Big data and the dairy cow:
Factors affecting fertility in UK herds

CHRISTOPHER DAVID HUDSON BVSc DCHP MRCVS

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

July 2015

Abstract

Routinely collected herd management data in a variety of formats were collated from 468 dairy herds, and novel objective measures of data recording quality were developed and applied. This revealed that there was a substantial amount of variation in data quality between herds, and the vast majority of herds failed to meet the threshold level for at least one of the data quality measures used. Analysis of trends in reproductive performance across the herds with good quality fertility event recording suggested that their fertility was generally declining through the first half of the 2000s, but there was some evidence that improvements in submission rate were beginning to reverse this decline in the later years studied (up to 2007).

Associations between reproduction and two endemic diseases common in dairy cattle (mastitis and lameness) were explored using multilevel discrete time survival modelling, and probabilistic sensitivity analysis (PSA) used to contextualise and illustrate the results. In both cases, statistical modelling revealed significant and sizeable associations between disease events and reproductive outcomes at lactation level. However, simulation and application of PSA showed that a herd's incidence rate of either disease was highly unlikely to influence its overall reproductive performance to a clinically relevant degree when other inputs to herd fertility were also considered.

Factors associated with the proportion of serves leading to a pregnancy (pregnancy rate) were explored using multilevel logistic regression modelling. This revealed that relatively little of the variation in herd pregnancy rate is explainable by routinely recorded milk recording data (including constituent concentration in early lactation as well as daily and lactation yields). A large amount of the unexplained variation was revealed to be at herd level and very little at cow level, suggesting that investigation of herd management practices associated with pregnancy rate would be rewarding.

Acknowledgements

Completing a PhD alongside a teaching post was always going to be a major challenge, and I'm immeasurably grateful to those who have helped to make it possible. Foremost, I would like to thank my primary supervisor, Professor Martin Green, for all his support and guidance and for introducing me to the worlds of veterinary epidemiology and Bayesian statistics. My second supervisor, Dr Andrew Bradley, provided extremely valuable advice and guidance, especially in the early part of the project (and when I was deciding whether or not a PhD was a good idea!). I also owe a great deal to my other colleagues, both at the University of Nottingham School of Veterinary Medicine and Science and elsewhere, who have helped in various ways. In particular, Professor Jon Huxley and Dr James Breen have provided invaluable advice, support and solace; I'm also extremely grateful to John Remnant for covering some of my teaching to give me time to prepare this thesis.

Without the help of the veterinary surgeons and consultants who facilitated my access to the data, this project would not have been possible. They are too numerous to name individually, but they know who they are (and in most cases have already been repaid in beer or other alcoholic beverages).

I am grateful to the University of Nottingham and to Quality Milk Management Services and Sum-IT Computing for their various roles in funding my studies, and hope that I have repaid their respective investments in the way that they would have hoped!

Finally, a massive thank you to Charlotte and (for the last year at least) Thomas for their support, and for putting up with me working at times when I should really have been doing other things!

Contents

Figures

Tables

Chapter 1 Introduction

1.1 Background

1.1.1 The dairy industry in the UK and worldwide

The dairy industry in the UK currently supports just under 15,000 dairy farmers, with a total of just under 1.8 million adult dairy cows (DairyCo, 2013). Whilst consumption of liquid milk per capita of population in the UK has declined slightly over the last 10 years (DEFRA, 2013), a trend echoed in a high proportion of other developed nations (Kearney, 2010), there has been an increase in UK consumption of other dairy products. Globally however, the picture is rather different, with a marked increase in liquid milk demand seen across many developing nations. Most notably, consumption of dairy products in China, India and Brazil has increased very substantially over the last decade (Wang and Li, 2008). These trends should also be seen in the light of the concurrent increase in world population, which in many cases is growing most rapidly in the nations where dairy consumption is increasing steeply (Lutz et al., 2001). The increase in population size has also led to a renewed focus on food security, defined by the World Health Organisation (1996) as "…access to sufficient, safe, nutritious food to maintain a healthy and active life". A key component of food security is food availability: the ability of food production systems to meet demand from the population. This is considered important both at global level, and also at national level, where "food security" is commonly taken to refer to a nation's ability to satisfy the food needs of its own populace.

However, the world's resources are clearly not infinite, and any increase in food production must come at a minimum resource cost, both in terms of system inputs (such as land, fuel and fertilizer) and waste outputs (such as greenhouse gases) (Steinfeld et al., 2006). The term "sustainable intensification" has recently come into use to describe the process required to meet ever-increasing requirements for food whilst minimising the drain on natural resources and reducing or mitigating any environmental impacts.

The increasing demand for milk in the global marketplace has led to an increase in the potential for the UK to export dairy products (currently and historically the UK has been a net importer of all of the major dairy products except liquid milk) (DairyCo, 2014), and a concurrent potential for UK farmgate milk prices to be increasingly influenced by global milk demand and prices. This market is notoriously volatile, and it is considered highly likely that price volatility in the UK milk market will increase in coming years.

Dairy farming businesses are also exposed to substantial volatility in production costs; most notably feed, fuel and fertilizer. The costs of home-grown and purchased feeds are highly influenced by climatic conditions. This means that extreme weather events, which have become increasingly common in the UK over the last decade (Jones et al., 2013), have the potential to cause substantial increases in the cost of production for a dairy business which can persist for prolonged periods of time. In addition, any increase in demand for agricultural products or by-products with the potential for use as either human or animal food (e.g. cereals) resulting from global population expansion will exert further upward pressure on feed costs.

Collectively, these factors are incentivising UK dairy farms to operate as efficiently as possible, minimising both the financial and the resource cost per unit of production. This will make dairy businesses as robust as possible to short- or medium-term changes in input or output price, and maximise the ability of the UK industry to respond to emerging global opportunities.

1.1.2 Dairy cow reproduction and efficiency

In view of this, it is critical to understand the key drivers of efficiency for a dairy business. One of these is reproductive performance, and maintaining fertility is important across all types of dairy system. This is because of the change in a cow's milk yield through a lactation: typically there is a peak in production at around 40-60 days after calving, followed by a steady decline through the rest of the lactation (Figure 1-1). The shape of the lactation curve varies substantially with different feeding systems, genetics, milking frequency, parity (Tekerli et al., 2000) and a variety of other factors. Since it is only possible to induce another peak lactation by calving the cow, minimising the interval from one calving to establishment of the next pregnancy tends to increase the proportion of a cow's lifetime which is spent in the early part of lactation when feed conversion is most efficient (Britt et al., 2003) and minimise time spent either at the less productive "stale" end of lactation or dry (non-lactating).

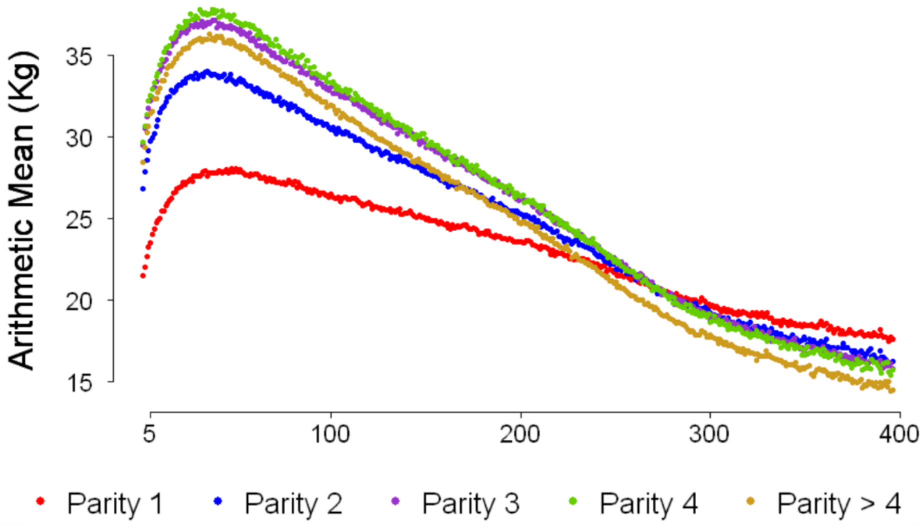

Figure 1-1 Mean lactation curve shapes across parities; horizontal axis shows days after calving
Derived from a dataset of UK dairy herds undertaking milk recording using a commercial provider (National Milk Records) during 2005-6. Taken from Madouasse (2009).

In reality, this is an over-simplification, and there are competing reasons not to shorten a cow's calving interval (the number of days between successive calvings) beyond a certain level. For example, it is almost universal practice to allow a cow a minimum length of dry period (often 30-40 days) before the next calving (predominantly because very short or absent dry periods are associated with large reductions in milk yield in the next lactation (Steeneveld et al., 2013)): therefore very short calving intervals in a cow with a very flat lactation curve could result in an increase in the proportion of the cow's life spent dry leading to a decrease in average lifetime daily production which is large enough to overcome the increase in productivity due to increased time spent in early lactation. The "tipping point" at which these effects are cancelled out is highly controversial, but it is widely accepted that a very small proportion of the UK national herd calve at intervals which are unprofitably short and that there is substantial scope to improve efficiency through better reproductive management.

Reproductive performance also has the potential to influence a herd's efficiency by affecting the proportion of the herd which are culled each year. Culling rates due to failure to conceive are notoriously difficult to estimate (as cows are frequently removed from the herd for a combination of reasons rather than a single reason), and will also be influenced by management system. However, reports of typical rates of culling for reproduction vary from 6 to 14% of the herd per year (Compton and McDougall, 2010; Hudson et al., 2010; Mee, 2004; Refsdal, 2007). Reducing the failure to conceive culling rate by optimising herd fertility allows either a reduction in the overall turnover rate of the herd (which in turn reduces the number of youngstock needed to maintain herd size) or allows more elective culling (for example of less productive individuals in the herd). Either of these outcomes is likely to lead to increased efficiency.

Reproductive performance therefore influences not only the profitability of an individual farm business (Evans et al., 2006; Gonzalez-Recio et al., 2004; LeBlanc, 2007), but also the sustainability of the industry as a whole. If herds are producing milk more efficiently as a result of good fertility, then this will minimize the quantities of system inputs (such as land area, purchased feed, fuel and fertilizer) required to produce a given volume of milk. Similarly, increased efficiency tends to minimize the environmental impact of the industry, both through increased feed conversion efficiency in lactating cows and by minimising the number of replacements which need to be reared (Garnsworthy, 2004).

1.1.3 Fertility management in dairy herds

There are a number of key management decisions which determine how reproduction is managed in a dairy herd. Perhaps foremost is whether natural or artificial insemination (or a combination of the two) is used. Where exclusively natural service is employed and a bull or bulls remain in the same group as the cows eligible to become pregnant, fertility management is entirely focussed on avoiding reproductive disease, encouraging rapid return to ovarian cyclicity in the cows and maximising the proportion of serves leading to a pregnancy (pregnancy rate). In a herd where artificial insemination (AI) is used there is an additional requirement for a system to detect cows in oestrus, so that they may be inseminated. This is traditionally achieved by simple observation of the eligible cows for behavioural signs of heat, but over the past 20 years this has become more challenging, and alternative or additional approaches (ranging from the use of simple stick-on mounting detectors to activity monitoring and telemetry systems) are now extremely common in the UK.

Another major management decision affecting reproductive management is whether the herd is to calve all year round, or whether there is to be a defined calving season. Through the 1990s and 2000s the UK saw a decline in the proportion of seasonally calving herds; the concurrent decline in reproductive performance (Esslemont and Kossaibati, 2002) may have been partly responsible for this as effective seasonal calving requires good fertility. During the past decade there has been a resurgence in seasonal calving, mostly driven by adoption of extensive, low-input systems such as those pioneered in New Zealand. Under such a system there is substantial pressure for a large proportion of the herd to calve within a short period of time in early spring, as the time when most of the herd are at peak lactation (and being re-bred) needs to coincide with that of maximal grass quality and growth rate. Seasonal calving herds in the UK typically breed for a period of 12-18 weeks, and cows not conceiving within this time period are either removed from the herd or retained for re-breeding a year later (severely reducing their profitability). In year-round calving herds, there is less pressure on generating pregnancies within a specified period of time.

Another key decision in a herd's breeding policy is the length of the voluntary waiting period (VWP). This is defined as the number of days after calving at which a cow is first considered eligible for service. The VWP serves two main purposes: preventing excessively short calving intervals (as the shortest possible interval is determined by the earliest serve) and enhancing the likely success of the first serve (by allowing time to elapse after calving for the effects of peri-parturient negative energy balance and bacterial uterine contamination to diminish). Clearly, increasing the VWP will always increase the herd’s calving index, so it is important to balance the relative importance of the factors affecting this choice.

1.2 Measuring and improving reproductive performance

1.2.1 Optimising fertility

The first step in improving fertility in a dairy herd is to understand where interventions should be targeted. Measurement of reproductive performance is discussed in more detail in Section 1.2.2, and in the majority of cases it is best to focus either on improving the proportion of oestrus events which are detected (in herds using predominantly AI) or improving the proportion of inseminations which lead to a pregnancy (in herds using either AI or natural service) (Hudson et al., 2012). In herds using mostly AI, it is common for heat detection to be the most appropriate target for intervention (Mawhinney and Biggadike, 1998). Determinants of the efficiency of heat detection can be divided into those affecting the cows’ ability to express oestrus and those affecting the ability of the systems of the farm to detect it.

The intensity and duration of expression of oestrus has been in decline both in the UK and worldwide over the past 20 years (Yoshida et al., 2009), and a study in 2005 found that 44% of cows either failed to express any behavioural signs of heat, or showed only brief signs which occurred overnight (Roelofs et al., 2005). A variety of both herd-level and cow-level factors have been shown to affect the duration and intensity of oestrus expression. Cow-level factors include a variety of diseases: the literature on associations between mastitis and lameness (two of the most common endemic diseases in dairy cows) is reviewed in detail in Chapter 4 and Chapter 6, but other events such as uterine bacterial disease are also considered important (Sheldon and Dobson, 2003). Increased milk yield has also been associated with reduced intensity of oestrus behaviour (Lopez et al., 2004). Herd-level factors mostly relate to the cows' environment. Factors such as floor surface (Platz et al., 2008), housing design and availability of loafing space (Pennington et al., 1985) and ambient temperature have all been implicated in the decline in expression of oestrus. The number of other cycling cows in the group, and time budget can also play a role here (Roelofs et al., 2010), with reduced opportunities for expression of heat in herds with prolonged milking times.

The ability of the farm's system to detect oestrus events is mainly determined by which methods of heat detection are being used. Detection by observation alone can be relatively efficient, but this is highly dependent on the time allocated to observation (Holman et al., 2011) and the signs taken to indicate oestrus (Van Vliet and Van Eerdenburg, 1996). Heat mount detectors (usually in the form of either paint applied to the tail head or an adhesive pad with a coloured indicator) are a simple, low cost and commonly adopted way to augment the efficiency of visual detection. The efficacy of such aids appears highly variable, with different studies reporting heat detection rates varying from 35% (Holman et al., 2011) to 95% (Xu et al., 1998). It is plausible that some of this variation is explained by variation in cows' ability to express oestrus in the various environments studied. Use of activity monitoring systems is also becoming more widespread in the UK; again there is considerable variation in the reported rates of heat detection achievable with this technology, but generally heat detection rates in the range 50 – 80% are commonly reported in the literature (Firk et al., 2003; Holman et al., 2011; Roelofs et al., 2005; Statham, 2012). In practice, it appears that a multifaceted approach to heat detection is most effective, combining more than one of these approaches with good use of information (for example, to predict which cows are due in oestrus on which dates) and appropriate use of veterinary examination and treatment.

Pregnancy rate, defined as the proportion of serves which lead to a pregnancy, is the other key driver of overall reproductive performance. Broadly, this is less amenable to manipulation than oestrus detection. Nutrition has a major influence on pregnancy rate, especially the degree of exposure of the cow to negative energy balance in early lactation. More prolonged or more severe periparturient negative energy balance has been shown to reduce pregnancy rate via a variety of mechanisms (Butler, 2001; Fenwick et al., 2008; Leroy et al., 2005; Villa-Godoy et al., 1990; Wathes et al., 2007). Unfortunately, negative energy balance can be very challenging to manage, partly because it is influenced by a large number of factors which can have large variations over short time-periods and are often not obvious (for example, changes in forage quality or daily dry matter intake). Reduced pregnancy rates have also been associated with a variety of different disease processes. Classically herd-level infectious diseases such as bovine viral diarrhoea (McGowan et al., 1993; Vanroose et al., 1998), infectious bovine rhinotracheitis (Vanroose et al., 1997) and leptospirosis (Dhaliwal et al., 1996) have been considered most important in this context. However, the influence of endemic diseases such as mastitis and lameness has also been studied (see Chapter 4 and Chapter 6 for a full review), and uterine bacterial disease has also been associated with reduced pregnancy rates (Gilbert et al., 2005; LeBlanc et al., 2002; McDougall, 2001).

Other factors associated with reduced pregnancy rates include high genetic merit for milk production (Veerkamp et al., 2001), high temperature-humidity index (Ravagnolo and Misztal, 2002) and mycotoxin ingestion (Whitlow and Hagler, 1999). In herds using AI, it is also important to maintain good semen preparation, insemination technique and timing, along with accurate heat detection (avoiding serves to “false positive” oestrus events). For herds using natural service, number and fertility of bulls as well as venereal disease must also be considered.

1.2.2 Recording reproductive data

In order to monitor fertility effectively, a minimum level of data recording is required. Essentially, this consists of basic cow data (such as identity and parity) alongside accurate recording of (as a minimum) calving, insemination and pregnancy diagnosis events. Additional detail (for example, which operator or bull performed each insemination) or extra events (such as disease occurrences and the results of any veterinary reproductive examinations) are also useful but are not always available. This information can be captured in a variety of ways:

- Via on-farm computer software: Many software packages are available which allow farmers to record the required individual cow data. The data recorded in this way are usually relatively up to date (as it can be entered in real time by farm staff), and most systems allow a high level of flexibility in which events are logged and how much detail is recorded. However, this method is only suited to farmers with a reasonable degree of technological competence, and the large variety of software available can make analysis in a consistent fashion more difficult.
- Via a bureau recording system: This involves an external organisation storing and managing the herd's data. Often the service provided includes input of data from a paper format into a computer recording system, potentially making this route more accessible for some farmers. In the early days of computer recording in the UK, this service was commonly provided by veterinary practices, but as the cost of computing power has decreased and the use of computers in day-to-day life has become more prevalent, there has generally been much less demand and it is now uncommon. This option preserves the potential for recording a high level of detail, as well as allowing easy and consistent analysis, but brings additional potential for errors and has a cost to the farmer.
- Via a milk recording organisation (MRO): This is essentially similar to the option above, with a milk recording organisation acting as the "bureau". In the UK, there are various ways in which this system operates, depending mainly on which MRO a farmer chooses to

use, and what level of service they pay for. Commonly, reproductive event records are collected from a farm diary or other paper recording system by a milk recording technician at each test day (usually monthly) and entered into a central database. Alternatively, events may be entered on farm via a web portal (similarly to on-farm software), or the MRO may opt to supply the farmer with their own recording software. The first of these options is highly accessible to farmers (as it requires very little input from them), and is probably currently the most widely used route in the UK. Data from MROs are easily accessible, available in a consistent format and securely stored. However, the main limitation of this system is the restriction on which events can be recorded (typically limited to calving, serve and pregnancy diagnosis) and the lack of the option to record detail about each event. Data quality also tends to be very variable.

1.2.3 Monitoring reproductive performance

Regular monitoring of reproductive performance is a critical part of managing a dairy herd, and understanding patterns in reproductive data should be a cornerstone of improving a herd's performance. However, the ability to undertake this successfully is heavily dependent on the data recording described in Section 1.2.2. There are essentially two groups of metrics which are commonly used to measure the reproductive performance of a herd:

- Interval-based measures (such as calving interval/index and calving to first serve interval) measure the number of days between a given pair of events; the distribution of these intervals across the population of interest is then commonly summarised using a measure of central tendency (e.g. mean or median). This is the more traditional approach to fertility monitoring, and has the advantage that the calving interval and its component parts clearly represent the outcome which fertility management aims to influence. However, there are problems with this approach to monitoring: firstly, it requires the intervals for a given cohort to be summarised in some way. The most robust way to evaluate the distribution of such intervals is by visualisation using a histogram, but this is cumbersome when several such intervals need to be monitored. The mean interval is commonly used, but interval data are almost invariably heavily right-skewed, so the mean is unlikely to be the measure which best represents the population. Conversely, there is a valid argument that the degree of skewness is itself important to measure, as any positive outliers are likely to be highly economically significant. There is substantial danger in using either measure without also assessing dispersion of individual intervals around the central estimate. A further problem is that the second of the events has to have occurred in order to calculate the interval. For example, if the cohort of interest is defined by the date of the first event (e.g. cows calving in a particular calendar year), then only those for which the second event has occurred will be included in any assessment of interval-based measures. This introduces an inherent downward bias in the intervals (as the cows with shorter intervals will be included earlier in time); this is lessened by increasing the "lag" period between the time of evaluation and the timeframe used to select the cohort, but this increases the retrospectiveness of the analysis (see Fetrow et al. (2007), Breen et al. (2009) and Hudson et al. (2012) for a fuller discussion). There is also potential for misleading results where some individuals never complete an interval; for example, cows which are culled for failure to conceive will never contribute to the herd's calving index.

- Proportion- or rate-based measures (such as pregnancy rate and first serve submission rate) represent an alternative approach, involving calculation of the proportion of cases in which a given criterion is met (for example, the proportion of inseminations leading to a pregnancy [pregnancy rate] or the proportion of cows receiving a first insemination within 24 days of becoming eligible [first serve submission rate]). Analysis of rates can be less retrospective, and temporal allocation of events is more accurate (i.e. the rate for a given period more accurately reflects performance during that period), allowing better monitoring of trends through time. However, rate-based measures inherently use binary outcomes, so provide less detailed information than interval-based measures, and care must be taken that the denominator population for calculation of the rate or proportion does not become too small: for example, when calculating pregnancy rate by month, in small herds there may be a very small number of inseminations in a given month (Fetrow et al., 2007).

A herd's overall reproductive performance can be monitored using either approach. Calving interval and calving to conception interval are commonly used, as are rate-based measures such as the 100-day in calf rate (proportion of eligible cows which have re-conceived 100 days after the previous calving) and fertility efficiency (proportion of eligible cows becoming pregnant in each 21 day period, also known as 21-day pregnancy rate or pregnancy risk) (de Vries et al., 2005; Fetrow et al., 2007; Hudson et al., 2012). In the case of the interval measures, it is important to ensure that failure to conceive rate is monitored alongside whichever interval measure is chosen.

Heat detection can also be measured in either way, with calving to first serve interval (de Vries and Risco, 2005; Norman et al., 2009; Refsdal, 2007) and inter-service intervals both widely used interval-based measures, and first serve and return to serve submission rate also common (Brownlie et al., 2013; Cornou et al., 2014; Fetrow et al., 2007). Generally, the rate-based measures are preferable for ongoing routine monitoring (as they are less retrospective and have better temporal allocation), but evaluation of the distribution of intervals on a periodic basis gives valuable additional information and is useful in the context of long-term benchmarking (where accurate comparisons are more important and retrospectiveness is less of a problem).

Success of service is commonly measured using a proportion-based approach, with the proportion of inseminations leading to a pregnancy representing the key measure. This is known either as "pregnancy rate" or "conception rate" (Brownlie et al., 2013; Norman et al., 2009). The former term best represents the outcome being measured (i.e. establishment of pregnancy as determined by pregnancy diagnosis or subsequent calving), but the latter is much more commonly used, both in the UK and worldwide. The term “pregnancy rate” will be used to represent the proportion of serves leading to a pregnancy throughout this thesis.

1.3 “Big data” and the dairy industry

1.3.1 What are “big data”?

The term “big data” has emerged over the past decade to describe a variety of concepts and applications related to managing and generating value from large datasets. One of the most commonly cited attempts to define the meaning of the term derives from a report from the analytics company Gartner (Laney, 2001) and describes three elements common to big data:

- Volume: the term relates to using “large” datasets, although the absolute size of datasets considered large is vary variable, and is often considered relative to computing power at the time
- Variety: big data includes use of data from various sources which can be integrated together to provide greater insight. This often includes combining quantitative data with text-based or qualitative data
- Velocity: information in such datasets often accumulates quickly, and development of a platform for integrating, storing and analysing such data in an ongoing way is often considered a key component

A fourth “v” often included in later iterations of this definition is “veracity”, promoting the concept that, with the increased size and heterogeneity of the available data comes greater need for awareness of the data quality and better methods for auditing this and “cleaning” data.

In spite of the nomenclature, application of big data principles is often independent of the absolute size of dataset being analysed (van Rijmenam, 2013). Indeed, the ongoing advancement of computer technology means that datasets which 10 years ago were considered too large to handle or derive value from can now be analysed without specialist hardware. Defining precisely what constitutes big data has been problematic for some time, but features such as deriving value from complex datasets to drive decision making, maximising use of routinely collected and stored data through alternative analytical applications and linking heterogeneous datasets from varied sources are common features (Ward and Barker, 2013). Markowetz (2014) outlines a structure where multiple sources of data are linked together, restructured to a common format and stored in a central location before analysis using a bespoke analytical platform. Big data is commonly also taken to include the idea of predictive (as opposed to descriptive) analysis; this is often used to describe how big data differs from traditional business intelligence or analytics (which tended to be more focussed on describing trends and features in data). The term "big data" originated amongst businesses, but has more recently been widely adopted by the scientific community (Boyd and Crawford, 2012; Lynch, 2008); Figure 1-2 shows the increase in the number of publications mentioning the phrase in either title or abstract between 2005 and 2013.

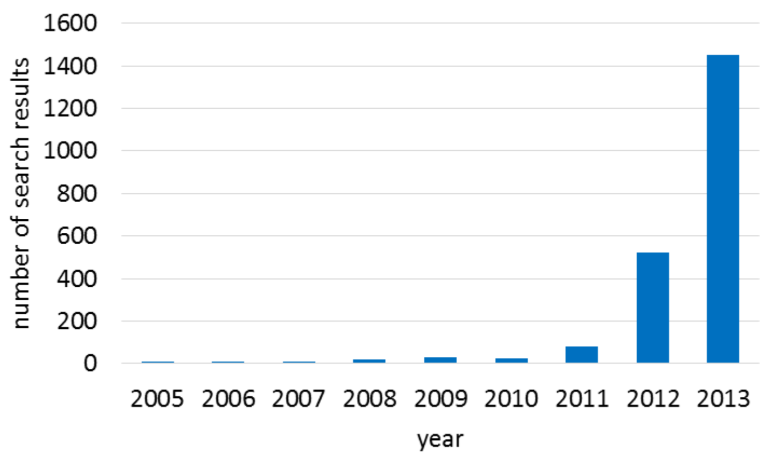

Figure 1-2 Trend in number of publications relating to "big data"
Number of results per year returned by a Scopus search for the phrase "big data" in either article title or abstract across journals and conference proceedings; searched 19th March 2014

Perhaps the most well-known examples of application of big data techniques come from the world of global business. Retailers in particular have been quick to capitalize on the potential of sales and other data. This has led to phenomena such as data-driven marketing, where promotion of products to individuals can be tailored according to the individual's purchase history. Purchase data are also used to optimise efficiency of the supply chain, allowing accurate predictions of sales of particular products from given outlets, such that supply can be matched accurately to projected demand. The retail supply chain (and logistics more generally) provides another good example, with real-time location information of delivery vehicles used to maximise efficiency.

The concepts of big data are also being adopted in the world of medicine, especially in nations where there is some form of widespread computer health record system such as the USA (Dilsizian and Siegel, 2014). Here, the availability of large volumes of patient clinical data opens up new avenues of research opportunities, as well as the potential for disease and risk prediction at patient level to bring a new era of personalised medicine. Medicine also provides one of the better known examples of using back-end search engine data as a source of information to which big data techniques can be applied. Ginsberg et al. (2009) describe and demonstrate the usefulness of detection of regional influenza epidemics by monitoring Google search queries, and show that level of influenza activity in each region of the USA was accurately predicted by the frequency of specific query requests. Social media (such as Twitter and Facebook) increasingly generate textual data which can be used in a similar way, as well as its routine use in delivering personalised marketing.

1.3.2 Dairy herd data: An underused resource?

There are many parallels in data utilisation between the current situation in the UK dairy industry and the state of large retail organisations 10-15 years ago. As herd sizes have increased, the popularity of computer recording systems has strengthened, and clinical experience in the field suggests that there are large numbers of herds for which detailed, complete and robust datasets are being generated. From the point of view of the herd manager, the data are being collected predominantly to assist with day-to-day cow management; in larger herds it is very difficult to organise the running of the herd without some form of computer record system. However, it is also readily used for monitoring performance in the herd, and for identifying potential areas for improvement. This has become an increasingly important aspect of the role of the farm animal veterinary surgeon in recent years, and the upswing in interest and capability in this field amongst the veterinary profession has created a further driver for farmers to maintain good electronic records, as they are able to see the benefits to their businesses of routine performance monitoring and management.

However, it is highly likely that substantially more value can be extracted from these data in a research context. Such datasets represent a rich resource for large-scale retrospective epidemiological studies across a wide range of subject areas, from endemic and infectious disease to production and fertility. There have been some good examples of exploitation of this resource in the UK, but these have used only data which is owned and stored centrally by MROs (Hanks and Kossaibati, 2010; Madouasse et al., 2010a, 2010b). This has the advantage of being easily accessible and continuously updated, but limits the depth of information available; the majority of herds included in such databases in the UK record little disease information and the quality of recording of reproductive events is extremely variable. Datasets from on-farm computer recording systems tend to be much more detailed and quality of record keeping is often better, but such data tends to be harder to access. Use of these datasets requires both a mechanism for data collection, as well as more advanced skills in data engineering, as there are a wide variety of common on-farm software systems which store data in very different formats. Whilst such data has been used for some small-scale regional studies in the UK (Kerby, 2009), these barriers have precluded use of such data on a wider scale and it represents a largely untapped resource.

1.4 Stochastic modelling and probabilistic sensitivity analysis

A wide variety of traditional and new techniques can be applied to analysis of the large, retrospective datasets, but in some cases more sophisticated and robust analytical techniques can yield results which are harder for the end user of the research to interpret and understand. For example, a large number of factors are known to affect the various elements of reproduction in cattle (e.g. return to cyclicity, heat detection, pregnancy rate), but in most cases it is still unclear what improvement in a herd's overall reproductive performance might be expected if one of the factors affecting reproduction were to be manipulated, given that other factors will have some natural variation over time. One potential route by which these relationships could be clarified in a quantitative way is stochastic modelling.

1.4.1 Concepts in stochastic modelling

There is no widely accepted strict definition of a stochastic model, but broadly the term is applied to a simulation-based approach to allow for uncertainty in the outcomes of an algorithm or system. Generally, inputs to an algorithm or calculation (which itself is deterministic in nature, such that a given set of inputs lead predictably to a fixed given output) are drawn at random from probability distributions, and the algorithm used to convert the set of drawn inputs to an output value (or a set of outputs). This process is repeated a large number of times, with each repeat usually referred to as an iteration or realisation, generating a dataset of input values with their corresponding outputs (Frey and Patil, 2002; Helton, 2008). Statistical analysis of this dataset can then be used to explore relationships and variability in both the inputs and the outputs of the system. This can be contrasted with a deterministic "what-if" or scenario analysis (which in many contexts has been superseded by stochastic approaches). In a deterministic scenario analysis, the algorithm or calculation is used to generate a set of outputs for a single set of inputs, often those considered to be the most likely, or for a small number of alternative sets of inputs.

This process of repeated simulation using inputs drawn from specified probability distributions is commonly known as a Monte Carlo simulation, owing to its similarity to observing results from repeated plays of casino games (Metropolis, 1987). Again there is considerable variation in the specific definition of this term (with some authors reserving it for application of stochastic simulation techniques to specific mathematical problems (Ripley, 2009)). The technique was first developed in order to solve complex mathematical problems (Metropolis and Ulam, 1949), classically high order differential equations, and notable early applications involved simulations during the development of nuclear weapons during the 1940s. However, the flexibility of the technique has made it attractive across a wide variety of fields and it stochastic modelling is now used across many areas of science and business. An early problem was that simulation-based methods tend to be relatively computationally intensive, so early use tended to be limited to simpler problems. Over the past twenty years, the reduction in the cost of processing power (Lloyd, 2000), and its parallel increase in availability have led to much more widespread use.

There are two key contexts in which stochastic models are commonly used; these can be thought of as alternative “modes” of stochastic modelling. Firstly, the technique can be used in a research context, whereby the simulation model is developed and run by the researchers, the results analysed and published, and this new knowledge used to inform further research and applied in the relevant practical field. There are many examples of this in fields as diverse as physics (Baeurle, 2009; MacGillivray and Dodd, 1982), engineering (Lamprou et al., 2013; Tsekouras and Koutsoyiannis, 2014), systems biology (Wilkinson, 2009) and public health (Forastero et al., 2010). In the field of agriculture, there were early reports of development of stochastic models to represent dairy herd management (Sørensen et al., 1992), but widespread adoption has been slow and additional applications have only been reported much more recently (Geary et al., 2012; Hockey and Morton, 2010; Hutchinson et al., 2013; Lu et al., 2013; Shalloo et al., 2004).

The second context in which stochastic modelling techniques are applied is in “close to user” decision support. In this context, simulation is used in a case-by-case way (commonly using existing research or knowledge to inform choice of input distributions and build the deterministic algorithm linking inputs to outputs) to evaluate the potential impact of uncertainty on choices or decisions. A large number of commercially available tools, mostly based on existing spreadsheet platforms, have been developed to facilitate this process (with Pallisade @RISK, and Oracle Crystal Ball perhaps the best known). This technique is very widely adopted in business, especially in the financial sector. Here, use of stochastic models to evaluate alternative investment opportunities and to aid in risk management is very common (Evans and Jones, 2009). There is clearly some overlap between these two contexts, and in many cases construction of a simulation model for use in a research environment can lead to its development for use in case-by-case decision support. This represents a key way in which research can drive changes in industry practice, and there is massive scope for stochastic models of complex systems (such as dairy herd reproduction) to allow multiple sources of research knowledge to be integrated and made accessible to decision makers.

A controversial aspect of stochastic modelling is the user’s choice of input distributions. In many cases this involves an element of subjectivity, and can have a major impact on the conclusions drawn from the model. In many cases there is no formal or quantitative knowledge to inform choice of input distributions, and elicitation of expert opinion is commonly used (Budnitz et al., 1998; Cooke and Goossens, 2004). Inputs are commonly defined either as uniform distributions (so that a range of plausible values for a given input is specified, and there is an equal chance of drawing any value from the range at each iteration) or according to some formal probability distribution. Commonly used probability distributions include beta and triangular distributions (Audigé and Beckett, 1999; Heller et al., 2011), as these are relatively simple to specify (requiring only a range of plausible values and a “most likely” central value). In most cases, distributions which are constrained to a range of values are used (for example, as opposed to a normal distribution, where any real value can potentially be drawn). Correlations between input parameters can also be specified, by drawing inputs from joint multivariate distributions (Nelson et al., 2010).

1.4.2 Probabilistic sensitivity analysis

Probabilistic sensitivity analysis (PSA) is an application of stochastic modelling devised for studying the relative importance of different inputs into a complex system (Oakley and O'Hagan, 2004). A stochastic model is constructed to represent the process to be studied, input distributions defined and the model run over an appropriate number of iterations. Output values from each iteration are stored along with the input values drawn for that iteration. The distribution of the output values across the iterations, and the correlations between specific inputs and any output of interest can then be analysed, providing a way to evaluate the relative extent to which different model inputs affect outcome (Helton et al., 2006; Oakley and O'Hagan, 2004).

Prior to the advent of PSA, alternative strategies were used to select combinations of simulation input values for sensitivity analysis. The simplest example of this is to select the value for each input which is considered most likely, and run the simulation once using this set of inputs (the deterministic scenario analysis approach referred to in Section 1.4.1). Clearly, this has a number of limitations, and in most practical scenarios, this particular set of inputs (which may represent a situation which is perfectly "average" in every respect) is highly unlikely to occur in practice. Alternatively, the extreme values considered likely or plausible for each input can be identified, and the model run using combinations of these extremes. This provides an idea of the most extreme outputs possible given plausible input ranges, but gives no information about the relative frequency with which such extreme outputs are likely, or of the most likely outputs (Frey and Patil, 2002). Modifications of the "design of experiments" (DOE) approach have also been used to select sets of inputs for simulation models. Generally, this approach is used over small runs of a simulation, and classical DOE theory (Anderson and McLean, 1974; Jacquez, 1998) used to select sets of inputs which explore combinations of possible or plausible inputs in the most efficient manner. This small set of simulation runs is then often used to fit a metamodel; a simpler equation estimating outputs from a given set of inputs which represents the more complex simulation process and can be validated against it and used to explore relative importance of inputs and make predictions (Kleijnen and Sargent, 2000; Vonk Noordegraaf et al., 2003). This approach is substantially more computationally economical that PSA, as a much smaller number of simulated iterations is required, but it can be more difficult to represent specific independent or joint distributions of inputs. This approach has previously been applied to population-level animal disease control (De Vos et al., 2006; Vonk Noordegraaf et al., 2002). The increase in computer processing power over the past decade has made PSA a relatively more attractive option; although it is considerably more computationally intensive, it is a more flexible and conceptually simpler approach.

PSA was originally developed as a tool for cost-effectiveness analysis in medicine, and it is in this field that the term is most commonly encountered (although there are many descriptions of a similar process in the literature which do not use the term PSA). In this context, PSA has been extensively employed to evaluate the likely cost-benefit of various public health screening and preventive programmes (Anderson et al., 2006; Gillies et al., 2008). Speigel et al. (2003) describes use of PSA to evaluate two alternative pharmacological options for management of human chronic arthritis. A novel treatment (a selective cyclooxygenase-2 inhibitor, or coxib) was compared to the current standard treatment of a non-selective non-steroidal anti-inflammatory drug; essentially the newer treatment was considered to have a preferable side-effect profile but with increased cost. In this scenario, it was not possible to perform a robust deterministic cost-benefit analysis, as there was uncertainty both in the research findings relating to the safety and efficacy of both products, as well as in wider factors such as treatment costs and patient factors. In order to measure the cost-effectiveness of each strategy across a wide variety of potential scenarios, probability distributions were defined for each uncertain input parameter. In this case, triangular distributions were used for all inputs. A simple decision tree model was used to convert the set of input values drawn at each iteration of the simulation into an outcome, which was stored along with the input draws. This analysis revealed that the cost per additional quality-adjusted life year under the coxib treatment was strongly likely to be above commonly used decision thresholds (such that the new strategy would not be considered cost-effective), but that this outcome was most sensitive to the price of the new product. PSA is so ubiquitous in the field of health economics that some national policy bodies (such as the National Institute for Health and Clinical Excellence in the UK) now require use of PSA to demonstrate cost-effectiveness before new treatment or prevention strategies are adopted (Andronis et al., 2009).

Probabilistic sensitivity analysis is also finding applications outside of public health. Steinbach et al. (2012) describe the use of PSA to evaluate the potential cost-effectiveness of introducing traffic calming measures across various urban scenarios, allowing quantification of uncertainty in the cost benefit outcomes. PSA has also been used in a veterinary context, with a very early report describing its use to assess cost-benefit of alternative diagnostic measures and treatment strategies in bovine respiratory disease complex (Detilleux, 2004); but simulation-based publications in the agricultural and veterinary literature have more commonly used restricted sets of inputs based on one of the approaches mentioned above (Lu et al., 2013; Nielsen et al., 2013). More recently, there have been some veterinary examples of application of PSA or similar techniques. Heller et al. (2011) describe the development of a stochastic model to represent the acquisition and transmission of methicillin-resistant *Staphylococcus aureus* in the canine population and its use within a PSA framework (although the term “PSA” is not used) to explore and rank the relative importance of the different inputs into the model. Other studies using a partial or full PSA approach (usually without reference to the term “PSA”) have also been published more recently, with examples relating to economics of sexed semen

(Hutchinson et al., 2013), impact of specificity in automated heat detection systems (Hockey and Morton, 2010) and economics of lameness (Bruijnis et al., 2010).

1.5 Conclusions

Dairy herd reproduction would appear to be a useful potential application of stochastic modelling and PSA: it is a complex multilevel system where events are nested within lactations, nested within cows, nested within herds. A very large number of factors are known to influence fertility at each of these levels, but the relative importance of these different inputs to the system is often unclear. Improving understanding of the key drivers of fertility performance, and their likely relative importance is critical if good quality decisions about potential interventions to improve herd fertility are to be made. This in turn is essential to ensure that dairy farming has an economically and environmentally sustainable future, both in the UK and elsewhere.

This study aims to apply some of the techniques developed during the “big data” revolution to routinely-collected dairy herd management data, in order to quantify recent trends in and current level of reproductive performance in UK dairy herds and to explore factors affecting this. Chapter 2 describes the collection, restructuring and auditing of the dataset, along with the development and implementation of novel, objective measures of the quality of recording of reproductive data. Chapter 3 presents descriptive statistics on the level of and long-term trends in reproductive performance in this sample of herds, as well as additional background information about the herds.

The subsequent three chapters aim to explore the association between common endemic disease events in dairy cows and fertility. Chapter 4 describes the use of multilevel discrete time survival modelling within a Bayesian framework, to test the hypothesis that clinical and/or subclincal mastitis is associated with depressed reproductive performance in affected cows. In Chapter 5, development of a stochastic simulation model is presented. This simulation model is then used as the basis for a PSA to contextualise the results of Chapter 4 and determine the potential for a herd's level of clinical and subclinical mastitis to affect herd-level reproductive performance. Chapter 6 describes the application similar techniques to investigate the association between clinical lameness events and reproductive performance, both at individual cow level (using multilevel discrete time survival modelling) and at herd level (using stochastic simulation and PSA). Both mastitis and lameness are extremely common in dairy herds, and their incidence rates are very variable across herds (Barker et al., 2010; Bradley et al., 2007; Whitaker et al., 2004). Although there is a body of existing evidence associating these diseases with changes in reproductive performance, large-scale studies are very rare and there is little or no evidence to support likely changes in a herd's reproductive performance resulting from improved management of these diseases.

Chapter 7 aims to evaluate the degree to which milk yield and constituent concentration in early lactation predict pregnancy rate. Milk constituent concentrations in particular are commonly used in the UK and elsewhere as proxy measures of energy balance. As energy balance is known to have a strong association with pregnancy rate this aspect of the study aimed to demonstrate the extent of this relationship and the validity of such proxy measures as well as to explore the proportion of variation in herd pregnancy rate which is explained by routinely captured milk recording data.

