore seem to suggest that milk constituent concentrations in early lactation do not predict energy balance at lactation level to a clinically useful extent in this sample of herds. This was despite the correction of these variables to account for variation introduced by DIM at time of test day and seasonality; Figure 1 (and the interactive online equivalents) shows that this variation is substantial, and implies that use of uncorrected values is likely to be considerably less useful. For example, even within the typical sampling window of the first test day of lactation, FPR (the most commonly used proxy measure) would be expected to vary from below 1.2 for a cow sampled at 5 DIM in early August to over 1.3 for a cow sampled at 23 DIM in February. However, it is also useful to remember that there are other reasons why these predictors would perhaps be expected to explain little of the herd-year variation in CR: for example, the possibility that milk constituents vary mostly at cow level within herds, and the large number of other factors known to affect CR (in general, the more factors affect an outcome, the smaller the proportion of outcome variance explained by any individual factor).

The inclusion of 305-day adjusted lactation yield as well as daily yields at first and second test days of lactation allowed the effects of overall level of production, and shape of the lactation curve through early lactation to be evaluated together. Broadly, higher levels of production (as measured by 305-day yield) were associated with lower CR, although for both first lactation and older cows very low yields were associated with a decreased CR. This apparently novel finding could plausibly be a result of a very low lactation yield acting as a marker of some (unrecorded or unmeasurable) disease event which had an impact on both production and fertility. For a given level of 305-day yield, CR was generally higher in lactations where daily yield rose steeply post-calving; this is in agreement with previous work in this field (Cook and Green, 2016), and measures based on yield at first test day or related to peak production have previously been suggested as markers of successful transition and early lactation health (Nordlund and Cook, 2004). Clearly, these findings do not imply a directly causal relationship between 305-day yield and CR (for example, because events occurring after conception may influence 305-day yield, and pregnancy itself is associated with a reduction in milk yield), and parameter estimates for the same model without the terms relating to 305-day yield are included as an Appendix. However, it is expected that events occurring well after peak lactation are likely to have a small role in determining 305-day yield in most herds, and the effect of pregnancy on daily yield is relatively small and only measurable in mid to late gestation (Coulon et al., 2010), although other studies have found larger effects (van Amburgh et al., 1997). Taking this into consideration, inclusion of 305-day yield to represent overall level of production and provide better insight into other factors (such as parity) having accounted for this was felt to be useful.

In this study, inseminations in parity 1 (i.e. first lactation animals) were associated with a lower CR compared to other lactation numbers (once the other predictors in the model are accounted for). Several previous studies have reported higher CR in parity 1 (Gröhn and Rajala-Schultz, 2000; Cook and Green, 2016); including one using a smaller subset of the same data as was used in the current work (Hudson et al., 2012). Indeed, in a simple univariable analysis of the dataset used in this study, parity 1 animals have a higher mean CR (44%) than those in later parities (41%, 40%, 39% and 35% for parities 2 to >5 respectively). This suggests that the relationship between CR and parity is confounded by other predictors accounted for in the model. Since many of the other variables in the model reported here were also included in other studies which found first lactation animals to be more fertile than older individuals (for example, in [Hudson et al., 2012]), it is more likely that a novel element of the model reported here which was not included in previous work sheds new light on the relationship. Sequential removal of model terms revealed that the representation of 305-day milk yield was key – when this was accounted for using separate polynomial terms for parity 1 and parity >1 (represented in Figure 3), a lower predicted CR for parity 1 was observed. This suggests that previous work may have found higher CR in first lactation animals because these animals have lower milk yields, and lower milk yields have tended to be associated with increased fertility. Where yield is accounted for in a more complex way, it becomes clear that first lactation animals tend to have a lower CR than would be expected given their level of production. Clear potential explanations for this finding exist: cows in the first lactation are usually amongst the least dominant animals in a group, so are more likely to have restricted access to any limited resources (for example, where feed or water space is limited). It is also possible that this association is only present in early lactation: this study used inseminations at <100 DIM, whilst previous studies often cover different time periods.

Model predictions for CR across herd-years were used to explore how observed variation in CR is accounted for by the various elements of the model. Herd-years were used as the units in this case partly as these would represent the way in which such data is often assessed in the field, and because herd-years were not directly included in the model as a random effect (as was, for example, herd). Collectively, all of the fixed effect predictors in the model (see Table 2) explained just over 22% of the observed variation, with less than 1% accounted for by the predictors relating to early lactation milk constituent concentration and around 7% by predictors relating to milk yield (both overall level of production and shape of early lactation curve). This reinforces the suggestion that early lactation milk constituents are not likely to be predictive of energy balance to a clinically useful extent.