A better understanding of the factors associated with reproductive performance could help target resources towards more profitable areas, both in terms of individual farm decision making and in terms of research funding allocation. This will facilitate improvement in reproductive performance, which in turn is critical in the context of the global need for sustainable intensification of dairy farming, with optimal cow fertility providing a potential route whereby level of production can be increased without additional resource cost.

Chapter 2 Data collection, restructuring and auditing

2.1 Data collection

A group of 20 veterinary surgeons throughout England and Wales (Figure 2-1) were contacted in September 2009 to request copies of datasets from dairy herds under their care which were considered to have a high standard of data recording. The veterinary surgeons within the group were practitioners acknowledged as having a special interest in performance monitoring in dairy herds; and were routinely using data from the herds to monitor performance and disease. This represented a non-probabilistic convenience sample, but this was deemed appropriate as high quality data were required for analysis and it was not possible to acquire this using a true probabilistic sampling method. There is a clear trade-off between potential introduction of selection bias (as a result of using data only from well-recorded herds) and the ability to draw meaningful conclusions from analysis (which would be decreased if herds with poorer recording were included). Use of such a convenience sample means that, although results for this sample of herds were more likely to be robust (as events appear well recorded), it is more difficult to generalise the results of the analysis to the wider population of herds.

A total of 468 herd datasets were received; just over half of these were from the central databases of milk recording organisations, and the remainder form a variety of on-farm software recording packages. Of the latter, the vast majority came from four packages: Interherd (Interagri/NMR; UK), Uniform Agri (Uniform BV; Netherlands), Total Dairy and Dairymate (Sum-IT; UK). No stipulations were made as to the timeframe over which data were recorded; each herd dataset comprised data for that herd from the time at which they began recording using that method until the time at which they were submitted. Similarly, no other specifications were made about herd eligibility for the study; for example herds were not excluded on the basis of breed, calving pattern or management system. No additional information about the source herds was requested. Datasets were anonymized on submission, with herds allocated an identification number according to the order in which their data arrived. Animal identification (such as official ear tag numbers and pedigree names) were overwritten with new values based on the sequence number of the herd. No personal data (or any data which could be used to identify individual businesses, people or animals) were stored as part of the project.

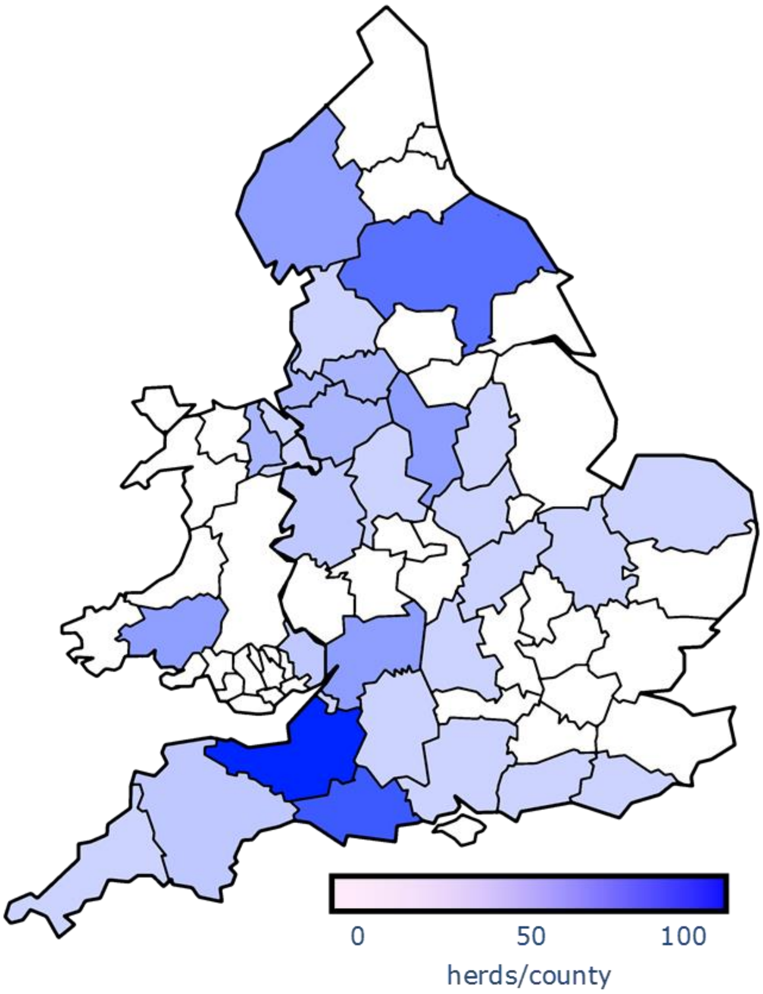

Figure 2-1 Location of herds from which datasets were submitted
Colours represent the number of herds submitted from each county across England and Wales

2.2 Data restructuring and quality assessment

2.2.1 Initial data handling

Datasets were converted to a consistent format; in most cases this was carried out using TotalVet (Sum-IT; UK), an analysis programme designed to use data from a wide variety of different sources. Data were stored as individual files from each herd in Access 2010 (Microsoft Corp.) file format. Within each file, separate linked tables stored information on each cow, each lactation and each event or milk recording test for each cow in that herd. Table 2.1 provides an outline of the structure of the files. Because each part of the project required different elements of data, often structured in different ways, no single amalgamated master version of the dataset was created. Instead, the collection of individual herd dataset files were used as the basis for data extraction for each section of the study, producing a bespoke dataset containing only the data required for that analysis (see Section 2.2.4).

Table 2.1 Structure of database file used to store each herd dataset

Table	Example fields
Cows	Animal ID (UK eartag format ID with prefix based on herd study ID and suffix allocated sequentially to animals in the herd)
	Date of birth
	Date of entry to herd
	Date of first calving
	Date of exit
	Exit reason/code
Lactations	Lactation ID (animal ID suffixed with lactation number, separated by underscore)
	Animal ID
	Lactation number
	Start calving date (date of the calving beginning the lactation)
	Date of conception
	Drying off date
	End calving date (date of the calving beginning the next lactation)
	305-day adjusted lactation milk yield
Events	Animal ID
	Lactation ID
	Event sequence number
	Date of event
	Event name
	Remarks/result/categories (dependent on data source)

2.2.2 Data quality assessment

Initial measurement of data quality was carried out across all the datasets for the years 2007-8, in order to assess the level of data recording across a selection of apparently well-recorded herds. Two of the veterinary surgeons submitted data from all herds under their care which undertook milk recording, as they were unsure which they considered to have sufficiently good data quality to be useful. The 88 datasets submitted in this way were included in the data collation process described in Section 2.2.4 (such that their data were included in all analyses for time periods within which the herd met the recording quality criteria). However, they were excluded from the descriptive analysis of data quality presented in this section, as they are essentially drawn from a different population (all herds undertaking milk recording) to the rest of the datasets (herds considered to have good quality records).

A set of novel measures of the quality of recording of reproductive events was developed and applied. These were designed to detect datasets where event data were missing either at random or in a systematic fashion. The main events of interest were calving and serve records: calving records alone would provide information on overall fertility performance, but serve records would be required to evaluate the component parts of fertility analysis (submission rate and pregnancy rate, see Section 1.2.3). Clinical experience in the field suggests that, in most cases, calving events are better recorded than serve events (as a calving event is required to begin a new lactation, so missing calving records typically prompt reminders either in data capture systems for milk recording organisations, or in on-farm software); but it was considered important to evaluate both. Since reproductive data for the first breeding of a heifer's life are typically very poorly recorded (and fertility in these animals is clearly separate to that in the milking herd), these data were not evaluated in this section of the study or used in any subsequent sections.

Two key metrics were employed to detect datasets with systematically missing data. The first was designed to reveal a pattern of under-recorded serve events in data from milk recording organisations. Where fertility data are collected at milk recording test days (which typically occur at monthly intervals, or occasionally less frequently), in some cases only the most recent serve for each cow is recorded at each data collection visit. This means that a maximum of one serve event per month is recorded for each cow. As the typical oestrus cycle length of a cow is 18-24 days, it is quite possible for cows to have two serves between milk recording test days (and in situations where the sensitivity of heat detection is high but the specificity low, such that there are a large proportion of inter-service intervals of less than 18 days, this may occur in a relatively high proportion of lactations). In order to assess this, the proportion of serve events which were the second of a pair of serves for the same cow between consecutive milk recording test days was calculated. It was anticipated that this would be very close to zero in datasets where this pattern of under-recording was a feature. Another common systematic under-recording error is failure to record unsuccessful serves. This pattern of errors can arise where the perception of the farmer or recorder is that the most important reason for recording serve events is to predict the date of the next calving. This leads to a falsely high pregnancy rate (proportion of serves leading to a pregnancy), so the apparent pregnancy rate in each herd was used to highlight datasets where this feature was potentially present.

In order to discover herds where serve events were under-recorded at random, the proportion of calving events where there was no corresponding serve record was calculated. A serve was considered to correspond to a calving event when it occurred 266-296 days before the calving; a range set to cover the normal gestation periods of common dairy breeds and their beef breed crosses (McGuirk et al., 1999, 1998). A low proportion was considered potentially indicative of under-recording of serve events (including unrecorded natural serves). The ratio of serve events to calving events was also calculated, again aiming to identify datasets where under-recording of serve events was a potential problem.

For each of these measures, the distribution of results across all the herds for the period 2007-8 was assessed in order to assign a threshold value for each metric. These distributions are shown in Figure 2-2. In some cases there was a clear cluster of herds which had a very extreme value for a given data quality measure; for example there was a group of herds where pregnancy rate was greater than 0.85. In other cases, herds were more evenly distributed (for example, in the proportion of serves which was the second of a pair between milk recording test days). In these cases it was more difficult to set a cut-off at which data quality could be considered suspect, and where there was no calculable expected value or existing evidence to guide expectation, clinical judgement was used to set a threshold. Thresholds are shown in Figure 2-2.

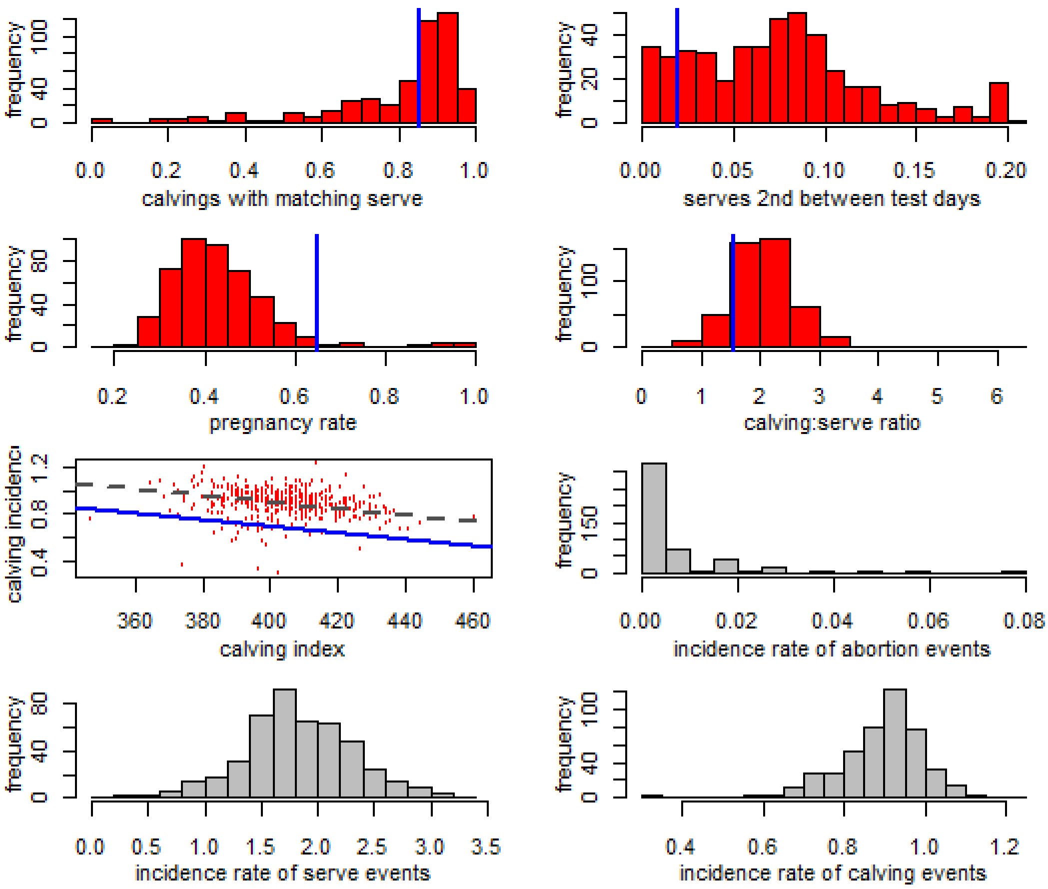

Figure 2-2 Distributions of measures of fertility data quality across 380 herds for 2007-8

Blue lines represent the chosen threshold levels for apparent good data quality. Charts coloured red show measures which were used to exclude herds or herd-years from analysis for at least one part of the study; charts coloured grey show further measures not used. Definitions of the measures are given below:

- *calvings with matching serve: proportion of calving events where a corresponding serve event was recorded (266-296d earlier)*
- *serves 2nd between test days: proportion of serve events which were the second of a pair between milk recording test day*
- *pregnancy rate: proportion of serves leading to a pregnancy*
- *calving:serve ratio: ratio of calving events to serve events*
- *calving incidence: incidence rate of calvings (events per cow-year)*
- *calving index: mean days between successive calvings*
- *incidence rates are shown as events per cow-year.*

In order to identify potential under-recording of calving events, the incidence rate of calvings (i.e. the number of calvings per cow per year) was calculated for each herd. However, as a low rate of calvings could alternatively reflect poor reproductive performance, this parameter was assessed in parallel with the herd calving index using a scatterplot (Figure 2-2). A line was added to the plot to represent the expected relationship of incidence rate and calving index, assuming herd size remained constant (such that the line had a gradient of 1/365 and crossed the point where calving index = 365 and incidence rate = 1, representing a herd where each cow calved once per year). To identify herds where further investigation of under-recording of calving events was required, a further line was added using an intercept at 90% of the "expected" line. Herds below this line were investigated further.

Recording quality for various disease events was also included in the initial assessment. This took the form of a simple herd incidence rate calculation for a number of different events over the years 2007-8. The results of this analysis are shown in Figure 2-3.

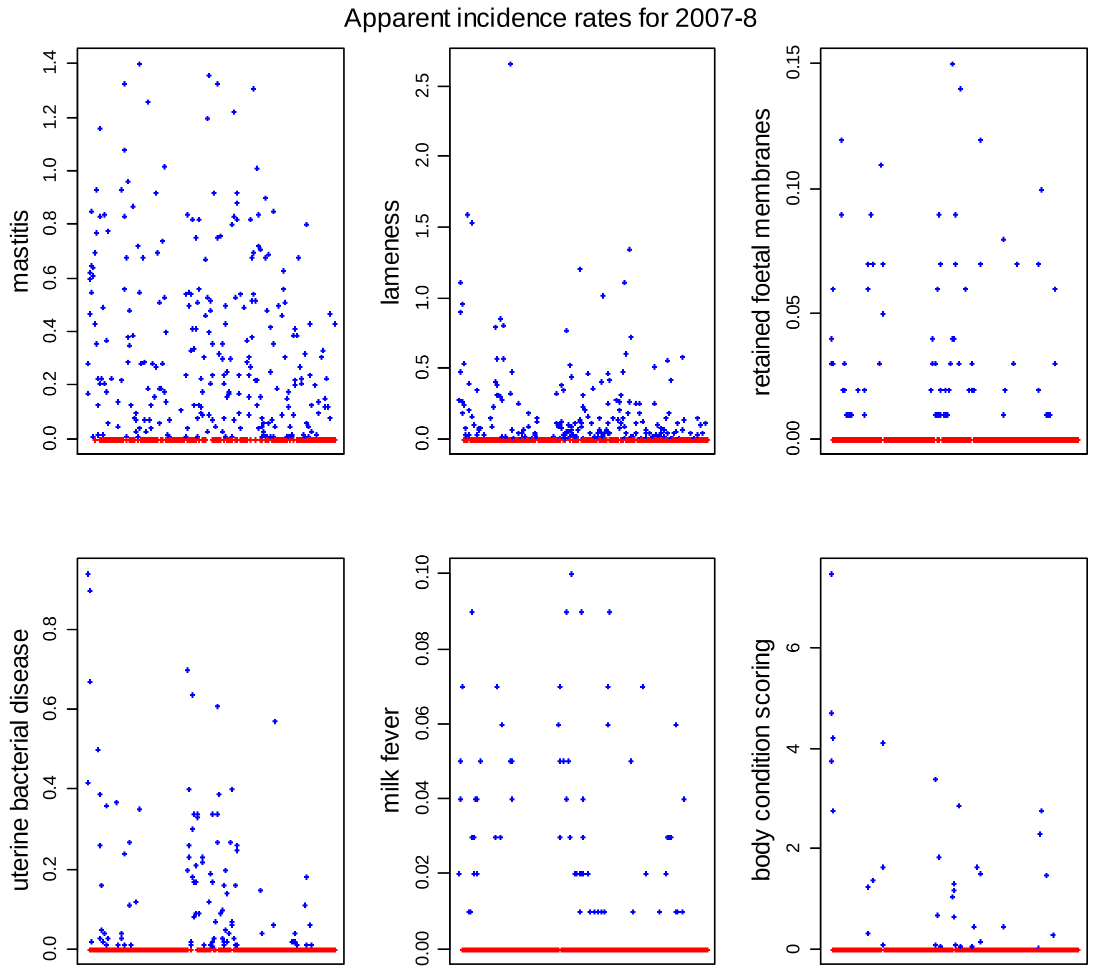

Figure 2-3 Data quality measures for recording of disease events and body condition scores
Each plot shows the distribution of calculated incidence rates across herds for the event indicted on the y-axis. Each blue dot represents a herd with a non-zero incidence rate, while each red dot represents a herd with an incidence rate of zero. Herds are spaced horizontally according to their study identification number. Incidence rates are shown as events per cow-year.

In addition to the quantitative measures of data quality, some manual checks were also employed for every dataset. The sequence of events in a random selection of 25 lactations was assessed to ensure that there was no evidence of missing events or anomalies. For example, extremely long gaps between serve events in a lactation with a very long calving interval (>500 days) were considered potentially suggestive of a missing abortion event. Where there was one or more anomalous event sequence in the 25 lactations, a further 25 were assessed, and if there was more than one anomalous sequence from the 50 lactations then a note was made against the dataset (such that it could be excluded from analyses if considered appropriate). Changes in apparent incidence rate of disease and serve events over time was also assessed visually: where there were very large fluctuations in rate between consecutive months, or where there were periods of time where recording apparently stopped (and where such variation was not explained by the herd's calving pattern), this was noted and the dataset excluded from analyses where aberrant recording of that disease or of fertility events was important.

Assessment of data quality was carried out separately for each part of the project (as different aspects of herd records were used for each, see later chapters). This involved measurement of the core (fertility-related) measures of data quality described above for each herd in each calendar year from 1999 to 2008, so that data were used from all herd-years where recording quality met the criteria. Additional data quality measures (dependent on the additional event or testing data to be used) were applied for each chapter; these are described in the Materials and Methods section of the chapter.

2.2.3 Assessment of voluntary waiting period

In order to maximise the accuracy of analysis of herd reproductive performance (Chapter 3), the voluntary waiting period (VWP) was evaluated for each dataset. In most cases, there was no indication within the dataset of the herd's VWP policy. One dataset format (recorded using Interherd software) included a value for the herd's VWP, but this was set to the software default value for the vast majority of herds submitted in this format, and even in herds where this had been changed it was difficult to be sure that this reflected current herd policy. Therefore, VWP was assessed by examination of the distribution of calving to first serve intervals in the herd for the years 2007-8, on the basis that this would allow evaluation of the earliest time after calving at which cows were being served (and therefore considered eligible).

Two approaches were used to derive a value for VWP from this distribution: visual assessment and calculation of a fixed percentile of the intervals. Visual assessment relied on the author's clinical experience to assign an appropriate VWP to the distribution, generally selecting a value where there was a clear increase in the number of first serves being given at a particular interval after calving. In some herds, a consistent policy is very rigorously applied, and in these cases it is simple to identify a VWP. However, it is very common for a smaller number of cows to be served during the last few days of the VWP; and in some herds there is little clear evidence of any policy at all. Figure 2-4 shows an example of this: the herd from which the left hand graph is derived clearly serve very few cows at less than 36 DIM but there are a substantial number of serves in each bin of the histogram above 36 DIM. In contrast, the right hand graph comes from a herd where there is less of a consistently applied policy, with small numbers of intervals at very low DIM and a gradual increase in the number of intervals in each bin from around 30 DIM onwards.

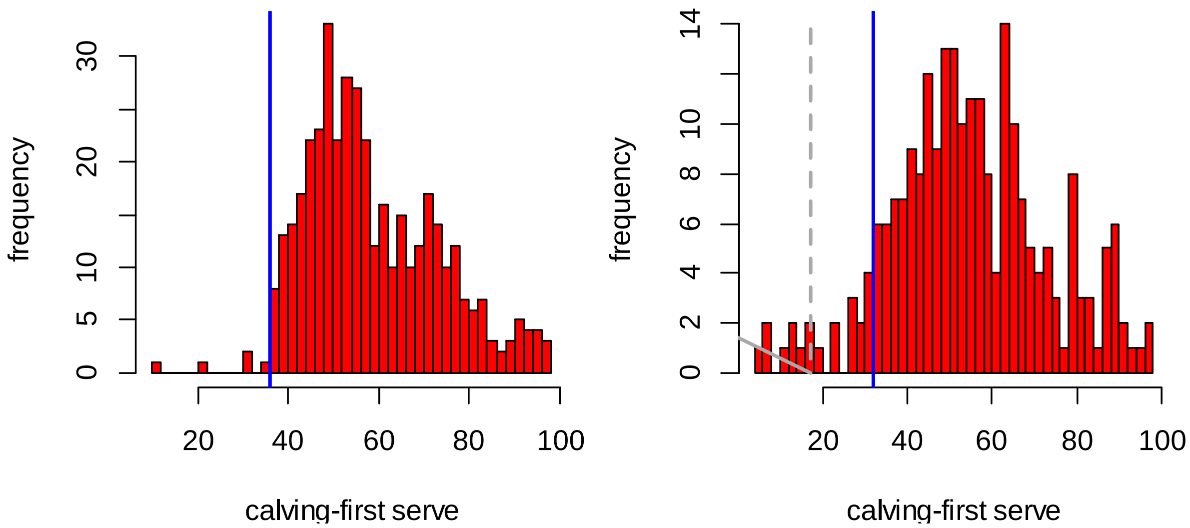

Figure 2-4 Distributions of calving to first serve intervals from two example herds
Distributions of calving to first serve intervals for a herd with a clear (left) and a much less clear (right) voluntary waiting period policy. Visually assigned values for VWP are shown as blue lines, and on the right hand graph the dashed grey line shows the second percentile of the intervals (for the left hand graph, this gave the same result as visual assessment, such that the lines overlie each other).

In order to explore the possibility of automatically deriving an objective and consistent value for VWP from a given distribution of calving to first serve intervals, a fixed percentile of the distribution was calculated. In order to identify which percentile appeared to best represent the herd policy, a series of percentiles was calculated (from one to five at integer intervals) from a subset of the first 50 herds submitted, and these compared to the visually assessed VWP for each herd (as shown for second percentile in Figure 2-5). The second percentile (i.e. the calving to first serve interval where two per cent of the cows had shorter intervals and 98% longer intervals) appeared to correlate best with a manually assessed VWP, so this was calculated for each herd dataset. The distributions of VWP values calculated using each method, and the correlation between values derived using the two methods are shown in Figure 2-5. Detailed examination of datasets where the second percentile value was very different (in almost all cases much smaller) to that assigned by visual assessment suggested that these were herds with little consistency in VWP, and that the manually assigned value better reflected true herd policy. The visually assigned VWP was used for the analysis of herd reproductive performance described in Chapter 3.

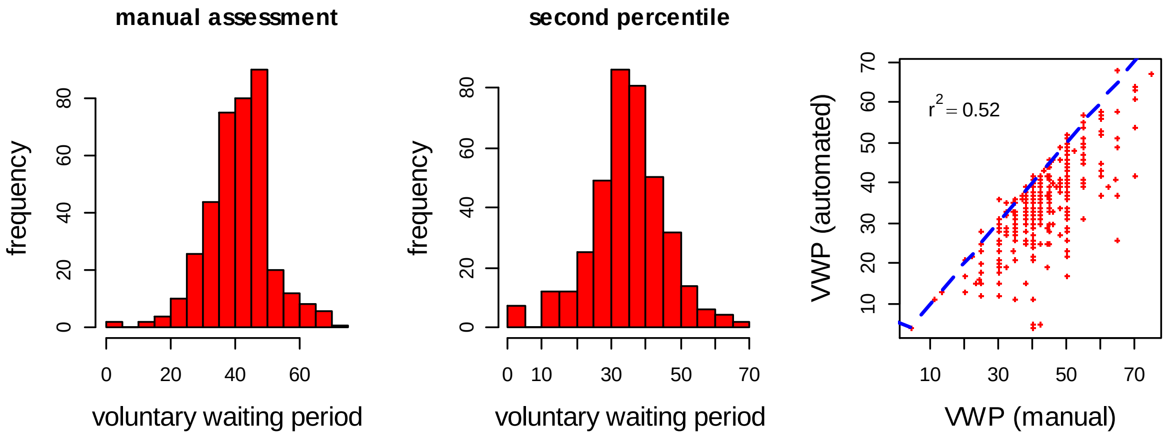

Figure 2-5 Apparent voluntary waiting periods
Distributions of apparent voluntary waiting period (VWP) across herds for the years 2007-8, with values assigned by visual evaluation(left hand graph) or calculation of the second percentile (central graph) of the distribution of calving to first serve intervals. The right hand graph shows the correlation between visually assessed and automatically (by calculation of the second percentile) assessed VWPs across herds.

2.2.4 Restructuring data for analysis

In order to generate a dataset to be used for each analysis within the project, data were collated from every herd in the dataset from lactations beginning in 1999 to 2008. The structure of the collated data varied depending on the analysis to be carried out: for example, the units (or "lines") of data represented periods of risk within a cow's lactation in some analyses (for example, as described in Sections 4.2.2.1 and 6.2.2), and serve events in others (for example, as described in Sections 4.2.2.2 and 7.2.3). Detail was added to each unit of data, depending on what was required for the analysis (for example, whether a pregnancy resulted during the risk period, or a binary indicator of occurrence of a disease event within a time window relative to the risk period). The data structure used for each analysis is described in detail in each chapter. Visual Basic for Applications (Microsoft Corp.) was used within Microsoft Access 2010 (Microsoft Corp.) to automate the process of retrieving the required elements of data from each dataset in turn and aggregating them to generate a single table containing data from every herd. This process could be applied to the complete set of herd datasets, or to any subset thereof, as the automated process was set up to collate data from every file within a specified physical or network location. This collated data was then used to apply the measures of reproductive event recording quality described in Section 2.2.2, along with any additional quality measures required for that element of the study, to each herd's data in each calendar year. The results of this were used to filter out units of data from the main aggregated table, such that only data from herd-years which met the data quality criteria were carried forward to the next stage. Further data quality measures were then applied at lactation level. For example, lactations with unresolved outcomes (i.e. the cow had not either re-calved or left the herd) were discarded. Data from the remaining lactations were then exported from Microsoft Access as a comma separated text file. For most analyses, further data restructuring, such as creation of additional variables or variable recoding, was carried out in R 2.14.2 (R Core Development Team, 2010), using the base, "plyr" (Wickham, 2011) and "reshape" (Wickham, 2007) packages, before the data were exported to analysis software as described in each chapter.

2.2.5 Comparison of herd subsets used in each chapter

The subsets of herds used in each chapter of the thesis were compared, in order to evaluate the potential for increased selection bias in chapters where smaller subsets were required to capture recording of additional data. Figure 2-6 illustrates the differences in distribution of a variety of parameters across these subsets.

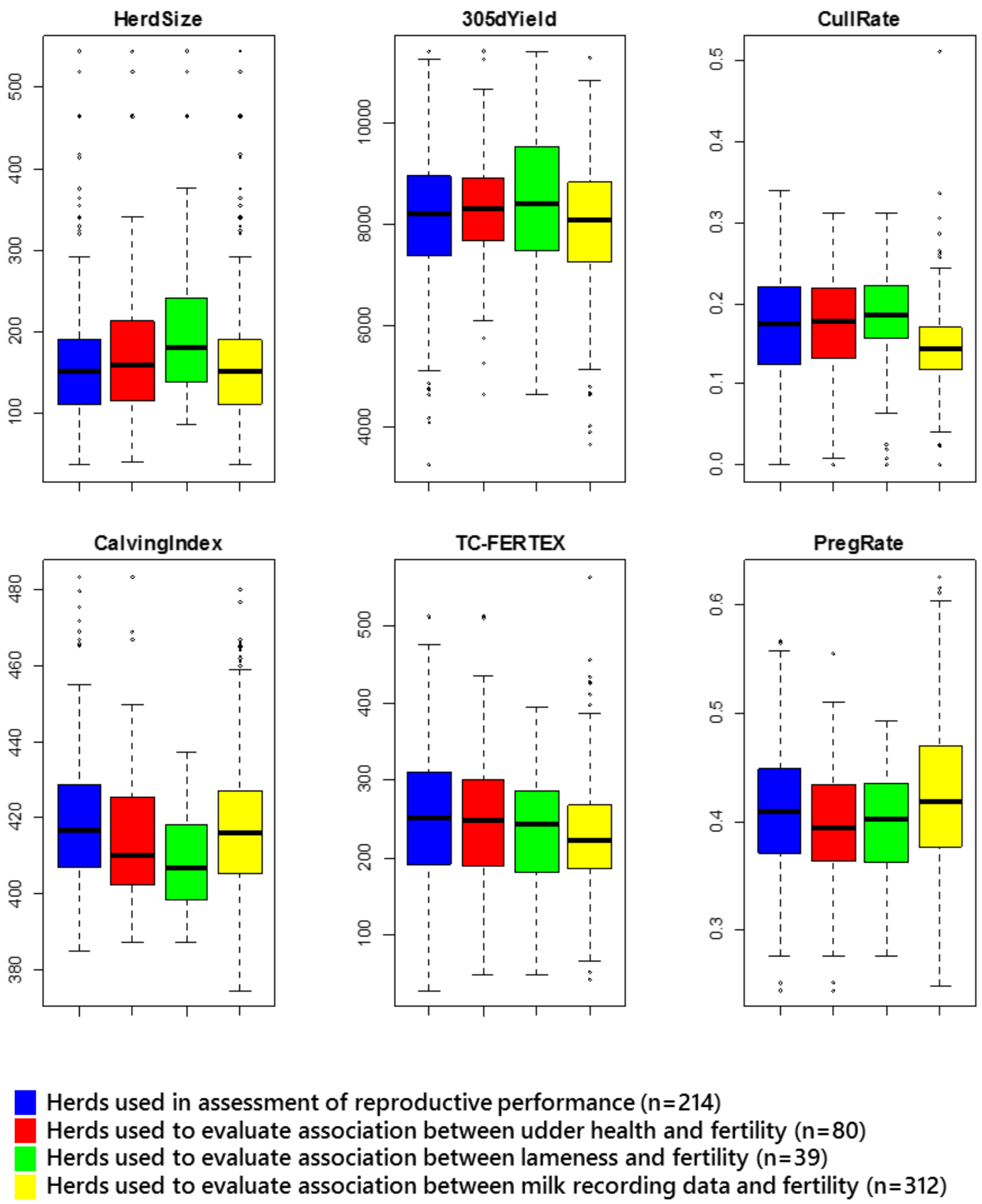

Figure 2-6 Distributions of herd parameters for herds used in each section of the study.
Boxplots showing distribution of herd size, 305d milk yield, culling rate, mean calving index, TC-FERTEX (a cost based measure of reproductive performance and wastage) and pregnancy rate; across the herds used in Chapter 3 (blue), Chapter 4 (red), Chapter 6 (green) and Chapter 7 (yellow) of the thesis.

Distributions for most parameters are very similar; in general the subset of herds used to evaluate the association between lameness and fertility (Chapter 6) was the most different from the other subsets. This is likely because this was the smallest subset of herds, and therefore had greater potential for selection bias and sampling variation. These herds tended to be larger and slightly higher producing; to have slightly higher culling rates and lower calving indices but very similar TC-FERTEX scores. TC-FERTEX is a cost-based measure which combines calving index and overall culling rate, by comparing each to a target (here set at 365 days for calving index and 5% for culling rate) and applying a unit cost to deviations from these targets (set at £2.50 per day on calving index and £1000 per cull) to produce a cost per cow per year (see Section 3.2 for further detail). This suggests that the smaller subset of herds had similar overall levels of reproductive performance, but differed in their culling policy (such that infertile cows were more likely to be culled than to re-calve with an extended interval).

There were also some differences between the two largest herd subsets. This is most noticeable in the lower culling rate amongst the herds used to evaluate associations between milk recording data and fertility (Chapter 7) compared to those used for assessment of reproductive performance in Chapter 3. This due largely to a difference in the way these were calculated: for the smaller dataset performance was measured across the whole herd after herd-level data quality screening (such that data from anomalous lactations could be included provided the herd-year met the data quality criteria), whilst for the larger subset culling rate was estimated directly from the data used for model building (which had undergone additional data quality screening). When culling rate in the smaller subset is calculated in the same way as in the larger subset, the distribution of culling rate is almost identical in both. Pregnancy rate was also slightly higher in the data used for Chapter 7; this is likely to be due to the increased use of data from earlier in the recording period for this work. This was enabled by increased automation of the data quality screening process in the later work, which increased ability to use shorter or intermittent periods of good recording as well as allowing data to be screened over a longer period of time); Chapter 3 demonstrates a trend for pregnancy rates to decrease over time.

Collectively, this suggests that surprisingly little additional selection bias is revealed when higher quality data is required for analysis – the smaller subsets of herds with better data quality are broadly very similar in characteristics to the wider subset of herds. It is therefore relatively unlikely that this additional selection bias will have altered the results in the relevant chapters. It is also important to understand that additional selection bias will only make an important difference to results if the relationship being examined is different in sample of herds used for analysis. For example, it is evident that the herds used for the analysis of lameness and reproduction (Chapter 6) tended to be larger than the wider population. However, this is only important if the relationship between fertility and lameness is very different in larger herds.

2.3 Discussion

This chapter provides an outline of an approach to measurement of the quality of recording of reproductive and other events in dairy herd data. There is very little pre-existing work in this field; and assessment of data quality in large scale epidemiological studies is often not described, or is described only in very vague terms (Berends et al., 2008; Bruun et al., 2002; Lavon et al., 2011). Clearly, in smaller scale studies it tends to be possible for the experimenters to have a greater degree of involvement in and control of the recording of event data, but where the number of herds sampled is of a similar order to that described here, this becomes impossible. However, such data are widely used in clinical veterinary practice to monitor herd performance and to inform management decisions, and artefacts due to poor recording could have a large impact on the results of such analysis.

Designing robust and objective measures of data quality will always present a major challenge, and there is no definitive way to identify datasets where under-recording has occurred. Whilst the standard inter-related structure of reproductive events within a lactation (beginning with a calving, followed by one or more serves, and ending with the subsequent calving or exit from the herd) allows some degree of objective measurement for fertility data, records for individual clinical events are much harder to assess. There is no perfect method to determine whether a dataset with very few records for a given disease represents under-recording, or simply a herd with a genuinely very low incidence of disease. It is acknowledged that the data quality measures described in this chapter will not distinguish perfectly between herds with substantial recording errors and herds with good records but unusual data patterns, but the application of the data quality controls described at both herd and lactation level will minimise the impact of recording errors on study results and thresholds for data quality were generally set to maximise quality of the data used in analysis even if this meant excluding a large number of herds.

It is important to note that many of the datasets which failed to meet quality criteria would still potentially be useful for monitoring herd performance in practice. For example, the measures of missing serve data would potentially exclude herds where a bull is grouped with non-pregnant cows and pregnancy achieved by natural service. In this situation, a high proportion of natural serves usually go unrecorded but if regular pregnancy diagnosis is carried out it is still possible to undertake useful performance monitoring (for example, by measuring the proportion of eligible cows becoming pregnant per unit time). Many herds employ a "sweeper" bull system, whereby cows are eligible for artificial insemination for a set period after the VWP elapses, after which time they join a group running with a bull. In this system, there are frequently unrecorded serves which lead to a calving, an indicator used directly here to exclude herds. In a clinical setting, it would be possible to analyse data for the cows eligible for artificial insemination separately. However, these methods proved usable in identification of a set of herds in which data recording quality was unlikely to influence analytical results; this is a key requirement for the application of "big data" techniques.

The distribution of VWP across herds also provides useful information (Figure 2-5). The majority of herds appear to be using VWPs of between 30 and 50 days, although there are a number with either very short or very long VWP values. Using visual assessment to assign VWP, there are very few herds with a VWP shorter than 25 days. Short VWPs are often an effective way to minimise the average calving interval in a herd, especially where reproductive performance is not strong; although cows conceiving at less than 40 DIM will be due for drying off at 260-280 DIM (with a 280 day gestation period and 40-60 day dry period). This is likely to be relatively inefficient for these individual lactations, and cows at this stage of lactation often still have relatively high milk yields at drying off. This increases the risk of intra-mammary infection during the dry period (Green et al., 2007), as well as making drying off practically more difficult. However, where reproductive performance is poor, relatively few cows will conceive this early, and these disadvantages in a small number of lactations may be outweighed by shifting the distribution of calving intervals for the whole herd to the left, leading to an overall improvement in efficiency of production.

There were also some herds with VWP longer than 50 days (up to 80 days). Long VWPs are employed to avoid the problems described in the previous paragraph, and to maximise first serve pregnancy rate. However, they are usually only justified where reproductive performance is exceptionally strong (such that cows become pregnant at a very rapid rate once their VWP is over). The relationship between VWP and reproductive performance is described in Section 3.3. The inability to determine a way to evaluate herd VWP without human intervention could represent an example of the limitations of the “big data” approach; inevitably there are some aspects of data restructuring which are less amenable to automation.

This chapter has outlined some approaches to measuring the quality of recording of reproductive event data in dairy herds, and described the level of data quality across a sample of herds considered by their veterinary surgeons to have good records. Methods based on “big data” concepts were used to develop an automated approach to collating and restructuring data from multiple herds into a standard format for statistical analysis. These data were used in subsequent chapters to assess the current state of reproductive performance in these UK herds, and to explore factors associated with fertility performance.

Chapter 3 Fertility performance in UK dairy herds

3.1 Introduction

It is clear that successful management of cow fertility is critical to success in running a profitable dairy enterprise (LeBlanc, 2007), and it is well recognised that fertility performance in the majority of UK dairy herds is sub-optimal. Dairy cow fertility is multifactorial, and is often very challenging to manage (Lucy, 2001). This increases the importance of maximising use of the available data in order to target interventions and changes as accurately as possible: as with all aspects of herd health and production management, the over-riding objective is to identify the areas of management where the most achievable changes have the greatest certainty of producing the biggest benefits (Green et al., 2012). As discussed in Section 1.2.1, monitoring and measuring herd fertility is a key first step in improving performance. This relies on an understanding of what level of performance is achievable, in order to highlight areas where improvement is feasible and likely to be profitable.

Internationally, there have been a number of studies evaluating reproductive performance across samples of herds. These are summarised in Table 3.1. Methods used to calculate reproductive indices and terminology used to describe them varies substantially between different nations, and differences in prevalent management systems also influence some of the measures. For example, the balance between calving index and fertility culling rate (see Section 1.2.3) will be different between nations where seasonal calving is common (e.g. New Zealand and Ireland) and those where year-round calving is more usual (e.g. USA and Canada). This means that making comparisons between nations can be challenging, and using these studies to inform performance monitoring in the UK is potentially dangerous.

Table 3.1 Summary of studies examining dairy herd reproductive performance worldwide
After Hudson et al. (2012)

Parameter	Country	Year	Value	Reference
Calving index	USA	2000	429d	deVries and Risco (2005)
	Ireland	2000	395d	Mee (2004)
	Norway	2005	383d	Refsdal (2007)
	USA	2006	422d	Norman et al. (2009)
	UK	2009	426d	National Milk Records plc (2009)
Failure to conceive rate	Ireland	2000	14%	Mee (2004)
	New Zealand	2002-4	9.0-10.2%	Compton and McDougall (2010)
	Norway	2005	6%	Refsdal (2007)
8-week in calf rate[1]	New Zealand	2002-4	78-83%	Compton and McDougall (2010)
6-week in calf rate[2]	Australia	1996-8	63%	Morton (2003)
100 day in calf rate[3]	Australia	1996-8	53%	Morton (2003)
Calving to first serve interval	USA	2001	104d	deVries and Risco (2005)
	Norway	2005	86d	Refsdal (2007)
	USA	2006	86d	Norman (2009)
Pregnancy rate[4]	USA	2006	30%	Norman (2009)
First serve pregnancy rate[4]	Australia	1996-8	49%	Morton (2003)
	USA	2006	31%	Norman (2009)

In contrast, information about fertility performance in UK dairy herds is scarce. This is partly because sources of reproductive data are limited: some milk recording organisations hold very large centralised databases containing basic fertility information, but herd data from these sources are inevitably of variable quality. Typically, calving events are relatively well recorded in such data, and this allows at least an estimation of calving index as a measure of overall fertility performance. The most recent estimate of mean calving index from such a dataset

[1] Defined as percentage of cows eligible for serve at the start of the breeding season that had conceived by 8 weeks into the breeding season.
[2] Defined as percentage of cows eligible for serve at the start of the breeding season that had conceived by 6 weeks into the breeding season.
[3] Defined as percentage of cows eligible for serve that had conceived by 100 days into lactation.
[4] Defined as percentage of serves leading to a pregnancy

(National Milk Records, 2009) was 426 days. However, whilst this type of figure is useful, it is based on data of uncertain quality and allows only a crude estimate of overall fertility performance. Alternatively, smaller datasets consisting of data subjectively considered to be of good quality have been evaluated (Kerby, 2009) – these give more reliable and detailed information, but selection bias in collection of such data is potentially a major issue, and it is difficult to make inferences from such work about fertility performance in the wider UK dairy cow population.

Perhaps the most recent detailed evaluation of fertility performance on a national basis in the UK was by Esslemont and Kossaibati (2002), describing the results of analysis of data from 52 herds from around the UK over the course of 10 years between 1989 and 1999. This suggested that dairy cow fertility was declining, and that both submission and pregnancy rates were contributing to this. Lack of accurate national fertility performance data can make it difficult for farmers, veterinary surgeons and other advisors to see the fertility performance of a herd in context and understand realistic scope for improvement. This study therefore aims to describe the current level of and trends in reproductive performance in a large group of well-recorded dairy herds from across England and Wales. Whilst it is accepted that results from this work may not generalise to the wider population of UK dairy herds, this is still likely to be useful to practitioners as it is only possible to measure reproductive performance accurately in herds such as these. This means that this sample may well be representative of the herds in which veterinary surgeons and other advisors are monitoring fertility, and as such provides useful information to contextualise a particular herd's performance.

3.2 Materials and Methods

This study used the subset of the data described in Chapter 2 of herds with good recording of reproductive data (see Chapter 2 for details of data quality audit). Rejection of herds which had evidence of missing serve or calving events (missing either at random or systematically) in every calendar year resulted in a dataset from 214 herds (rejecting 46% of the datasets from on-farm or bureau recording sources and 57% of datasets derived directly from milk recording organisation databases). It is worth noting that most of these datasets would be usable in some way in practice, but would need to be treated critically and possibly investigated further.

The fertility parameters detailed in Table 3.2 were calculated for each lactation, with values summarised across lactations in each herd beginning in each calendar year (i.e. for each "herd-year") from 2000 to 2007. As data were available for each herd at least as far as the end of September 2009, this means that there was a known outcome (based on subsequent calving date) for each lactation included in this evaluation: a cow calving at the end of 2007 (i.e. the last lactation to be included in the analysis) would have had more than 300 days to conceive plus time for a gestation period and recorded calving. Parameters were summarised either as proportions (e.g. pregnancy rate, the proportion of serves leading to a pregnancy) or using the mean or median value for the herd-year (for interval-based measures such as calving index). As overall reproductive performance is reflected by a combination of calving index and failure to conceive rate, a modification of the FERTEX score (Esslemont and Kossaibati, 2002) was used to combine these. This involves comparing each outcome to a target value (set at 365 days for calving index and 5% for failure to conceive rate), then multiplying the difference between the target value and the observed value for each herd year by a unit cost for each outcome (set at £2.50 per day for calving index and £1000 per failure to conceive cull, based on an updated version of the approach used by Esslemont and Kossaibati (2002)). These costs are then added to generate a total cost per cow per year. This method may be an over-simplified way to estimate a herd's true "recoverable cost", but this approach at least provides a robust way of generating a single reproductive outcome to make useful between-herd summary comparisons.

Table 3.2 Reproductive parameters measured and their definitions

Parameter (unit)	Definition
Calving index (days)	Number of days from the calving beginning the lactation to the subsequent calving, with the values for lactations within a herd-year summarised by either mean or median
Failure to conceive rate (%)	Proportion of the lactations in which a cow received at least one serve but which did become pregnant during the lactation (using absence of a first serve in a lactation to identify cows culled for non-fertility reasons)
FERTEX score (£/cow/year)	Combination of calving index and failure to conceive rate in the form of potential recoverable fertility cost per cow per year (£), see text for further detail
100-day in calf rate (%)	Proportion of lactations where the cow became pregnant at 100 days or less after the calving beginning the lactation
Calving to first serve interval (days)	Number of days from the calving beginning the lactation to the first serve, with values for lactations within a herd-year summarised either by mean or median
First serve submission rate (%)	Proportion of lactations where the cow received a first serve within 24 days of the end of the herd's estimated voluntary waiting period
Returns submission rate (%)	Proportion of inter-service intervals falling in the range 18-24 days
Pregnancy rate (%)	Proportion of serves leading to a pregnancy

A value was calculated for each parameter shown in Table 3.2 in each herd-year; for interval-based parameters both the mean and median interval were calculated. In total, data were available from 1306 herd-years across the 214 herds. The distribution of values for each measure of fertility performance for lactations beginning in 2007 was represented using a histogram (to demonstrate current levels of performance and degree of variability), and the mean, median, inter-quartile range and 95% coverage interval calculated. To evaluate trends in the main reproductive parameters over time, the distribution of each parameter across herds in each calendar year was summarised by calculating the median, inter-quartile range, 80% and 95% coverage intervals, and these were represented on a separate line graph for each parameter.

In order to assess correlations between the various fertility indices a scatterplot matrix was generated. As there were a large number of herd-years relative to the variability in the values for many of the fertility indices, high density heatmap scatterplots were used, whereby the density of points in a given area of the plot is represented by colour intensity. Traditional scatterplots were also produced, with fitted Loess-smoothed trend lines. The Spearman-rank correlation coefficient for each relationship was calculated, and represented on each scatterplot; this non-parametric measure of correlation was used as most parameters showed a marked skew in their distributions. Analysis was carried out and plots produced out using R 3.0.2 (R Core Development Team, 2010). A further scatterplot matrix was generated to evaluate associations between herd parameters (e.g. herd size and level of production) and fertility parameters.

3.3 Results

The distributions of fertility parameters across herds for lactations beginning in the year 2007 are shown in Figure 3-1. The median herd had a mean calving index of 418 days, with a median of 398 days and a 100 day in calf rate of 35%. Interval measures at cow-level (e.g. distribution of cow calving intervals within each herd) appeared to be positively skewed in almost all herds, with mean value generally around 20 days longer than median. This is not unexpected, and is commonly seen in practice, with a relatively small number of cows extending mean intervals. Most parameters showed a very wide spread of values; representing a large difference in performance between the best and worst herds.

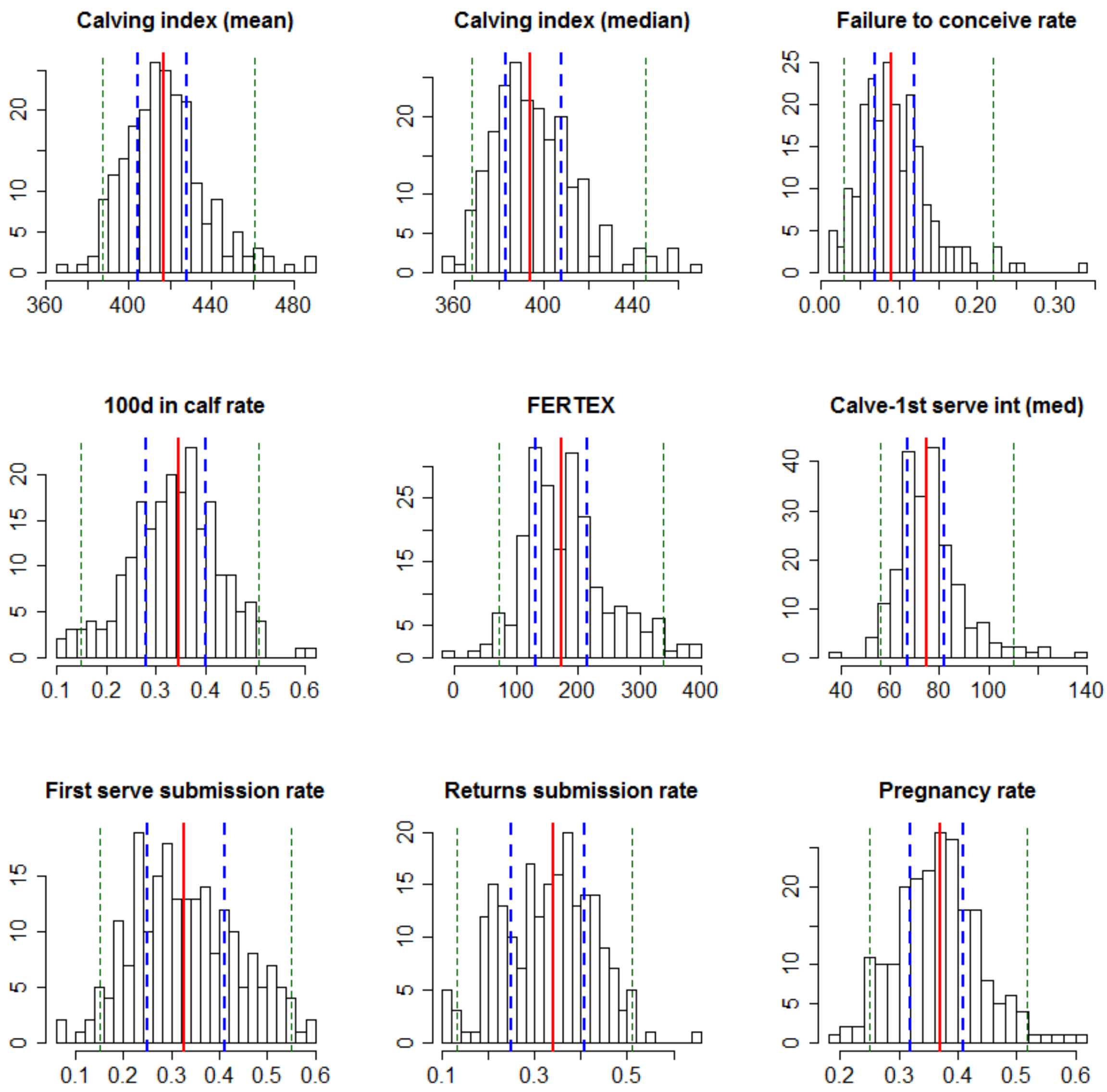

Figure 3-1 Distributions of fertility parameters across 214 herds for lactations beginning in 2007.
Distributions represented as frequency histograms (vertical axes represent frequency). Vertical red lines represent the median value across herds, dashed blue lines the inter-quartile range and dashed green lines the 95% coverage interval.

Trends in reproductive performance over the period studied are shown in Figure 3-2. The general trend in the measures of overall fertility (calving index, failure to conceive rate, FERTEX and 100 day in calf rate) shows a deterioration over most of the study period, which appears to be reversed in the last year examined (2007). The difference between performance in 2006 and that in 2007 was significant (Mann-Whitney test p-value < 0.05; bootstrap hypothesis test p-value < 0.01) for FERTEX score, but not for 100-day in calf rate.

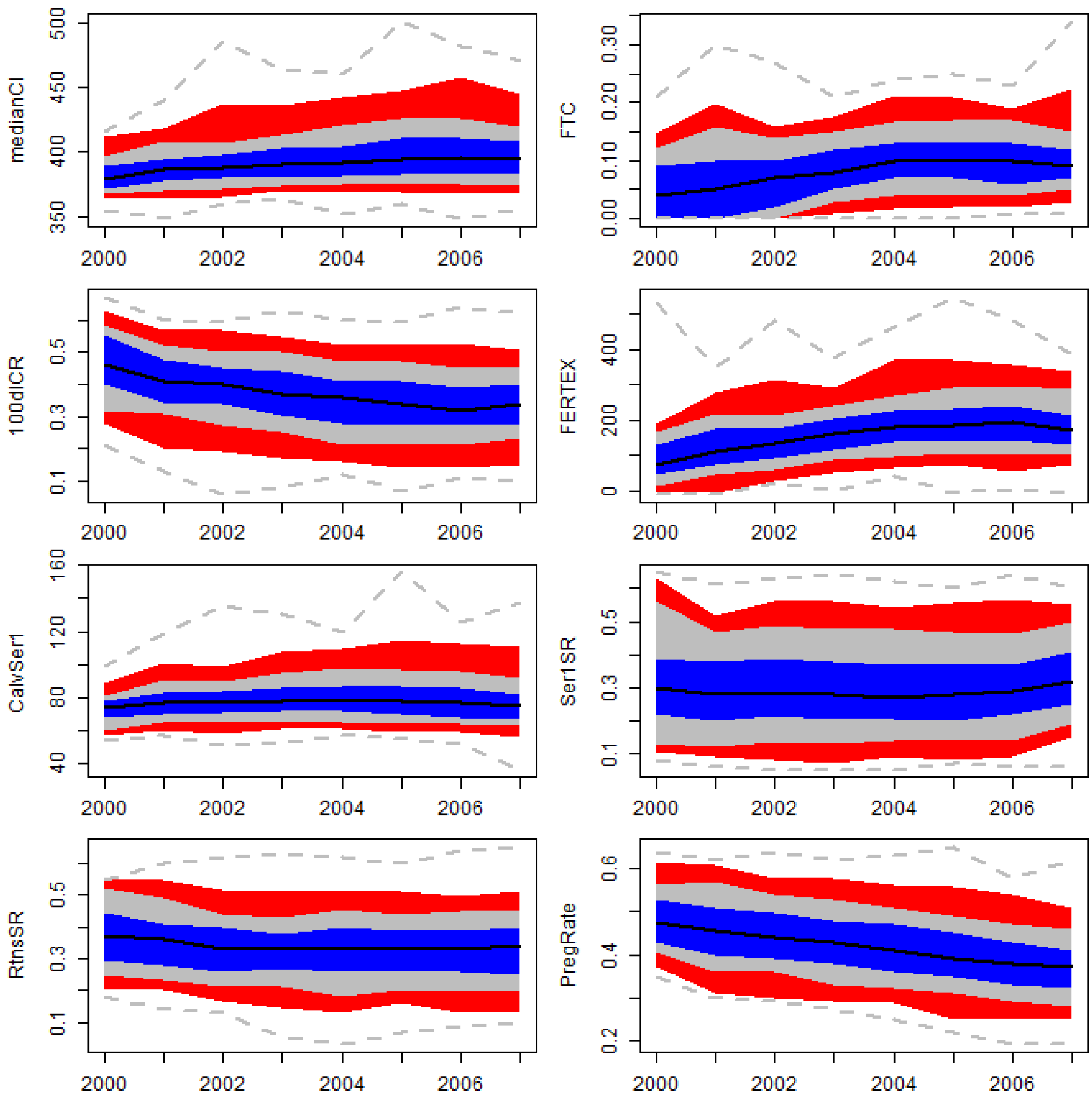

Figure 3-2 Trends in fertility performance over the years 2000 to 2007

Black lines show median value across herds for each parameter in each year; shaded blue areas show the inter-quartile range, shaded grey areas the 80% coverage interval (10th percentile to 90th percentile), shaded red areas the 95% coverage interval and dashed grey lines the range. CI = calving index, FTC = failure to conceive, 100dICR = 100 day in calf rate, Calv-1stSer Int = calving to first serve interval, Ser1 = first serve, SR = submission rate, Rtns = returns to serve; PregRate = pregnancy rate.

The measures of heat detection (first serve and returns submission rates, calving to first serve interval) show a similar pattern, with deteriorating performance over 2000 to 2005, with a reversal in the decline in 2006 and 2007. The difference between first service submission rate in 2006 and that in 2007 was significant (Mann-Whitney test p-value < 0.01; bootstrap hypothesis test p-value < 0.05), unlike the difference in return submission rate. Pregnancy rate deteriorated in a uniform linear fashion over the full period, with the median value declining from 47% in 2000 to 37% in 2007.

These results tentatively suggest that the long-term decline in fertility may have halted. Results for 2007 appear to indicate that overall performance may now be improving, with a statistically significant improvement in FERTEX score between 2006 and 2007. This appears to be mainly driven by an improvement in heat detection, especially for first serves (demonstrated by significant improvement in first serve submission rate between 2006 and 2007) which is large enough to override the continued downward trend in pregnancy rate. There is little change in the variability of the parameters between herds over the study period, with the magnitudes of inter-quartile ranges and 80% coverage intervals generally remaining similar for each parameter over time.

Figure 3-3 shows correlations between the various fertility parameters across all the herd-years. Amongst the indicators of overall reproductive performance, 100-day in calf rate was very strongly correlated with median calving index (Spearman rank correlation coefficient, r_s = -0.91) and strongly correlated with both mean calving index and FERTEX score. There was very little correlation between the failure to conceive rate and either mean or median calving index. FERTEX score was moderately correlated with all three measures of heat detection (calving to first serve interval, first serve and returns submission rates; magnitude of r_s 0.3 – 0.4). Calving to first serve interval and first serve submission rate were moderately strongly correlated, and a weaker correlation existed between both of these first service heat detection measures and returns submission rate. There was very little correlation between any of the heat detection measures and pregnancy rate.

Very few notable correlations were revealed between basic herd information and measures of fertility performance (Figure 3-4). Interestingly, herd 305-day adjusted milk yield showed poor correlations with all of the overall reproductive parameters (magnitude of all $r_s < 0.25$), and only a weak to moderate correlation with pregnancy rate (r_s = -0.31). Herd size was not correlated to a meaningful extent with any of the reproductive parameters; in fact it was best correlated with other basic herd information (being positively correlated with milk yield, r_s = 0.30, and negatively correlated with voluntary waiting period, r_s = -0.25). Voluntary waiting period had a weak to moderate correlation with calving to first serve interval (r_s = 0.39), but was poorly correlated to other parameters.

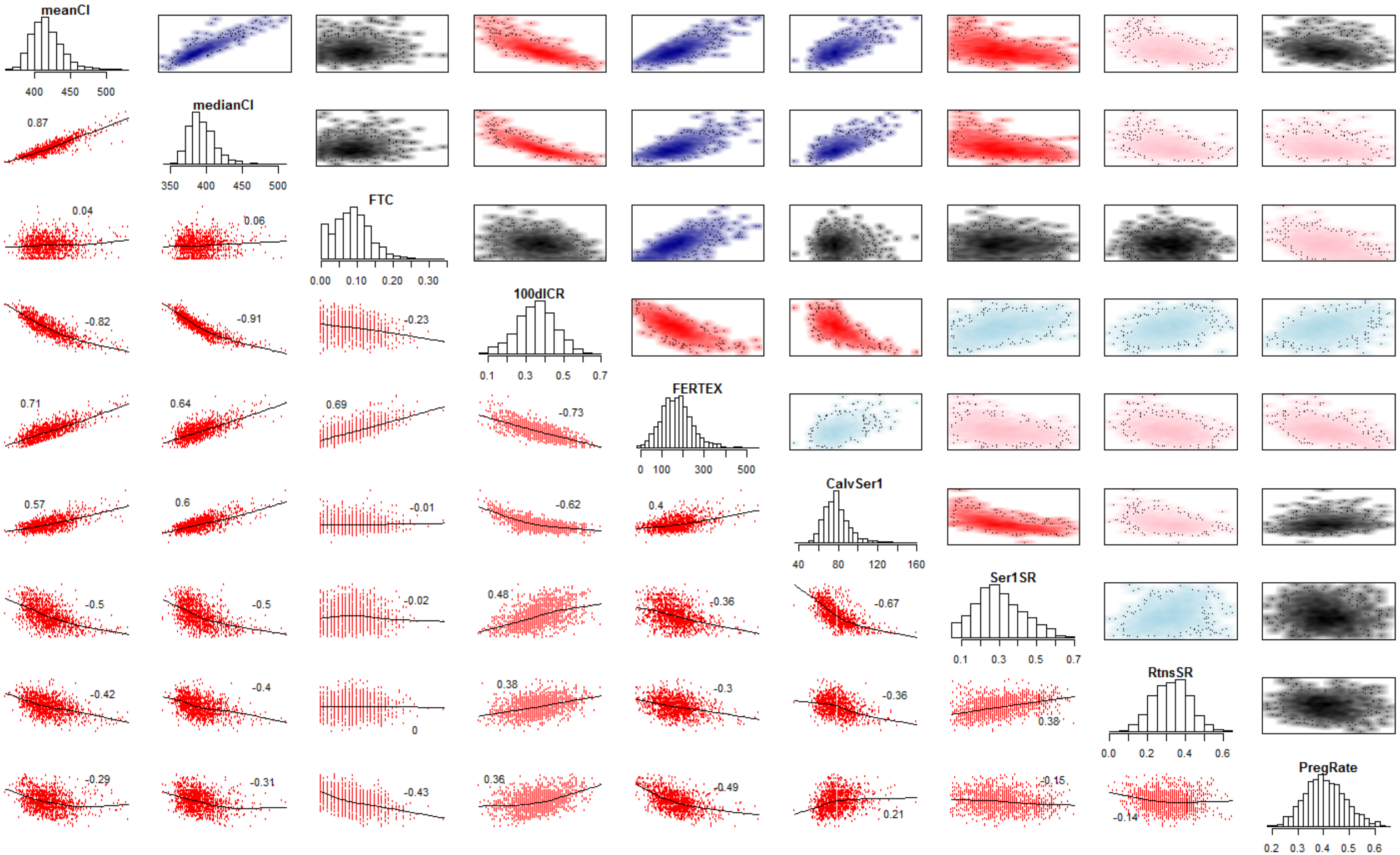

Figure 3-3 Correlation matrix for fertility parameters

Histograms on the diagonal show the distribution of each parameter across herd-years. The lower-left half of the plot shows standard scatterplots with fitted Loess-smoothed lines and the Spearman rank correlation coefficient (r_s) for each relationship. The upper-right half shows heatmap scatterplot: colour intensity shows point density at that point on the plot, colours denote degree of correlation (dark blue: $r_s > 0.6$, light blue: $0.6 > r_s > 0.4$, black: $0.4 > r_s > -0.4$, pink: $-0.4 > r_s > -0.6$, red: $-0.6 > r_s$). meanCI = mean calving index; medianCI = median calving index; FTC = failure to conceive rate; 100dICR = 100d in calf rate; CalvSer1 = median calving to first serve interval; Ser1SR = first serve submission rate; RtnsST = returns submission rate; PregRate = pregnancy rate

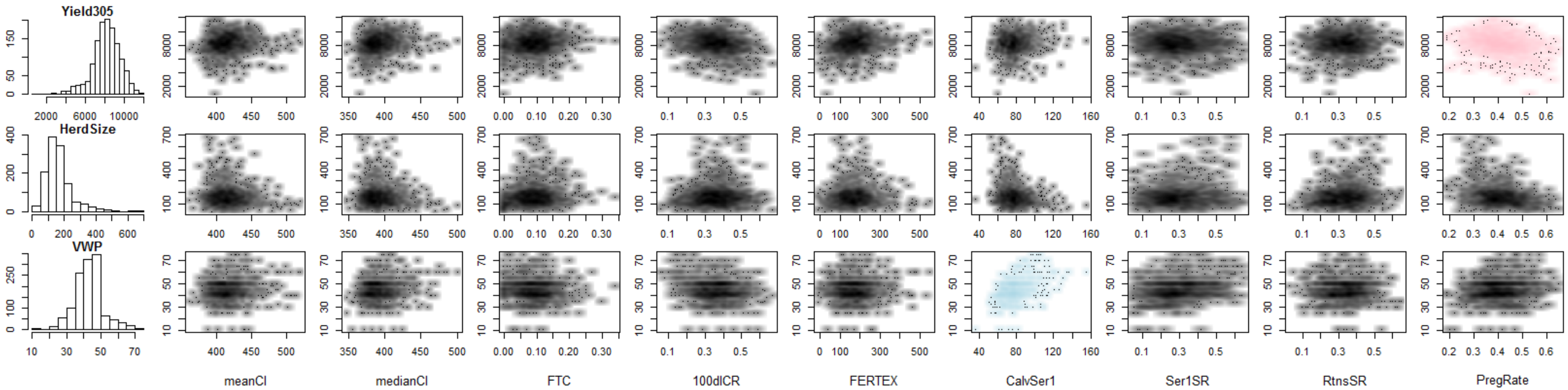

Figure 3-4 Correlations between additional herd information and fertility parameters

Histograms on the left show distribution of each additional herd variable (with titles denoting the variables). High density scatterplots on each row show the relationship between each additional herd information variable (y-axis) and each fertility parameter (x-axis); colour intensity shows point density at each point on the plot. Colours denote value of the Spearman rank correlation coefficient for each relationship (dark blue: $r_s > 0.6$, light blue: $0.6 > r_s > 0.4$, black: $0.4 > r_s > -0.4$, pink: $-0.4 > r_s > -0.6$, red: $-0.6 > r_s$).Yield305 = 305 day adjusted milk yield, HerdSize = herd size, VWP = estimated voluntary waiting period, meanCI = mean calving index; medianCI = median calving index; FTC = failure to conceive rate; 100dICR = 100d in calf rate; CalvSer1 = median calving to first serve interval; Ser1SR = first serve submission rate; RtnsSR = returns submission rate; PregRate = pregnancy rate

3.4 Discussion

These results suggest that there is substantial variation in the level of reproductive performance across these herds, with the bottom 25% of herds having a FERTEX score of greater than £205/cow/year while the top 25% have a value of less than £150/cow/year. Although this may not represent true recoverable cost in these herds, the degree of variation in net cost is likely to be representative. The mean value of the mean calving index across the herds was 418 days, relatively close to the most recently published figure of 426 days from a wider dataset (National Milk Records, 2009).

In order to assess overall fertility performance, calving index should be evaluated in conjunction with a measure of culling for failure to conceive (FTC). In this study, the proportion of cows served which did not re-calve was used. This has previously been described as a proxy for FTC culling rate (Esslemont and Kossaibati, 2002), on the basis that cows which received a first serve have not been identified as culls for non-fertility reasons, and therefore cows which receive a first serve but do not subsequently conceive during that lactation are culled for fertility reasons. A more accurate estimate could potentially be gained using recorded reasons for culling, but culling reasons were only consistently recorded in a very small proportion of datasets, and it is often difficult to assign a single predominant reason for a particular cull. In the small number of herds where this was a usable alternative, calculation of FTC culling rate using reasons for cull provided very similar estimates to the method used for the main analysis.

Submission rates were poor, with median first serve and return to serve submission rates both around 33%. These values are well below proposed target figures of 75% and 60% respectively (Breen et al., 2009; Hudson et al., 2012). However, submission rates are relatively amenable to manipulation, and there is some suggestion that the downward trend in submission rates over the past few years is beginning to reverse (Figure 3-2). These rates are perhaps even lower than would have been expected, and it is worth considering why this may be. Esslemont and Kossaibati (2002) did not directly measure first serve submission rate, but did measure returns submission rate using a similar technique to this study. The value for 1997/98 (the last calving period included) was 48%, which is considerably higher than the earliest value for the current dataset (calvings in 2000, mean returns submission rate 37%). However, reported overall performance from this previous study was also better (mean calving index for calvings in 1997/8 was 390 days) than that in the current dataset (mean calving index for calvings in 2000, 404 days). It is unlikely that performance had declined this severely over such a short period of time, so selection bias towards high performing herds may have been stronger in the earlier study, making a direct comparison difficult to interpret. There are, however, some potential reasons why the current analysis may have produced artifactually low estimates of submission rates.

In terms of first serve submission rate, calculation of this measure is reliant on use of the correct voluntary waiting period (VWP) for each herd. When dealing with a large number of herds, this represents a significant problem: some datasets from on-farm or bureau recording software systems contained an apparently appropriate VWP; but these were a tiny minority. Calving to first serve intervals for each herd were therefore individually examined in order to derive a reasonable estimate of VWP for each herd (see Section 2.2.3), which was then used to calculate the first serve submission rate for that herd. It is impossible to be sure that the VWP values used accurately reflected farm policy, and there may have been changes in VWP over time which were not accounted for. A similar problem exists for seasonally calving herds, which may again have resulted in cows being included in the calculation before the herd's breeding period began. Additionally, some farms may not have a consistent policy regarding first serves (e.g. those who prefer not to serve cows who are yielding above a certain level), which makes accurate estimation of submission rate impossible. Return to serve submission rate is also unlikely to reflect directly the proportion of heats detected, as extended inter-service intervals may be due to embryonic or foetal death as well as failure to detect an oestrus.

Pregnancy rate is the other major determinant of fertility performance. Mean pregnancy rate for lactations beginning in 2007 was 37%, which again is perhaps lower than would be expected, with an inter-quartile range of 32% to 41% (meaning that 25% of herds had a pregnancy rate lower than or equal to 32%). Pregnancy rate has long been recognised to be in steady decline, and this study provides support for this (Figure 3-2), with a decrease of 10% in median pregnancy rate over the seven year period examined. As calculation of pregnancy rate is relatively simple, it is likely that this is representative of the true proportion of serves leading to a pregnancy in these herds. However, it is important to remember that the apparent deterioration in pregnancy rate could be influenced by an improvement in recording of serve events (although if this is the case the most recent figures will be the most accurate).

Correlations between the various indicators of herd fertility (Figure 3-3) revealed strong correlations between 100-day in calf rate and most of the other measures of reproductive performance. This suggests that 100-day in calf rate may be a useful headline parameter for routine monitoring of reproductive performance, as it is much less retrospective than the other overall measures which it would appear to predict. Amongst measures of heat detection, there was only a moderately weak correlation between first serve and returns submission rates; suggesting that there are many herds where heat detection is good for first serves and poorer for returns to serve, and many herds for which the reverse is true. This is potentially significant as these could be herds where more focus on the current “weak” area would lead to easy improvements in performance. Correlations between submission and pregnancy rates were weaker still, suggesting that an improvement in the rate at which heats area being detected does not inevitably reduce pregnancy rate.

Figure 3-4 shows that herd level of production is surprisingly poorly correlated with any of the reproductive measures; with a weak correlation to pregnancy rate the largest association. Many previous studies have demonstrated (often large) associations between increased milk yield and decreased reproductive performance. There are a number of potential reasons why these were not apparent here: it is possible that yield is more important at cow level than at herd level, or that factors confounding this relationship make it impossible to detect in a univariate analysis. Herd size was very poorly correlated with any of the reproductive indices, although larger herds tended to have higher milk yields and shorter voluntary waiting periods. The latter is difficult to explain, but could be related to an increased awareness of the need to manage fertility efficiently in bigger enterprises.

The apparent reversal in the declining trend in fertility performance in this sample of UK dairy herds is encouraging: although the improvement in performance was only apparent between lactations beginning in 2006 and 2007, the magnitude of the change was relatively large and the large number of herds examined makes it less likely that the improvement was due to sampling variation. It is clear from Figure 3-2 that this improvement has largely been driven by an improvement in submission rates, both to first and subsequent serves. This may be the result of an enhanced awareness over the past few years of the importance of heat detection as a key determinant of fertility performance. This may have led to an improvement through increased use of heat detection aids (from simple heat-mount detectors to more complex activity monitor systems), or an improvement in the use of information (e.g. increasing use of on-farm software allowing better awareness of cows due to come into oestrus). An increase in average herd size may also have played a role here, with more scope for specialization within farm staff. Increasing herd size may also increase the size of the sexually active group, potentially making heat detection easier (Hurnik et al., 1975). Whilst it is extremely heartening that the apparent improvement in heat detection has finally begun to show dividends in terms of reversing declining overall fertility performance, it is worth bearing in mind that there is still huge scope for further improvement here in the vast majority of units. Improvement will only continue if efforts with heat detection persist.

The trend in pregnancy rate is less positive, with a steady decline over the entire study period. This is likely to become more and more significant in the future. Over the past few years there has (rightly) been much focus on improving submission rates, often with little regard to pregnancy rate, which has sometime been seen almost as “beyond control”. Submission rates are much more easily influenced, and changes can be implemented with a greater degree of certainty of a larger improvement compared to those aimed at improving pregnancy rate. There is strong logic behind this approach, and in the majority of cases targeting recommendations at improving submission rates is likely to have a bigger overall impact than attempting to improve pregnancy rates. Additional encouragement can be taken from the fact that the improvement in heat detection over the past couple of years does not appear to have been associated with an appreciable acceleration in the decline in pregnancy rate.

However, it is important to understand that the relative importance of pregnancy rate as a determinant of overall performance increases as pregnancy rate itself decreases. This can be illustrated by considering the effect of a 10% decrease in pregnancy rate on herds with different initial pregnancy rates. For a herd with an initial pregnancy rate of 50%, a 10% decrease will mean each cow will require an average of 0.5 extra serves to become pregnant (from 2 serves at 50% pregnancy rate to 2.5 serves at 40% pregnancy rate). Cows in herd starting with a 35% pregnancy rate will require an average of just over one additional serve should pregnancy rate fall by the same 10% (from just under 3 serves at 35% pregnancy rate to 4 serves at 25%). Heat detection is also important here, as this will influence the number of days added to a cow’s calving interval for each additional serve required. While maintaining the (apparently successful) drive to improve submission rates, it is important to preserve awareness of pregnancy rate and the factors that influence it.

This chapter has demonstrated a medium-term deterioration in reproductive performance across a large sample of UK dairy herds, but also provides evidence that improvements in heat detection may be beginning to reverse this trend whilst pregnancy rates continue to decline. There is opportunity at the current time to arrest the decline in pregnancy rate before it becomes a limiting factor in a large number of herds. Understanding the factors affecting fertility in UK dairy cows, and their relative importance at herd level, will help optimise reproductive performance in the future. The next three chapters consider the potential for two very common endemic diseases (mastitis and lameness) to influence a herd's reproductive performance.

Chapter 4 Udder health and fertility: A discrete time survival analysis

4.1 Introduction

In view of the scope for improvement in fertility performance described in Chapter 3, it is increasingly critical to improve understanding of the factors that influence reproductive performance. This encompasses both herd-level management factors (such as methods of husbandry, feeding and oestrus detection) and individual cow-level factors (such as disease events, milk yield and genetics). Mastitis is one of the most common clinical disease events in dairy cattle, with the most recent estimate of incidence rate in the UK between 50 and 70 cases per 100 cow-years (Bradley et al., 2007). As a condition associated with inflammation and pain (Kemp et al., 2008), it is reasonable to hypothesize that mastitis may have a negative effect on cow fertility.

In a previous study, Barker (1998) evaluated reproductive performance in dairy cows in a single herd having a case of clinical mastitis (CM) during early lactation compared to unaffected herdmates, and found that CM before a positive pregnancy diagnosis was associated with a significantly longer interval from calving to conception. This work was extended by Schrick (2001), who confirmed this finding and reported that subclinical mastitis before first serve was also associated with longer intervals to conception. Similar results were later found by several groups using similar approaches (Ahmadzadeh et al., 2009; Gunay and Gunay, 2008; Nava-Trujillo et al., 2010; Santos et al., 2004).

A major drawback of such an approach (i.e. comparison of a population of cows which experienced mastitis with a population which did not) is that it makes it difficult to account fully for factors potentially confounding the relationship between mastitis and fertility. For example, it is plausible that cows with higher milk yields are more likely to develop CM (Windig et al., 2005) and also more likely to have poor reproductive performance (Nebel and

McGilliard, 1993). Alternative methods of data analysis better able to account for such factors include construction of statistical models to predict reproductive outcomes, with or without the inclusion of random effects terms to improve correction for unmeasured or unrecorded factors. Such techniques have been extensively employed in this area, and studies have revealed associations between CM and increased odds of abortion (Risco et al., 1999), abnormal length inter-service intervals (Moore et al., 1991), and failure to become pregnant after a serve (Hertl et al., 2010; Loeffler et al., 1999). Other work has identified associations between subclinical mastitis as measured by increased individual cow somatic cell count (ICSCC) and increased odds of embryonic loss (Moore et al., 2005), abortion (Pinedo et al., 2009) and failure to become pregnant to first serve (Pinedo et al., 2009).

Although there is convincing evidence that both CM and ICSCC can have negative associations with fertility, there is a lack of studies where the association between reproduction and both CM and ICSCC is explored. Furthermore, only the most recent work (Hertl et al., 2010) has evaluated the importance of the timing of CM in a sophisticated way, and this study was performed on a limited number of similar herds. Indeed, much of the work in this area has been carried out in intensively-managed, high producing herds in the USA, and it is unclear how the results of such studies generalise to production systems in other parts of the world with more modest outputs. Perhaps most importantly, the association between CM and ICSCC and the likelihood of a cow being served has not been thoroughly investigated. Early work in the field suggested that a detectable relationship existed, with cows experiencing early lactation CM showing increased calving to first serve intervals compared to control cows (Barker et al., 1998; Schrick et al., 2001). More recent work has tended to focus on the associations between CM or ICSCC and the probability of pregnancy as a result of a given serve.

The current study therefore aims to evaluate the relationship between clinical and subclinical mastitis and reproductive performance across a large number of UK dairy herds, using multilevel discrete time survival modelling.

4.2 Materials and Methods

4.2.1 Data Collection and Restructuring

Initial data collection and restructuring is described in Chapter 2. After removal of herds that failed to meet the inclusion criteria for mastitis and fertility data quality in any of the years examined, 105 herd datasets remained from the 468 initially submitted. Computational limits (maximum size of dataset supported by the software used) determined that 80 datasets could be used for model building, and these were randomly selected from the 105. Basic statistics describing the 80 herds are provided in Table 4.1. Herds were not excluded on the basis of predominant breed, although the vast majority of herds were mainly Holstein or Holstein-Friesian. No data regarding specific management practices in each herd (e.g. use of fixed time insemination, frequency of milking etc.) were available. Clinical mastitis cases were diagnosed according to the normal practice of the herds.

Table 4.1 Descriptive statistics for the 80 herds used in model building

			Percentiles					
	Mean	**Min.**	**10%**	**25%**	**50%**	**75%**	**90%**	**Max.**
Herd size	209	48	105	133	176	232	339	672
305d milk yield (kg)	8,254	4,182	6,941	7,592	8,227	9,151	9,725	11,008
Calving index (days)	417	354	394	404	416	428	437	476
Culling rate (%/year)	22.1	11.5	15.9	18.8	21.7	24.5	28.2	42.8
IRCM (cases/cow-yr)[5]	0.55	0.15	0.23	0.34	0.50	0.70	0.90	1.37
BMSCC (x10^3/ml)[6]	205	77	147	169	204	233	269	367

[5] IRCM: Incidence rate of clinical mastitis (given in cases per cow-year at risk)
[6] BMSCC: Bulk milk somatic cell count, calculated from milk recording data

Each dataset was screened for duplicate event records (including records of multiple cases of CM within seven days of each other, and serves within three days of each other), and duplicate events removed. Further data quality audit was performed at lactation level, and unsuitable lactations discarded (where there were insufficient milk recording test days or where there was no serve date corresponding to the calving that ended the lactation). In order to facilitate construction of a discrete time survival model, data were then amalgamated and restructured into a format where each unit of data represented a two-day “risk period” in every lactation between 20 and 220 days in milk (DIM), such that each lactation could contribute a maximum of 100 units. Two alternative choices of risk period duration (20 days and two days) were evaluated and produced substantively similar results. The shorter risk period was chosen as it allowed a more accurate and detailed assessment of the effect of timing of udder health events. Cows were censored after culling, death, sale or pregnancy occurred.

Occurrence of a serve was recorded as a binary event in each two-day risk period, and occurrence of a cow becoming pregnant was calculated for each risk period. Pregnancy was determined to have occurred where a calving was recorded 266-296d after a serve; this range was designed to cover the normal range of gestation periods for the common dairy breeds and their crosses (McGuirk et al., 1999, 1998). For each risk period, a number of potential explanatory variables were also calculated; these are listed in Table 4.2. CM variables were binary (i.e. CM occurred or did not), with a separate code for "ineligible" (used when the CM variable referred to a timeframe before the lactation began). ICSCC variables were grouped into six categories, as shown in Table 4.3, in order to explore the potential effect of magnitude of ICSCC as well as apparent presence or absence of an intramammary infection (IMI) as defined by a simple threshold. Categorising the ICSCC variables also allowed retention in the dataset of risk periods where there was no test day in the timeframe referred to by one or more of the ICSCC variables. Other categorical variables (such as parity and year) were re-coded as necessary to avoid categories with very small numbers of risk periods: for example, the parity variable was re-coded so that all animals of parity five or above were grouped into a single category. The final dataset consisted of 2,338,025 risk periods from 39,590 lactations in 21,068 cows from 80 herds.

Table 4.2 Variables (with variable type) calculated at each level of data for each risk period

Risk period level	
Served	Binary (served or not served)
Becomes pregnant	Binary (becomes pregnant or does not)
DIM[7] at start of risk period	Continuous
Season of risk period	Categorical (Jan-Mar/ Apr-Jun/ Jul-Sept/ Oct-Dec)
CM[8] 71-90d after risk period	Binary (case of CM recorded or not)
CM 57-70d after risk period	Binary (case of CM recorded or not)
CM 43-56d after risk period	Binary (case of CM recorded or not)
CM 29-42d after risk period	Binary (case of CM recorded or not)
CM 15-28d after risk period	Binary (case of CM recorded or not)
CM 8-14d after risk period	Binary (case of CM recorded or not)
CM 1-7d after risk period	Binary (case of CM recorded or not)
CM during risk period	Binary (case of CM recorded or not)
CM 1-7d before risk period	Binary plus N/A[9] (case of CM recorded or not)
CM 8-14d before risk period	Binary plus N/A (case of CM recorded or not)
CM 15-28d before risk period	Binary plus N/A (case of CM recorded or not)
CM 29-42d before risk period	Binary plus N/A (case of CM recorded or not)
CM 43-56d before risk period	Binary plus N/A (case of CM recorded or not)
CM 57-70d before risk period	Binary plus N/A (case of CM recorded or not)
ICSCC[10] 91-120d before risk period	Categorical - see Table 4.3
ICSCC 61-90d before risk period	Categorical - see Table 4.3
ICSCC 31-60d before risk period	Categorical - see Table 4.3
ICSCC 15-30d before risk period	Categorical - see Table 4.3
ICSCC 8-14d before risk period	Categorical - see Table 4.3
ICSCC 0-7d before risk period	Categorical - see Table 4.3
ICSCC 1-30d after risk period	Categorical - see Table 4.3
Lactation level	
Year in which lactation began	Categorical (2003 or earlier/ 2004/ 2005/ 2006/ 2007 /2008)
Parity of cow	Categorical (1/2/3/4/>4)
CM at 0-14 DIM	Binary (case of CM recorded or not)
305d adjusted milk yield ('000 kg)	Continuous

[7] DIM: days in milk
[8] CM: clinical mastitis
[9] N/A category was used where the timeframe to which the variable referred was outside of the lactation.
[10] ICSCC: individual cow somatic cell count

Herd level	
Herd size	Continuous
Herd cull & death rate	Continuous
Herd average 305d adjusted milk yield ('000 kg)	Continuous

Table 4.3 Categorisation of individual cow somatic cell count explanatory variables

Category	Description
1	0-20,000 cells/ml
2	21,000 - 60,000 cells/ml
3	61,000 - 99,000 cells/ml
4	100,000 - 199,000 cells/ml
5	200,000 - 399,000 cells/ml
6	> 399,000 cells/ml
N/A	Cow not eligible (timeframe to which variable referred was outside of the lactation) or no milk recording within timeframe

4.2.2 Statistical Analysis

4.2.2.1 Probability of pregnancy during a risk period and potential explanatory variables (model 4.1)

Discrete time survival analysis was used to evaluate the association between the probability of a cow becoming pregnant in a given risk period and CM, ICSCC and other potential explanatory variables (Yang and Goldstein, 2003). A three-level hierarchical model was used to account for correlations within the data, with risk periods nested within cows nested within herds. A four-level structure was also evaluated (with lactations within cows as an extra level), but was unsuitable due to the high proportion of cows which only contributed a single lactation - a fixed effect for parity was forced into the model to account for this. The model specification took the form shown in Equations 4.1-4.3.

$$\mathrm{Preg}_{tij} \sim \mathrm{Bernoulli}(\mathrm{mean}=\mu_{tij})$$

$$\ln\left(\frac{\mu_{tij}}{1-\mu_{tij}}\right)=\alpha+\beta_1 \mathrm{lnDIM}_{tij}+\beta_2\left(\mathrm{lnDIM}_{tij}\right)^2+\boldsymbol{\beta}_3\mathbf{X}_{tij}+\boldsymbol{\beta}_4\mathbf{X}_{ij}+\boldsymbol{\beta}_5\mathbf{X}_j+u_{ij}+v_j \quad (4.1)$$

$$v_j \sim \text{normal distribution } (0, \sigma_v^2) \quad (4.2)$$

$$u_{ij} \sim \text{normal distribution } (0, \sigma_u^2) \quad (4.3)$$

where t represents a two-day risk period and i and j the i^{th} cow in the j^{th} herd; μ_{tij} the fitted probability of Preg_{tij} (the outcome of the i^{th} cow in the j^{th} herd becoming pregnant during risk period t); lnDIM_{tij} the natural logarithm of days in milk at the beginning of risk period t; α the regression intercept; β_1 and β_2 the coefficients for the terms representing days in milk; $\mathbf{X}_{tij}$ the vector of risk period level covariates and $\boldsymbol{\beta}_3$ the corresponding vector of coefficients for covariates $\mathbf{X}_{tij}$; $\mathbf{X}_{ij}$ the vector of cow-level covariates and $\boldsymbol{\beta}_4$ the corresponding vector of covariates of coefficients $\mathbf{X}_{ij}$; $\mathbf{X}_j$ the vector of herd-level covariates and $\boldsymbol{\beta}_5$ the corresponding vector of coefficients of covariates $\mathbf{X}_j$; u_{ij} the random effect to reflect variation between individual cows and v_j the random effect representing variation between herds, with σ_u^2 and σ_v^2 the variances of the normal distributions of the respective random effects terms.