The remaining variation in herd-year CR was split relatively evenly between the herd-level random effect and the bottom level model residuals (i.e. the variation not explained by any elements of the model). This suggests that a large proportion of variation in CR is attributable to factors which are relatively consistent within herd over time, but which were not measured in this dataset, or indeed are not measurable. This could cover a wide range of factors (including environmental and feeding management, disease status and insemination related factors), and it is notable that the association between these unmeasured herd-level factors and herd-year CR is several times larger than that between CR and milk yield. The cow-level random effect term explained a negligible amount of variation in herd-year CR, suggesting that unmeasured factors that are consistent within cow across inseminations and parities are unimportant as drivers of herd CR.

The use of predictions across herd-years also serves as an example of the value of carrying out further analysis to explain and contextualise the results of (especially logistic) regression analysis. Conventional presentation of model results as odds ratios alone (Table 2) would be difficult to interpret in this situation. In part this is because of the complexity of the model –

interaction terms and non-linear representations of continuous predictors are inherently non-intuitive to interpret in a numerical format. Additionally, the intuitive tendency to interpret odds ratios as relative risks would also be a problem in this case – as the overall risk of pregnancy resulting from an insemination is relatively large, the odds of a successful serve are substantially different from the probability, and odds ratios will tend to exaggerate effect size (Davies et al., 1998). There are a number of approaches which can be useful to aid interpretation of such models – in addition to those reported here, population attributable risk (for example, Peeler et al., 2000) and stochastic simulation modelling (for example, Hudson et al., 2015) can be highly useful.

It is relevant to consider the potential for misclassification of the outcomes of inseminations with the methods used in this study. Subsequent calving date was the main determinant of the insemination success, this was largely due to the source of the data: use and recording of pregnancy diagnosis was according to each herd's usual practice, so was highly variable between herds. Clearly, this approach has potential for misclassification of outcome in both directions – for example, where two inseminations occur close together, or where a cow aborts. Although the rules for determining the outcome in this study were designed to minimise such errors, some misclassification is still possible. However, the alternative approach of relying more heavily on pregnancy diagnosis records also has potential for misclassification, and would also have led to the exclusion of a large number of inseminations, plausibly in such a way that would introduce substantial bias (for example, it is possible that a higher proportion of inseminations with no pregnancy diagnosis outcome are unsuccessful). Even if classification errors were evenly distributed, this would still have potential to influence the results of the study, generally by reducing the size of estimated coefficients and shifting variance from herd- and cow-level towards the bottom (unexplained) level. This is a feature inherent in such large-scale, retrospective studies.

Some features of the statistical modelling approach used in this study also merit discussion. As with all regression modelling, there were a number of somewhat subjective choices to be made during the model building process (such as interaction terms and non-linear representation of continuous explanatory variables). In such cases, a balance needs to be struck between model complexity, informativeness and the biological questions being explored. Whilst formal statistical methods balancing model fit against degree of complexity exist, and were used to some extent here (such as deviance information criterion), the potential for overfitting also needs to be considered (Babyak, 2004). Use of internal cross-validation here helped to provide some evidence that overfitting had not occurred, as well as providing some indication of potential out-of-sample predictiveness. MCMC was used for final estimation of the reported model parameters. This has a number of advantages over conventional methods, including generally more robust parameter estimates for multilevel models (Browne and Draper, 2006) and a more intuitive "Bayesian" interpretation of results than is the case for frequentist methods. For example, this approach produces a full posterior distribution for each model parameter, allowing probabilistic statements about results (such as "it is 95% probable that the true value for this parameter is between X and Y") without relying on an understanding of the concept of long-run repetition. However, MCMC is substantially more computationally intensive than conventional methods. Evaluation of chain behavior should be a standard aspect of parameter estimation using MCMC: this study presents a robust approach to dealing with unexpected behavior of MCMC chains.

CONCLUSIONS

This study demonstrates that measures based on early lactation milk constituent concentrations are unlikely to predict herd-level CR to a clinically useful extent, even when corrected for potential nuisance factors such as season and DIM at sampling. A relatively sophisticated representation of milk yield (accounting both for overall level of yield and shape of lactation curve) was much more predictive of CR, but still accounted for only around 7% of observed herd-year variation. After accounting for milk yield in this way, predicted CR was highest in parities 2 and 3. Unmeasured effects which were consistent at herd level (represented by a herd-level random effect) accounted for over 40% of the variation, and further investigation into the herd-level factors explaining this would be highly valuable.