Model building was carried out in MLwiN version 2.20 (Rasbash et al., 2010), using iterative generalized least squares (Rasbash et al., 2009) for initial parameter estimation. Final parameter estimates were then generated using Markov chain Monte Carlo (MCMC) with Gibbs sampling in MLwiN (Browne, 2009) using a burn-in chain length of 5,000 and monitoring chain length of 20,000 iterations. Diffuse prior distributions (functionally equivalent to a normal distribution with a very large variance for fixed parameters and a uniform distribution for scalar variances (Browne, 2009, 1998)) were specified for model parameters. Estimate traces for each parameter were visually assessed to ensure that satisfactory convergence had occurred. Use of MCMC for parameter estimation had the advantage of producing parameter estimates that are likely to be more reliable (Browne and Draper, 2006), as well as providing an indication (the deviance information criterion, DIC) of model fit (Spiegelhalter et al., 2002).

Initial model building was by forward selection: explanatory variables were added to the model one at a time (within the described model framework), and retained in the model if the 95% interval of highest posterior density (HPD) of the estimated coefficients for at least one of the variable's categories did not cover zero. Discarded variables were then individually reintroduced to the model, and retained if they satisfied the criteria described above. First order interactions between explanatory variables were considered only if the interaction was held to be of potential clinical importance. This was considered important for model parsimony: inclusion of a large number of interaction terms could easily have led to a very complicated model, but one not giving any extra information that was likely to be of use in practice. None of the tested interaction terms were retained in this model. The possibility that the magnitude of associations between udder health and reproductive outcomes varied from herd to herd was also considered: this was represented by evaluating models with random slopes (Rasbash et al., 2009) for the udder-health-related explanatory variables with the largest coefficients. This led to two candidate models: one including and one excluding random slopes. Of these, the model with the lowest DIC was selected (Spiegelhalter et al., 2002), a lower DIC representing a better combination of model fit and complexity. In this case, the model without random slopes was selected.

The potential for overparamaterization in this type of model was recognised during the model selection process: relative magnitudes of the odds ratios for ICSCC variables (as these were considered to have the greatest potential for correlation) were compared to the patterns in the raw data, and found to be similar. Further evidence that overparamaterization had been avoided was provided by the assessment of model fit using posterior predictions (see below), and the good convergence behaviour of the MCMC chains during final parameter estimation.

4.2.2.2 Probability of pregnancy conditional on a serve and potential explanatory variables (model 4.2)

In order to improve understanding of the relationship between the outcome variable (probability of pregnancy during a given two-day risk period) and the explanatory variables, a second model was constructed. The dataset used to construct this model was a subset of the main dataset described above, containing only risk periods where a serve occurred (so that each unit of data represented a serve). The outcome variable now represented the probability of a cow becoming pregnant to a given serve. This model took a similar form to that described in Equations 4.1 - 4.3, with the exception that a third rather than a second order polynomial term was required to represent stage of lactation. Model building was carried out as described in Section 4.2.2.1.

4.2.2.3 Model assumption checking and assessment of model fit

A simulation-based posterior prediction procedure was used to assess model fit (Gelman et al., 1996; Green et al., 2009). Posterior predictions were generated by simulation using WinBUGS version 1.4 (Spiegelhalter et al., 2003). For computational reasons, it was not possible to generate predictions for each risk period in the dataset for model 4.1, so subsets of 100,000 risk periods were randomly selected. As the dataset for model 4.2 was substantially smaller, it was possible to use every unit of data in this dataset to produce full posterior predictions from this model. Predictions were generated over 10,000 iterations, using the equation:

$$\mathrm{PredPreg}_{tij} \sim \text{Bernoulli probability}\left(\mathrm{P}_{tij}\right)$$

$$\ln\left(\frac{\mathrm{P}_{tij}}{1-\mathrm{P}_{tij}}\right) = \alpha + \beta_1 \ln\mathrm{DIM}_{tij} + \beta_2\left(\ln\mathrm{DIM}_{tij}\right)^2 + -\boldsymbol{\beta}_3\mathbf{X}_{tij} + \boldsymbol{\beta}_4\mathbf{X}_{ij} + \boldsymbol{\beta}_5\mathbf{X}_j + u_{ij} + v_j \qquad (4.4)$$

where P_{tij} is the predicted probability of a cow becoming pregnant during risk period t in the i^{th} cow in the j^{th} herd, $PredPreg_{tij}$ is a draw from a Bernoulli distribution with probability P_{tij}, and all other parameters are as described for Equations 4.1 – 4.3.

In order to assess model fit, units were grouped by variable or risk period (such as herd and parity of the cow, and stage of lactation and season of the risk period), and predictions summarised across the groups. The summary predictions were compared with the observed proportions of units in each group where a pregnancy occurred. The number of observations in each group varied from 438 (number of risk periods where CM occurred during the risk period in the 100,000 risk period subset used for predictions from model 4.1) to more than 10,000 (for example for each parity and season group). Where observed values fell outside the 95% credible interval for the predicted value, further investigation was instigated and steps taken to improve model fit in this region by restructuring of the covariate categories or investigation of interaction terms, so that for the final models the 95% credible interval for each group covered the observed proportion.

4.2.2.4 Posterior predicted relative risks

In order to present results graphically as relative risks (which have a more intuitive interpretation than odds ratios), further posterior predictions were carried out. These were performed using the method described above. Each udder health related variable was considered in turn. For binary explanatory variables (i.e. those relating to CM), the subset of risk periods from the dataset where the value of the explanatory variable was equal to one was selected. Predicted probability of pregnancy was calculated for each of these risk periods over 10,000 iterations as described in Equation 4.4. Predictions were then repeated with the value of the variable under consideration set to zero. For categorical predictor variables (i.e. those relating to ICSCC), the same process was followed for each category of the variable in turn. At each iteration, the total number of predicted pregnancies was calculated

(SumPredPreg) where the covariate was set to one and where it was set to zero. The predicted relative risk (PredRR) for each iteration was then calculated as shown in Equation 4.5:

$$\text{PredRR} = \frac{\text{SumPredPreg (covariate} = 1)}{\text{SumPredPreg (covariate} = 0)} \tag{4.5}$$

The distributions of the predicted relative risks were summarised across the 10,000 iterations as medians and 95% credible intervals.

4.3 Results

4.3.1 Probability of pregnancy during a given time period

A total of 29,237 pregnancies occurred during the 2,338,025 two-day risk periods under analysis. The mean probability of a pregnancy occurring during any given risk period was therefore 0.0125: this would correspond to a probability of 0.131 of a cow becoming pregnant during a given 21-day period. A total of 10,096 of the 39,590 lactations (0.255) contained at least one case of clinical mastitis during the part of the lactation (i.e. 20-220 DIM) used for analysis.

Of the explanatory variables not directly related to udder health (i.e. those included as potential confounding factors), parity, season, year and lactation 305 day milk yield were associated with the probability of a cow becoming pregnant during a risk period. The association between the outcome variable and lactation milk yield was negative (Table 4.4). Seasons 1 and 4 (October to March inclusive) were associated with higher probabilities of pregnancy than the summer months (April to September inclusive). Parities 1,2 and 3 were not different from each other, but parities 4 and greater than 4 were associated with progressively lower probabilities of pregnancy occurring. Parameter estimates for Model 4.1 are shown in Table 4.4: estimates for odds ratios (OR) were calculated by exponentiation of the estimate for each coefficient at each MCMC iteration and calculating the median value and area of 95% highest posterior density (HPD) across all the iterations. Clinical mastitis was associated with the largest reduction in the probability of pregnancy when the case of CM was close to the two-day risk period being evaluated - CM 1-7 days before the risk period had the lowest odds ratio (indicating the largest reduction in the odds of the cow becoming pregnant at that risk period: OR = 0.58, 95% HPD interval 0.50 – 0.66), with CM during the risk period (OR = 0.68, 95% HPD interval 0.53 – 0.85) and CM 1-7 days afterwards (OR = 0.78, 95% HPD interval 0.68 – 0.87) the next lowest.

Table 4.4 Parameter estimates for discrete time survival model with outcome representing the probability of a cow becoming pregnant during a 2 day risk period (model 4.1)

				95% HPD[11] interval	
Model terms	**N**	**Coefficient**	**Odds ratio**	**Lower**	**Upper**
Intercept	2338025	-37.5		-37.6	-37.4
ln[12](days in milk)	2338025	14.2		14.1	14.2
$(\ln(\text{days in milk}))^2$	2338025	-1.47		-1.48	-1.47
Parity 1	611466		Reference		
Parity 2	528926		1.03	1.00	1.07
Parity 3	409107		0.98	0.94	1.02
Parity 4	286885		0.91	0.87	0.95
Parity >4	501641		0.73	0.70	0.77
Centred 305d milk yield ('000 kg)	2338025		0.92	0.92	0.93
Season 1: January - March	640512		Reference		
Season 2: April - June	525806		0.87	0.84	0.90
Season 3: July - September	496246		0.78	0.75	0.81
Season 4: October - December	675461		1.02	0.99	1.05
Year: 2003 or earlier	500673		Reference		
Year: 2004	349349		0.95	0.91	0.99
Year: 2005	408004		0.86	0.82	0.89
Year: 2006	479072		0.88	0.84	0.91
Year: 2007	521664		0.86	0.82	0.89
Year: 2008	79263		0.83	0.77	0.91
No CM[13] 15-28d before	2118515		Reference		
CM 15-28d before	61175		0.88	0.81	0.95
NA[14] for CM 15-28d before	158335		0.23	0.14	0.35
No CM 1-7d before	2307361		Reference		
CM 1-7d before	30664		0.58	0.49	0.66
No CM during risk period	2329414		Reference		
CM during risk period	8611		0.68	0.52	0.85
No CM 1-7d after	2308091		Reference		
CM 1-7d after	29934		0.78	0.68	0.87
No CM 8-14d after	2308312		Reference		
CM 8-14d after	29713		0.87	0.77	0.97
No CM 15-28d after	2280642		Reference		
CM 15-28d after	57383		0.86	0.78	0.93
No CM 29-42d after	2281647		Reference		
CM 29-42d after	56378		0.85	0.77	0.92
No CM 43-56d after	2283316		Reference		

[11] HPD: highest posterior density
[12] ln: natural logarithm
[13] CM: clinical mastitis
[14] NA: not applicable – see Tables 1 and 2

CM 43-56d after	54709	0.88	0.81	0.96
No CM 57-70d after	2285476	Reference		
CM 57-70d after	52549	0.85	0.77	0.93
ICSCC[15] category 1 1-30d after	374291	Reference		
ICSCC category 2 1-30d after	759648	0.96	0.93	1.00
ICSCC category 3 1-30d after	299930	0.94	0.89	0.98
ICSCC category 4 1-30d after	293112	0.92	0.87	0.96
ICSCC category 5 1-30d after	173920	0.84	0.79	0.89
ICSCC category 6 1-30d after	195594	0.77	0.72	0.82
No ICSCC 1-30d after	241530	0.90	0.86	0.95
ICSCC category 1 8-14d before	94172	Reference		
ICSCC category 2 8-14d before	190130	0.90	0.84	0.97
ICSCC category 3 8-14d before	71761	0.89	0.81	0.97
ICSCC category 4 8-14d before	69392	0.88	0.80	0.96
ICSCC category 5 8-14d before	41247	0.95	0.85	1.05
ICSCC category 6 8-14d before	47919	0.95	0.85	1.06
No ICSCC 8-14d before	1823404	0.93	0.88	0.99
ICSCC category 1 31-60d before	285182	Reference		
ICSCC category 2 31-60d before	588530	0.95	0.92	0.99
ICSCC category 3 31-60d before	222427	0.97	0.92	1.01
ICSCC category 4 31-60d before	217363	0.93	0.88	0.97
ICSCC category 5 31-60d before	129870	0.94	0.89	1.00
ICSCC category 6 31-60d before	156561	0.91	0.86	0.97
No ICSCC 31-60d before	738092	0.88	0.84	0.91
ICSCC category 1 91-120d before	140716	Reference		
ICSCC category 2 91-120d before	279156	0.98	0.93	1.03
ICSCC category 3 91-120d before	103605	0.95	0.89	1.02
ICSCC category 4 91-120d before	100249	0.99	0.91	1.05
ICSCC category 5 91-120d before	61998	0.95	0.87	1.03
ICSCC category 6 91-120d before	76983	0.87	0.80	0.94
No ICSCC 91-120d before	1575318	1.06	1.01	1.11

The duration of time for which CM was associated with decreased risk of pregnancy was much greater after the risk period compared to before the risk period: a case of CM 57-70 days after the risk period was still associated with a decrease in risk of pregnancy at the risk period, while no such association was found for predictor variables relating to CM cases more than 28 days before the risk period. The magnitude of association between probability of pregnancy occurring during a risk period and the ICSCC variables was generally smaller than that with the CM variables, and generally increased in size with an increase in ICSCC categories (such that

[15] ICSCC: individual cow somatic cell count

categories representing higher ICSCC values were generally associated with a greater decrease in the probability of pregnancy). These results are illustrated using posterior predictions of relative risks (accounting for the overall likelihood of pregnancy occurring during a risk period) in Figure 4-1 and Figure 4-2.

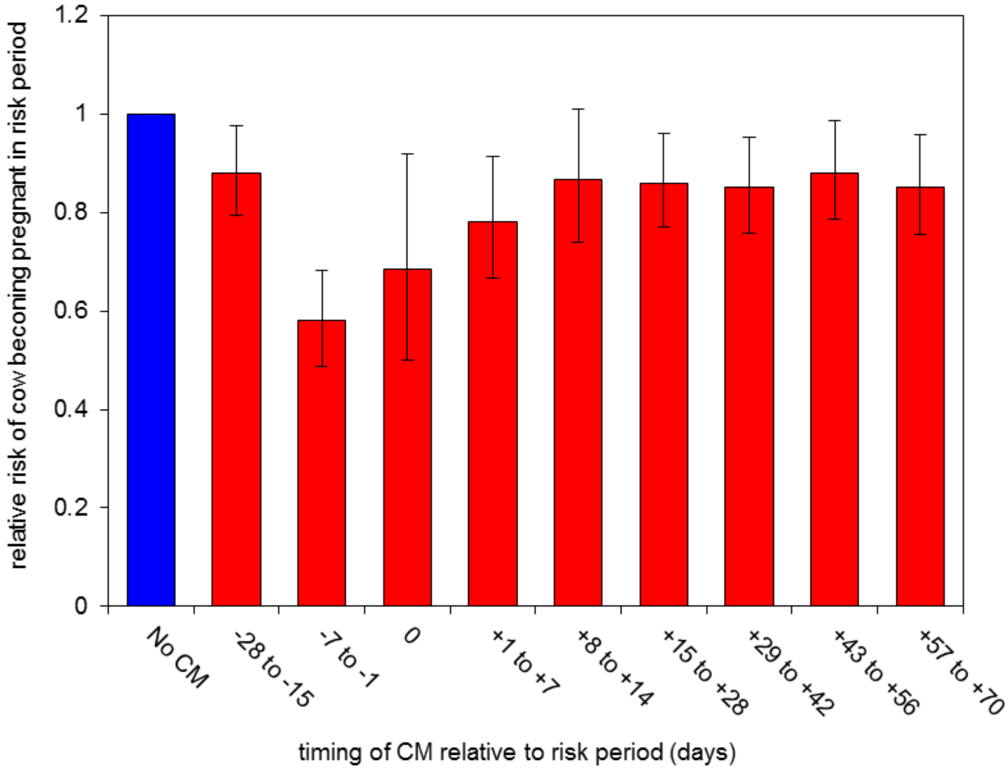

Figure 4-1 Association between the predicted relative risk of pregnancy at a given risk period and CM
Error bars represent the 95% credible interval for each predicted relative risk.

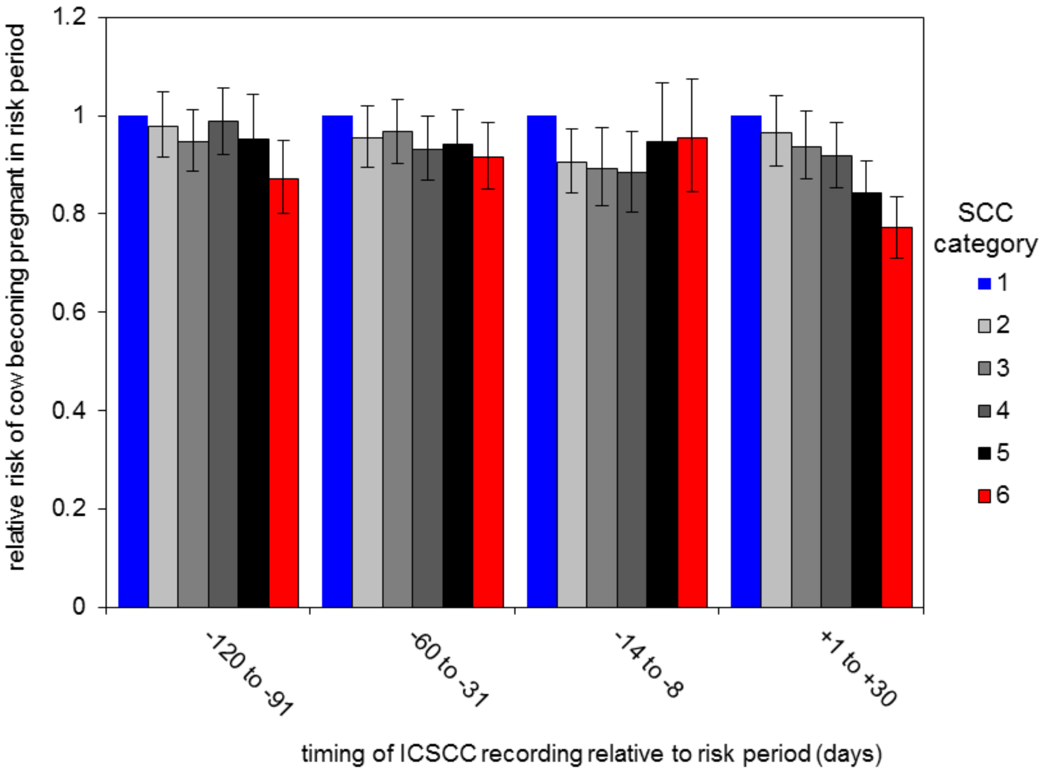

Figure 4-2 : Association between the predicted relative risk of pregnancy at a given risk period and ICSCC
SCC categories are defined in Table 4.2. Error bars represent the 95% credible interval for each predicted relative risk.

4.3.2 Probability of a given serve resulting in a pregnancy

A total of 85,482 serves occurred in the dataset: these formed the units of data for Model 4.2. A total of 29,237 pregnancies resulted from these serves: the overall pregnancy rate (i.e. the proportion of serves that led to a pregnancy) was therefore 34.2%. The median pregnancy rate at herd level was 35.8%, with an inter-quartile range of 31.2% - 40.5%.

Parameter estimates for Model 4.2 are shown in Table 4.5. The associations between probability of pregnancy and the explanatory variables not directly related to udder health were very similar to those described in Section 4.3.1. Broadly similar relationships with CM was also seen, although notably in this model CM at the time of serve had the largest negative

Table 4.5 Parameter estimates for logistic regression model with the outcome representing probability of a cow becoming pregnant as a result of a given serve (Model 4.2)

				95% HPD[16] interval	
Model terms	**n**	**Coefficient**	**Odds ratio**	**Lower**	**Upper**
Intercept	85482	-22.9		-23.1	-22.5
ln[17](days in milk)	85482	13.7		13.5	13.8
$(\ln(\text{days in milk}))^2$	85482	-2.76		-2.78	-2.74
$(\ln(\text{days in milk}))^3$	85482	0.188		0.186	0.19
Parity 1	22086		Reference		
Parity 2	20119		0.97	0.93	1.02
Parity 3	15224		0.96	0.91	1.01
Parity 4	10403		0.90	0.85	0.96
Parity >4	17650		0.73	0.70	0.77
Centred 305d milk yield (‘000 kg)	85482		0.91	0.90	0.92
Season 1: January - March	25342		Reference		
Season 2: April - June	19441		0.92	0.89	0.97
Season 3: July - September	15551		0.87	0.83	0.91
Season 4: October - December	25148		1.02	0.98	1.06
Year: 2003 or earlier	17060		Reference		
Year: 2004	12201		0.99	0.94	1.05
Year: 2005	14937		0.83	0.79	0.87
Year: 2006	17830		0.84	0.79	0.88
Year: 2007	20513		0.77	0.73	0.81
Year: 2008	2941		0.77	0.70	0.85
No CM[18] 15-28d before	83157		Reference		
CM 15-28d before	2181		0.90	0.81	0.99
NA[19] for CM 15-28d before	144		0.66	0.39	1.06
No CM 1-7d before	84655		Reference		
CM 1-7d before	827		0.75	0.63	0.89
No CM during risk period	85191		Reference		
CM during risk period	291		0.68	0.51	0.91
No CM 1-7d after	84401		Reference		
CM 1-7d after	1081		0.72	0.62	0.84
No CM 8-14d after	84367		Reference		
CM 8-14d after	1115		0.78	0.68	0.90
No CM 15-28d after	83380		Reference		
CM 15-28d after	2102		0.81	0.72	0.89
No CM 29-42d after	83488		Reference		

[16] HPD: highest posterior density
[17] ln: natural logarithm
[18] CM: clinical mastitis
[19] NA: not applicable – see Tables 1 and 2

CM 29-42d after	1994	0.82	0.74	0.92
No CM 43-56d after	83554	Reference		
CM 43-56d after	1928	0.87	0.78	0.97
No CM 57-70d after	83664	Reference		
CM 57-70d after	1818	0.83	0.74	0.92
ICSCC[20] category 1 1-30d after	12711	Reference		
ICSCC category 2 1-30d after	27896	0.97	0.93	1.02
ICSCC category 3 1-30d after	11528	0.91	0.86	0.96
ICSCC category 4 1-30d after	11254	0.90	0.85	0.96
ICSCC category 5 1-30d after	6378	0.82	0.76	0.88
ICSCC category 6 1-30d after	6825	0.74	0.69	0.80
No ICSCC 1-30d after	8890	0.89	0.83	0.95
ICSCC category 1 31-60d before	14275	Reference		
ICSCC category 2 31-60d before	27459	0.95	0.91	0.99
ICSCC category 3 31-60d before	9944	0.97	0.92	1.03
ICSCC category 4 31-60d before	9310	0.93	0.88	0.99
ICSCC category 5 31-60d before	5346	0.99	0.91	1.07
ICSCC category 6 31-60d before	6247	0.97	0.90	1.04
No ICSCC 31-60d before	12901	0.93	0.88	0.98

association with the probability of pregnancy (OR = 0.68, 95% HPD interval 0.49 – 0.89), followed by CM during the 1-7 days after the serve (OR = 0.72, 95% HPD interval 0.62 – 0.83), with CM during the 1-7 days before the serve the third largest association (OR = 0.75, 95% HPD interval 0.64 – 0.90). Fewer ICSCC predictor variables were retained in this model compared to model 4.1: ICSCC recordings at 31-60 days before and 1-30 days after the serve were the only timings that remained in the final model. ICSCC at 31-60 days before the risk period also showed a different relationship between ICSCC category and outcome: ICSCC categories 2 and 4 were associated with a lower probability of pregnancy compared to category 1 (the reference category), while categories 3, 5 and 6 were not different to category 1. These results are summarized as predicted relative risks in Figure 4-3 and Figure 4-4.

[20] ICSCC: individual cow somatic cell count

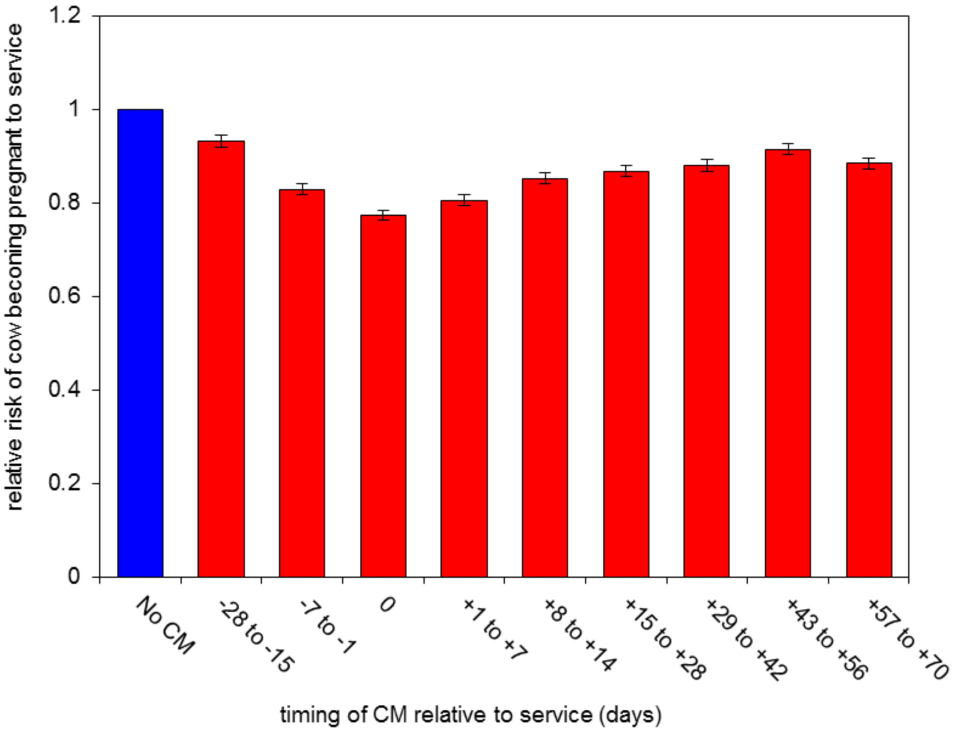

Figure 4-3 Association between the predicted relative risk of pregnancy at a given serve and CM
Error bars represent the 95% credible interval for each predicted relative risk.

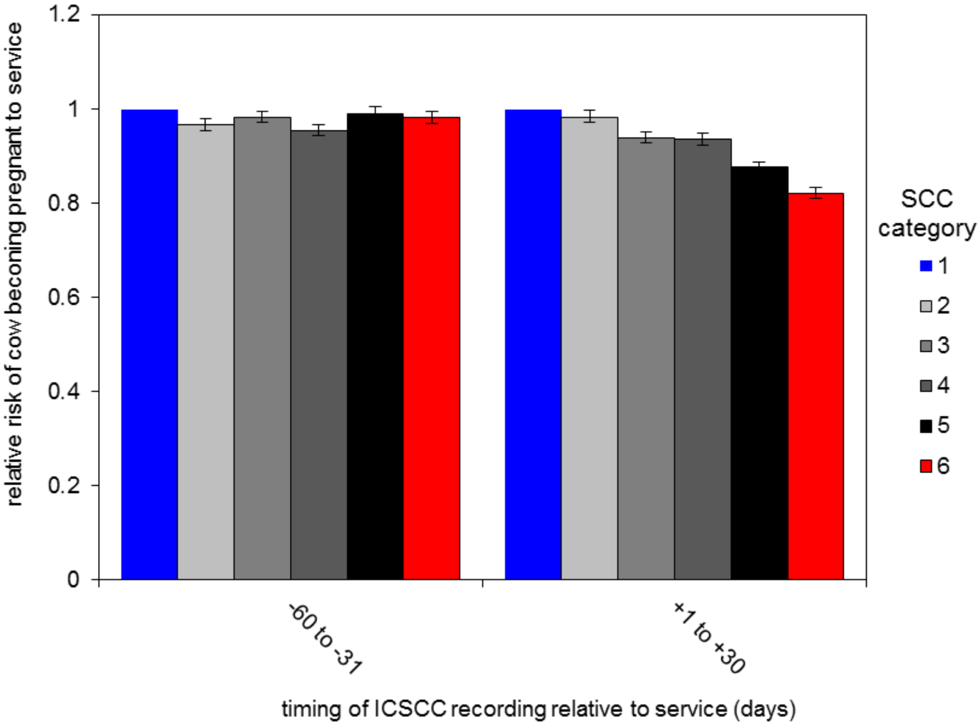

Figure 4-4 Association between the predicted relative risk of pregnancy at a given serve and ICSCC
SCC categories are defined in Table 4.2. Error bars represent the 95% credible interval for each predicted relative risk.

4.3.3 Model checking

The predicted probability of pregnancy for risk periods at different stages of lactation is shown in Figure 4-5, along with the observed proportion of risk periods at each stage where pregnancy occurred. The 95% credible interval for the predicted probability at each stage is also shown, and covers the observed value at all points. Full posterior predictions for the model predicting probability of a serve leading to a pregnancy are demonstrated in a similar manner in Figure 4-6, showing predicted and observed pregnancy rates through lactation. Overall probability of a pregnancy occurring during a risk period and overall pregnancy rate were also predicted for each herd; in every case the observed value for each herd fell within the 95% credible interval of the prediction for that herd.

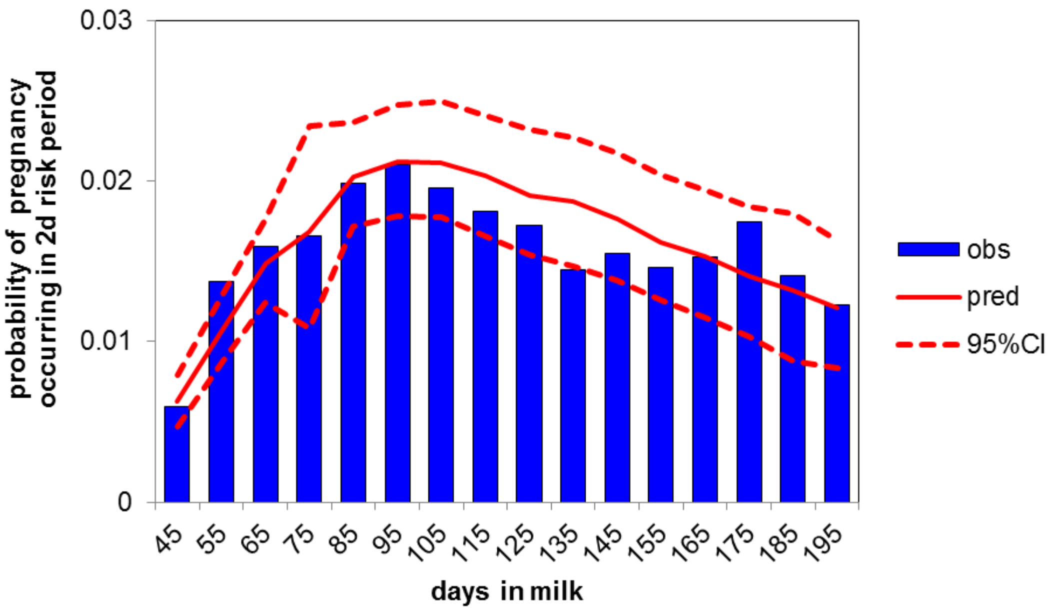

Figure 4-5 Predicted and observed probability of pregnancy by DIM
Dotted lines show the 95% credible interval (CI) for each prediction.

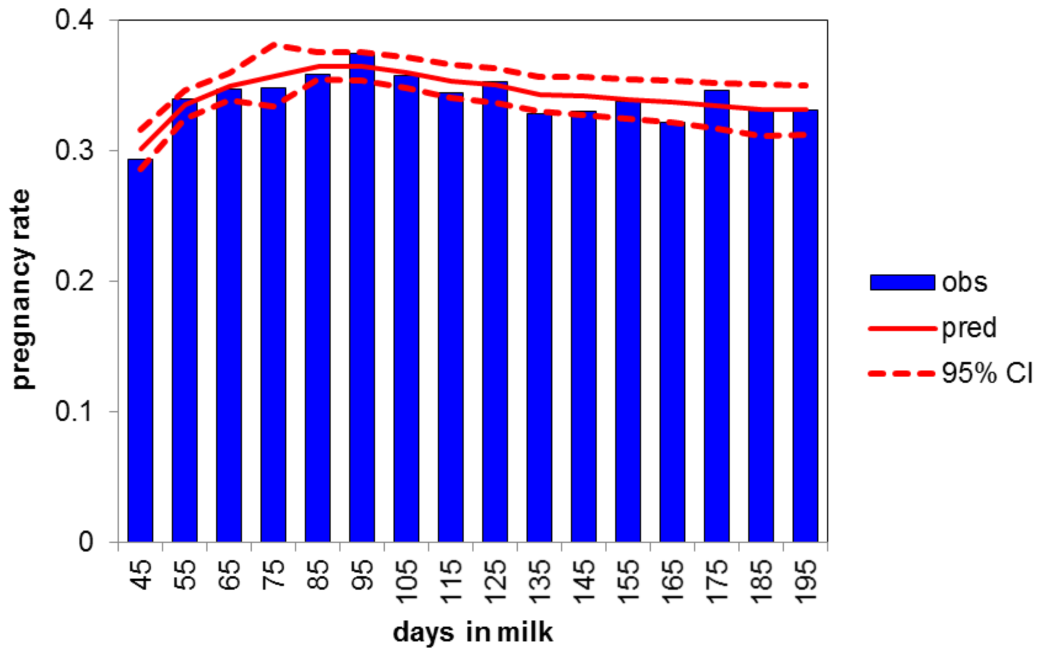

Figure 4-6 Predicted and observed probability of pregnancy to a given serve by DIM
Dotted lines show the 95% credible interval (CI) for each prediction.

In order to verify model fit further, predicted probabilities and observed proportions were also calculated for each parity and each season, lactations where yield was high (greater than 12,000 kg) or low (less than 5,500 kg) and for various categories of the CM and ICSCC variables. In each case the observed proportion of cases in which a pregnancy occurred lay within the coverage of the 95% credible interval for the model prediction. Therefore, posterior predictions suggested that model fit was good.

4.4 Discussion

This study supports previous work in this field suggesting that both CM and ICSCC can have substantial associations with fertility performance. In terms of overall fertility performance (model 4.1), CM appeared to be associated with poorer reproductive outcomes over a wide span of time. A case of CM was associated with a lower probability of the cow becoming pregnant from 10 weeks before the case to 4 weeks afterwards (with the exception of 8-14 days after the case, where no association was detected). A substantial sized and additive relationship with subclinical mastitis (as measured by ICSCC) was also observed, although the sizes of these associations were generally smaller compared to those with CM at similar timing. Generally, recordings in higher SCC categories tended to be associated with decreased reproductive performance.

Having found associations between CM and ICSCC variables and overall reproductive performance, model 4.2 was constructed in an attempt to improve understanding of how these associations were mediated. Reproductive performance is effectively dependent on two factors: the likelihood of an eligible cow being served (which depends on factors such as expression and detection of estrus, and post-parturient return to ovarian cyclicity) and the likelihood that a serve will lead to a pregnancy (the pregnancy rate). Therefore the construction of a second model in which the outcome represented pregnancy rate was useful.

The same CM-related variables were retained in both models; so the same timings of CM were associated with decreased pregnancy rate as with reduced overall fertility performance. However, there was a difference in the relative magnitudes of associations between outcome and the various CM variables between models. In model 4.2 (predicting probability of a serve leading to a pregnancy), CM at the time of serve was associated with the largest reduction in pregnancy rate, with a generally decreasing magnitude of effect size for timings of CM up to 70 days post-serve. In model 4.1, the largest association was seen where CM occurred at one to seven days before the risk period, with a smaller association with CM during the risk period, and a broadly decreasing effect size of CM variables further in the future. This suggests that a major component of the association between CM one to seven days earlier and the chance of a cow becoming pregnant is a reduced chance of her being served. This could clearly be as a result of a management decision not to serve a cow which had recently had a case of CM, but could also be related to suppression of ovulation or expression of estrus in cows which have recently had CM. If heats where the cow was not served were recorded accurately in the dataset, it would be possible to distinguish between these possibilities, but such events are rarely recorded consistently in UK herds.

No previous work has accurately evaluated the importance of timing on the associations between udder health and overall reproductive performance; recent existing work in this field has tended to focus on pregnancy rate as an outcome. The results of this study support previous work which suggested that CM in the period shortly before and shortly after first serve has a negative relationship with fertility (Barker et al., 1998; Gunay and Gunay, 2008; Santos et al., 2004; Schrick et al., 2001). The current study has shown a slightly longer duration of association between CM and reproductive performance compared to most existing research. Previous work evaluating overall fertility (as opposed to pregnancy rate) as an outcome has tended to categorise CM as occurring either before first serve, between first serve and pregnancy diagnosis or after a positive pregnancy diagnosis (Barker et al., 1998; Schrick et al., 2001). Previous studies have tended not to reveal a significant difference in the 'CM after pregnancy diagnosis' group compared to cows with no CM during the lactation. This could be because this crude categorisation of the timing of CM would group together CM cases at any stage of lactation after pregnancy diagnosis - thus including cases at the end of lactation (which would be unlikely to exert any influence of fertility) in the same category as cases as early as 28 days after the first serve. This could tend to mask the effect of CM in the period around and shortly after the stage at which pregnancy diagnosis is commonly undertaken. Alternatively, it is possible that grouping lactations in this way and comparing reproductive performance at lactation level between groups fails to account for confounding variables which may suppress the relationship between CM and fertility.

The associations revealed between subclinical mastitis and overall reproductive performance also broadly support earlier work, while providing novel information regarding the importance of the timing of subclinical mastitis. Schrick et al. (2001) demonstrated an extended calving to conception interval in cows which had subclinical mastitis (diagnosed by bacteriological sampling) before first serve, but found no significant effect of subclinical mastitis either between first serve and positive pregnancy diagnosis or later in lactation. The current study used ICSCC as a proxy for infection status, and as occurrences of CM are also included in the model, the apparent effects of ICSCC will represent the association between subclinical mastitis and reproductive outcome. In contrast to Schrick (2001), the current study found an association between reproductive performance and ICSCC status in the month following the risk period. In fact, ICSCC at this time had the largest association with reproductive performance compared to any timing of ICSCC before the risk period.

The results also provide support for existing work demonstrating an association between clinical or subclinical mastitis decreased pregnancy rate (Ahmadzadeh et al., 2009; Gunay and Gunay, 2008; Hertl et al., 2010; Loeffler et al., 1999; Moore et al., 1991; Perrin et al., 2007; Pinedo et al., 2009). This is the first study to specifically examine the effect of timing of subclinical mastitis relative to serve, and interestingly demonstrated that subclinical mastitis present at between one and 30 days post-serve was associated with the largest decrease in pregnancy rate. As might be expected, the magnitude of the relationship tended to increase with increasing ICSCC. The odds of a serve leading to a pregnancy were reduced by around 18% where an ICSCC of between 200,000 and 399,000 cells/ml was recorded at <31 days post-serve, while the odds were reduced by almost 26% when the ICSCC was >399,000 cells/ml. The magnitudes of these relationships are comparable with those with CM very close to the serve date - only CM at the time of serve was associated with a much greater decrease in pregnancy rate. This provides evidence that subclinical as well as clinical mastitis has a clinically significant relationship with reproductive outcome.

A number of potential mechanisms have been proposed to explain the effect of udder health on reproductive performance. These are comprehensively reviewed by Hansen (2004), but broadly encompass detrimental impact of inflammatory mediators on ovarian follicular function (Herath et al., 2007; Williams et al., 2008), intra-uterine embryonic survival (Soto et al., 2003) and the balance of luteolytic versus luteotrophic prostaglandins post-conception (Hockett et al., 2000; Neuvians et al., 2004). Thus, mechanisms exist to explain the effect of IMI both before and after the risk period or serve on the probability of establishment of pregnancy. In this study, IMI before serve was generally associated with a smaller effect on fertility, suggesting that effects on oocyte quality are perhaps less important compared to the other suggested mechanisms.

One of the drawbacks of the approach taken in this study was the sampling strategy employed to collect data. Where non-probabilistic sampling techniques are used, it is important to be cautious in generalising from the results of the research to the wider population. In this case, strong sampling bias towards herds with better kept records was present. It is plausible that such herds will, for example, tend to be larger, more carefully managed and more intensive in production compared to the population of UK dairy herds as a whole (and therefore will not constitute a representative sample). However, it is important to view this in a biological as well as a statistical context: this study aimed to evaluate the associations between udder health and reproductive performance at cow level. Although these relationships may be different in different types of herd, differences at cow level are likely to be smaller. This study also illustrates the potential for use of routinely recorded dairy herd data in research: whilst this is relatively commonplace for centrally held data (such as that retained by dairy herd improvement or milk recording organisations), such data is typically less rich and less robust than that managed on farm using computer software by a large number of farmers. Use of methods commonly used to analyse “big data” in industrial and operational research allows maximum value to be derived from such heterogeneous data.

In conclusion, this study demonstrates clear associations between both clinical and subclinical mastitis and depressed reproductive performance. IMIs appear to have a negative but variable relationship with cow fertility over a very long period of time, and subclinical mastitis can in some situations have a magnitude of relationship size similar to CM. This provides extra impetus to implement strategies to control mastitis at herd level as well as giving greater insight into factors affecting fertility. A key implication of better understanding of the association between udder health and reproduction is the opportunity for clearer insight into the impact of this at herd level. Application of the results of this study at herd level is not straightforward; and further work is required to allow interpretation of this chapter's results to guide decision making in the field. In the next chapter, a simulation-based approach is used to explore the potential for a herd's clinical and subclinical mastitis to impact on its level of reproductive performance.

Chapter 5 Udder health and fertility: Development of a simulation model to aid interpretation of results

5.1 Introduction

It is clear that the reproductive performance of a dairy herd is a complex, multi-factorial system: although detailed knowledge exists about many specific elements of this system, it can be difficult to evaluate how such knowledge fits together to determine the overall outcome. For instance, Chapter 4 describes an addition to the number of recent publications demonstrating associations between a cow's udder health and the probability of her conceiving to a specific insemination, or during a given period of lactation (Hertl et al., 2010; Lavon et al., 2011), but the likely importance of this at herd level is unclear. For decision makers, it remains difficult to evaluate the potential improvement in a herd's reproductive performance which might be expected if udder health on the farm is improved. This makes it difficult to use the existing knowledge about associations between udder health and fertility to prioritise interventions to improve a herd's reproduction. For example, it is difficult for a farmer or advisor to compare the likely benefit to fertility which would be associated with decreasing the herd's incidence rate of clinical mastitis compared to improving oestrus detection.

Probabilistic sensitivity analysis (PSA, see Section 1.4.2) is a tool which can be used to improve understanding of which inputs to a complex system (such as reproductive performance in a dairy herd) are most able to perturb its outcome. It has clear potential for application in this context, allowing the predicted impact of changing different herd fertility 'inputs' (such as level of mastitis, or submission rate) to be directly compared. In this chapter, PSA is used to evaluate the relative importance of different model inputs where minimal assumptions are made about the distribution of input parameters (i.e. under conditions of extreme uncertainty): that is, all values within a specified range are equally likely to be drawn at each iteration. The aim was to evaluate the potential scope for change in a herd's reproductive performance which could result from an improvement in intramammary infection status, relative to the other factors which affect fertility.

5.2 Materials and Methods

5.2.1 Discrete time survival model

This study was based on the statistical model outlined in Chapter 4 (Section 4.2.2.1), which describes reproductive performance in dairy cows by predicting the probability that a given cow will become pregnant in each consecutive 2-day risk period throughout lactation. Explanatory variables significantly associated with this outcome (Table 4.4) were used as the input parameters for the simulation model described here.

5.2.2 Distributions of simulation input variables

The distributions of the simulation input parameters are described in Table 5.1. Independent uniform distributions were selected for all herd-level inputs, covering ranges considered likely to encompass true values for the vast majority of UK herds. Although these distributions were not intended to represent the true "real world" distributions of the inputs, ranges were selected so that evaluation was carried out across the full range of plausible herd level scenarios. These scenarios were treated as equally likely by assigning a uniform probability across the range for each input parameter. The input parameters for each lactation, and for each risk period within the lactation, were mostly dependent on herd level inputs, so were drawn from appropriate distributions based on the relevant herd level parameter (Table 5.1).

Table 5.1 Input parameters used at each level of simulation along with distributions from which inputs were drawn

Input variable	Type	Input distribution
Herd level		
Submission rate (proportion of eligible cows inseminated every 21d)	Continuous	Uniform (0.1, 0.8)
Pregnancy rate (proportion of inseminations leading to a pregnancy)	Continuous	Uniform (0.1, 0.6)
Herd average 305d milk yield (kg)	Continuous	Uniform (3000, 12500)
Proportion of herd which are first lactation	Continuous	Uniform (0.1, 0.4)
Herd incidence rate of clinical mastitis (cases per cow-year of risk)	Continuous	Uniform (0.15, 1.7)
Proportion of clinical mastitis cases originating from dry period infection	Continuous	Uniform (0.1, 0.9)
Proportion of cows beginning lactation with ICSCC >200k	Continuous	Uniform (0.02, 0.4)
Proportion of cows moving from ICSCC <200k to >200k between milk recording test days	Continuous	Uniform (0.02, 0.25)
Proportion of cows moving from ICSCC >200k to <200k between milk recording test days	Continuous	Uniform (0.05, 0.45)
Cost per day of extension of calving index (£)	Continuous	Uniform (1.2, 4.2)
Cost per cow culled for failure to conceive (£)	Continuous	Uniform (550, 1750)
Lactation level		
Lactation number	Categorical (1, 2, 3, 4, >4)	Multinomial, based on proportion of herd in lactation 1
305d milk yield (kg)	Continuous	Beta, centred on herd average with standard deviation 1.5k
Risk period level		
Season (quarter of year)	Categorical (1, 2, 3, 4)	Multinomial for season at calving
Occurrence of CM 15-28d before risk period	Binary	Yes/No
Occurrence of CM 1-7d before risk period	Binary	Yes/No
Occurrence of CM during risk period	Binary	Yes/No
Occurrence of CM 1-7d after risk period	Binary	Yes/No
Occurrence of CM 8-14d after risk period	Binary	Yes/No
Occurrence of CM 15-28d after risk period	Binary	Yes/No
Occurrence of CM 29-42d after risk period	Binary	Yes/No
Occurrence of CM 43-56d after risk period	Binary	Yes/No
Occurrence of CM 57-70d after risk period	Binary	Yes/No
ICSCC 1-30d after risk period	Binary	(<=200k, >200k)

A further study explored the possibility that choice of input distributions and correlations between the input parameters would affect the outcome of the simulation. Distributions of the input parameters for each of the 80 herds in the original dataset (Section 4.2.1) were evaluated. Assessment of the univariate distribution of each parameter in turn showed that the ranges of the parameters across herds were very similar to those chosen for the uniform input distributions shown in Table 5.1, and that many of the inputs did not appear normally distributed. As it was plausible that all inputs were jointly correlated in a complex fashion (and clear that few approximated a normal distribution), attempting to fit a parametric multivariate distribution to the data was considered inappropriate. Instead, a non-parametric approach was taken, whereby the simulation exercise was repeated using the observed joint distribution of the parameters across the herds was used as simulation inputs, so that at each iteration of the simulation the set of observed input parameters for one of the 80 herds was used as the input for the simulation model. This process was also repeated using the joint distributions of input parameters observed for each herd-year (i.e. for each herd in each year) in the dataset (n=435).

5.2.3 Simulation model

The structure of the simulation model is represented diagrammatically in Figure 5-1. Simulation was carried out in Microsoft Excel 2010 (Microsoft Corp.), using Visual Basic for Applications (Microsoft Corp.) for process control. A total of 50,000 herds were simulated, with each one consisting of 200 lactations. The first step in simulating a herd was to draw the herd level input parameters from their distributions (Table 5.1) before simulating the first lactation in the herd (again, beginning by drawing the lactation level inputs from relevant distributions). Next, a simulated udder health history was generated for the lactation (Figure 5-2).

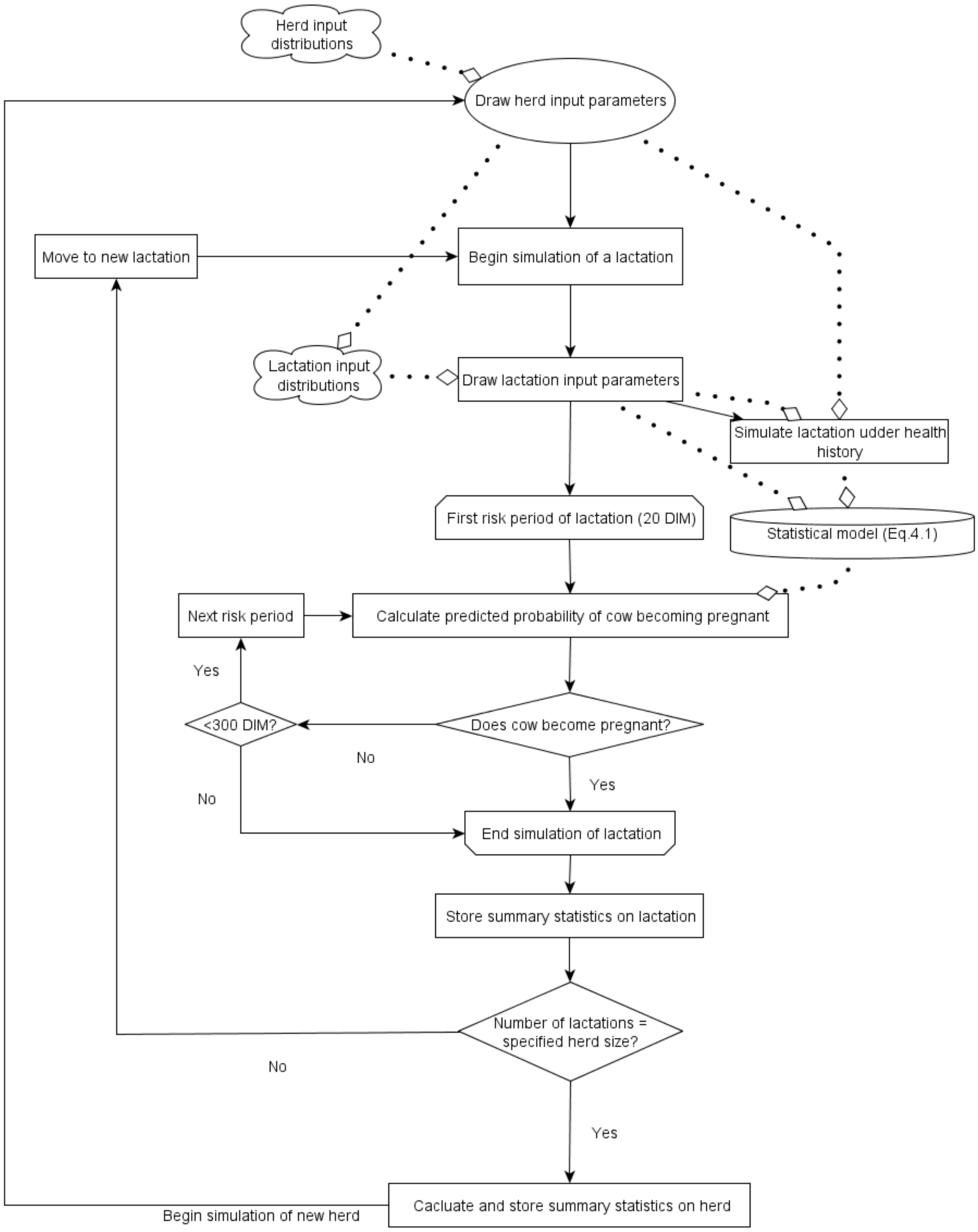

Figure 5-1 Overview of the simulation model process
Solid black lines indicate process flow, and dotted lines indicate that information from the source of the line is used in the step of the process to which the line leads (denoted by a diamond).

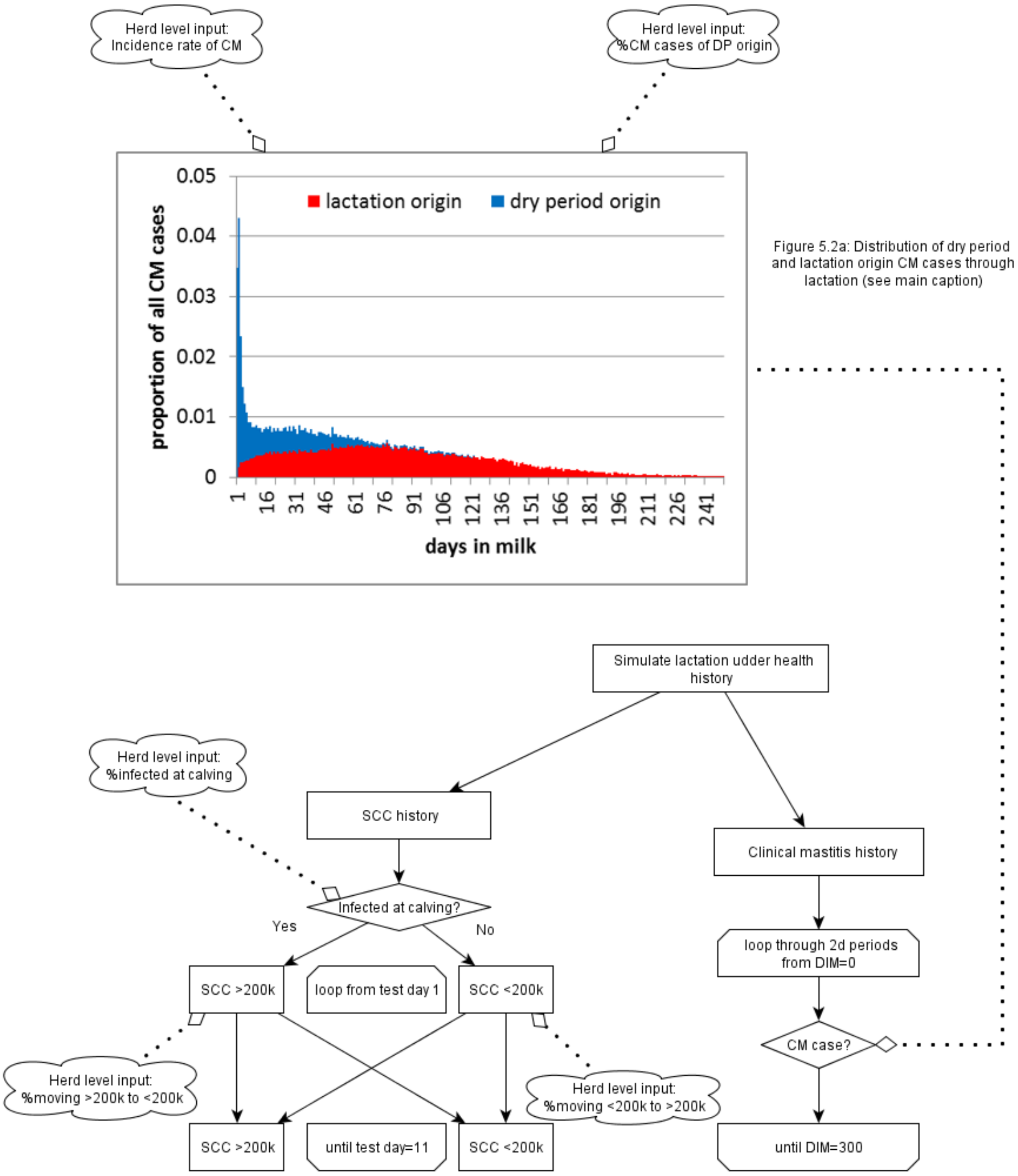

Figure 5-2 Process for simulation of udder health history through a lactation
Solid black lines indicate process flow, and dotted lines indicate that information from the source of the line is used in the step of the process to which the line leads (denoted by a diamond). Figure 5.2a shows proportion of clinical mastitis cases from full dataset by days in milk, split into likely dry period versus lactation origin using data from Green et al. (2002).

A CM history for the lactation was simulated based on two herd-level input parameters: the incidence rate of CM, and the proportion of CM cases resulting from intramammary infection during the dry period. In order to use these parameters to predict occurrence of CM as a binary event for each two-day risk period, a value for the number of DIM at each case of CM was extracted from the 80-herd dataset described in Section 4.2.1: this determined the distribution of cases of CM over the course of lactation. A total of 67,994 cases of CM were included in this analysis. Data from Green et al. (2002) were then used to attribute the proportion of cases at each two-day period through lactation as either dry period or lactation origin, with a very high proportion of cases in early lactation being attributed to the dry period (Figure 5-2a), and a very high proportion of cases in late lactation attributed as lactation origin. These results were then used to calculate the proportion of all dry period origin cases and of all lactation origin cases which occurred at each two-day risk period. For each herd simulated, the input parameters were used to determine the separate incidence rates for dry period and lactation origin CM (by multiplying the overall incidence rate by the proportion of cases of dry period origin). This allowed prediction of the probability of the occurrence of either dry period origin or lactation origin CM at each two-day risk period during the lactation: the simulation model then assigned events by drawing from a binomial distribution based on the calculated probability of CM at each risk period.

In order to simulate ICSCC history, it was assumed that the cow would have a first milk test day of the lactation at a random stage within the first 30 DIM (so that DIM at first test day was drawn from a uniform distribution between 0 and 30), and would have test days at regular 30 day intervals after this. ICSCC was treated as a binary variable, such that the cow could occupy one of two states; infected (ICSCC>200k) or uninfected (ICSCC<200k). The herd-level input parameters were then used to determine the cow's status at the first recording of lactation (a draw from a binomial distribution with probability equal to the overall proportion of cows with a first ICSCC of lactation >200k), and the likelihood that her status will change at each subsequent test day.

The fixed effects part of the statistical model described in Section 4.2.2.1 was then used to calculate the probability of pregnancy occurring at each 2-day risk period of the lactation (based on the input parameters for that herd, lactation and risk period). This probability was then adjusted to account for additional marginal (i.e. unexplained by model input parameters) variation in the herd's submission rate (proportion of eligible cows served every 21 days) and pregnancy rate (proportion of serves leading to a pregnancy).

A binary outcome for pregnancy in each 2-day risk period was then drawn from a binomial distribution based on this adjusted probability, with repeated risk periods simulated until either pregnancy or 300 days in milk (DIM). The reproductive outcome of the lactation was recorded using two variables: a binary outcome representing whether the cow reached 300 DIM without becoming pregnant, and if this was not the case then also the number of DIM at which pregnancy occurred. This information was stored along with the input parameters for the lactation, and simulation of the next lactation begun. This process was repeated until the 200 lactations making up the herd were complete, at which point the mean number of DIM to pregnancy (i.e. calving to conception interval) and the proportion of lactations where the cow reached 300 DIM without becoming pregnant were calculated over the herd and stored, along with the herd input parameters. These two measures were combined by comparing each to a selected baseline value (65 days for calving to conception interval and 0% for 300 day failure to conceive rate), applying a cost per unit deviation from the target (with unit cost for each represented as herd-level input parameters) and summing the total cost per cow to create a modified 'FERTEX' (mFX) score for each herd (Esslemont and Kossaibati, 2002). The baseline values for calving to conception and failure to conceive at 300 DIM were set at very low levels to avoid herds which performed better than the baseline level (and therefore had negative mFX scores). Although this mFX score represented an appropriate single outcome measure for this study, the absolute value of mFX score for each simulated herd would therefore not reflect true recoverable loss due to infertility (although changes in mFX score would be realistic). Simulation of the next herd was then begun.

5.2.4 Analysis of results

Summary data for each of the 50,000 simulated herds were exported to R 2.14.2 (R Core Development Team, 2010) for analysis. The associations between each herd-level input parameter and the outcome (mFX score) were initially explored using high-density scatterplots. As the mFX scores were strongly positively skewed (as expected with a cost-based outcome), Spearman rank correlation coefficients were calculated for the relationships between mFX score and each input. Multiple regression, with the natural logarithm of herd mFX score as the outcome variable, was used to partition variance in mFX score between the herd input parameters, and to predict the effect of changes in each individual parameter on herd mFX score. In order to represent these results graphically as a tornado plot, the regression model was used to predict change in mFX score where each input parameter in turn was increased from the median value of its input distribution by a value representing 25% of the range of the distribution while the other inputs were held at their median values. This allowed evaluation of the change in outcome (mFX score) when each input parameter is altered by a comparable amount, allowing visualisation of relative effect size.

5.3 Results

5.3.1 Univariate analysis

High density scatterplots showing the associations between each herd-level input parameter and the herd mFX score (with higher mFX scores indicating poorer overall performance), along with the Spearman rank correlation coefficient (r_s) for each relationship are shown in Figure 5-3. The association between herd submission rate and mFX score is the most striking, with a clear "funnelling" of points in the bottom right hand corner of the graph, indicating that herds with high submission rates (especially over 50%) have a much narrower range of mFX scores, with a much stronger concentration around the lower mFX scores (i.e. better reproductive performance). The relationship between pregnancy rate and mFX score shows a similar but less marked funnelling pattern, with a high concentration of herds in the bottom right corner of the graph (where a very good pregnancy rate drives strong overall performance), but a larger range of mFX scores amongst herds with high pregnancy rates compared to those with high submission rates. Spearman rank correlation coefficients for both relationships were around -0.6, suggesting a moderate to strong negative correlation.

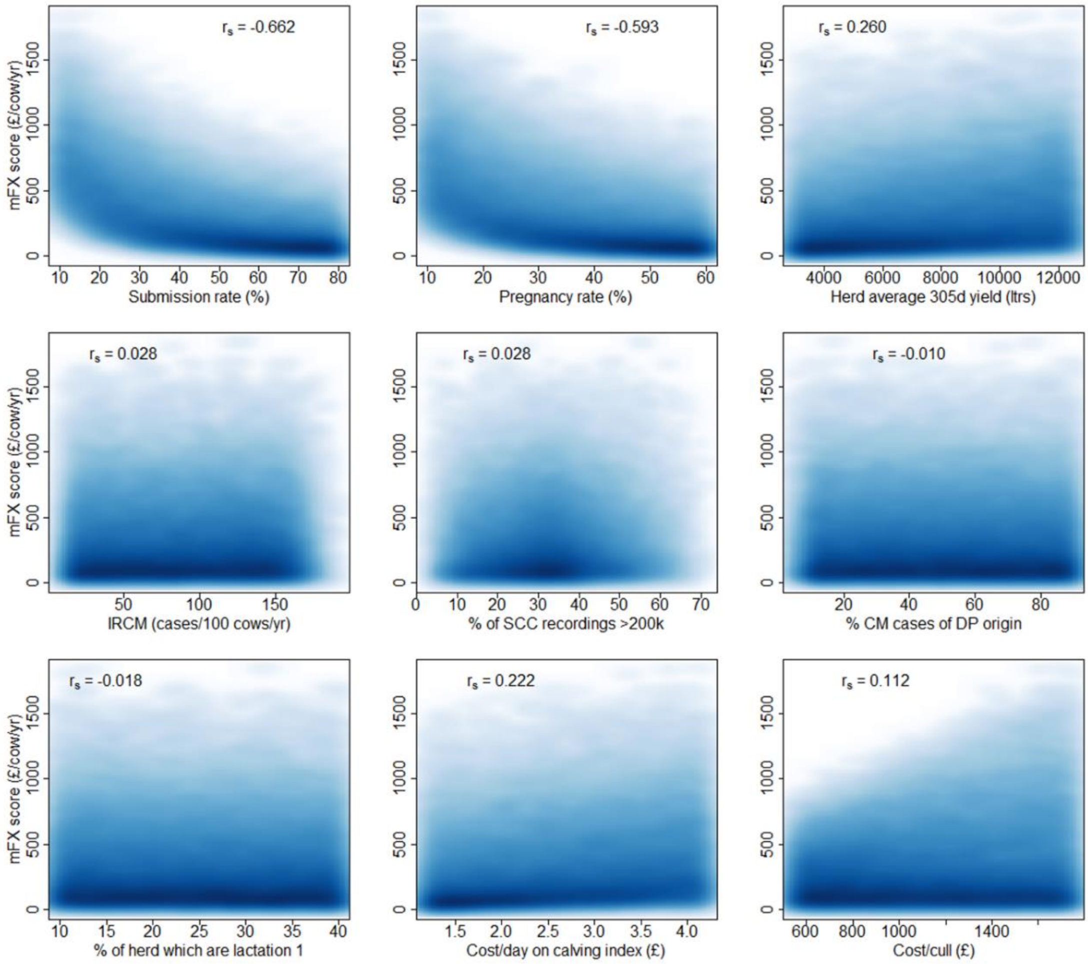

Figure 5-3 Associations between overall fertility outcome and herd-level input variables
Darker colours indicate higher densities of points. r_s = Spearman rank correlation coefficient. mFX: modified FERTEX score, representing overall herd fertility outcome; IRCM: Incidence rate of clinical mastitis; SCC: Somatic cell count; CM: clinical mastitis; DP: dry period.

Herd average 305 day milk yield showed a different relationship with mFX score: although the area of highest point density shows a general trend for mFX scores to increase with milk yield, there is a wide distribution of mFX scores across the full range of milk yields. The shallow gradient of the area of highest density also suggests that the magnitude of influence of milk yield on overall fertility performance is likely to be relatively small. The Spearman rank correlation coefficient of 0.26 suggests that the correlation between the two variables is moderately weak.

The high-density scatterplots showing relationship between the udder-health-related input parameters and mFX score show no obvious correlations, with point clouds assuming a square appearance and no evident trend in the line of highest point density.

5.3.2 Multiple regression analysis

The results of variance partition by regression analysis are shown in Table 5.2. Each line of the table shows the proportion of variation in mFX score explained by each input parameter, after accounting for the variation explained by the other input parameters. It is clear that submission rate (42.9% of total variance) and pregnancy rate (35.2% of total variance) collectively account for the vast majority of variance in the outcome.

Table 5.2 Partition of variance in modified herd FERTEX score (mFX) between input parameters

Input parameter	% variance explained
Submission rate	42.9%
Pregnancy rate	35.2%
305d yield	7.4%
IRCM	0.1%
% SCC recordings >200k	0.1%
% CM cases which are DP origin	<0.1%
% of herd in lactation 1	<0.1%
Cost per day on calving index	5.5%
Cost per cull	1.3%
Total	92.5%

The predicted effects of changes in inputs are represented graphically as a tornado plot in Figure 5-4. Changing submission or pregnancy rate is expected to have a large impact on overall reproductive performance, with a move from median (45%) to upper quartile (62.5%) submission rate predicted to generate a saving of more than £85 per cow per year: this can be seen as relative “room for investment” in this area, and provides an important example of the potential of this type of modelling to inform decision making at farm level. Cost per additional day on calving index and average 305-day adjusted milk yield were associated with smaller changes in mFX score, and cost per cull predicted to lead to a slightly smaller change again. Udder-health-related inputs are predicted to have little impact on overall reproductive performance.

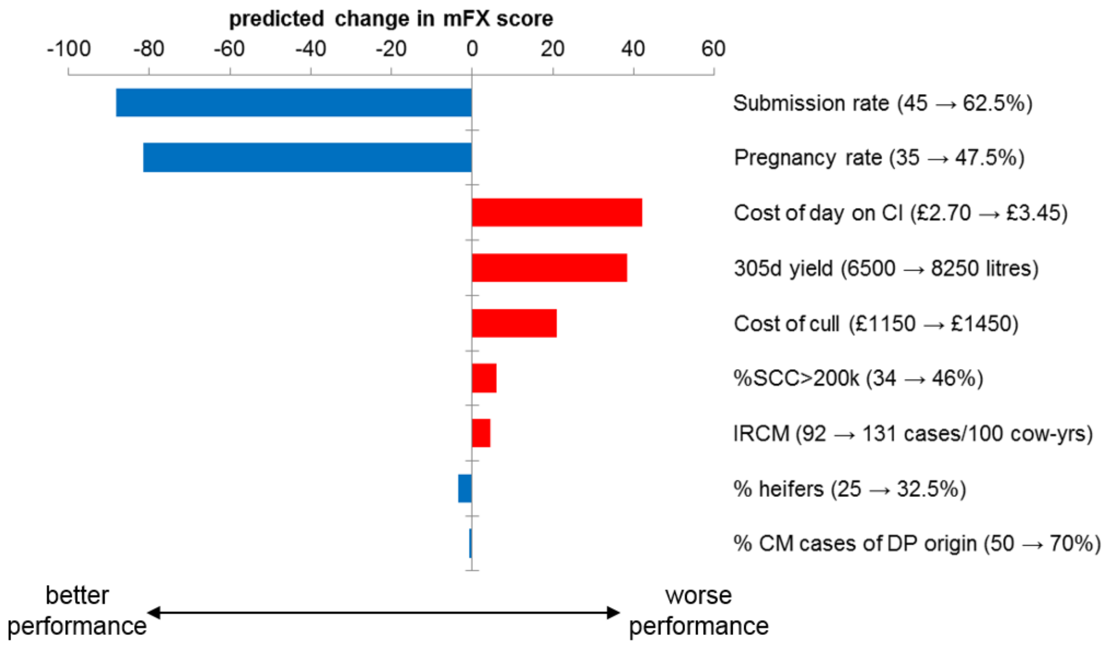

Figure 5-4 Predicted effect of an equivalent increase in each input parameter on overall fertility
Tornado plot showing the predicted effect of increasing each input parameter in turn by a value representing 25% of the range of its input distribution from the median value, while the other input parameters are held at their population medians. The input parameters are listed on the right hand side of the graph, and the change in each input (from median to upper quartile) is given in parentheses. For example, the top bar shows that the predicted effect of moving from a submission rate of 45% (the median of the input distribution for this parameter) to 62.5% (the upper quartile of the input distribution) would be a decrease of just under £90/cow/year in herd mFX score. NB: For the proportion of recordings where SCC>200k parameter (which was the only input not drawn directly from a uniform distribution), the change in the parameter (+12.4%) represented 25% of the 95% coverage interval of the distribution of this parameter.

The low degree of association between udder health parameters and reproductive performance is demonstrated further in Figure 5-5: Figure 5-5a and Figure 5-5b show the distributions (as kernel density plots) of mFX scores for herds with extremely high or low values for IRCM or proportion of ICSCC recordings >200k respectively. It is clear that the two lines on each figure follow a very similar shape, demonstrating that herds at either extreme of the input distribution for udder health parameters have very similar ranges of reproductive performance. By contrast, Figure 5-5c shows the distributions of mFX scores for herds with extremely high and extremely low submission rates: here it is clear that herds with high submission rates have a much tighter distribution of mFX scores centred on a much lower mFX score compared to low submission rate herds.

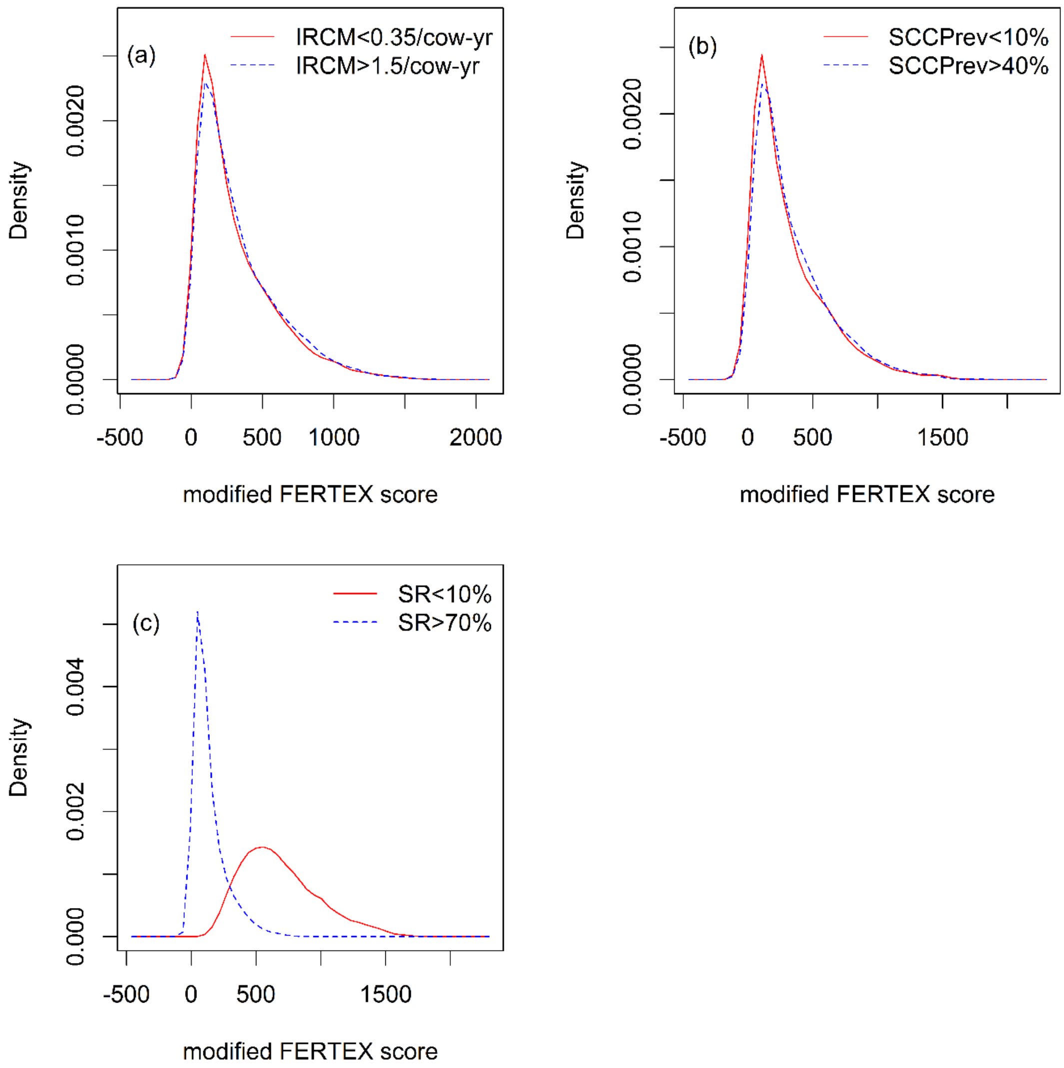

Figure 5-5 Kernel density plots for simulated herds with extreme input parameter values
The kernel density plots show distribution of modified FERTEX score for herds with extreme values for: Figure 5-5a (top left) incidence rate of clinical mastitis (IRCM) in cases/cow-year: IRCM<0.35 cases/cow-year, solid line; IRCM>1.5 cases/cow-year, dotted line. Figure 5-5b (top right) proportion of somatic cell count recordings >200k (SCCPrev): proportion <10%, solid line; proportion >40% dotted line. Figure 5-5c (bottom left) submission rate (SR): submission rate <10%, solid line; submission rate >70%, dotted line

5.3.3 Alternative input distributions

Repeating the simulation and analysis using the observed joint input distributions from the original dataset (instead of those described in Table 5.1) affected the results of the univariate analyses, but multivariate regression analyses produced similar results to those generated using independent uniform input distributions. Although the regression coefficients for both udder health related input parameters increased slightly (and the predicted effect of IRCM became the larger of the two), the predicted effect of changes in these parameters remained much smaller than the predicted effects of changes to the key drivers of mFX score. The tornado plot (equivalent to Figure 5-4) generated using the observed joint input distributions of herd-years from the dataset was almost identical plot to that derived using independent uniform input distributions. It therefore appears that the choice between these alternative input distributions would not have a substantial impact on the biological interpretation of the results of this study, and the results reported are those derived from the independent uniform input distributions.

5.4 Discussion

Recent work has demonstrated that clinical mastitis around the time of insemination is associated with a reduction in the probability of pregnancy to the insemination of between 20 and 80% (Hertl et al., 2010), and that elevated ICSCC can be associated with reductions in the order of 20% (Lavon et al., 2011). However, although these effect sizes intuitively appear quite large and are broadly consistent with earlier work in the area (Loeffler et al., 1999; Pinedo et al., 2009; Schrick et al., 2001), interpreting their likely impact at herd level has been difficult owing to the large number of other factors that influence the relationship between mastitis and reproduction (for example, the frequency and distribution of CM cases and elevations of ICSCC throughout lactation). Specifically, these results did not give farmers or veterinary surgeons any indication of the potential to improve a herd's reproduction by maximising udder health.

Here, development of a simulation model and its use within a PSA framework have revealed that improvements in udder health at herd level are highly unlikely to lead to useful improvement in herd fertility performance under the vast majority of plausible scenarios.

Therefore, given the variability in udder health performance typically observed in UK dairy herds (represented by the ranges chosen for the distributions of the input parameters), it is highly unlikely that improving a herd's udder health (either in terms of CM or SCC) would lead to a detectable improvement in the reproductive performance of the herd. The study also confirmed that the marginal effects of submission rate and pregnancy rate (after accounting for effects of other model inputs, such as milk yield) are key drivers of performance, and gave an indication of the potential room for investment in these areas.

The use of a regression model to analyse the simulation results provides a very basic analogue to the construction of a metamodel for the simulation (Kleijnen and Sargent, 2000; Vonk Noordegraaf et al., 2003). This concept is commonly used in the field of commercial operational research, where a simpler model is developed and validated to represent the results of a more complex system under simulation. Here, the regression model was used simply as a way to partition variance and explore the relative importance of inputs to the simulation model, but the regression model can also be used to predict the outputs for a given set of input values (in a similar way to that used to construct the tornado plot shown in Figure 5-4).

Use of stochastic modelling (and associated techniques such as PSA) is becoming increasingly commonplace in a variety of areas. Essentially, such models have two main applications. Firstly, they can be used in a research setting (e.g. for PSA) to evaluate likely importance of different model inputs across a variety of possible scenarios. Results of such research can then be used to inform clinical guidance, as well as prioritising promotion of existing knowledge and allocation of resources towards future research. Clinical decision making in human medicine presents an excellent example here, with PSA widely adopted for cost-effectiveness studies informing blanket clinical guidelines (Andronis et al., 2009). Secondly, stochastic modelling can be used on a case-by-case basis, whereby simulation using a model can be used to evaluate likely outcomes for a specific real-life scenario under alternative potential strategies or interventions. Risk management in business (especially the financial sector) presents perhaps the best example of this process: for example, use of such tools is extremely common for evaluation of alternative investment opportunities. It is easy to see excellent uses for both of these approaches in clinical veterinary medicine (especially in farm animal practice, where decisions regarding potential interventions at herd level are common). Recently, there has been more interest in both applications of stochastic modelling to herd-level (Giordano et al., 2012; Hockey and Morton, 2010) and cow-level (Cabrera, 2012) management decisions in dairy farms, but it is often considered that such methods are too complex and cumbersome to be widely employed by farmers or their advisors (Walster, 2012). However, the simulation model in this paper was deliberately developed in a software environment that would allow for development of customised decision support tools, based on the approach described, which could be widely distributed and used within the industry.

Whilst PSA is a robust and well established technique, a common criticism is that unjustified assumptions are made about parameter input distributions. In this case PSA was being used to evaluate dairy herd reproduction as a system and assess which input parameters are most able to perturb the system: effectively this represented simulating hypothetical herds across as wide a range of plausible situations as possible. This is the reason uniform distributions were used for the input parameters: although these clearly do not reflect the distributions of the same parameters across real life herds, they allow the relative importance of each parameter to be evaluated across a wide variety of possible scenarios. The udder health inputs are a good example of this: here CM and SCC history through each lactation were simulated independently. In reality, these are both driven by an underlying latent variable (the true intramammary infection status through lactation), which is difficult to evaluate and therefore to simulate realistically. However, as the overall effects of CM and SCC appear to be very small, this is not likely to have made a substantive difference to the results of this study. In this case, it also appeared that using independent input distributions did not lead to a different conclusion than that reached using the observed joint distributions from the original data.

This study has found that the association between herd intramammary infection status (as measured by CM and ICSCC) and herd-level reproductive performance is likely to be weak under the vast majority of plausible scenarios, despite the relatively large association sizes at lactation and serve level revealed by previous work and used as model inputs. In this example, development of a stochastic model and PSA were found to be useful tools to aid understanding of dairy herd reproduction as a system. Importantly, this work has also provided a model structure that can be extended and built upon in future research. In Chapter 6, this approach is applied to the relationship between clinical lameness events and reproductive outcomes.

Chapter 6 Associations between clinical lameness events and reproductive performance

6.1 Introduction

Lameness is one of the most common endemic diseases in the modern dairy herd, with reported prevalence in the UK at over 35% (Barker et al., 2010), and has previously been associated with depressed reproductive performance in affected cows compared to unaffected controls (Alawneh et al., 2011; Garbarino et al., 2004; Machado et al., 2010; Melendez et al., 2003). However, a very high proportion of previous studies have been carried out using either a single herd or a small number of herds, and those deriving data from wider populations have often failed to detect an association (Loeffler et al., 1999; Sogstad et al., 2006), as did the most recent study in UK dairy cows (Peake et al., 2011). Alongside this, a very wide variety of other factors are known to affect cow fertility. Therefore the clinician wishing to improve a herd's reproductive performance needs to interpret this research evidence in the context of the other influences on fertility when deciding how much weight should be given to control of lameness to improve reproduction.

This study focuses on the relationship between a time-to-event outcome (in this case, the time between parturition and subsequent conception in a dairy cow) and a disease event (in this case lameness). Techniques for analysis of such data have evolved over the years, and this specific field has seen publications evaluating this relationship in a univariate way (Peake et al., 2011) using Kaplan-Meier survival analysis, and in a multivariate framework, using various modifications of the Cox proportional hazards model (Alawneh et al., 2011; Hernandez et al., 2001; Machado et al., 2010). However, accounting appropriately for time-dependent variables (for example, accounting for the possibility that a case of lameness may affect probability of conception within a specific frame of time around the case) using such approaches can be challenging, and model assumptions can be difficult to satisfy and are not always tested (Bellera et al., 2010).

Another approach is discrete time survival analysis (Singer and Willett, 1993; Steele, 2003), where the dataset is amplified into smaller units of time for each individual animal and logistic regression is used to predict the probability of the outcome of interest at each time-point. This method is substantially more flexible, and more easily incorporates statistical advances such as multilevel regression using random effects to account for hierarchical clustering within data (Rasbash et al., 2009; Steele, 2003) (for example, of cows within herds), and Markov chain Monte Carlo sampling for parameter estimation within a Bayesian framework (Browne, 2009). However, results from this type of analysis can be difficult to interpret, especially at the population level. For example, such analysis may yield an estimated odds ratio for the association between a lameness event and the probability of conception occurring during a given period of time, but there is no intuitive way to interpret the likely importance of this at the population level. In this context, on-farm interpretation is very important, because decision makers (e.g. a dairy herd's manager or veterinary clinician) need to be able to estimate the potential improvement in a herd's reproductive performance that could result from a reduction in lameness in order to conduct a cost benefit analysis for intervention.

In this study, the association between clinical lameness events and reproductive performance was evaluated using routinely collected management data from a group of dairy herds. The aim of the study was to explore the usefulness of simulation-based techniques as an aid to interpret the clinical significance of a discrete time survival model evaluating association between disease events and reproductive performance at herd level. This chapter outlines the application of the methods used to explore relationship between udder health and fertility in Chapter 4 and Chapter 5 to clinical lameness, both at individual cow and at herd level.