REFERENCES

Archer, S.C., C.D. Hudson, and M.J. Green. 2015. Use of stochastic simulation to evaluate the reduction in methane emissions and improvement in reproductive efficiency from routine hormonal interventions in dairy herds. PLoS ONE 10:e0127846. doi:10.1371/journal.pone.0127846.

Babyak, M.A. 2004. what you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. Psychosom. Med. 66:411. doi:10.1097/01.psy.0000127692.23278.a9.

Blavy, P., M. Derks, O. Martin, J.K. Höglund, and N.C. Friggens. 2016. Overview of progesterone profiles in dairy cows. Theriogenology 86:1061–1071. doi:10.1016/j.theriogenology.2016.03.037.

Browne, W.J. 2009. MCMC Estimation in MLwiN v2.20. Centre for Multilevel Modelling, University of Bristol.

Browne, W.J., and D. Draper. 2006. A comparison of Bayesian and likelihood-based methods for fitting multilevel models. Bayesian Anal. 1:473–513.

Butler, W.R. 2001. Nutritional effects on resumption of ovarian cyclicity and conception rate in postpartum dairy cows. BSAS Occasional Publication 133–146.

Butler, W.R. 2003. Energy balance relationships with follicular development, ovulation and fertility in postpartum dairy cows. Livest. Prod. Sci. 83:211–218. doi:16/S0301-6226(03)00112-X.

Cook, J.G., and M.J. Green. 2016. Use of early lactation milk recording data to predict the calving to conception interval in dairy herds. J. Dairy Sci. 99:4699–4706. doi:10.3168/jds.2015-10264.

Coulon, J.B., and B. Rémond. 1991. Variations in milk output and milk protein content in response to the level of energy supply to the dairy cow: A review. Livest. Prod. Sci. 29:31–47. doi:10.1016/0301-6226(91)90118-A.

Coulon, J.B., L. Pérochon, and F. Lescourret. 2010 Modelling the effect of the stage of pregnancy on dairy cows’ milk yield. Anim. Sci. 60:401-408. doi: 10.1017/S1357729800013278

Davies, H.T.O., I.K. Crombie, and M. Tavakoli. 1998. When can odds ratios mislead? Br. Med. J. 316:989–991. doi:10.1136/bmj.316.7136.989.

Duffield, T.F., D.F. Kelton, K.E. Leslie, K.D. Lissemore, and J.H. Lumsden. 1997. Use of test day milk fat and milk protein to detect subclinical ketosis in dairy cattle in Ontario. Can. Vet. J. 38:713.

Geishauser, T.D., K.E. Leslie, T.F. Duffield, and V.L. Edge. 1998. An evaluation of protein/fat ratio in first DHI test milk for prediction of subsequent displaced abomasum in dairy cows. Can. J. Vet. Res. 62:144–147.

Gröhn, Y.T., and P.J. Rajala-Schultz. 2000. Epidemiology of reproductive performance in dairy cows. Anim. Reprod. Sci. 60:605–614. doi:10.1016/S0378-4320(00)00085-3.

Hanks, J., and M. Kossaibati. 2016. Key performance indicators for the UK national dairy herd. University of Reading, Reading, UK.

Heuer, C., Y.H. Schukken, and P. Dobbelaar. 1999. Postpartum body condition score and results from the first test day milk as predictors of disease, fertility, yield, and culling in commercial dairy herds. J. Dairy Sci. 82:295–304. doi:10.3168/jds.S0022-0302(99)75236-7.

Hudson, C., J. Breen, A. Bradley, and M. Green. 2010. Fertility in UK dairy herds: Preliminary findings of a large-scale study. Cattle Pract. 18:89–94.

Hudson, C.D., A.J. Bradley, J.E. Breen, and M.J. Green. 2012. Associations between udder health and reproductive performance in United Kingdom dairy cows. J. Dairy Sci. 95:3683–3697. doi:10.3168/jds.2011-4629.

Hudson, C.D., A.J. Bradley, J.E. Breen, and M.J. Green. 2015. Dairy herd mastitis and reproduction: Using simulation to aid interpretation of results from discrete time survival analysis. Vet. J. 204:47–53. doi:10.1016/j.tvjl.2015.01.024.

Jorritsma, R., T. Wensing, T.A.M. Kruip, P.L.A.M. Vos, and J.P.T.M. Noordhuizen. 2003. Metabolic changes in early lactation and impaired reproductive performance in dairy cows. Vet. Res. 34:16. doi:10.1051/vetres:2002054.