6.2 Materials and Methods

6.2.1 Data Collection and Restructuring

This study used data from the subset of herds described in Chapter 2 which demonstrated consistent recording of clinical lameness events (i.e. treatment of lame cows). This was assessed by calculating the overall incidence rate of lameness in each herd-year and by evaluating changes in the number of cases recorded in each calendar month (using a similar approach to that described for clinical mastitis in Section 4.2.1).Detail regarding each event (for example, which limb was affected and the diagnosis made) was not evaluated: all clinical lameness events were treated as equal. Where two lameness events were recorded for the same cow within 7 days, the second was removed (since both treatment records would have been likely to reflect the same disease event). Table 6.1 shows descriptive information for the herds used.

Table 6.1 Summary statistics of basic herd information for 39 herds with consistent clinical lameness records

			Percentiles			
	Mean	**Minimum**	**25%**	**50%**	**75%**	**Maximum**
Herd size	243	88	153	202	292	669
Cull rate (%/year)	22	13	20	22	25	31
305 day adjusted milk yield (litres)	8329	4776	7366	8266	9566	11008
Incidence rate of clinical lameness (cases/cow-year)	0.40	0.10	0.22	0.30	0.41	1.88

Data were restructured into a format where each unit (line) of data was a two-day period during each lactation between 20 and 220 days after parturition (days in milk, DIM) where the cow was “at risk” of becoming pregnant (lactations were censored after culling, death, sale or conception occurred). For each of these two-day risk periods, a binary variable was used to represent whether the cow became pregnant during the risk period. Clinical lameness records were used to determine whether a case of lameness was recorded at a variety of different time-frames relative to each risk period (see Table 6.2). Additional variables at both lactation level (e.g. parity of cow, lactation 305-day adjusted milk yield) and risk period level (e.g. DIM at beginning of risk period, month and year of risk period) were calculated for each risk period (Table 6.2). Where necessary, categorical variables were recoded to avoid categories containing small numbers of risk periods/lactations (e.g. animals of parity 5 or above were grouped as a single category). This generated a dataset consisting of 1,247,677 risk periods from 21,913 lactations in 12,515 cows from 39 herds. Initial data collation and restructuring was carried out using Microsoft Access 2010 (Microsoft Corp.), with further restructuring and variable calculation carried out using R 2.14.0 (R Core Development Team, 2010).

Table 6.2 Potential explanatory variables calculated for each risk period

Variable	Level	Variable type
Parity (lactation number)	Lactation	Categorical (>4 recoded as single group)
305-day lactation milk yield	Lactation	Continuous
Year in which lactation began	Lactation	Categorical (<2003 recoded as single group)
DIM at start of risk period	Risk period	Continuous
Season of risk period	Risk period	Categorical (Jan-Mar, Apr-Jun, Jul-Sep, Oct-Dec)
Lame 71-100d before risk period	Risk period	Binary (lameness case recorded or not)
Lame 43-70d before risk period	Risk period	Binary (lameness case recorded or not)
Lame 15-42d before risk period	Risk period	Binary (lameness case recorded or not)
Lame within 14d of risk period	Risk period	Binary (lameness case recorded or not)
Lame 15-42d after risk period	Risk period	Binary (lameness case recorded or not)
Lame 43-70d after risk period	Risk period	Binary (lameness case recorded or not)
Lame 71-100d after risk period	Risk period	Binary (lameness case recorded or not)

6.2.2 Discrete-time survival analysis

A multilevel discrete-time survival model (Yang and Goldstein, 2003) was constructed to evaluate the association between the probability of a cow becoming pregnant during a two-day risk period (the outcome) and the potential explanatory variables described in Table 6.2. A three-level hierarchical structure (with risk periods nested within cows nested within herds) was used to account for correlations between risk periods from the same cow and cows from the same herd.

The model took the standard form:

$$\mathrm{Preg}_{tij} \sim \mathrm{Bernoulli}(\mathrm{mean}=\mu_{tij})$$

$$\ln\left(\frac{\mu_{tij}}{1-\mu_{tij}}\right)=\alpha+\beta_1 \mathrm{lnDIM}_{tij}+\beta_2\left(\mathrm{lnDIM}_{tij}\right)^2+\boldsymbol{\beta}_3\mathbf{X}_{tij}+\boldsymbol{\beta}_4\mathbf{X}_{ij}+u_{ij}+v_j \quad (6.1)$$

$$v_j \sim \mathrm{N}\ (0, \sigma_v^2) \quad (6.2)$$

$$u_{ij} \sim \mathrm{N}\ (0, \sigma_u^2) \quad (6.3)$$

where t represents a two-day risk period and i and j the i^{th} cow in the j^{th} herd; μ_{tij} the fitted probability of Preg_{tij} (the outcome of the i^{th} cow in the j^{th} herd becoming pregnant during risk period t); lnDIM_{tij} the natural logarithm of DIM at the beginning of risk period t; α the regression intercept; β_1 and β_2 the coefficients for the terms representing days in milk; $\mathbf{X}_{tij}$ the vector of risk period level covariates and $\boldsymbol{\beta}_3$ the corresponding vector of coefficients; $\mathbf{X}_{ij}$ the vector of cow-level covariates and $\boldsymbol{\beta}_4$ the corresponding vector of coefficients; u_{ij} the random effect to reflect variation between individual cows and v_j the random effect representing variation between herds, with σ_u^2 and σ_v^2 the variances of the normal distributions of the respective random effects terms.

Model building and final parameter estimation was carried out using MLwiN v2.20 (Rasbash et al., 2010). Model building and selection used the approach described in Chapter 4, with Markov chain Monte Carlo (MCMC) sampling used for final parameter estimation (Browne, 2009) and retention in the model of variables where the 95% area of highest posterior density (HPD) for the variable's coefficient did not cover zero. Biologically plausible first order interaction terms were tested, and retained in the model only if their inclusion made a substantial difference to parameter estimates for coefficients of the main effects. Inclusion of herd-level random effects (slope variation) for the lameness-related model terms was also tested, to account for the possibility that the association between lameness and reproductive performance could vary between herds. These were again retained in the model only if they altered parameter estimates for main effects by more than 1%, or if between-herd variation was large relative to mean effect size (such that the variance of the herd-level random effect for the variable was more than 20% of the mean/overall effect).

Model sensitivity analysis revealed that the parameters of interest were not sensitive to choices made during data restructuring and model building (e.g. choice of risk period duration, choice of function to represent DIM or selection of timeframes for lameness events). Simulation-based posterior predictions were used to evaluate model fit as described in Chapter 4, by subsetting the data in a variety of ways, using the model to predict probability of pregnancy for each risk period in the subset and checking that the observed proportion of risk periods where pregnancy occurred lay within the 95% coverage interval of the predicted risk. Model results were illustrated as relative risks using a similar prediction-based approach (as described in Section 4.2.2.4). Posterior predictions were carried out in R v2.14, using MCMC chains exported from MLwiN.

6.2.3 Probabilistic sensitivity analysis

In order to explore the relationship between herd reproductive performance and the incidence rate of lameness at herd level, a simulation model was developed. The aim of this part of the study was to evaluate the results of the discrete time survival analysis in a wider context to assess its potential usefulness to inform clinical on-farm management decisions.

6.2.3.1 Simulation model structure and process

The outline structure of the simulation model is the same as described in Chapter 5 (Figure 5-1). The model was constructed in Microsoft Excel 2010 (Microsoft Corp.), using Visual Basic for Applications (Microsoft Corp.) for process control. The explanatory variables in the final discrete-time survival model became input parameters for the herd-level simulation model, which was used to simulate 50,000 herds of 200 lactations each. Simulating a herd first involved drawing the herd-level input parameters (e.g. the herd's mean 305-day adjusted milk yield and incidence rate of clinical lameness) from the distributions shown in Table 6.3. Simulation of the first cow-lactation in the herd was then commenced by drawing the lactation-level inputs (e.g. the parity of the cow) from the relevant distributions and simulating a clinical lameness history for the lactation. The latter was accomplished by using the distribution of DIM of all clinical lameness events from the original dataset (Figure 6-1) to assign a probability that a lameness event would occur at each two-day risk period through a lactation in a herd with a given overall lameness incidence rate. The discrete-time survival model described in the previous section was used to calculate the predicted probability of pregnancy occurring during each two-day risk period given the input parameters for that herd, lactation and risk period. This probability was adjusted to account for the herd's overall ("background") level of submission rate and pregnancy rate (i.e. the variation in these parameters not explained by lameness, milk yield or other model inputs), but it is important to note that any association between lameness events and submission or pregnancy rate will be represented as effects from the discrete time survival model.

Table 6.3 Input parameters used for each level of simulation

Input variable	Level	Input distribution
Submission rate	Herd	Uniform, range 10-80%
Pregnancy rate	Herd	Uniform, range 10-60%
Herd average 305d milk yield	Herd	Uniform, range 3000-12,500 litres
Proportion of herd in first lactation	Herd	Uniform, range 10-40%
Incidence rate of lameness	Herd	Uniform, range 0.1-1.5 cases/cow-year
Cost per extra empty day	Herd	Uniform, range £1.20-£4.20
Cost per failure to conceive cull	Herd	Uniform, range £550-£1750
Parity/lactation number	Lactation	Discrete, based on proportion of herd in first lactation
305d lactation milk yield	Lactation	Beta, centred on herd average with standard deviation of 1,500 litres; adjusted for parity
Days in milk	Risk period	As described in text
Lame 43-70d before risk period	Risk period	Binary, as described in text
Lame within 14d of risk period	Risk period	Binary, as described in text
Lame 43-70d after risk period	Risk period	Binary, as described in text
Lame 71-100d after risk period	Risk period	Binary, as described in text

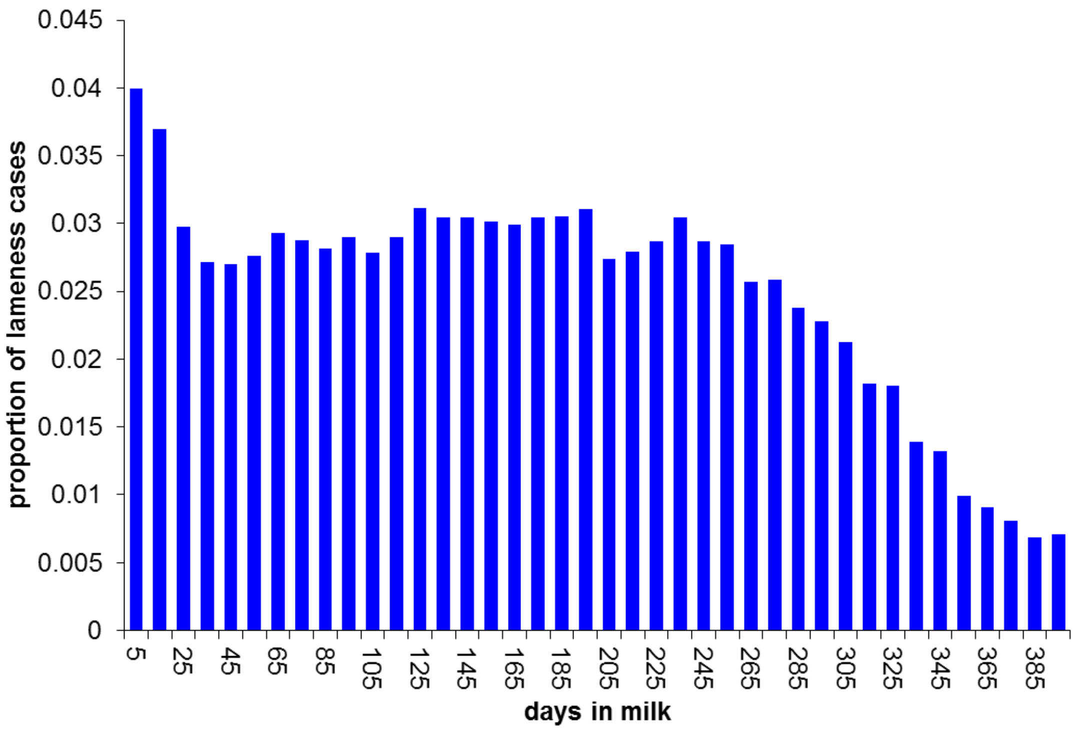

Figure 6-1 Distribution of lameness cases observed by days in milk

A binary outcome to represent whether or not the cow became pregnant during the risk period was drawn from a binomial distribution based on this calculated probability. Repeated risk periods were simulated for each cow, until she either became pregnant or reached 300 DIM (a point at which farmers would commonly elect to remove cows from the herd if not pregnant), at which time simulation of the next lactation was begun. When 200 lactations had been simulated, the herd was considered complete. The mean number of DIM at pregnancy and the proportion of lactations ending without a pregnancy in each herd were stored along with the herd-level input parameters before beginning simulation of the next herd.

6.2.3.2 Simulation model inputs

The input distributions for each parameter were selected based on the authors' clinical experience, such that the ranges would be expected to cover the majority of UK dairy herds (Table 6.3). Uniform distributions were specified for all herd-level inputs, so that every potential scenario was equally likely to be selected. This was not used to represent the true distributions of these parameters across herds; the objective was not to speculate on which situations might occur most commonly, but to evaluate the potential impact of all different lameness incidence rates across as wide a variety of herd scenarios as possible. Some of the lactation-level inputs were drawn from non-uniform distributions so that the architecture of each simulated herd was realistic (so, for example, the milk yield for a lactation was drawn from a beta distribution parameterised such that a cow was likely to draw a lactation yield close to the herd average, and there was a smaller chance of drawing a yield much further from the average), as described in Table 6.3.

6.2.3.3 Simulation model outputs and analysis

A single herd-level outcome was devised to represent reproductive performance for each simulated herd (to allow evaluation of associations between this and the various input parameters). The mean number of DIM at pregnancy and the proportion of cows reaching 300

DIM without conceiving were combined using a modification of the method of Esslemont and Kossaibati (2002) to produce a “modified FERTEX” (mFX) score. This involved comparing each value to a pre-set target (set at 60 days for mean DIM at pregnancy and zero for proportion of cows reaching 300 DIM without conceiving), and applying a unit cost to the difference from target for each. Since the cost of a culled cow and an additional empty day are widely acknowledged to vary from herd to herd, these were considered as herd-level inputs, and each drawn randomly for each herd from the distributions described in Table 6.3. The mFX score for each simulated herd was therefore a cost-based single measure of overall fertility performance (so that higher performing herds had lower mFX scores and vice versa).

Results from the simulations were analysed initially by illustrating associations between herd-level input parameters and mFX scores graphically using high-density scatterplots. Spearman rank correlation coefficients were calculated for the association between each herd-level input and mFX score (a non-parametric measure of correlation was selected as the mFX scores were positively skewed). Multiple regression (with the natural logarithm of mFX score as the outcome) was used to partition variance in mFX score between the various herd-level inputs. The resulting regression model was also used to predict the effect on mFX score of increasing each individual input in turn from the middle of its input distribution to the upper quartile so that results could be displayed graphically as a tornado plot (a standard approach for presentation of PSA results).

6.3 Results

There were a total of 16,706 pregnancies from the 1,247,677 risk periods in the dataset, so that 1.34% of risk periods resulted in a pregnancy (corresponding to around 14% of cows becoming pregnant during each 21 day oestrous cycle). Of the 22,319 lactations in the dataset, 4,360 involved at least one case of lameness (corresponding to a lactational first case incidence rate of 19.5%).

6.3.1 Discrete-time survival analysis

Table 6.4 shows the parameter estimates for the regression model derived to predict the probability of pregnancy resulting during a two-day risk period. The predictor variables not directly associated with lameness showed very similar associations to those described in Chapter 4, with probability of pregnancy peaking at around 110 DIM, decreasing with increasing 305-day adjusted milk yield and lower predicted probabilities of pregnancy for cows in higher parities and during the months April to September. Clinical lameness events during four different time frames relative to the two-day risk period showed associations with the probability of pregnancy during the risk period. The largest association was seen when a lameness event was recorded within 14 days of the risk period, when the odds of pregnancy were reduced by almost 25% (odds ratio [OR] 0.76, area of 95% highest posterior density [HPD] 0.69 – 0.84). Lameness events recorded 43 to 70 days before, 43 to 70 days after and 71 to 100 days after a risk period were all associated with a reduction in the odds of pregnancy during the risk period of around 15% (ORs 0.85, 0.88 and 0.86 respectively; areas of 95% HPD 0.76 – 0.95, 0.80 – 0.98 and 0.79 – 0.95 respectively). These associations are represented as posterior predicted relative risks in Figure 6-2. Posterior predictions were also used to demonstrate that model fit was good. For each subset of data tested, the observed proportion of risk periods where pregnancy occurred fell within the 95% area of HPD of predicted risk for that subset (Figure 6-3).

Table 6.4 Parameter estimates for discrete time survival model
HPD 2.5% and HPD 97.5% represent the lower and upper bounds (respectively) of the 95% area of highest posterior density for each parameter.

Model term	n	coefficient	odds ratio	HPD 2.5%	HPD 97.5%
Intercept	1247677	-40.1		-40.3	-39.9
ln DIM	1247677	15.4		15.3	15.4
(ln DIM)^2	1247677	-1.62		-1.62	-1.61
Parity 1	325621		*Reference*		
Parity 2	288951		1.06	1.00	1.11
Parity 3	223118		0.98	0.92	1.03
Parity 4	153753		0.95	0.89	1.01
Parity >4	256234		0.76	0.72	0.81
Year: 2002 or earlier	148578		*Reference*		
Year: 2003	86158		1.00	0.92	1.09
Year: 2004	147847		0.90	0.83	0.97
Year: 2005	216142		0.93	0.86	1.00
Year: 2006	313278		0.86	0.80	0.92
Year: 2007-8	335674		0.90	0.83	0.97
Season 1: Jan-Mar	332357		*Reference*		
Season 2: Apr-Jun	278139		0.90	0.86	0.94
Season 3: Jul-Sept	266050		0.74	0.70	0.78
Season 4: Oct-Dec	371131		1.00	0.96	1.04
Centred 305d yield (x1000kg)	1247677		0.92	0.91	0.93
No lameness 70-43d before	1219868		*Reference*		
Lameness case 70-43d before	27809		0.85	0.76	0.95
No lameness within 14d	1207760		*Reference*		
Lameness case within 14d	39917		0.76	0.69	0.84
No lameness 43-70d after	1207155		*Reference*		
Lameness case 43-70d after	40522		0.88	0.80	0.97
No lameness 71-100d after	1203737		*Reference*		
Lameness case 71-100d after	43940		0.86	0.79	0.95

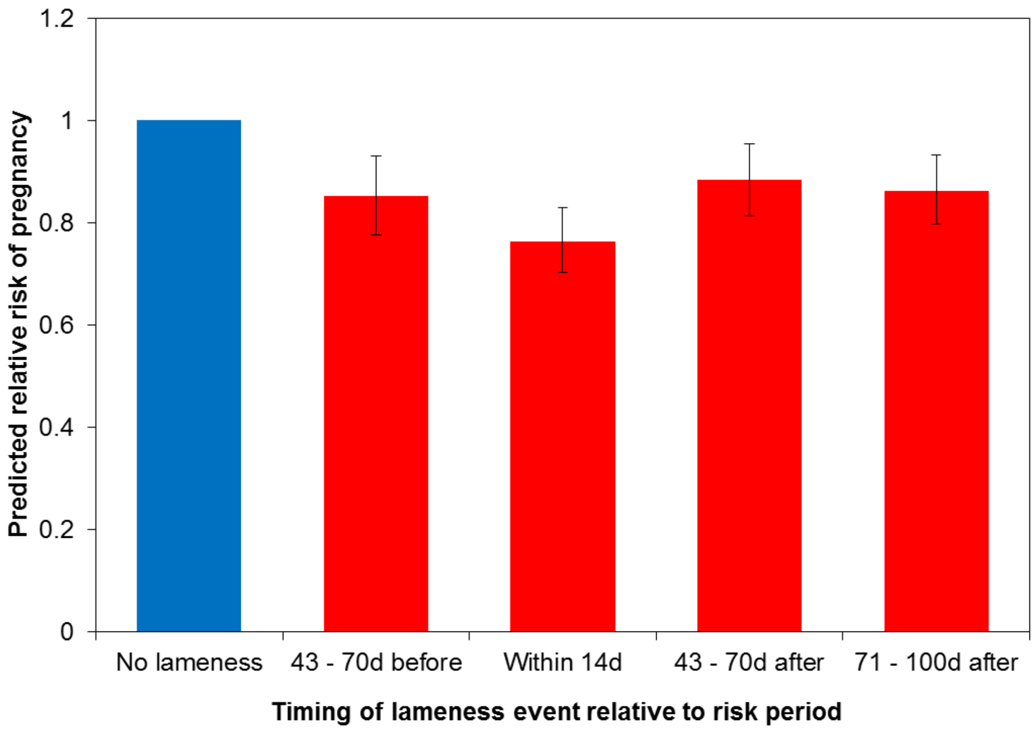

Figure 6-2 Association between predicted relative risk of pregnancy and clinical lameness.
Error bars represent the 95% credible interval for each predicted relative risk.

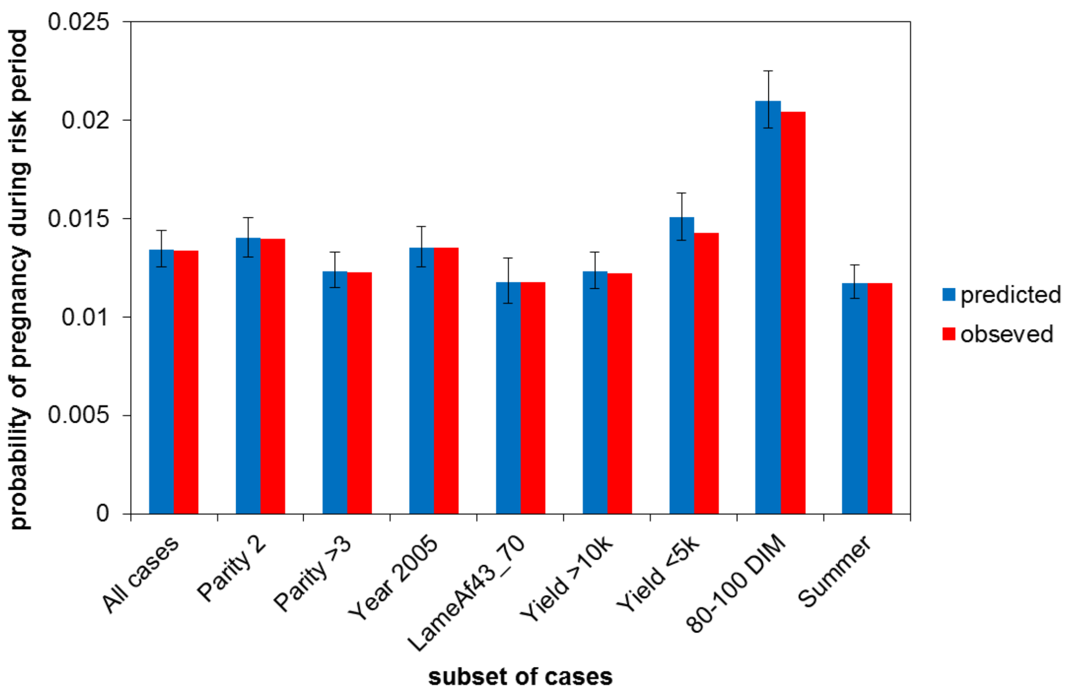

Figure 6-3 Predicted and observed risk of pregnancy across various categories
Predicted absolute risk of pregnancy (blue bars) at risk periods in various categories (x-axis) compared to the observed proportion of risk periods in that category where a pregnancy occurred (red bars). Error bars represent the 95% credible interval for each predicted risk.

6.3.2 Probabilistic sensitivity analysis

6.3.2.1 Univariate analysis

Univariate analysis of PSA results is presented using high-density scatterplots in Figure 6-4. These show that, as in Chapter 5, a herd's "background" level of submission and pregnancy rate were the individual inputs with the strongest influence on overall herd fertility performance, with both being moderately strongly correlated with herd mFX score (Spearman rank correlation coefficient -0.65 for submission rate and -0.59 for pregnancy rate). The herd incidence rate of clinical lameness had no clear relationship with mFX score, with a Spearman rank correlation coefficient of 0.028 and the scatterplot showing a square appearance with no clear trend in the area of highest point density.

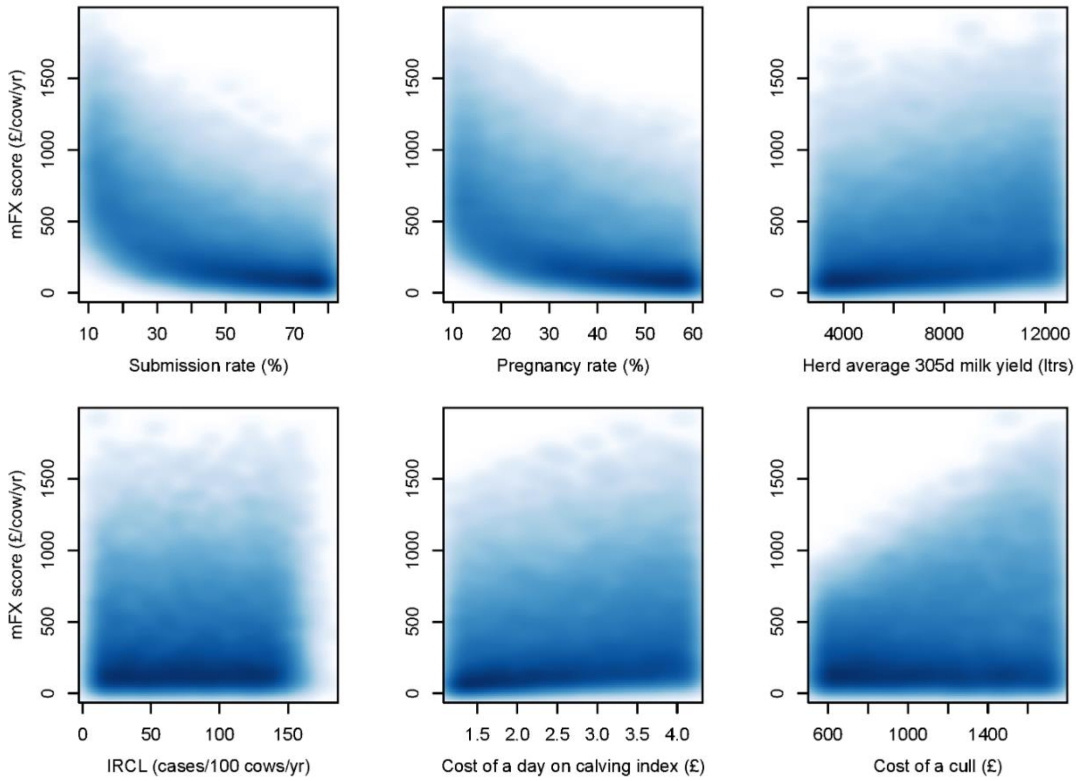

Figure 6-4 Associations between simulation inputs and overall herd-level reproductive performance
High density scatterplots showing the association between each simulated herd's reproductive performance (represented by modified FERTEX score, mFX, y-axis) and selected simulation input variables. Darker colours indicate areas of higher point density, IRCL: incidence rate of clinical lameness

6.3.2.2 Multivariate analysis

Analysis of the simulation results in a multivariate framework allows visualisation of results from the discrete time survival model in a clinical context. Table 6.5 shows that the herd's "background" level of submission and pregnancy rate explained the vast majority of the variation in herd mFX score, with 75% of overall variance explained by these two input parameters. It is important to remember that these inputs represent the marginal effect of between-herd variation in these aspects of fertility performance after the other model inputs have been accounted for (so, for example, a herd's "background" pregnancy rate would reflect its insemination success rate after accounting for any effects of milk yield, age structure and level of lameness).

Table 6.5 Multiple regression derived partition of variance in modified FERTEX scores

Input parameter	Proportion of variance explained
Submission rate	41.4%
Pregnancy rate	34.2%
305-day adjusted lactation milk yield	8.9%
Cost per additional day on calving interval	5.7%
Cost per failure-to-conceive cull	2.0%
Incidence rate of clinical lameness	0.1%
Proportion of herd in lactation 1	0.0%

Figure 6-5 shows the predicted change in herd mFX score which would result from a herd increasing each input parameter in turn from the middle of the range of the input distribution by 25% of the total range while the other inputs remain at the population median. For example, the top line on the plot shows that an increase in submission rate from the median value of the range of distribution for this input (45%) to the value representing the lower boundary of the upper quartile of the range (62.5%) would be expected to result in a decrease in mFX score (i.e. an improvement in overall reproductive performance) of around £100/cow/year. Increasing the herd's incidence rate of lameness cases from 80 to 115 cases/100 cow-years would be expected to increase herd mFX score by just over £5/cow/year.

Therefore, a reduction in lameness incidence of 35 cases/100 cow-years (which would represent a large improvement, and may require substantial financial and time investment from the farmer) would be expected to lead to the same degree of improvement in fertility as an increase in submission rate of less than 1% (a small change, which would be expected to require substantially less investment).

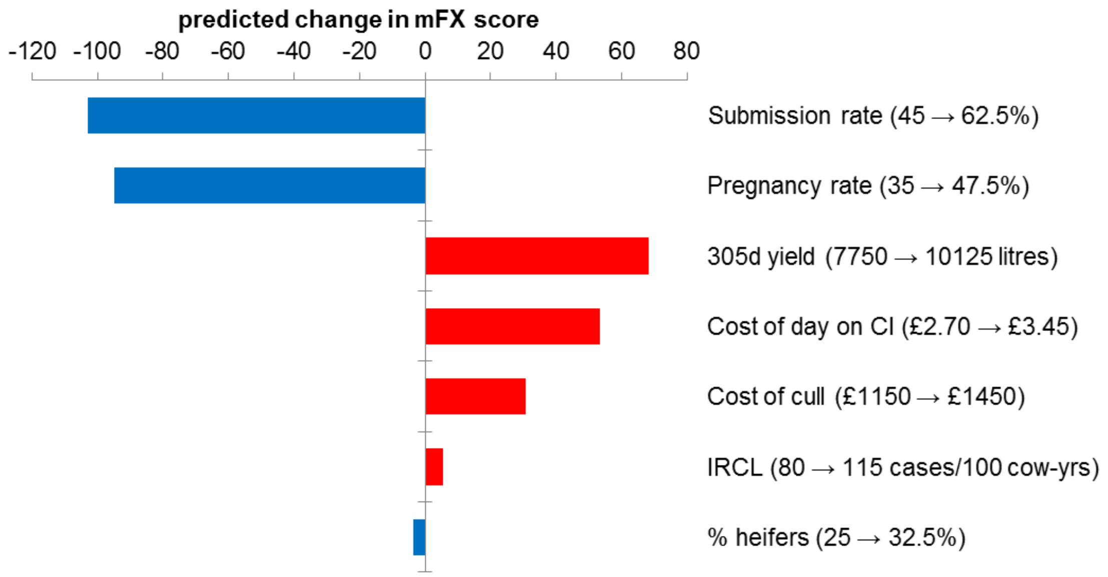

Figure 6-5 Predicted effect of an equivalent increase in each input parameter on overall reproductive performance
Tornado plot showing the predicted effect of increasing each input parameter in turn by a value representing 25% of the range of its input distribution from the median value, while the other input parameters are held at their population medians. The input parameters are listed on the right hand side of the graph, and the change in each input (from median to upper quartile) is given in parentheses. For example, the top bar shows that the predicted effect of moving from a submission rate of 45% (the median of the input distribution for this parameter) to 62.5% (the upper quartile of the input distribution) would be a decrease of just over £100/cow/year in the herd's modified FERTEX (mFX) score.

6.4 Discussion

This study showed that relatively large associations between clinical lameness events and reproductive performance could be demonstrated at the level of a risk period within lactation (for example, occurrence of a lameness case within 14 days of a risk period was associated with a 25% reduction in the risk of the cow becoming pregnant during the risk period, Figure 6-2). However, PSA revealed that a herd's incidence rate of lameness was highly unlikely to make a significant contribution to its overall level of reproductive performance (either through reduction in submission rate or in pregnancy rate) when other factors affecting fertility were also taken into account.

There is substantial variation in the conclusions of existing work evaluating the association between lameness and reproductive performance. A variety of previous studies have found associations between decreased fertility and either clinical lameness events (Alawneh et al., 2011; Bahonar et al., 2009; Hernandez et al., 2001) and/or identification of lameness through visual gait assessment (Bicalho et al., 2007; Hernandez et al., 2005). In contrast, other studies have failed to reveal such an association (Loeffler et al., 1999; Peake et al., 2011; Sogstad et al., 2006): notably there seems to be a tendency for studies involving larger numbers of herds to fail to identify significant associations. Many of the pre-existing papers in this area describe studies involving less than five herds (and most use a single herd); the notable exceptions to this are Loeffler et al. (1999) (43 herds) and Sogstad et al. (2006) (112 herds), neither of which found significant associations between lameness events and reproductive outcomes. It is biologically plausible that any effect of lameness on reproductive performance will vary between herds (for example, due to the variation in the predominant causes of lameness in each herd and variation in the effectiveness of management of lame cows). The current study used data from 39 herds, but from a much larger number of cows compared to previous work. The possibility of between-herd variability in the association between lameness events and fertility was explored here using herd-level random effects terms for the explanatory variables related to lameness. This revealed relatively little between-herd variability in effect within this group of herds.

Some of the variability in previous published results may also be related to the way in which reproductive outcomes were measured: this study revealed significant associations between lameness events and the probability of pregnancy over a specific window of time relative to the lameness case, but when results were used to evaluate this within a PSA framework it transpired that lameness incidence rate was unlikely to influence overall herd reproductive performance. This means that previous studies focussing on particular categories or timings of lameness event and/or reproductive outcome may have been more likely to generate significant findings than those using broader categories or timeframes.

This study illustrates the usefulness of simulation-based techniques (such as PSA) to aid interpretation and contextualisation of model results. The approach described provides a potential route for researchers to facilitate better understanding of the results of their work and how they should be interpreted in a clinical context. This in turn can enhance research impact, and accelerate change in clinical practice. Although this example describes application of PSA to help interpret the results of a discrete time survival analysis, the technique would be equally applicable to other types of complex model, and to other analyses based on logistic regression. In logistic regression, the model coefficients themselves can be difficult to interpret. Classically the coefficients are exponentiated to produce odds ratios (as shown in Table 6.4), but odds ratios themselves can be misleading because humans intuitively tend to think in terms of risk or probability rather than odds (and these can be quite different, especially where the risk is close to 0.5). This topic has been extensively explored in the medical literature (Bland and Altman, 2000; Davies et al., 1998; Zhang and Yu, 1998), where results of such analyses must be interpreted by clinicians, some of whom may have a limited understanding of statistical methods. It is possible to convert an odds ratio to a relative risk for more intuitive interpretation (as shown in Figure 6-2), but where decisions are to be made at population level these can also be difficult to interpret. For example, in this case the relative risks would have been hard to interpret without a method to incorporate the likely range of herd-level lameness incidence rates and the distribution of lameness events through lactation. Here, the results from the discrete time survival model alone (along with some of the pre-existing literature) may have encouraged clinicians to place too much emphasis on control of lameness to improve herd-level reproductive performance.

This study provides another example of the usefulness of simulation-based techniques such as PSA as an extension of statistical modelling to help illustrate model results in an intuitive way within a clinical veterinary context. In this example, while there are associations between lameness events and reproductive performance at specific time-points, it is unlikely that a herd's incidence rate of lameness will have a substantial impact on herd fertility. This does not mean that lameness control is not important: lameness has well defined and significant impacts on both animal welfare and productivity (Huxley, 2013). Rather, our analysis suggests that herd lameness control is unlikely to lead to a clinically significant improvement in overall reproductive performance in the majority of situations.

Chapter 7 Associations between fertility and routine milk recording data

7.1 Introduction

The relationship between energy balance and reproduction is well established, and a commonly posited reason for the medium-term decline in dairy cow fertility is that increasing milk yields have made energy balance more challenging to manage (Lucy, 2001). A period of negative energy balance (NEB, defined as a situation where a cow's daily energy intake cannot meet her total energy requirement) is extremely common during early lactation in modern dairy cows (de Vries and Veerkamp, 2000; Jorritsma et al., 2003), but exposure to prolonged or severe NEB has frequently been associated with decreased fertility. This can happen when the cow is exposed to NEB very early in lactation (well before she is eligible to be bred), or when NEB occurs around the time of insemination. There are a number of proposed mechanisms which may be responsible for this relationship. Exposure of developing ovarian follicles to biochemical conditions associated with NEB leads to a reduction in the quality of the oocyte as well as a decrease in the quality of the corpus luteum produced after ovulation (Leroy et al., 2008, 2005; Roth et al., 2001). Post-ovulation levels of progesterone tend to increase with increasing number of ovulations after calving (Villa-Godoy et al., 1990), and progesterone concentration is associated with embryo survivability (Butler, 2001). As the timing of resumption of cyclicity is associated with the degree of early lactation NEB (Garnsworthy et al., 2008), this represents a further potential link. Energy balance has also been shown to affect the speed of post-calving endometrial repair (Wathes et al., 2007). Most of these studies demonstrate or support associations between energy balance and pregnancy rate (i.e. the proportion of serves resulting in a pregnancy); there is substantially less evidence linking energy balance to intensity of oestrus expression. It is therefore likely that the majority of the influence of energy balance on reproductive performance is mediated through pregnancy rate, rather than submission rate (i.e. the proportion of eligible cows inseminated per unit time).

Monitoring energy balance in early lactation is a key component of any herd monitoring programme. However, although numerous approaches to measuring energy balance exist, all of them have significant drawbacks. Body condition scoring has traditionally been the main tool, but this technique is inevitably slightly subjective, and inter- and intra-observer repeatability is a problem (Ferguson et al., 1994; Kristensen et al., 2006). Additionally, there has recently been increased awareness that condition scoring measures degree of subcutaneous fat deposition (Roche et al., 2009): it may be that intra-abdominal fat deposits vary more with energy balance, and these are not well correlated with subcutaneous fat. Measurement of metabolites in blood is considered to represent the gold standard in monitoring energy balance (Cooper, 2011). Both beta-hydroxybutyrate (a ketone body produced under conditions of negative energy balance) and non-esterified fatty acids (the form in which mobilised body fat is transported in the circulation) can be measured in peripheral blood, and abnormally elevated levels have been associated with decreased production (Duffield et al., 2009; Edwards and Tozer, 2004), impaired reproductive performance (Ospina et al., 2010; Walsh et al., 2007) and a variety of peri-parturient diseases (Duffield et al., 2009; LeBlanc et al., 2005). However, sampling of every cow has generally been impractical, meaning that inferences about herd prevalence of subclinical ketosis were often made on sub-samples of the herd, which introduced an additional level of variation. More recently, the availability of relatively inexpensive hand-held ketometers has made widespread sampling more practical.

Despite this, there is still substantial interest in proxy measures of energy balance, primarily those derived from routinely collected milk recording data. These are generally based on the principle that milk protein concentration tends to be reduced under conditions of negative energy balance (Coulon and Rémond, 1991) whilst milk butterfat can be increased where body fat is being mobilised. Simply measuring either protein of butterfat concentration in early lactation is generally not useful because of the additional effect of dilution of milk constituents with yield, so most measures make some attempt to account for this. The most commonly used proxy measure is the ratio of butterfat to protein (fat:protein ratio, FPR). The rationale behind this measure is that FPR will increase with either an increased butterfat or a decreased protein concentration, and that using the ratio between the two accounts for the effect of milk yield on both.

Many studies have demonstrated associations between elevated FPR at first milk recording of lactation and impaired fertility (Heuer et al., 1999; Loeffler et al., 1999; Podpečan et al., 2008) or increased disease risk (Geishauser et al., 1998; Heuer et al., 1999). However, there is little clear evidence that links FPR (or other milk recording proxy measures of energy balance) directly to subclinical ketosis. A study by Duffield et al. (1997) demonstrated that FPR had an optimal sensitivity and specificity for detecting subclinical ketosis of 58% and 69% respectively, and concluded that this would limit its use in practice. A smaller UK study also failed to detect clinically useful associations (Cooper, 2011). Despite this, reports based on FPR are available in all the major UK performance monitoring software packages, and anecdotal reports suggest that it remains in widespread use in the UK and elsewhere.

A potential drawback with the use of measures such as FPR as proxies for NEB is that milk butterfat and protein concentrations are known to vary with season, and with days in milk (DIM) at the time of sampling. Whilst monitoring of FPR is commonly restricted to the first test day of a cow's lactation, the cow's first test day could occur as early as 2-5 DIM (although cows are usually not recorded in the first week after calving) or as late as 40 DIM even in herds recording on a regular monthly basis. It is plausible that this introduces a large degree of explainable variation in FPR which is not routinely accounted for when this is used in practice.

Seasonal changes are likely to be most marked in herds where nutrition varies markedly through the year, for example where early lactation cows graze over the summer months and are housed through the winter (a common approach in UK herds). To assess whether correcting for these factors influenced the usefulness of this technique, Madouasse et al. (2010a) used a large dataset from UK dairy herds to evaluate relationships between a cow's milk recording information at the first two test days of lactation and her calving to conception interval. The resulting model included several terms relating to measures of milk quality and yield, and was reported to be predictive of the probability of conception across seven discrete time intervals from 20 to 145DIM.

As discussed earlier, It is likely that the aspect of fertility most influenced by early lactation energy balance is pregnancy rate (i.e. the probability of a given serve leading to a pregnancy) rather than submission rate (the proportion of eligible cows served per unit of time). Since the current dataset included reliable serve and serve-outcome data (unlike the data used in the earlier UK work), the aim of this study was to correct early lactation milk recording data for DIM and season at the time of sampling, and to evaluate associations between these parameters and the probability of a serve leading to a pregnancy. In an attempt to explain more of the variation in pregnancy rate, other variables from the dataset (such as lactation number and milk yield) were tested in the same model.

7.2 Materials and Methods

7.2.1 Data restructuring and quality

Collection, initial auditing and restructuring of the dataset is described in Chapter 2. For this study, data quality measures intended to select herds with good quality milk recording and fertility data were applied. Data quality measurement was initially made at herd-year level, so that herds contributed data from each year in which they met the criteria. These included the proportion of calvings for which no corresponding serve event was recorded and the proportion of lactations with “unresolved” outcomes (i.e. the cow had not left the herd or recalved at least two years after the calving beginning the lactation); herds with pregnancy rates to first or all serves which were considered biologically implausible (greater than 65%) were also excluded.