Kadarmideen, H.N., R. Thompson, and G. Simm. 2000. Linear and threshold model genetic parameters for disease, fertility and milk production in dairy cattle. Anim. Sci. 71:411–419. doi:10.1017/S1357729800055338.

Leroy, J., T. Vanholder, A.T.M. Van Knegsel, I. Garcia-Ispierto, and P.E.J. Bols. 2008a. Nutrient prioritization in dairy cows early postpartum: mismatch between metabolism and fertility? Reproduction in Domestic Animals 43:96–103.

Leroy, J.L., G. Opsomer, A. Van Soom, I.G. Goovaerts, and P.E. Bols. 2008b. Reduced fertility in high-yielding dairy cows: are the oocyte and embryo in danger? Part I: The importance of negative energy balance and altered corpus luteum function to the reduction of oocyte and embryo quality in high-yielding dairy cows. Reprod. Domest. Anim. 43:612–622. doi:10.1111/j.1439-0531.2007.00960.x.

Loeffler, S.H., M.J. de Vries, and Y.H. Schukken. 1999. The effects of time of disease occurrence, milk yield, and body condition on fertility in dairy cows. J. Dairy Sci. 82:2589–2604. doi:10.3168/jds.S0022-0302(99)75514-1.

Madouasse, A., J.N. Huxley, W.J. Browne, A.J. Bradley, I.L. Dryden, and M.J. Green. 2010. Use of individual cow milk recording data at the start of lactation to predict the calving to conception interval. J. Dairy Sci. 93:4677–4690. doi:10.3168/jds.2010-3235.

McGuirk, B.J., I. Going, and A.R. Gilmour. 1998. The genetic evaluation of beef sires used for crossing with dairy cows in the UK - 1. Sire breed and non-genetic effects on calving survey traits. Anim. Sci. 66:35–45. doi:10.1017/S135772980000881X.

McGuirk, B.J., I. Going, and A.R. Gilmour. 1999. The genetic evaluation of UK Holstein Friesian sires for calving ease and related traits. Anim. Sci. 68:413–422. doi:10.1017/S1357729800050414.

Morton, J. 2011. InCalf Fertility Data Project 2011. Dairy Australia.

Nordlund, K.V., and N.B. Cook. 2004. Using herd records to monitor transition cow survival, productivity, and health. Vet. Clin. N. Am.-Food A. 20:627–649. doi:10.1016/j.cvfa.2004.06.012.

Norman, H.D., J.R. Wright, S.M. Hubbard, R.H. Miller, and J.L. Hutchison. 2009. Reproductive status of Holstein and Jersey cows in the United States. J. Dairy Sci. 92:3517–3528. doi:10.3168/jds.2008-1768.

Ospina, P.A., D.V. Nydam, T. Stokol, and T.R. Overton. 2010. Associations of elevated nonesterified fatty acids and β-hydroxybutyrate concentrations with early lactation reproductive performance and milk production in transition dairy cattle in the northeastern United States. J. Dairy Sci. 93:1596–1603. doi:10.3168/jds.2009-2852.

Peeler, E.J., M.J. Green, J.L. Fitzpatrick, K.L. Morgan, and L.E. Green. 2000. Risk factors associated with clinical mastitis in low somatic cell count British dairy herds. J. Dairy Sci. 83:2464–2472. doi:10.3168/jds.S0022-0302(00)75138-1.

Podpečan, O., J. Mrkun, and P. Zrimšek. 2008. Diagnostic evaluation of fat to protein ratio in prolonged calving to conception interval using receiver operating characteristic analyses. Reprod. Domest. Anim. 43:249–254. doi:10.1111/j.1439-0531.2007.00895.x.

R Core Development Team. 2010. R: A language and environment for statistical computing. r foundation for statistical computing, Vienna, Austria.

Rasbash, J., C. Charlton, W.J. Browne, M. Healy, and B. Cameron. 2010. MLwiN Version 2.2. Centre for Multilevel Modelling, University of Bristol, UK.

Remnant, J.G., M.J. Green, J.N. Huxley, and C.D. Hudson. 2015. Variation in the interservice intervals of dairy cows in the United Kingdom. J. Dairy Sci. 98:889–897. doi:10.3168/jds.2014-8366.

Roche, J.R. 2006. The effect of nutritional management of the dairy cow on reproductive efficiency. Anim. Reprod. Sci. 96:282–296. doi:16/j.anireprosci.2006.08.007.

Spiegelhalter, D.J., N.G. Best, B.R. Carlin, and A. van der Linde. 2002. Bayesian measures of model complexity and fit. J. Royal Stat. Soc - Ser. B 64:583–616.

van Amburgh, M.E., D.M. Galton, D.E. Bauman, R.W. Everett. 1997. Management and economics of extended calving intervals with use of bovine somatotropin. Livest. Prod. Sci. 50: 15-28. doi:10/1016/S0301-6226(07)00069-9.

Villa-Godoy, A., T.L. Hughes, R.S. Emery, L.T. Chapin, and R.L. Fogwell. 1988. Association between energy balance and luteal function in lactating dairy cows. J. Dairy Sci. 71:1063–1072. doi:10.3168/jds.S0022-0302(88)79653-8.

de Vries, M.J., and R.F. Veerkamp. 2000. Energy balance of dairy cattle in relation to milk production variables and fertility. J. Dairy Sci. 83:62–69. doi:10.3168/jds.S0022-0302(00)74856-9.

Wathes, D.C., M. Fenwick, Z. Cheng, N. Bourne, S. Llewellyn, D.G. Morris, D. Kenny, J. Murphy, and R. Fitzpatrick. 2007. Influence of negative energy balance on cyclicity and fertility in the high producing dairy cow. Theriogenology 68:S232–S241. doi:16/j.theriogenology.2007.04.006.

TABLES AND FIGURES

Table 1. Potential explanatory variables at each level of data used to build a logistic regression model with the outcome of insemination success or failure.

		Distributional characteristics	
Variable	**Representation**	**Insemination level**[1]	**Herd-year level**[2]
Insemination level			
Outcome	Binary	Mean = 0.4	0.42 (0.17–0.70)
Days in milk at insemination	Polynomial (order <4)	Mean = 72d, SD =17d	73 (63–87)
Month of insemination	Categorical; months as individual categories		
Inter-service interval	Categorical (<18d, 18-24d, 25-35d, 36-48d, >48d, NA[3])	[0.02, 0.12, 0.04, 0.027, <0.01, 0.78]	18-24d: 0.10 (0.00-0.27) NA: 0.82 (0.52-1.00)
Lactation level			
305-day lactation yield	Centred around population mean, polynomial (order <4)	Mean = 8,230 litres, SD = 2,167 litres	8,067 (4,944-10,433)
Month of calving	Categorical; months as individual categories		
Lactation number	Categorical (1, 2, 3, 4, 5+)	[0.26, 0.23, 0.18, 0.12, 0.20]	Lact 1: 0.25 (0.00-0.54) Lact 5+: 0.20 (0.00-0.46)
Butterfat % at recording 1	Linear, standardised for DIM and day of year at recording (such that 0 represents expected population mean given DIM and day of year, and a 1 unit change represents one population standard deviation)		-0.02 (-0.80-1.03)
Butterfat % at recording 2			-0.05 (-0.79-1.12)
Protein % at recording 1			0.01 (-0.65-1.21)
Protein % at recording 2			-0.02 (-0.79-1.35)
Lactose % at recording 1			0.00 (-0.71-0.98)
Lactose % at recording 2			-0.02 (-0.78-098)
Fat:protein ratio at recording 1			-0.05 (-0.82-0.87)
Fat:protein ratio at recording 2			-0.04 (-0.88-0.94)
Daily yield at recording 1			-0.04 (-1.48-0.95)
Daily yield at recording 2			-0.04 (-1.58-0.98)

[1] Distributional characteristics across inseminations in the dataset – means and standard deviations are reported for continuous variables, and proportion in each category (reported in the order the categories are listed in the "Representation" column) for categorical variables.

[2] Herd-year level distributional characteristics are median and 95% coverage interval for the means of each continuous variable for each herd-year (i.e. the first value represents the median herd-years, and the numbers in brackets the range covering 95% of herd-years). For categorical variables, the variation in proportion of the herd in certain categories is reported in the same way.

[3] NA in the inter-service interval category indicates the first insemination of a lactation.

Table 2. Parameter estimates from a multivariable logistic regression model with the outcome representing pregnancy resulting from a given insemination.