In order to manage herds with different milk recording test intervals, individual cow recordings at 5 to 35 DIM were considered to represent a typical “first” test day of lactation, and those at 35-65 DIM a typical “second” test day (as most herds have test days at monthly intervals). Where a lactation contained more than one recording within one or both of these windows, the nearest to the centre of the window was selected: for example, if a cow had test days at 7 and 25 DIM, the latter would be selected to represent her “first” recording of lactation, as it is nearer to the centre of the window for “first” recordings (at 20 DIM). This process also excluded all milk recordings at less than 5 DIM, because cows are not usually recorded this early in lactation (as there is still a high proportion of colostrum at this time). The proportion of all lactations including a test day for each window was calculated at herd-year level, and herd-years with low proportions excluded (on the basis that this was likely to reflect variable or abnormal recording intervals, or frequent recording errors or missing cows). There were 1,493 herd-years from 312 herds which met these quality criteria.

Further data quality criteria were applied at lactation level, with exclusions based on missing fertility (e.g. lactations ending in calving where no serves were recorded, or with unresolved outcomes) or milk recording (e.g. missing butterfat, lactose or protein percentage or milk yield within the windows described above) data. A total of 165,715 lactations met these criteria. From these lactations, each serve event at less than 100 DIM represented a unit (line) of data for the logistic regression model (Section 7.2.3). This model used serves from early lactation only, as it was considered likely that early lactation energy balance would have the most profound effect on these serves.

7.2.2 Correcting milk recording data

Milk recording data from each of these lactations were used to perform regression analyses of each milk recording parameter (daily yield, FPR and butterfat, protein and lactose percentages) against DIM and day of the year for that recording test day. FPR was calculated for each recording event by dividing butterfat percentage by protein percentage. Distributions for each variable were evaluated, and recording events with an outlying observation for any variable removed from the dataset. Scatterplots for mean and standard deviation (SD) of each variable by DIM and by day of the year were produced; from these it appeared that polynomial functions would represent changes in both mean and SD of each variable with DIM, and trigonometric functions would be required to represent the cyclical variation with day of the year. A normal-outcome linear regression model was built using R 3.0.0 (R Core Development Team, 2010) for each milk recording variable in turn, with DIM and day of year of recording as predictor variables. The models took the form:

$$y = \beta_0 + \beta_1 DIM + \beta_2 DIM^2 + \ldots + \beta_6 DIM^6 + \beta_7 sin\left(\frac{2\pi Day}{365}\right) + \beta_8 cos\left(\frac{2\pi Day}{365}\right) \qquad (7.1)$$

where y represents the outcome variable, β_0 the overall intercept, β_n the other model coefficients, and *DIM* and *Day* represent days in milk and day of year of the recording event respectively. Interaction terms between the DIM and Day terms were also tested. Terms were retained in the model where the 95% confidence interval for the estimate of the coefficient did not cover zero, and Q-Q plots used to evaluate distribution of model residuals (with outliers or influential points removed when revealed).

The resulting models were used to generate predicted values for each milk recording parameter for each milk recording event (based on DIM and day of year for that event). A similar approach was used to produce a predicted SD for each parameter for each recording event. Each parameter at each recording event was then corrected by subtracting the predicted value from the observed value and dividing by the predicted SD for that parameter (such that the corrected values for each parameter were centred around zero and scaled to have a standard deviation of 1). High density scatterplots for each corrected parameter were then used to confirm that the corrected values did not vary systematically with either DIM or day of year. These corrected values were then used as predictor variables in the models described in Section 7.2.3.

7.2.3 Logistic regression analysis

7.2.3.1 Model building

A multilevel logistic regression model (similar to that described in Section 4.2.2.2) was constructed to predict the probability that a given serve would result in a pregnancy. The model took the form:

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

$$\ln\left(\frac{\mu_{ijkl}}{1-\mu_{ijkl}}\right) = \alpha + \boldsymbol{\beta}_1 \mathbf{X}_{ijkl} + \boldsymbol{\beta}_2 \mathbf{X}_{jkl} + \mathrm{u}_{kl} + \mathrm{v}_l \quad (7.2)$$

$$\mathrm{v}_l \sim \text{normal distribution } (0, \sigma_v^2) \quad (7.3)$$

$$\mathrm{u}_{kl} \sim \text{normal distribution } (0, \sigma_u^2) \quad (7.4)$$

where *i* represents a given serve in lactation *j* of cow *k* in herd *l*. Preg_{ijkl} represents the binary outcome of serve *i* leading to a pregnancy with fitted probability μ_{ijkl}, α the overall intercept, $\mathbf{X}_{ijkl}$ the vector of serve-level predictor variables with $\boldsymbol{\beta}_{ijkl}$ the corresponding vector of coefficients, $\mathbf{X}_{jkl}$ the vector of lactation-level predictor variables and $\boldsymbol{\beta}_{jkl}$ the corresponding vector of coefficients. v_l represents the herd-level random effect (with variance σ^2_{v}) and u_{kl} the cow-level random effect (with variance σ^2_{u}). A four-level structure (with lactations within cows as the additional level) was rejected because of the large number of lactations contributing only a single serve.

Model building was carried out in MLwiN version 2.29 (Rasbash et al., 2010), with iterative generalised least squares methods used for initial parameter estimation during model building. The potential predictor variables included in the model building process are listed in Table 7.1.

Table 7.1 Potential predictor variables included in the model building process

Variable	**Representation**
Lactation level	
Butterfat at recording 1	Linear, corrected for DIM and day of year at test day
Butterfat at recording 2	Linear, corrected for DIM and day of year at test day
Protein at recording 1	Linear, corrected for DIM and day of year at test day
Protein at recording 2	Linear, corrected for DIM and day of year at test day
Lactose at recording 1	Linear, corrected for DIM and day of year at test day
Lactose at recording 2	Linear, corrected for DIM and day of year at test day
Fat:protein ratio at recording 1	Linear, corrected for DIM and day of year at test day
Fat:protein ratio at recording 2	Linear, corrected for DIM and day of year at test day
Daily yield at recording 1	Polynomial (order <3) , corrected for DIM and day of year
Daily yield at recording 2	Polynomial (order <3) , corrected for DIM and day of year
305-day adjusted lactation yield	Centred around population mean, polynomial (order <4)
Month of calving	Categorical: months as individual categories
Lactation number	Categorical: >4 recoded as single category
Serve level	
Days in milk	Polynomial (order <4)
Inter-service interval	Categorical: <18d, 18-24d, 25-35d, 36-48d, >48d, NA (first serve of lactation)
Month of serve	Categorical: months as individual categories

The model was built by forward selection, with terms retained in the model if the magnitude of the estimated coefficient was greater than double the estimated standard error. Univariate scatterplots of observed pregnancy rate versus each continuous predictor variable in turn were used to assess whether a linear or polynomial function was most appropriate to represent that predictor. Where a non-linear pattern was evident, a polynomial of appropriate degree (to represent the number of points of inflection) was considered to represent that term. Polynomial functions were retained in the model where parameter estimates for each term of the polynomial met the criteria above, or where their inclusion made a difference of at least 10% to at least one other parameter estimate in the model. First order interactions which were considered to be biologically plausible or important were tested (for example, an interaction between each milk recording covariate and the natural logarithm of DIM was tested to allow for the possibility that the effect of early lactation milk constituents would be different for serves at higher DIM), and retained where they met the criteria described above or where inclusion of the interaction term altered at least one other parameter estimate by at least 10%. To maximise model parsimony, categorical variables were recoded where several categories had similar coefficient estimates. For example, each of the months June to September had a very similar coefficient, and the remaining months were also similar to each other. Month was therefore recoded as a binary indicator for months June to September, compared to a reference category representing October to May. Herd-level random effect terms (random slopes) were tested for each main fixed effect (such that the effect of each variable was allowed to vary at herd level); such terms were retained in the model where they resulted in a change of at least 10% to at least one other parameter estimate or a Wald test for inclusion of the term had a p-value less than 0.1 (Rasbash et al., 2010). Only the binary indicator variable representing the summer months met these criteria for a herd-level random effect; as seasonal effects were not the key focus of the study and the variance of the herd-level random effect was very small compared to the central estimate, this random effect term was removed for model parsimony. Finally, all rejected predictor variables were re-tested in turn and retained if they met the criteria described above.

Final parameter estimation was carried out using Markov chain Monte Carlo (MCMC) methods in MLwiN (Browne, 2009), with 10,000 monitoring iterations used after a 2,000 iteration burn-in. Diffuse prior distributions (functionally equivalent to a normal distribution with a very large variance for fixed parameters and a uniform distribution for scalar variances (Browne, 2009, 1998)) were specified for model parameters. Chains for each parameter estimate were exported to R 3.0.0, and the CODA package (Plummer et al., 2006) was used for their analysis. Plots of estimate by iteration number were inspected for each parameter to ensure that satisfactory convergence had occurred. Areas of 95% highest posterior density were calculated for each parameter.

Three further (separate) models were built using similar methods, to explore different aspects of the relationship between early lactation milk constituents/yields and reproduction. The first was built using serve events occurring at between 100 and 200 DIM, in order to evaluate changes in these relationships in serves later in lactation. A further model was built using only serves at less than 100 DIM from lactations where a milk recording event occurred at 5-15 DIM; this was designed to explore the possibility that milk recording information may better reflect energy balance if captured very early in lactation. A final model was constructed in a discrete time survival framework (similar to those described in Sections 4.2.2.1 and 6.2.2), with units of data representing 21 day periods from each lactation beginning at 20 DIM and ending at conception, exit from the herd or 166 DIM. This was designed to replicate the model built by Madouasse et al. (2010a), and confirm that results were similar from this smaller subset of better-recorded herds.

7.2.3.2 Assessment of model fit

In order to assess model fit, posterior predictions were generated for each serve in the dataset, using the full MCMC chain for each parameter. Various subsets of the data were created; some were based on variables included in the final model (such as serves from second lactations, and serves at 30-50 DIM) and some based on variables from outside the model (such as serves from lactations where the first SCC recording was below the population mean, and serves from lactations beginning with a calving in April). For each subset of the data, the distribution of posterior predicted probability of pregnancy across all the serves in that subset was summarised as a mean and 95% coverage interval. This was then compared to the observed proportion of serves within that subset which resulted in a pregnancy (the observed pregnancy rate). Model fit was considered acceptable when the observed pregnancy rate fell within the 95% coverage interval of the predicted posterior distribution for all subsets examined; where this was not the case additional interaction terms or slope variation were explored during additional model building. Posterior predictions were carried out using R 3.0.0.

7.2.3.3 Illustration of model results using posterior predictions

In order to illustrate the results of the model, posterior predictions were generated for some out-of-sample example scenarios. In general, this involved producing a prediction from the model for a set of out-of-sample cases where all the variables were set at their population mean values and a single variable varied over a given range. For example, to illustrate the relationship between lactose concentration at first milk recording and pregnancy rate, predictions were made for a case where lactose at first recording varied over the range -2 to 2 while all other covariates remained at their population means. Since milk recording covariates had been standardised against expected mean and standard deviation (given DIM and day of year at that recording event), this range would cover two standard deviations either side of the mean. Where an interaction term was included in the model, predictions were generated over a range of one variable repeatedly for different values of the second variable. For example, to represent the interaction between lactose concentration at first recording and DIM at time of serve, predictions were generated for example cases where lactose varied from -2 to 2 whilst DIM was set at 50, 75 and 100, and all other variables were set at their population means. This generated three sets of predictions for the three values of DIM, with the range of lactose concentrations repeated in each set. Line plots were then generated for each set of predictions. For example, for lactose concentration at first recording, the line plot showed predicted pregnancy rate on the y-axis, lactose on the x-axis (over range -2 to 2) and had three lines representing predictions for serves at 50, 75 and 100 DIM.

For selected binary predictor variables, the population attributable risk for the variable was estimated. This involved generating full posterior predictions (again, using the full MCMC chains for each parameter) over a modified version of the dataset where all cases for which the variable of interest took the value one had that value reset to zero. This posterior distribution was then compared to that generated from the original unmodified dataset, to evaluate any expected change in population pregnancy rate were the effect of the variable of interest to be removed. For example, to illustrate the effect of the summer months on pregnancy rate, a posterior distribution was derived from a modified version of the dataset in which the indicator variable for summer was reset to zero for all serves. This distribution was then compared to that derived from the unmodified dataset.

In order to evaluate the degree to which the model would predict pregnancy rate in a clinically useful way, the dataset was divided into subsets by herd-year (i.e. into serves from a given herd in a given calendar year). Herd-years with less than 50 serves were removed from this dataset, and posterior predicted pregnancy rate generated for each remaining herd-year. These predictions were then compared to the observed pregnancy rates for each herd-year using scatterplots and Pearson correlation coefficients. In order to explore which predictors explained the variability in herd-year pregnancy rate, this process was repeated using different combinations of restricted elements of the model. For example, predictions were generated using the full fixed and random effects model, then without the cow-level random effect but retaining the herd-level effect, and finally from the fixed effects only. Comparison of r^2 values for the correlations between each set of predictions and the observed herd-year pregnancy rate allowed assessment of the proportion of variation explained by the complete model, and by each element of the model.

7.3 Results

7.3.1 Correcting milk recording data

Regression planes illustrating the relationship between early lactation milk recording parameters and DIM and day of year at the time of the recording event are shown in Figure 7-1. Concentrations of butterfat, protein and lactose all fell markedly over the first 30 DIM, during which time there was a steep increase in daily yield (in both mature cows and first lactation heifers). FPR also rose, but had a more marked and earlier peak in lactation than the nadir for either butterfat or protein percentage.

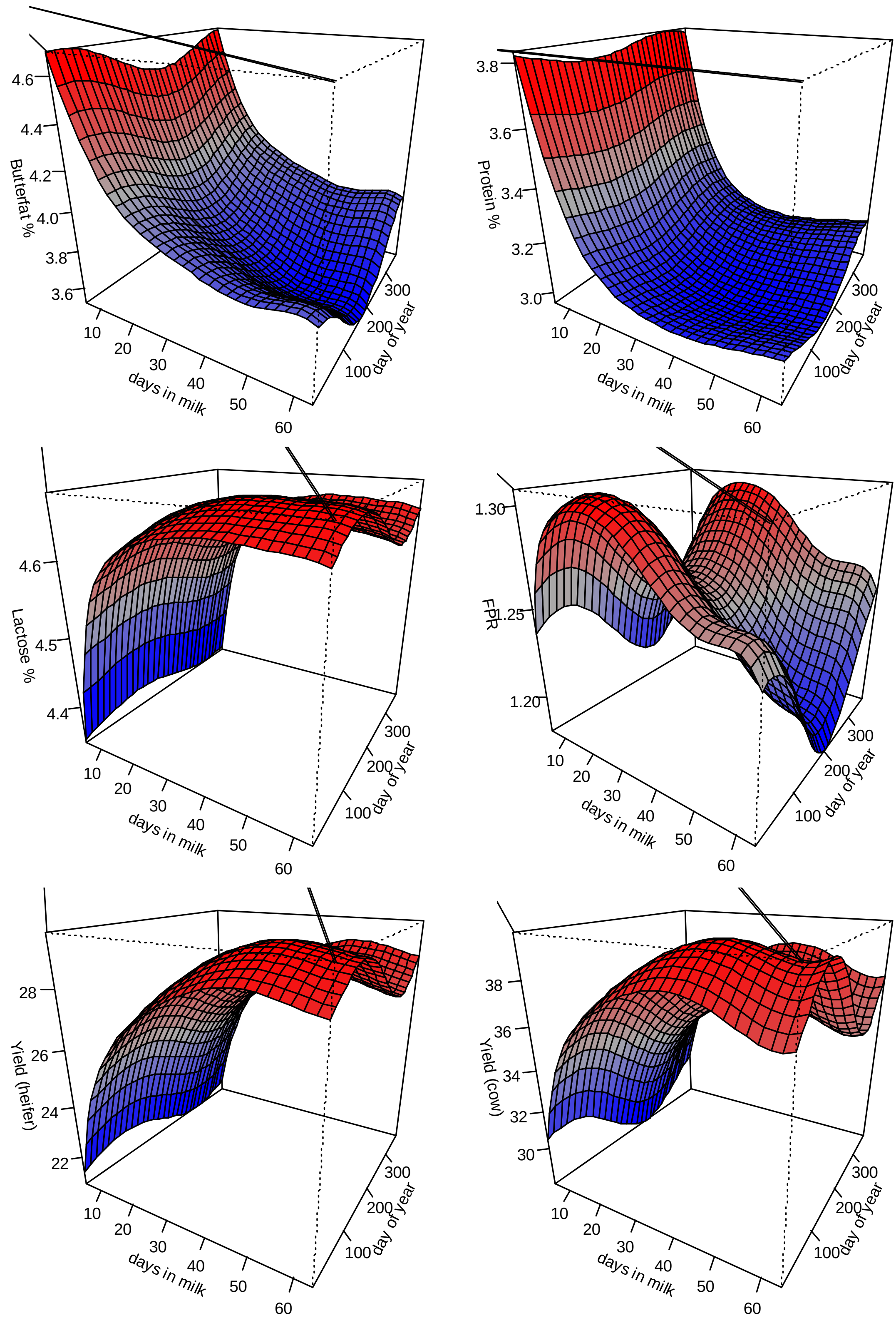

Figure 7-1 Regression planes demonstrating relationship between milk recording parameters and DIM/season
Vertical axes represent milk recording parameters, horizontal axes represent days in milk and day of year at the time of the recording event. Colours provide an additional representation of the vertical axis. Yields are in litres/day.

There was a marked seasonal pattern in daily yield, with an increase through winter to a peak in spring followed by a decrease through the summer months and a rise in autumn. There are converse trends in the concentrations of milk constituents, but this was most marked for butterfat percentage, which showed a major nadir in the spring. As a result, FPR was also heavily affected by season.

7.3.2 Logistic regression model

Terms and parameter estimates for the final model with outcome representing probability of a given serve leading to a pregnancy are shown in Table 7.2. Of the milk constituent terms tested (see Table 7.1), butterfat at first recording, protein at first and second recording and lactose at first and second recording had significant associations with the outcome. In the case of the latter three, there were interactions between the milk constituent term and DIM, such that the effect of the milk constituent varied with DIM at the time of the serve. Daily milk yield at both test days as well as lactation 305-day adjusted yield had significant associations with the outcome, and these relationships were complex, involving polynomial functions and interaction terms between the yield measures. Probability of a serve leading to a pregnancy decreased with subsequent years, and was lower in the summer months (similar to the effects described in Section 4.3.2). DIM also had a similar association with the outcome to that described in earlier models, with a steep increase through very early lactation followed by a relative levelling off around 70 DIM. In contrast to the findings in Chapter 4, in this model parity 1 was associated with the lowest pregnancy rate and parity 2 the highest. The interval to the preceding serve in the lactation (inter-service interval, ISI) also had a significant association with the outcome, with an abnormally short ISI (<18 days) associated with an approximately 50% reduction in the odds of a serve leading to a pregnancy. A serve after a "normal" length ISI (18-24 days) was associated with very similar odds of pregnancy as the first serve of a lactation (the reference category), while longer ISIs were associated with decreased odds of pregnancy.

Table 7.2 Model terms and parameter estimates for a logistic regression model with outcome representing probability of a pregnancy resulting from a given serve

		HPD interval	
Model term	**Odds ratio**	**2.5%[21]**	**97.5%[22]**
BF[23] at recording 1[24]	0.98	0.97	1.00
Pr[25] at recording 1	1.05	1.03	1.06
Pr at recording 2[26]	1.36	1.19	1.56
Lct[27] at recording 1	1.96	1.62	2.27
Lct at recording 2	0.60	0.50	0.69
(Pr at recording 2).(lnDIM[28])	0.93	0.90	0.96
(Lct at recording 1).(lnDIM)	0.87	0.84	0.90
(Lct at recording 2).(lnDIM)	1.13	1.09	1.17
Daily yield at recording 1	1.17	1.14	1.19
Daily yield at recording 2	1.25	1.22	1.28
(Yield at recording 2)^2	1.03	1.02	1.05
(Yield at rec'g 1).(Yld at rec'g2)	1.05	1.03	1.06
(Yield at rec'g 1).((Yld at rec'g 2)^2)	0.97	0.97	0.98
305d Yield	0.75	0.74	0.76
(305d Yield)^2	0.96	0.96	0.96
(305d Yield)^3	1.01	1.01	1.00
(305d Yield).(Parity 1)	0.86	0.84	0.87
((305d Yield)^2).(Parity 1)	1.03	1.03	1.04
((305d Yield)^3).(Parity 1)	1.00	1.00	1.01
DIM[29]	1.11	1.11	1.12
(DIM)^2	1.00	1.00	1.00
(DIM)^3	1.00	1.00	1.00
Parity 1	Reference		
Parity 2	1.55	1.49	1.62
Parity 3	1.52	1.46	1.59
Parity 4	1.40	1.34	1.47
Parity 5	1.15	1.11	1.20
ISI[30] NA (first serve)	Reference		
ISI<18d	0.49	0.45	0.53
ISI 18-24d	1.04	1.01	1.08
ISI 25-35d	0.81	0.76	0.86
ISI 36-48d	0.90	0.84	0.97
ISI>48d	0.75	0.65	0.87
Year <2003	Reference		

[21] 2.5% HPD: lower bound of the 95% interval of highest posterior density (HPD)
[22] 97.5% HPD: upper bound of the 95% interval of highest posterior density (HPD)
[23] BF: standardised butterfat concentration
[24] Recording 1: first recording test day of lactation
[25] Pr: standardised protein concentration
[26] Recording 2: second recording test day of lactation
[27] Lct: standardised lactose concentration
[28] lnDIM: natural logarithm of days in milk
[29] DIM: days in milk
[30] ISI: interservice interval

Year 2003	0.98	0.93	1.02
Year 2004	0.93	0.89	0.98
Year 2005	0.88	0.84	0.92
Year 2006	0.82	0.78	0.86
Year 2007	0.81	0.77	0.85
Year 2008	0.79	0.73	0.84
Summer (Jun-Sept)	0.84	0.82	0.86

During the model building process, it was evident that either butterfat percentage or FPR at first recording would remain in the model without the other, but failed to meet inclusion criteria (i.e. showed no evidence of a significant association with the outcome) when both were included in the same model. This behaviour was suggestive of a strong correlation between these two variables. In order to investigate this, a high density scatterplot showing butterfat percentage at first recording versus FPR at first recording was produced (Figure 7-2, the equivalent for protein percentage is included for comparison). There is a very strong linear relationship between butterfat concentration and FPR, with almost 75% of variation in first recording FPR explained by butterfat alone (Pearson $r^2 = 0.746$).

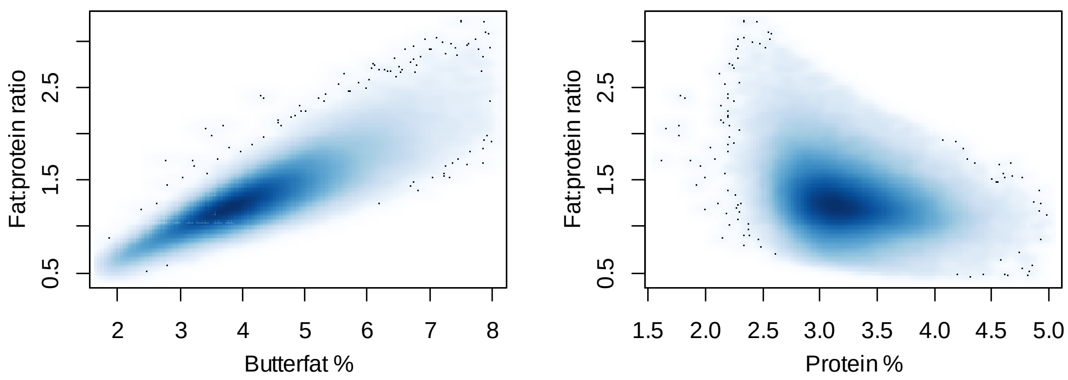

Figure 7-2 Association between fat:protein ratio and butterfat or protein percentage at first milk recording of lactation
High density scatterplots show point density by depth of colour (such that darker coloured regions represent areas with large numbers of data points overlying each other.

7.3.3 Assessment of model fit

Figure 7-3 shows observed and predicted (with 95% credible intervals) pregnancy rate across a variety of subsets of serves from the dataset. The observed overall pregnancy rate across all the serves was 40.0%, and the predicted overall rate 40.2% (95% credible interval 39.5% to 40.9%). In all cases, the observed pregnancy rate for that subset of serves fell within the 95% credible interval of the prediction, suggesting that model fit was acceptable.

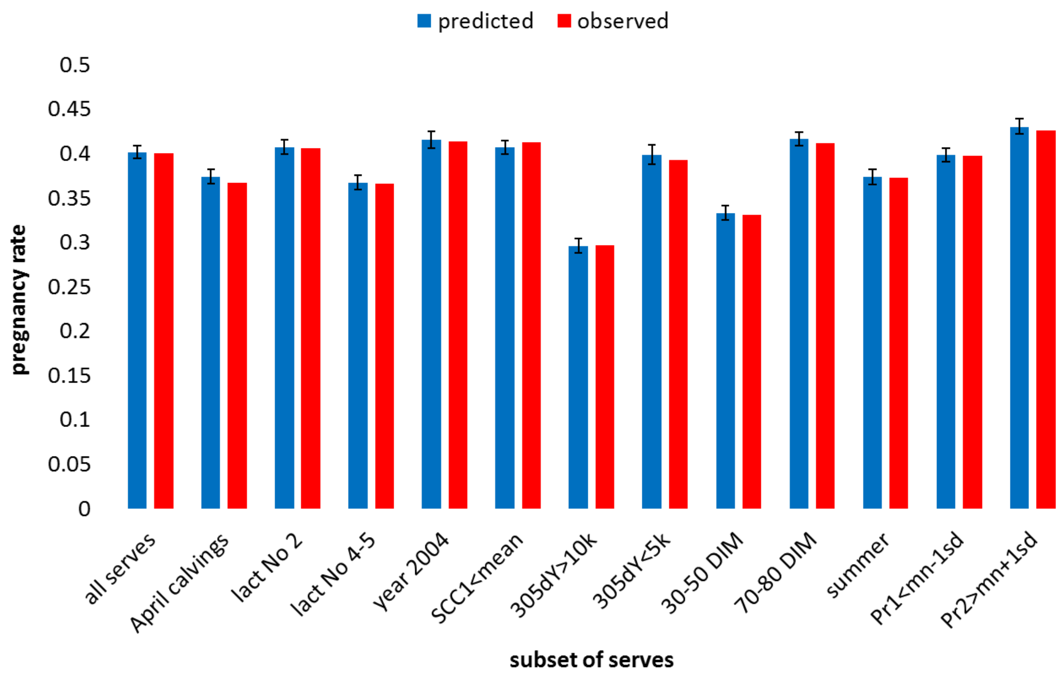

Figure 7-3 Predicted and observed pregnancy rate across a variety of subsets of serve events

April calvings: serves from lactations beginning with a calving in April; lact No: lactation number; SCC1<mean: serves from lactations where SCC at first test day was below population mean; 305dY>10k/305dY<5k: serves from lactations where the 305-day adjusted yield was greater than 10,000 litres and less than 5,000 litres respectively; 30-50DIM/70-80DIM: serves at 30-50 and 70-80 days in milk respectively, summer: serves during the months June – September; Pr1<mn-1sd: serves from lactations where corrected protein percentage at first test day was less than population mean minus 1 standard deviation; Pr2>mn+1sd: serves from lactations where corrected protein percentage at second test day was greater than population mean plus 1 standard deviation.

7.3.4 Forward predictions to illustrate model results

Figure 7-4 and Figure 7-5 show the results of forward predictions from the model using illustrative example scenarios. In Figure 7-4, each plot shows predictions to illustrate the association between pregnancy rate and a single milk constituent variable, with all other predictors fixed at their population mean values. For example, the top left plot shows how pregnancy rate would be predicted to vary as corrected lactose at first test day changes over the range -2 to 2 (i.e. two standard deviations either side of the population mean). The line colours represent predictions for serves at different DIM (to represent the interaction between the variables representing lactose at first test day and DIM). In this example, it is clear that higher lactose values at first test day are associated with a higher probability of pregnancy for a serve at 50DIM (with a predicted increase of around 8% over the full range of lactose concentrations), that this effect is still present but is reduced in magnitude for a serve at 75 DIM and that there is no relationship between lactose at first test day and probability of pregnancy to a serve at 100 DIM. The magnitude of the relationship between lactose at first test day and predicted probability of pregnancy for a serve at 50 DIM is the largest of all the scenarios explored (with the red line in the top left chart showing the steepest gradient). Generally, changes in milk constituent predictor variables from -2 to 2 (i.e. very large changes in constituent concentration) would be predicted to change pregnancy rate by a very small amount (mostly less than 5%).

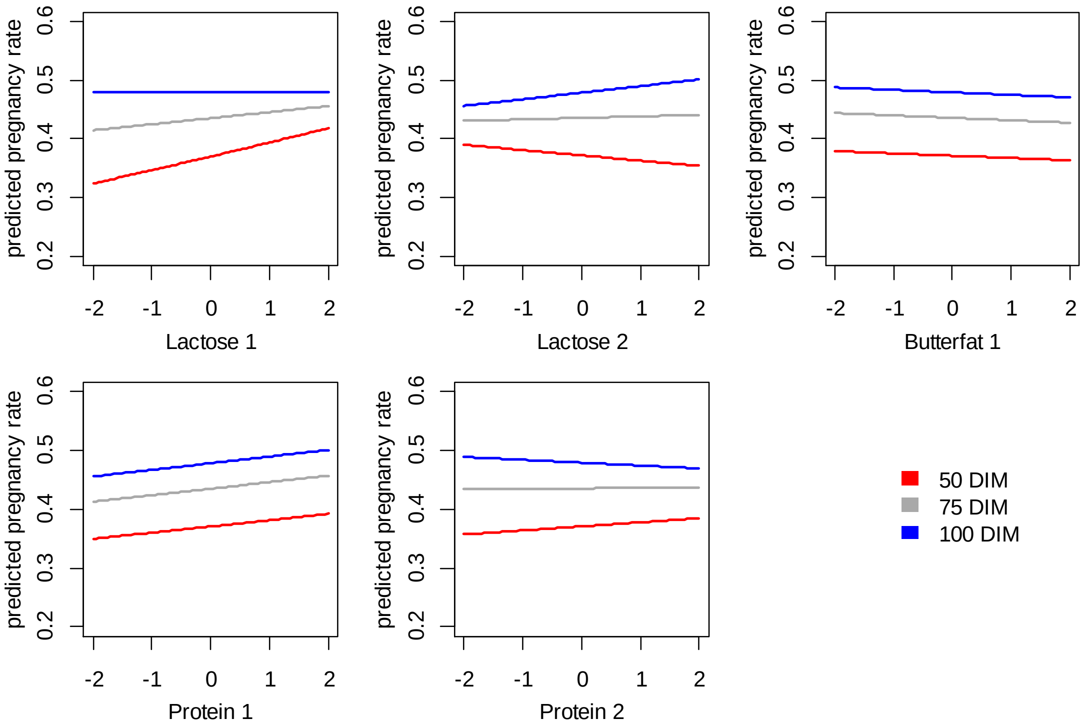

Figure 7-4 Forward predictions for example scenarios to illustrate relationships between pregnancy rate and early lactation milk constituents

Each plot shows predicted probability of pregnancy for a set of example serves 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 how the predicted probability varies across the range of that variable, and the line colours represent serves given at different stages of lactation. Numeric suffixes on variables refer to test days within lactation (e.g. Lactose 1 refers to corrected lactose concentration at the first test day of lactation).

Protein and butterfat concentrations at first test day had no interaction with DIM, so the plots for these variables show parallel lines. Predicted probability of pregnancy to a given serve falls with increasing butterfat concentration, but increases with increasing protein concentration. Protein and lactose concentration at second test day show a different pattern again; in each case the relationship observed at 50 DIM (probability of pregnancy increasing with increasing protein, and decreasing as lactose increases) is absent at 75 DIM (with grey lines on both plots relatively flat) and the direction of the relationship reversed at 100 DIM.

Figure 7-5 shows the relationships between pregnancy rate and the predictor variables relating to milk yield in a similar fashion. The left hand two plots illustrate the same relationship between daily yield at first and second test day and the outcome. Generally, predicted pregnancy rate increased with increasing daily yield at either first or second test day. The effect of yield at either test day was smaller where yield at the other test day was low (for example, the middle plot demonstrates that where yield at second test day is between -2 and -1, there is very little difference in predicted pregnancy rate when yield at first test day is set to -1 or 1).

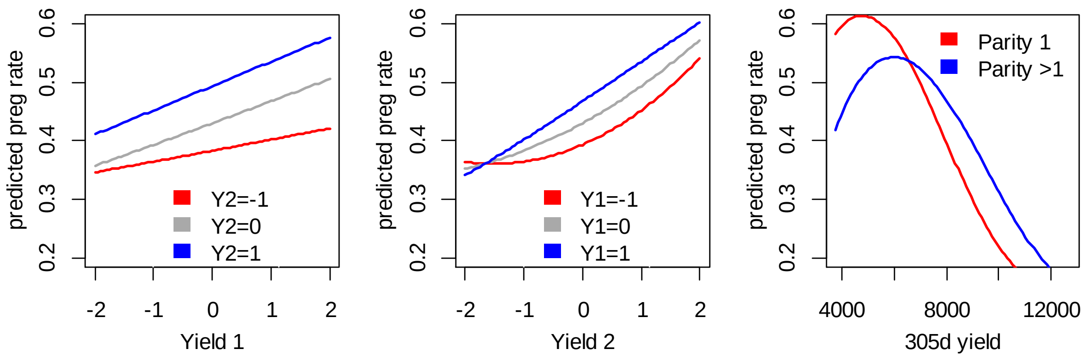

Figure 7-5 Forward predictions for example scenarios to illustrate relationships between pregnancy rate and milk yield

Each plot shows predicted probability of pregnancy for an example serve where all predictor values are set at their population means except for the variables indicated in the x-axis of the plot and in the plot legend. Each line shows how the predicted probability varies across the range of the variable indicated on the x-axis, for fixed values of the variable shown in the legend. Numeric suffixes on variables refer to test days within lactation (e.g. Yield 1 refers to corrected daily yield at the first test day of lactation). Y1: yield on second test day; Y2: yield on second test day.

There was a non-linear relationship between 305-day lactation yield and predicted pregnancy rate, with lower predicted pregnancy rates in lactations with very low yields, a peak in predicted pregnancy rate at around 5,000 litres for first lactation heifers ad 6,500 litres for multiparous cows, and a fall in predicted pregnancy rate as lactation yields were increased beyond this. Effect sizes for the yield variables are noticeably larger than for the milk constituent variables, with predicted pregnancy rate varying by more than 20% over the range of values for test day yields, and by over 30% over the range of values for 305-day lactation yield.

Figure 7-6 summarises distributions of full posterior predictions derived using alternate versions of the dataset, in order to illustrate population attributable risk of pregnancy for binary predictor variables. The red (left hand) bars illustrate the change in the population pregnancy rate (an increase of just under 1%) which would be expected if the effect of the summer months was removed. The blue (right hand) bars illustrate the change in pregnancy rate which would be expected if the effect of inter-service intervals of less than 18 days was removed (an increase of around 1.5%).

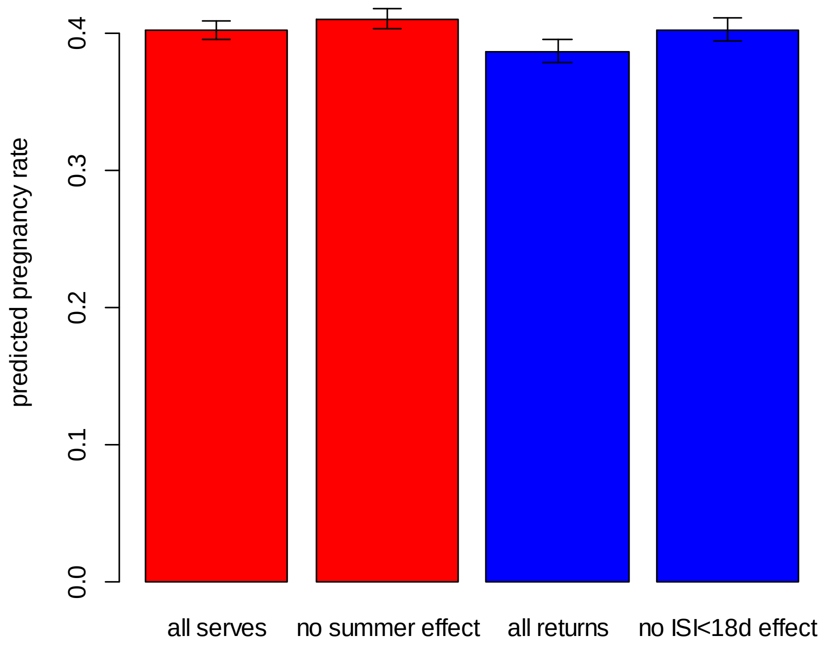

Figure 7-6 Attributable risk illustrations for season and inter-service interval
The red (left hand) pair of bars illustrates attributable risk for the effect of season, and the blue (right hand) pair the attributable risk for the effect of serves at short inter-service intervals (ISI). Each bar shows the predicted pregnancy rate across an alternate version of the dataset (with 95% credible intervals). The far left bar shows the prediction from the full, unedited dataset (the same prediction as the left hand bar in Figure 7-3), the next left shows a prediction from the same dataset when all cases have the "summer" indicator variable set to zero, the next bar a prediction from the unedited dataset excluding first serves, and the far right bar a prediction from this dataset when all cases have the indicator variable for an ISI<18d set to zero.

7.3.5 Variance partition using forward predictions

The proportion of variation in herd-year pregnancy rate explained by each element of the model is shown in Figure 7-7 and Figure 7-8. The full model (including herd- and cow-level random effects) explained almost 64% of variation, meaning that 36% of variation was unexplained and at the level of the individual serve. This figure was almost wholly unaltered by removing the cow-level random effect for the model, suggesting that there is negligible unexplained variability at cow level. Removing the herd-level random effect from the model reduced the variation explained very substantially, with the fixed effects model explaining just over 22% of the observed variation in herd-year pregnancy rate, implying that over 40% of the variation is unexplained by the fixed effects but occurs at herd level. This changed very little when predictor variables derived from milk constituents were not used to generate predictions (with these variables explaining around 0.5% of the observed variation). Finally, the model including only fixed effects for DIM, parity, season, ISI and year explained just over 15% of variation, suggesting that around 7% was explained by the predictor variables relating to milk yield (both 305-day lactation yield and daily yield at the first two test days of lactation).

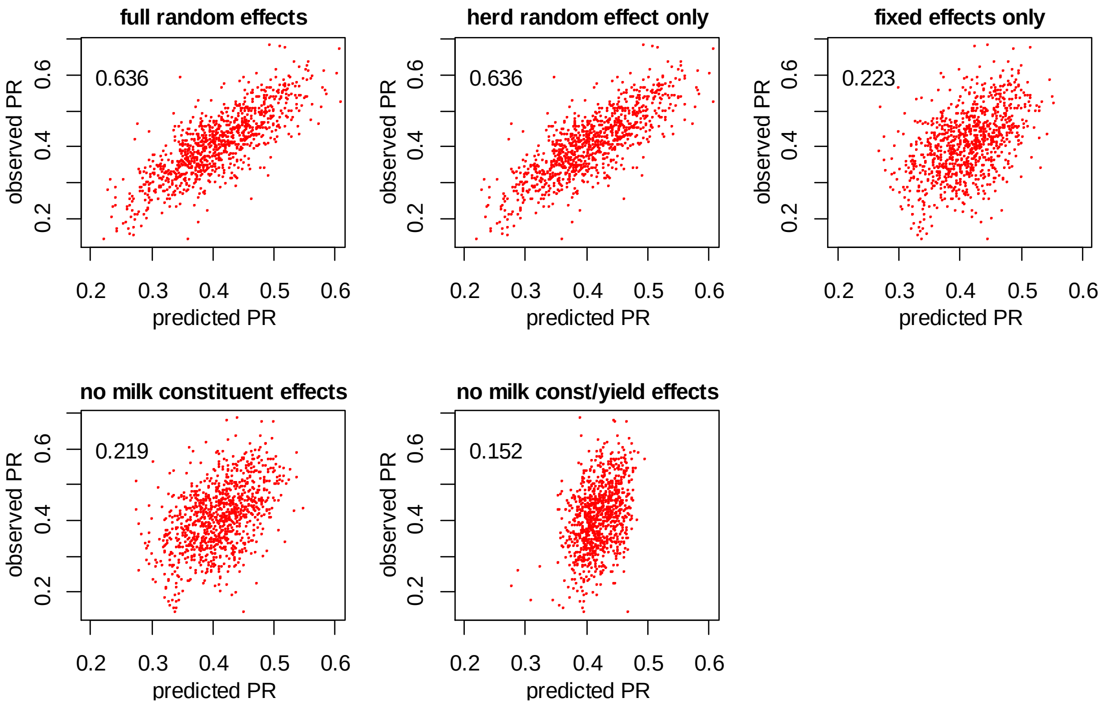

Figure 7-7 Predicted versus observed herd-year pregnancy rate using different model elements
Plot subtitles indicate which elements of the model were employed to generate the set of predictions shown. Text on the plots shows Pearson r^2 *values for each correlation.*

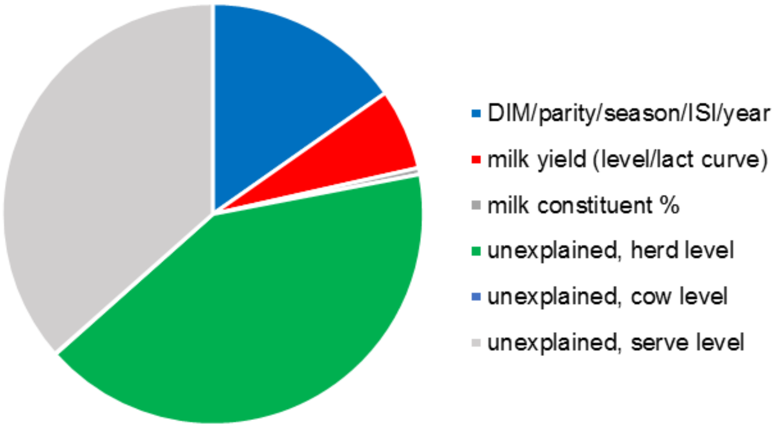

Figure 7-8 Proportion of observed variation in herd-year pregnancy rate explained by model elements

7.3.6 Additional models

Variables retained in the model and parameter estimates were broadly similar to those described above when the model was rebuilt using serves at 100-200 DIM, although the effect sizes for the variables reflecting milk recording information from first and second test days were generally smaller. This was especially true for daily yield at first and second test days; interestingly the variable representing 305-day lactation yield was not significantly associated with the outcome for this model. Again, results were broadly similar when only serves from lactations where the first test day occurred at 15 or fewer DIM. Here, the only notable change was a slight increase in the effect size of the variable representing butterfat at first test day; although this effect was still amongst the smaller of those from the milk recording variables. Expanding the dataset into 21 day periods to build a discrete time survival model with the outcome of a cow becoming pregnant in a given 21 day period resulted in a model broadly similar to that described by Madouasse et al. (2010a), with only relatively minor differences in parameter estimates and interaction terms.