Model term	Odds ratio	95% HPD[1] interval	
		Lower	Upper
Butterfat % at recording 1[2]	0.98	0.97	1.00
Protein % at recording 1	1.05	1.03	1.06
Protein % at recording 2	1.44	1.21	1.75
Lactose % at recording 1	1.59	1.26	1.97
Lactose % at recording 2	0.72	0.59	0.89
(Protein % at recording 2).(lnDIM[3])	0.92	0.88	0.96
(Lactose % at recording 1).(lnDIM)	0.91	0.86	0.96
(Lactose % at recording 2).(lnDIM)	1.08	1.03	1.14
Daily milk yield at recording 1	1.16	1.14	1.19
Daily milk yield at recording 2	1.25	1.22	1.28
(Yield at recording 2)^2	1.03	1.02	1.04
(Yield at recording 1).(Yield at rec’g 2)	1.04	1.03	1.06
(Yield at recording 1).((Yield at rec’g 2)^2)	0.97	0.97	0.98
305-day lactation yield (‘000 litres)	0.75	0.74	0.76
(305-day lactation yield)^2	0.96	0.96	0.96
(305-day lactation yield)^3	1.01	1.00	1.01
(305-day lactation yield).(Parity 1)[4]	0.86	0.84	0.88
(305-day lactation yield)^2.(Parity 1)[4]	1.03	1.03	1.04

[1] HPD: highest posterior density

[2] Butterfat, protein, lactose and yield values are standardised to account for DIM and day of year at sampling, such that a value of 0 would represent expected population mean (given DIM and day of year at sampling), with a unit change representing 1 population standard deviation away from mean. Odds ratios are for a 1 unit change in each variable, adjusted for all other terms in the model.

[3] lnDIM: natural logarithm of days in milk at insemination

[4] The relationship between 305-day yield and CR was very similar for all parity categories except parity 1, so for model parsimony only the interaction with this parity group was included.

(305-day lactation yield)^3.(Parity 1)[4]	1.00	1.00	1.01
DIM[5]	1.12	1.12	1.13
(DIM)^2	1.00	1.00	1.00
(DIM)^3	1.00	1.00	1.00
Parity 1	*Reference*		
Parity 2	1.54	1.48	1.61
Parity 3	1.52	1.46	1.59
Parity 4	1.40	1.34	1.47
Parity >4	1.15	1.10	1.20
ISI[6] N/A (first insemination)	*Reference*		
ISI <18 days	0.49	0.45	0.53
ISI 18-24 days	1.05	1.01	1.08
ISI 25-35 days	0.81	0.76	0.86
ISI 36-48 days	0.90	0.84	0.97
ISI >48 days	0.75	0.65	0.87
Summer (June – September)	0.84	0.82	0.86
Year <2003	*Reference*		
Year 2003	0.98	0.93	1.02
Year 2004	0.93	0.89	0.97
Year 2005	0.87	0.83	0.91
Year 2006	0.82	0.78	0.86
Year 2007	0.80	0.77	0.84
Year 2008	0.78	0.73	0.84

[5] DIM: days in milk at insemination
[6] ISI: inter-service interval (days since previous insemination)

FIGURE CAPTIONS

Figure 1: Regression surfaces illustrating relationship between test day milk parameters and DIM/season at sampling.

Figure 2: Predictions (from the model reported in Table 2) for example scenarios to illustrate relationships between conception risk (CR) and early lactation milk constituent concentrations. Each plot shows probability of pregnancy resulting from a set of example inseminations where all predictor values are set at their population means except for the variable indicated in the x-axis of the plot and days in milk (DIM). Each line shows the variation in predicted CR across the range of the variable, with line colors/types representing inseminations at different stages of lactation. Milk constituent variables are standardised (such that 0 represents population mean and 1 represents mean plus one standard deviation compared to expected value given season and DIM at test day), and numeric suffixes represent test day number.

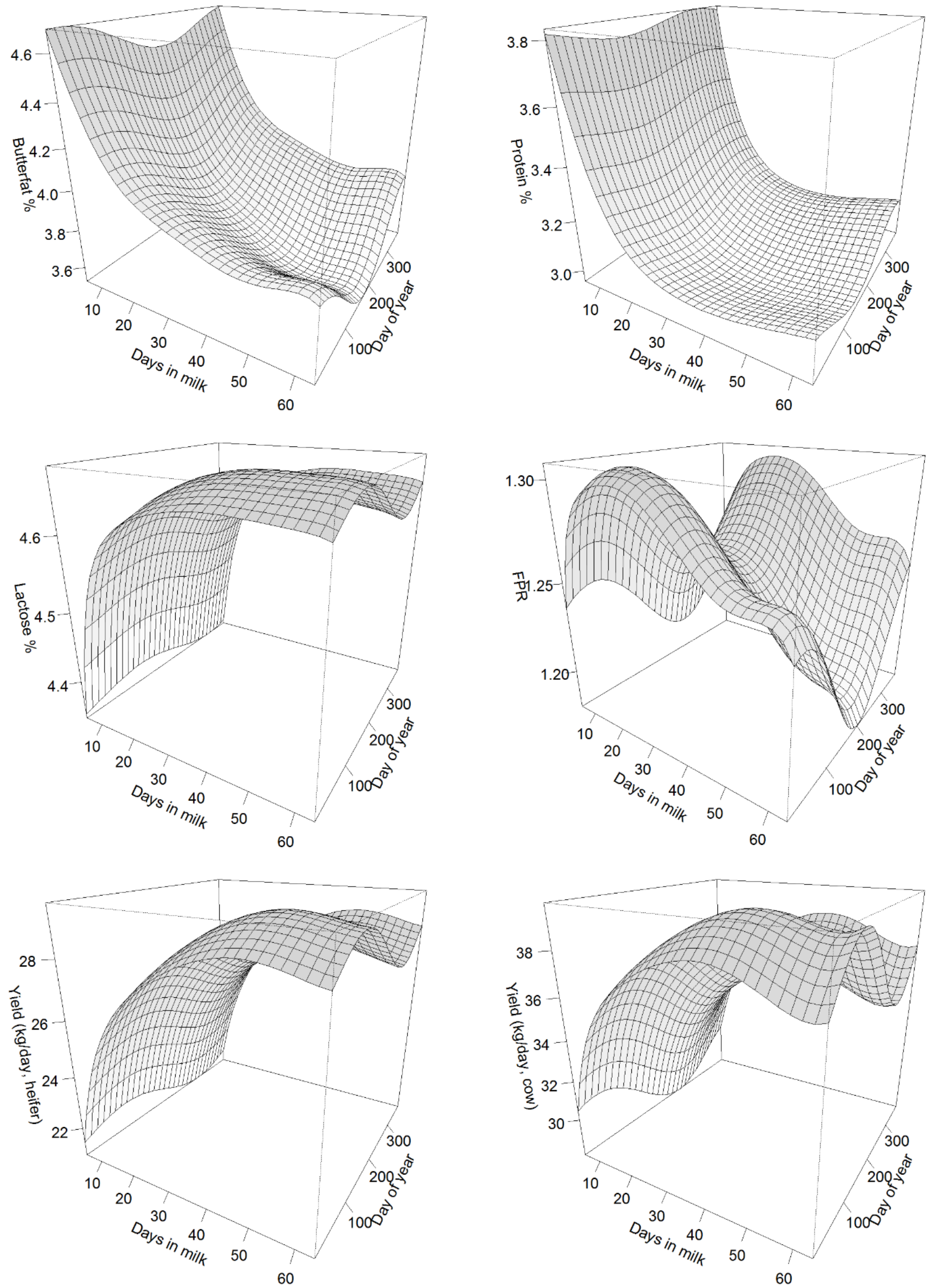
Figure 3: Predictions (from the model reported in Table 2) for example scenarios to illustrate relationships between conception risk (CR) and milk yield. Each plot shows probability of pregnancy resulting from a set of example inseminations where all predictor values are set at their population means except for the variables indicated in the plot legend and x-axis. The left-hand plot shows predicted CR across a range of 305-day lactation yields (with line color representing parity). The central and right-hand plots show the association between CR and daily yield at first and second test day (standardised such that 0 represents population mean and 1 represents mean plus one standard deviation compared to expected value given season and DIM at test day). Numeric suffixes represent test day number. Y1: yield at first test day; Y2: yield at second test day.

Figure 4: Regression surface illustrating the predicted relationship (from the model reported in Table 2) between conception risk (CR) and daily yield at first and second test day

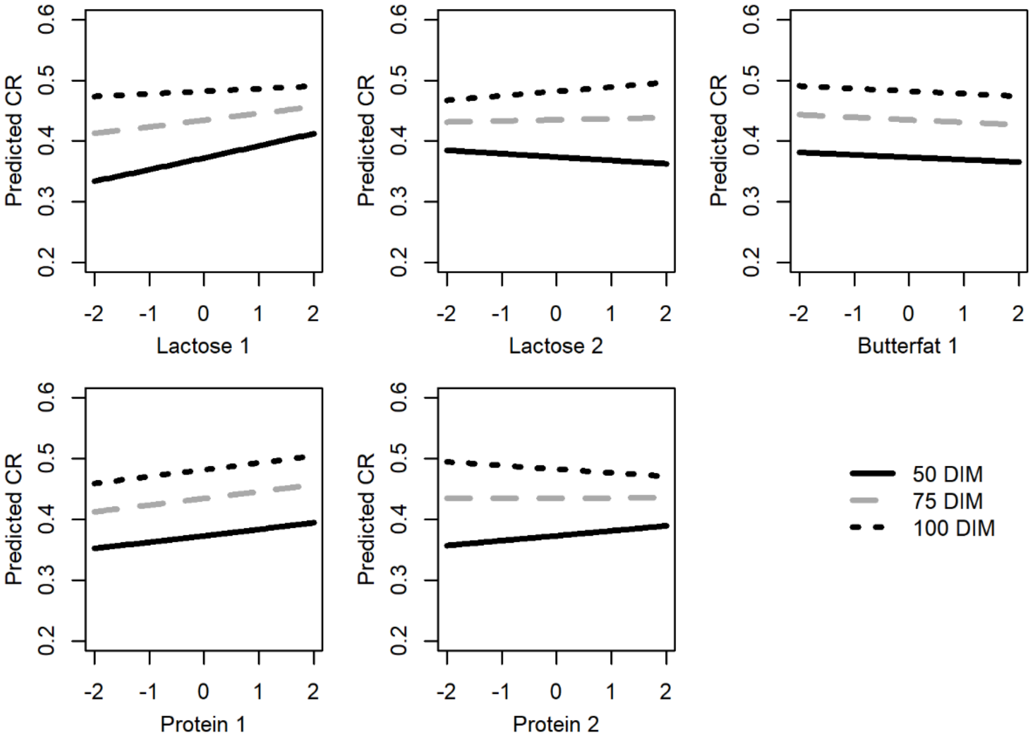
(standardised such that 0 represents population mean and 1 represents mean plus one standard deviation compared to expected value given season and DIM at test day).

Figure 5: a) – e) Predicted versus observed herd-year conception risk (CR) generated using different elements of the model reported in Table 2. Plot titles show which elements of the model were used to create each set of predictions, and plot text shows Pearson r^2 value for each correlation. f) Proportion of variance in herd-year CR attributable to each model element. Yield includes all variables representing milk yield; constituent% includes all variables based on milk constituent concentrations; other includes all other fixed effects in the model. Herd-level, cow-level and unexplained show residual variation at each level of the model.

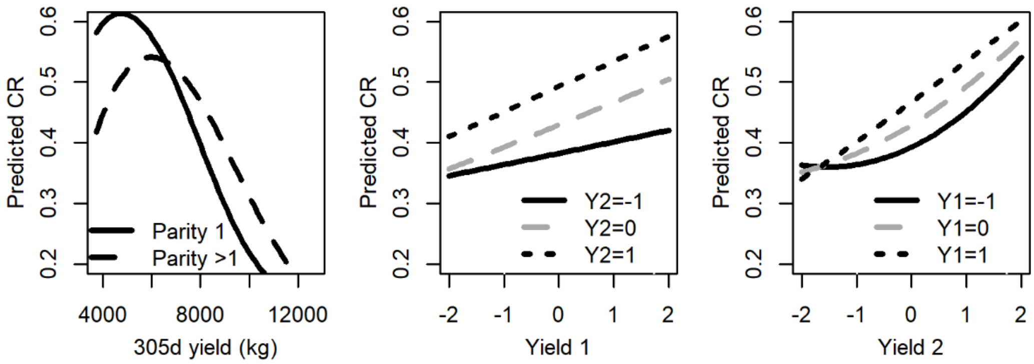
HUDSON, FIGURE 1



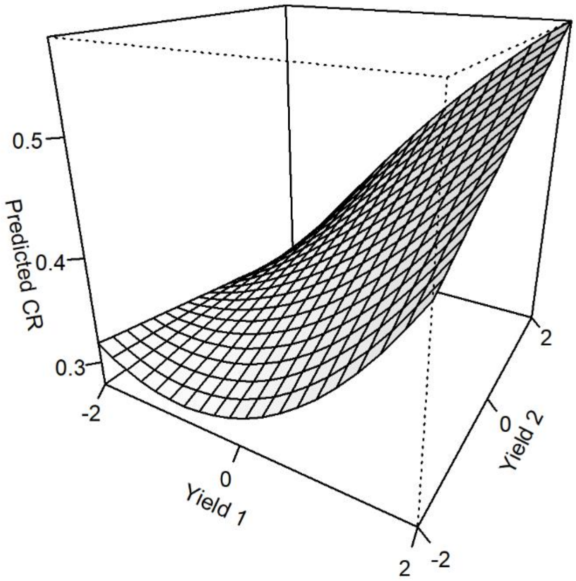
HUDSON, FIGURE 2



HUDSON, FIGURE 3



HUDSON, FIGURE 4



HUDSON, FIGURE 5