7.4 Discussion

The main aim of this study was to investigate relationships between milk recording data from early lactation and the probability of a given serve leading to a pregnancy. Although a number of significant relationships between fertility and milk constituent percentages were revealed, forward predictions demonstrated that the size of these relationships was generally small, and that they collectively accounted for only a very small proportion in the observed variation in pregnancy rate across herd-years. This is broadly in agreement with the only other sizeable study based on UK data (Madouasse, 2009; Madouasse et al., 2010a), where the model predicted less than 0.1% of the variation in herd reproductive performance in an external cross-validation dataset. However, a number of earlier studies from the USA (Kristula et al., 1995), the Netherlands (Heuer et al., 1999), Slovenia (Podpečan et al., 2008) found apparently stronger relationships between reproduction and milk constituent concentrations. As these studies examined the relationship at cow, rather than herd level, it is plausible that other factors affecting herd-level reproduction have a much greater impact, introducing a large amount of additional between-herd variation.

It is also plausible that management systems in the UK tend to introduce more variability in milk constituents (as, for example, summer grazing may be more common and is typically associated with large changes in milk butterfat and protein percentage, Figure 7-1), and that the additional variation in constituents introduced by these practices is large enough to mask any relationship with reproduction. This is the main reason that this study and that of Madouasse et al. (2010a) corrected the milk recording parameters for DIM and season at the test day, but it is possible that there are additional factors confounding this relationship which were not accounted for. It is also notable that the previous studies have used data from much smaller numbers of herds (at most 22) than either of the more recent UK studies. The small amount of variability in herd-year pregnancy rate accounted for by the milk constituent variables (Figure 7-7) suggests that either these variables represent early lactation energy balance very poorly in UK herds, and/or that early lactation energy balance is very weakly related to pregnancy rate. The weight of evidence for the latter relationship is extremely strong (Butler, 2001; Lucy, 2001; Roche, 2006), and the former explanation seems more likely and is supported by other recent UK work (Cooper, 2011).

The shapes of these relationships were broadly as expected, with probability of pregnancy generally increasing with increasing protein or lactose concentration and decreasing with increasing butterfat concentration; this is in accordance with previous work (Kristula et al., 1995; Madouasse et al., 2010a; Podpečan et al., 2008). There was some departure from the directions of these relationships for lactose at second test day for earlier serves and for second test day protein and later serves (Figure 7-4). Neither of these effects is readily explicable; the relationship between constituents at second test and pregnancy rate appears to be more variable with DIM at serve than those between constituents at first test day and pregnancy rate. Given that very few serves will occur close in time to the first test day (as this would commonly be within the voluntary waiting period), this is the reverse of what would be expected. However, the magnitude of effect size for these "reversed" relationships is very small, with only a few percent difference in predicted pregnancy rate over the full range of illustrated milk constituent concentrations (-2 to 2 standard deviations from the mean). It is possible that exploration of further ways to represent the interaction between these terms and DIM would have resulted in different results, but this is highly unlikely to have made a major difference to interpretation of the model as a whole.

The variables relating to milk yield showed relationships of greater magnitude with predicted pregnancy rate. The inclusion of variables representing daily yield at first and second test day (corrected for DIM and seasonal effect) along with a variable representing 305-day lactation yield allows associations between pregnancy rate and both the overall level of production and the shape of the lactation curve through early lactation to be explored. The 305-day yield had a strong and broadly negative relationship with predicted pregnancy rate; this is in accordance with a large body of existing work (Lucy, 2001; Nebel and McGilliard, 1993; Roche, 2006). However, in this case the relationship was non-linear, with pregnancy rates predicted to be lower in lactations with very low 305-day yields (Figure 7-5). This has not previously been demonstrated, but may reflect a very low lactation yield acting as an indicator of a "problem" lactation; for example where a high-consequence disease event (such as a left displaced abomasum) occurred. The size of this relationship was relatively large, with predicted pregnancy rate ranging from just over 20% to just over 50% over the range of illustrated yields for a multiparous cow, and a slightly greater range for a primiparous heifer. As would be expected, the peak in predicted pregnancy rate occurred at a lower 305-day yield in heifers compared to cows.

Daily yield variables for both first and second test day had a positive relationship with predicted pregnancy rate such that, for a given level of 305-day lactation yield, predicted pregnancy rate was higher where yield rose more quickly at the beginning of lactation. Again, this has not previously been demonstrated, but could plausibly represent the impact of a successful transition period (encompassing the end of the dry period, calving and the beginning of lactation) on the rate at which yield rises towards peak lactation. Good management of this period has long been recognised as a key determinant of performance through lactation (Drackley, 1999; Jorritsma et al., 2003). This could represent an indirect link to energy balance, whereby improved energy balance in early lactation could lead to an increased speed of yield increase as well as improved pregnancy rate. First test day yield compared to expected value has been used as an indicator of transition period success in the USA (Nordlund and Cook, 2004).

The categorical predictor variables included in the model also revealed some significant relationships with predicted pregnancy rate. The association between parity and reproduction was different in this model to that found in previous chapters (Sections 4.3 and 6.3.1); with a reduction in pregnancy rate associated with parity 1. This is likely to reflect the more sophisticated representation of milk yield used in this study; in the previous chapters the effect of 305-day yield was represented as a single linear relationship which applied to cows and heifers and it is plausible that the separate cubic functions used for cows and heifers in this model represent the data better. This suggests that, after accounting appropriately for differences in level of production, heifers tend to show slightly lower pregnancy rates than older cows. This is biologically highly plausible, as heifers are often lower in the herd dominance hierarchy and typically have limited access to any potentially restricted resources (for example, feed space), but the clear potential for lower yields in heifers to confound this relationship may explain why this is not widely recognised.

As in previous chapters, reproductive performance declined over the period studied, and was poorer through the summer months. Again, there are a number of reasons why the summer may be associated with poorer pregnancy rates; most notably decreased control of nutrition when cows are at grass in grazing herds, and the increased potential for heat stress where cows are housed through the summer. However, the attributable risk illustration for the summer variable suggests that only a very limited gain in population pregnancy rate would be achieved even if it were possible to remove this seasonal effect. Although there was a large association between short ISI and decreased probability of pregnancy, the low relative frequency of such intervals meant that only a modest gain in population pregnancy rate would be expected if this effect was eradicated. ISI of less than 18 days (i.e. less than the normal length of an oestrous cycle) are typically due to insemination when a cow is not in oestrus, so a decreased pregnancy rate to such serves is expected. It is important to remember that the "inaccurate" serve may be the one beginning, rather than ending the interval, so it would be expected that around half of such serves are to cows outside of oestrus. Prolonged ISIs were associated with a much smaller decrease in the probability of pregnancy; plausibly this could be related to a reduction in fertility in the heat after embryonic or foetal loss.

The model was substantially more predictive of herd-year pregnancy rate than that of Madouasse (2009) was of herd calving to conception interval. This supports the theory that pregnancy rate is the element of reproductive performance most influenced by factors which could also affect early lactation milk recording data (such as energy balance and transition management). This study also evaluated a wider range of potential predictors, so would be expected to explain more variation. However, just under a quarter of overall variation in herd-year pregnancy rate was explained by the fixed effects included in the model (Figure 7-8). The remaining unexplained variation was divided relatively equally between herd level and serve level variation, with very little of the variation explained by the cow-level random effect. This suggests that unmeasured factors acting at either the herd level (for example, management of energy balance and dry matter intake, insemination technique and herd infectious disease status) or at the level of the individual serve (such as daily climatic conditions and state of the cow's reproductive tract at the time of serve, as well as chance variation) are highly important in determining pregnancy rate, but that there is very little variation in intrinsic fertility from cow to cow. This is particularly interesting in view of the focus on improving the genetics of fertility in UK dairy cows during the past decade (Wall et al., 2003); suggesting that this may not lead to major gains in overall performance. Further investigation of herd-level factors not studied here which could explain some of this variability would seem potentially fruitful.

This study demonstrates that early lactation milk constituent concentrations have very small (although statistically significant) associations with pregnancy rate in UK dairy herds, even when corrected for DIM and season at test day. This in turn is strongly suggestive that such measures do not reflect early lactation energy balance to a clinically useful degree. This is an important finding, as use of metrics such as FPR at first test day is extremely common in the

UK dairy industry. Figure 7-1 provides strong evidence that this is not appropriate, as it is clear that there is very substantial variation in the expected or “average” FPR through the year as well as with DIM at test day, both of which will vary randomly for a given cow. Here, FPR at first test day had no significant association with pregnancy rate even when corrected for DIM and season, although it is important to bear in mind the very strong correlation between FPR and butterfat percentage (Figure 7-2), which was included in the final model but had a small magnitude of effect. It is possible that better results would be expected if correction for DIM and season at test day was carried out at herd level (so that historic data from the individual herd were used to generate expected milk recording values, rather than the full dataset), as this would reflect variation in management practices between herds (such as whether cows are grazed or housed through the summer). However, the relationships revealed in this study were so small that it is unlikely that correcting at herd level would make enough difference to develop a clinically useful proxy for energy balance.

Chapter 8 General discussion and conclusions

8.1 Discussion

This research used a large and heterogeneous dataset from dairy herds across England and Wales to assess the level of reproductive performance and explore factors associated with fertility. There is generally a balance to be struck in epidemiological studies between quantity and quality of data, and Chapter 2 describes the development and application of novel measures of data recording quality. This revealed that, even in herds considered by their veterinary surgeon to have good quality records, there is considerable variability in level of recording, and many herds failed to reach the threshold level for at least one measure of quality. Chapter 3 showed that reproductive performance in these herds was generally in decline over the first half of the 2000s, but also provided some evidence that improvements in submission rates were beginning to reverse this trend.

Significant and sizeable associations between occurrence of a case of clinical mastitis (CM) or an elevated somatic cell count (SCC) and reproductive outcomes were demonstrated in Chapter 4, but the development and use of a stochastic simulation model for probabilistic sensitivity analysis (PSA) described in Chapter 5 showed that a herd's udder health status was highly unlikely to influence its overall level of fertility performance under plausible conditions. Chapter 6 provided similar findings in the case of clinical lameness events: again sizeable and significant effects on reproduction were demonstrated at the level of a unit of time within lactation, but simulation work demonstrated that herd lameness incidence was highly unlikely to have a clinically relevant impact on herd reproductive performance under typical conditions. Chapter 7 explored factors affecting a specific element of the reproductive process, and demonstrated that relatively little of the variation in a herd's pregnancy rate is explainable with routinely recorded management and milk recording data, while a large amount of the unexplained variation is at herd rather than cow level (suggesting that herd rather than individual cow factors are key drivers).

The link between the results of the statistical modelling based on the collected data (Chapter 4 and Section 6.3.1) and those derived from use of simulation models within a PSA framework (Chapter 5 and Section 6.3.2) is critical to understanding of this project as a whole. It is important to remember that the simulation models used for PSA are explicitly based on the results of the statistical models derived from these data. For example, the simulation model described in Chapter 5 uses the statistical model from Chapter 4 to predict the outcome for each simulated unit in time in each lactation. The PSA demonstrated that, although CM or an elevated SCC may be important in the probability of pregnancy at a particular point in time, or as the result of a particular serve, it is highly unlikely that the herd's level of CM or prevalence of high SCC cows will influence its overall reproductive performance. This suggests that the frequency and timing of CM cases and elevations of SCC typically observed in these herds are such that the significant effects revealed by the statistical model are massively outweighed by the other influences on a herd's reproductive performance. In both simulation models (Chapter 5 and Section 6.3.2) the largest drivers of overall fertility were the herd's marginal submission and pregnancy rate, reflecting the variation in these aspects of their performance after accounting for variation explained by the statistical model. Regression analysis of the results of the simulations suggested that there is potentially a large amount of "room for investment" in these areas, with Figure 5-4 and Figure 6-5 showing substantial likely economic gains from moderate improvements in these inputs.

Employment of this technique to help interpret and contextualise the results of complex statistical analysis proved to be extremely useful in this study. As discussed in the individual chapters, the statistical model elements of this study often found similar degrees of association between endemic disease events and reproduction to those already reported in the literature. However, taking the additional step of building these into a simulation model and using PSA allowed these results to be seen in the context that is relevant to decisions being made in the field. Without the additional simulation work, it is easy to see how the results from the statistical models (along with pre-existing evidence) could lead to an over-estimation of the importance of endemic disease in determining fertility performance. As dairy farmers and their advisors must take decisions about preventive management at herd level, an understanding of the likely importance of the herd's disease incidence in determining reproductive performance represents a major benefit. Use of stochastic modelling explicitly as a way to illustrate and contextualise results from a statistical model is currently extremely rare in the veterinary literature, although there are an increasing number of wider examples of use to represent uncertainty in complex systems (Heller et al., 2011; Hockey and Morton, 2010; Hutchinson et al., 2013; Lu et al., 2013). As demonstrated in this thesis, such a deficiency could potentially lead to over-estimation of the relative importance of research findings in some cases.

Chapter 7 demonstrated an alternative approach to interpreting the results of a complex statistical model. Here, the model was used to make predictions over a range of example scenarios, and these predictions used to illustrate the model results graphically. This was particularly useful in this part of the study, as the model under consideration had a large number of continuous predictor variables, many of which had non-linear associations with the outcome variable and where there were many interaction terms. This is in contrast with the models described in Chapter 4 and Chapter 6, where the key predictor variables of interest were binary or categorical, such that results could be simply presented as predicted relative risks. The use of predictions to illustrate results from these types of model represents a major step towards increasing the potential impact of such research on clinical practice. Although complex multivariate multilevel statistical analysis often improve the robustness and generalizability of results, it can make results much harder to interpret, and illustrating results using predictions represents a way to capitalise on these advantages without confusing potential end-users of the research, who may not have detailed statistical knowledge.

The large number of herds involved in this project represents a major strength. The total number of herds submitted (468) represents almost 5% of the herds in England and Wales (DairyCo, 2013). Whilst a large proportion of these datasets did not contain sufficiently robust data on disease events to contribute to Chapter 4 or Chapter 6, the sample sizes for these chapters are still very large compared to the majority of the pre-existing work in those fields, and sample sizes for the other chapters are much larger still. There have been previous publications using large datasets for similar research (Bruun et al., 2002; de Vries and Risco, 2005; Madouasse et al., 2010a, 2010b; Sogstad et al., 2006), but these have often been derived from data held in a central database (such as the Dairy Herd Improvement Association database in the USA and the Norwegian Dairy Herd Recording System). The data in this study came from a variety of sources, largely because it was considered unlikely that there was no central database in the UK which would provide the completeness and detail of data recording required for this work. Milk recording organisations (MRO) are the only likely sources of UK data on this scale, but the format used to store these datasets restricts the events recorded and the level of detail, and experience in the field suggests that a very small proportion of herds record clinical event and fertility data reliably through their MRO. To the author's knowledge, there have been no previous descriptions in the veterinary literature of studies on this scale using data recorded in multiple formats including those captured at source using on-farm management software. This concept has substantial overlap with the ideas behind "big data" described in the Introduction, and several of the techniques used to collate and process the heterogeneous herd datasets are based on those developed to deal with industrial big data.

8.2 Study limitations

Whilst a substantial effort was made to ensure that only data of an appropriate level of quality was included in each analysis, the possibility exists that some data from poorly recorded sources could have been used. It is likely that this problem would have been greatest with the work using clinical lameness event recording (Chapter 6). Records of lameness treatment events have a number of problematic features which do not apply significantly to mastitis or milk recording data. Firstly, there is a large degree of subjectivity in the degree of lameness resulting in a treatment: some farmers will detect and treat cows as early as possible whilst others may tend to treat cows only when they reach a more severe degree of lameness. The former will clearly tend to record higher numbers of cases, but these are likely to be milder in nature. This problem can be overcome to a great extent with the additional integration of regular mobility scoring data, and this has commonly been employed in smaller studies evaluating the relationship between lameness and reproduction. However, there is an inherent degree of subjectivity in mobility scoring itself, and the requirement for this to be done in a consistent fashion has tended mean that this has to be done by the researchers (rather than using scores recorded by farmers), substantially limiting the number of herds which can be included in such studies. Regular herd mobility scoring has received substantial promotion within the UK dairy industry over the last few years, so it is plausible that it is becoming much more widespread amongst UK dairy farmers (not least because this is a common requirement of farm assurance schemes for producers supplying supermarkets). It is likely that there are now large numbers of herds with mobility scoring data covering a substantial period of time. The UK industry has adopted a very simple mobility score system, with the aspiration that this should be more repeatable between observers, but despite this there is a large potential for variability in the reliability of mobility scores recorded by farm owners, managers or staff (Schlageter-Tello et al., 2014). Temporal allocation of lameness events is the other major problem: because many causes of lameness in dairy cows are characterised by a relatively insidious and gradual onset, it is very difficult to assign a point at which the lameness case truly occurred, and again when using clinical treatment records this will be affected by between-farm variation in sensitivity of lameness detection and threshold for treatment. If anything, these effects would tend to artificially inflate the apparent effect of lameness on reproductive performance seen in Chapter 6, as only the more severe cases will be treated and recorded.

Another potential criticism of the data is that a convenience sample of herds was used instead of sampling using a true probabilistic method. Requesting data from herds considered to have good quality records clearly applies a selection bias to the study, and this makes it more difficult to generalise the results of this work to the wider population of dairy cows. This is only a problem if it is considered likely that there are systematic differences between well recorded and poorly recorded herds which could affect the relationships studied. For example, it is credible that larger or more intensively managed herds will have better record keeping. It is also plausible that such herds may have a different level of reproductive performance, or different incidence rates of clinical disease compared to the wider population. This would have an impact on the results presented in Chapter 3 (levels of reproductive performance across the herds), as it would be very difficult to draw conclusions about the standard of fertility across all UK herds from the results obtained from this sample (as discussed in Section 3.4).

However, the work presented in the other chapters would only be affected by this sampling bias if better-recorded herds were likely to show different relationships between reproduction and the potential factors affecting it. For example, it is quite plausible that these herds may have had higher incidence rates of CM (for example, the risk of CM has been shown to increase with level of production (Windig et al., 2005)). It is much less likely, however, that these herds will demonstrate a different relationship between reproduction and mastitis. A potential biological pathway for this could exist (for example, if higher yielding herds had a very different pattern of likely causal pathogens for mastitis, and if there was a pathogen effect on the relationship between CM and fertility), but this is much less likely. The statistical techniques employed in these analyses will also reduce the potential impact of this effect. Most notably, the investigation of herd-level random effects for variables representing clinical events, allowed the effect of an event on reproduction to vary between herds: if variations in pathogen pattern had a major impact on this relationship and if this varied substantially between herds then it would be accounted for in the model.

Assessment of the characteristics of these herds presented in the histograms in Figure 3-4 suggest that there is some difference between the sizes of the herds studied and those of the wider UK population. For example, the median herd size in 2007 was 158 amongst the sample of herds studied, whilst in the next year it was 112 across all UK herds (DairyCo, 2013). However, evaluation of the distribution of herd sizes from the sample shows that there are a large number of smaller herds within this sample, and the difference between the medians is not large. A similar but relatively smaller pattern is seen when comparing level of production, although this is more difficult to do rigorously because of the difficulties of estimating milk yields from herds with poor records.

A potential weakness of the stochastic modelling element of the project is the potential for the selected distributions of the simulation input variables to influence the outcome of the work. This is not an unreasonable concern; for although the selected input distributions were all uniform and jointly independent (such that no assumptions were made regarding the most likely value or combination of values within the specified range), the ranges of these distributions were chosen using the clinical experience of the author and colleagues in the field. The possibility that this would have a clinically important impact on the results of this analysis was explored in detail in Chapter 5, where analysis was repeated using input distributions, including joint multivariate distributions based on the observed values in the dataset. Although this had a noticeable but slight impact on the outcome of the univariate analyses of the simulation results (by influencing which parts of the input parameter space which were explored), it had negligible effect on the multivariate analysis (such as that illustrated in Figure 5-4 and Figure 6-5). This provides confidence that the key messages of these elements of the study are not likely to have been influenced by arbitrary choice of input distributions, but would hold under possible alternatives.

8.3 Potential future work

The work described in this thesis has produced two key legacies which will serve as the foundation for future research in this area. Firstly, the development of a flexible agri-informatics platform allowing automation of the process of data quality analysis, application of quality criteria and restructuring for research use has wide potential for future use. The ability to update analyses with new data quickly and easily provides a potential benefit, allowing analyses to be repeated and updated over time. Automation of data collection from source would greatly facilitate this, and remove the remaining labour intensive element of the process. This could be achieved, for example, through work in collaboration with providers of on-farm software, to develop a mechanism whereby consenting farms' data is automatically anonymised and submitted on a regular basis. This work is already underway with one software provider, and MROs provide an additional and more straightforward avenue for this. If regular automated data collection could be achieved, this would also allow an element of forecasting to be included (for example, using Bayesian updating, as is common in meteorology (Gouweleeuw et al., 2005; Raftery et al., 2005)). Continuing to use and develop the principles and techniques from the industrial big data revolution will allow maximum value to be derived from this routinely recorded data, and ensure that its huge resource potential is harnessed and used for impactful and robust research.

The second key legacy of the project is the development of a simulation framework representing dairy herd reproduction, and experience of its use for PSA. This structure has potential for a wide variety of additional and novel uses. In particular, there is a major need in the dairy industry for a better understanding of the associations between reproductive performance and profitability. This is critical to on-farm decision makers and advisors, as it informs the likely scope for investment in fertility performance to prove profitable. However, the approach to assigning a cost to fertility performance which is commonly used in the UK is based on various modifications of the FERTEX method, developed during the 1990s (Esslemont, 2003), as described in Chapter 3 (and used to combine calving index and failure to conceive rate to a single overall measure in Chapter 5 and Chapter 6). The basis of this calculation is extremely simple, and there is substantial debate over issues such as the degree to which yield and lactation curve shape may influence the cost of poor reproduction.

There is therefore a clear need for a better understanding of this area, and a simulation based approach would seem to offer a logical route to this. An additional advantage of using stochastic simulation is the potential to develop decision support tools which use simulation (either in part or wholly) to apply research results to a given real-life situation. Another major advantage is the possibility to integrate research from a wide variety of sources, including the potential effect of uncertainty in research findings. In the medium term, development of a decision support tool integrating as much existing knowledge as possible to predict the likely impact of given interventions on a herd's fertility (given the characteristics and current performance of the herd), represents a key goal. The outputs from this tool could also include production and profitability measures, which could be presented in a truly probabilistic way (for example, as a distribution of predicted cost benefit for a given intervention). This would allow managers and advisors to make evidence-based decisions, informed by the best existing knowledge and accounting for the attitude to risk of the potential investor. Such a tool would have potential for massive impact in the UK dairy industry and abroad.

8.4 Conclusions

This project has added to existing knowledge regarding factors affecting reproductive performance in dairy herds. Specifically, the relationship between reproduction and the most common endemic diseases has been evaluated using robust and sophisticated techniques, and a simulation-based approach adopted to illustrate and contextualise results. Despite significant associations between risk of pregnancy and lameness, clinical and subclinical mastitis, the influence of these disease on a herd's overall reproductive performance is likely to be very slight. Exploration of factors affecting pregnancy rate has reinforced the evidence that early lactation milk constituent concentrations have no clinically significant association with reproduction in UK herds, as well as revealing that between-herd variability appears much greater than that between cows.

In addition, the project has provided the building blocks for future research in this area. Development of a platform to automate the process of data quality assessment, collation and restructuring will greatly facilitate future studies. Similarly, the simulation model framework developed represents a highly useful tool for exploring dairy herd fertility, optimising use of existing knowledge and providing evidence-based decision support to farmers and their advisors.

The challenge facing the UK and global dairy industries is to increase production whilst minimising resource use and environmental impact, in order to promote national and global food security. Optimal management of reproduction is a key requirement if dairy farming is to meet this challenge, allowing maximally efficient production and minimising the requirement for replacements.

References

Ahmadzadeh, A., Frago, F., Shafii, B., Dalton, J.C., Price, W.J., McGuire, M.A., 2009. Effect of clinical mastitis and other diseases on reproductive performance of Holstein cows. Anim. Reprod. Sci. 112, 273–282.

Alawneh, J.I., Laven, R.A., Stevenson, M.A., 2011. The effect of lameness on the fertility of dairy cattle in a seasonally breeding pasture-based system. J. Dairy Sci. 94, 5487–5493.

Anderson, K., Jacobson, J.S., Heitjan, D.F., Zivin, J.G., Hershman, D., Neugut, A.I., Grann, V.R., 2006. Cost-effectiveness of preventive strategies for women with a BRCA1 or a BRCA2 mutation. Ann. Intern. Med. 144, 397–406.

Anderson, V.L., McLean, R.A., 1974. Design of Experiments: A Realistic Approach. CRC Press.

Andronis, L., Barton, P., Bryan, S., 2009. Sensitivity analysis in economic evaluation: an audit of NICE current practice and a review of its use and value in decision-making. Health Technol. Assess. 13, 1–45.

Audigé, L., Beckett, S., 1999. A quantitative assessment of the validity of animal-health surveys using stochastic modelling. Prev. Vet. Med. 38, 259–276.

Baeurle, S.A., 2009. Multiscale modeling of polymer materials using field-theoretic methodologies: a survey about recent developments. J. Math. Chem. 46, 363–426.

Bahonar, A.R., Azizzadeh, M., Stevenson, M.A., Vojgani, M., Mahmoudi, M., 2009. Factors affecting days open in Holstein dairy cattle in Khorasan Razavi province, Iran; A cox proportional hazard model. J. Anim. Vet. Adv. 8, 747–754.

Barker, A.R., Schrick, F.N., Lewis, M.J., Dowlen, H.H., Oliver, S.P., 1998. Influence of clinical mastitis during early lactation on reproductive performance of Jersey cows. J. Dairy Sci. 81, 1285–1290.

Barker, Z.E., Leach, K.A., Whay, H.R., Bell, N.J., Main, D.C.J., 2010. Assessment of lameness prevalence and associated risk factors in dairy herds in England and Wales. J. Dairy Sci. 93, 932–941.

Bellera, C.A., MacGrogan, G., Debled, M., de Lara, C.T., Brouste, V., Mathoulin-Pélissier, S., 2010. Variables with time-varying effects and the Cox model: some statistical concepts illustrated with a prognostic factor study in breast cancer. BMC Med. Res. Methodol. 10, 20.

Berends, I.M.G.A., Swart, W.A.J.M., Frankena, K., Muskens, J., Lam, T.J.G.M., van Schaik, G., 2008. The effect of becoming BVDV-free on fertility and udder health in Dutch dairy herds. Prev. Vet. Med. 84, 48–60.

Bicalho, R.C., Vokey, F., Erb, H.N., Guard, C.L., 2007. Visual locomotion scoring in the first seventy days in milk: Impact on pregnancy and survival. J. Dairy Sci. 90, 4586–4591.

Bland, J.M., Altman, D.G., 2000. The odds ratio. BMJ 320, 1468.

Boyd, D., Crawford, K., 2012. Critical Questions for Big Data. Inf. Commun. Soc. 15, 662–679.

Bradley, A.J., Leach, K.A., Breen, J.E., Green, L.E., Green, M.J., 2007. Survey of the incidence and aetiology of mastitis on dairy farms in England and Wales. Vet. Rec. 160, 253–258.

Breen, J.E., Hudson, C.D., Bradley, A.J., Green, M.J., 2009. Monitoring dairy herd fertility performance in the modern production animal practice. Cattle Pract. 17, 196–201.

Britt, J.S., Thomas, R.C., Speer, N.C., Hall, M.B., 2003. Efficiency of converting nutrient dry matter to milk in Holstein herds. J. Dairy Sci. 86, 3796–3801.

Browne, W.J., 1998. Applying MCMC methods to multilevel models (PhD Thesis). University of Bath, UK.

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

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

Brownlie, T., Morton, J., Heuer, C., Hunnam, J., McDougall, S., 2013. Reproductive performance of seasonal-calving, pasture-based dairy herds in four regions of New Zealand. N. Z. Vet. J. 62, 77–86.
Bruijnis, M.R.N., Hogeveen, H., Stassen, E.N., 2010. Assessing economic consequences of foot disorders in dairy cattle using a dynamic stochastic simulation model. J. Dairy Sci. 93, 2419–2432.
Bruun, J., Ersbøll, A.K., Alban, L., 2002. Risk factors for metritis in Danish dairy cows. Prev. Vet. Med. 54, 179–190.
Budnitz, R.J., Apostolakis, G., Boore, D.M., Cluff, L.S., Coppersmith, K.J., Cornell, C.A., Morris, P.A., 1998. Use of Technical Expert Panels: Applications to Probabilistic Seismic Hazard Analysis*. Risk Anal. 18, 463–469.
Butler, W.R., 2001. Nutritional effects on resumption of ovarian cyclicity and conception rate in postpartum dairy cows. BSAS Occas. Publ. 133–146.
Cabrera, V.E., 2012. A simple formulation and solution to the replacement problem: A practical tool to assess the economic cow value, the value of a new pregnancy, and the cost of a pregnancy loss. J. Dairy Sci. 95, 4683–4698.
Compton, C.W.R., McDougall, S., 2010. A longitudinal study of reproductive performance and management of 82 dairy herds in the Waikato region with differing policies on the routine use of induction of parturition. N. Z. Vet. J. 58, 175–183.
Cooke, R.M., Goossens, L.H.J., 2004. Expert judgement elicitation for risk assessments of critical infrastructures. J. Risk Res. 7, 643–656.
Cooper, R.L., 2011. An assessment of milk compositional indicators as a predictor for subclinical and clinical ketosis in UK dairy cattle (Diploma in Cattle Health and Production). Royal College of Veterinary Surgeons, London.
Cornou, C., Østergaard, S., Ancker, M.L., Nielsen, J., Kristensen, A.R., 2014. Dynamic monitoring of reproduction records for dairy cattle. Comput. Electron. Agric. 109, 191–194.
Coulon, J.B., Rémond, B., 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.
DairyCo, 2013. Dairy statistics: An insider's guide 2013.
DairyCo, 2014. Dairy market information [WWW Document]. URL http://www.dairyco.org.uk/market-information/#.UyCAhPl_uSk (accessed 3.12.14).
Davies, H.T.O., Crombie, I.K., Tavakoli, M., 1998. When can odds ratios mislead? BMJ 316, 989–991.
De Vos, C.J., Saatkamp, H.W., Nielen, M., Huirne, R.B.M., 2006. Sensitivity analysis to evaluate the impact of uncertain factors in a scenario tree model for classical swine fever introduction. Risk Anal. 26, 1311–1322.
De Vries, A., Risco, C.A., 2005. Trends and seasonality of reproductive performance in Florida and Georgia dairy herds from 1976 to 2002. J. Dairy Sci. 88, 3155–3165.
De Vries, A., Steenholdt, C., Risco, C.A., 2005. Pregnancy rates and milk production in natural service and artificially inseminated dairy herds in Florida and Georgia. J. Dairy Sci. 88, 948–956.
De Vries, M.J., Veerkamp, R.F., 2000. Energy balance of dairy cattle in relation to milk production variables and fertility. J. Dairy Sci. 83, 62–69.
DEFRA, 2013. Family Food 2012 - Publications - GOV.UK [WWW Document]. URL https://www.gov.uk/government/publications/family-food-2012 (accessed 3.12.14).
Detilleux, J.C., 2004. Javelin diagrams: applications in veterinary medical decision analysis. Vet. Res. 35, 617–624.
Dhaliwal, G.S., Murray, R.D., Dobson, H., Montgomery, J., Ellis, W.A., 1996. Reduced conception rates in dairy cattle associated with serological evidence of Leptospira interrogans serovar hardjo infection. Vet. Rec. 139, 110.

Dilsizian, S.E., Siegel, E.L., 2014. Artificial intelligence in medicine and cardiac imaging: harnessing big data and advanced computing to provide personalized medical diagnosis and treatment. Curr. Cardiol. Rep. 16, 441.

Drackley, J.K., 1999. ADSA Foundation Scholar Award. Biology of dairy cows during the transition period: the final frontier? J. Dairy Sci. 82, 2259–2273.

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

Duffield, T.F., Lissemore, K.D., McBride, B.W., Leslie, K.E., 2009. Impact of hyperketonemia in early lactation dairy cows on health and production. J. Dairy Sci. 92, 571–580.

Edwards, J.L., Tozer, P.R., 2004. Using Activity and Milk Yield as Predictors of Fresh Cow Disorders. J. Dairy Sci. 87, 524–531.

Esslemont, R.J., 2003. The costs of poor fertility and what to do about reducing them. Cattle Pract. 11, 237–250.

Esslemont, R.J., Kossaibati, M., 2002. DAISY Research Report No. 5: The costs of poor fertility and disease in UK dairy herds - Trends in DAISY herds over 10 seasons. University of Reading, UK.

Evans, G.E., Jones, B., 2009. The application of Monte Carlo simulation in finance, economics and operations management. In: 2009 WRI World Congress on Computer Science and Information Engineering, CSIE 2009. pp. 379–383.

Evans, R.D., Wallace, M., Shalloo, L., Garrick, D.J., Dillon, P., 2006. Financial implications of recent declines in reproduction and survival of Holstein-Friesian cows in spring-calving Irish dairy herds. Agric. Syst. 89, 165–183.

Fenwick, M.A., Fitzpatrick, R., Kenny, D.A., Diskin, M.G., Patton, J., Murphy, J.J., Wathes, D.C., 2008. Interrelationships between negative energy balance (NEB) and IGF regulation in liver of lactating dairy cows. Domest. Anim. Endocrinol. 34, 31–44.

Ferguson, J.D., Galligan, D.T., Thomsen, N., 1994. Principal Descriptors of Body Condition Score in Holstein Cows. J. Dairy Sci. 77, 2695–2703.

Fetrow, J., Stewart, S., Eicker, S., Rapnicki, P., 2007. Reproductive Health Programs for Dairy Herds: Analysis of Records for Assessment of Reproductive Performance. In: Current Therapy in Large Animal Theriogenology. pp. 473–489.

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.

Forastero, C., Zamora, L.I., Guirado, D., Lallena, A.M., 2010. A Monte Carlo tool to simulate breast cancer screening programmes. Phys. Med. Biol. 55, 5213.

Frey, C.H., Patil, S.R., 2002. Identification and review of sensitivity analysis methods. Risk Anal. 22, 553–578.

Garbarino, E.J., Hernandez, J.A., Shearer, J.K., Risco, C.A., Thatcher, W.W., 2004. Effect of lameness on ovarian activity in postpartum Holstein cows. J. Dairy Sci. 87, 4123–4131.

Garnsworthy, P.C., 2004. The environmental impact of fertility in dairy cows: a modelling approach to predict methane and ammonia emissions. Anim. Feed Sci. Technol. 112, 211–223.

Garnsworthy, P.C., Sinclair, K.D., Webb, R., 2008. Integration of Physiological Mechanisms That Influence Fertility in Dairy Cows. animal 2, 1144–1152.

Geary, U., Lopez-Villalobos, N., Begley, N., McCoy, F., O'Brien, B., O'Grady, L., Shalloo, L., 2012. Estimating the effect of mastitis on the profitability of Irish dairy farms. J. Dairy Sci. 95, 3662–3673.

Geishauser, T.D., Leslie, K.E., Duffield, T.F., Edge, V.L., 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.

Gelman, A., Meng, X.L., Stern, H., 1996. Posterior predictive assessment of model fitness via realized discrepancies. Stat. Sin. 6, 733–760.

Gilbert, R.O., Shin, S.T., Guard, C.L., Erb, H.N., Frajblat, M., 2005. Prevalence of endometritis and its effects on reproductive performance of dairy cows. Theriogenology 64, 1879–1888.

Gillies, C.L., Lambert, P.C., Abrams, K.R., Sutton, A.J., Cooper, N.J., Hsu, R.T., Davies, M.J., Khunti, K., 2008. Different strategies for screening and prevention of type 2 diabetes in adults: cost effectiveness analysis. Br. Med. J. 336, 1180–1191.

Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S., Brilliant, L., 2009. Detecting influenza epidemics using search engine query data. Nature 457, 1012–1014.

Giordano, J.O., Kalantari, A.S., Fricke, P.M., Wiltbank, M.C., Cabrera, V.E., 2012. A daily herd Markov-chain model to study the reproductive and economic impact of reproductive programs combining timed artificial insemination and estrus detection. J. Dairy Sci. 95, 5442–5460.

Gonzalez-Recio, O., Perez-Cabal, M.A., Alenda, R., 2004. Economic value of female fertility and its relationship with profit in Spanish dairy cattle. J. Dairy Sci. 87, 3053–3061.

Gouweleeuw, B.T., Thielen, J., Franchello, G., De Roo, A.P.J., Buizza, R., 2005. Flood forecasting using medium-range probabilistic weather prediction. Hydrol Earth Syst Sci 9, 365–380.

Green, M.J., Bradley, A.J., Medley, G.F., Browne, W.J., 2007. Cow, farm, and management factors during the dry period that determine the rate of clinical mastitis after calving RID B-7289-2008 RID A-8686-2010. J. Dairy Sci. 90, 3764–3776.

Green, M.J., Green, L.E., Huxley, J.N., Statham, J.M.E., Statham, S., 2012. Concepts in dairy herd health. In: Dairy Herd Health. CABI.

Green, M.J., Green, L.E., Medley, G.F., Schukken, Y.H., Bradley, A.J., 2002. Influence of Dry Period Bacterial Intramammary Infection on Clinical Mastitis in Dairy Cows. J. Dairy Sci. 85, 2589–2599.

Green, M.J., Medley, G.F., Browne, W.J., 2009. Use of posterior predictive assessments to evaluate model fit in multilevel logistic regression. Vet. Res. 40.

Gunay, A., Gunay, U., 2008. Effects of clinical mastitis on reproductive performance in Holstein cows. Acta Vet. Brno 77, 555–560.

Hanks, J., Kossaibati, M., 2010. Key performance indicators for the UK national dairy herd in 2010: A study of herd performance in 500 milk recording herds. University.

Hansen, P.J., Soto, P., Natzke, R.P., 2004. Mastitis and fertility in cattle - Possible involvement of inflammation or immune activation in embryonic mortality. Am. J. Reprod. Immunol. 51, 294–301.

Heller, J., Innocent, G.T., Denwood, M., Reid, S.W.J., Kelly, L., Mellor, D.J., 2011. Assessing the probability of acquisition of meticillin-resistant Staphylococcus aureus (MRSA) in a dog using a nested stochastic simulation model and logistic regression sensitivity analysis. Prev. Vet. Med. 99, 211–224.

Helton, J.C., 2008. Uncertainty and sensitivity analysis for models of complex systems, Lecture Notes in Computational Science and Engineering.

Helton, J.C., Johnson, J.D., Sallaberry, C.J., Storlie, C.B., 2006. Survey of sampling-based methods for uncertainty and sensitivity analysis. Reliab. Eng. Syst. Saf. 91, 1175–1209.

Herath, S., Williams, E.J., Lilly, S.T., Gilbert, R., Dobson, H., Bryant, C.E., Sheldon, I.M., 2007. Ovarian follicular cells have innate immune capabilities that modulate their endocrine function. Reproduction 134, 683–693.

Hernandez, J., Shearer, J.K., Webb, D.W., 2001. Effect of lameness on the calving-to-conception interval in dairy cows. J. Am. Vet. Med. Assoc. 218, 1611–1614.

Hernandez, J.A., Garbarino, E.J., Shearer, J.K., Risco, C.A., Thatcher, W.W., 2005. Comparison of the calving-to-conception interval in dairy cows with different degrees of lameness during the prebreeding postpartum period. J. Am. Vet. Med. Assoc. 227, 1284–1291.

Hertl, J.A., Grohn, Y.T., Leach, J.D.G., Bar, D., Bennett, G.J., Gonzalez, R.N., Rauch, B.J., Welcome, F.L., Tauer, L.W., Schukken, Y.H., 2010. Effects of clinical mastitis caused by gram-positive and gram-negative bacteria and other organisms on the probability of conception in New York State Holstein dairy cows. J. Dairy Sci. 93, 1551–1560.

Heuer, C., Schukken, Y.H., Dobbelaar, P., 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.

Hockett, M.E., Hopkins, F.M., Lewis, M.J., Saxton, A.M., Dowlen, H.H., Oliver, S.P., Schrick, F.N., 2000. Endocrine profiles of dairy cows following experimentally induced clinical mastitis during early lactation. Anim. Reprod. Sci. 58, 241–251.

Hockey, C.D., Morton, J.M., 2010. Use of a stochastic simulation model to assess effects of diagnostic specificity of systems for detecting ovulating cows on herd reproductive performance in year-round calving dairy herds. Anim. Reprod. Sci. 122, 189–199.

Holman, A., Thompson, J., Routly, J.E., Cameron, J., Jones, D.N., Grove-White, D., Smith, R.F., Dobson, H., 2011. Comparison of oestrus detection methods in dairy cattle. Vet. Rec. 169, 47–52.

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

Hudson, C.D., Kerby, M., Statham, J.M.E., Wapenaar, W., 2012. Managing herd reproduction. In: Dairy Herd Health. CABI.

Hurnik, J.F., King, G.J., Robertson, H.A., 1975. Estrous and related behaviour in postpartum Holstein cows. Appl. Anim. Ethol. 2, 55–68.

Hutchinson, I.A., Shalloo, L., Butler, S.T., 2013. Expanding the dairy herd in pasture-based systems: The role of sexed semen use in virgin heifers and lactating cows. J. Dairy Sci. 96, 6742–6752.

Huxley, J.N., 2013. Impact of lameness and claw lesions in cows on health and production. Livest. Sci. 156, 64–70.

Jacquez, J.A., 1998. Design of experiments. J. Frankl. Inst., Part Special Issue on Biomathematics 335, 259–279.

Jones, M.R., Fowler, H.J., Kilsby, C.G., Blenkinsop, S., 2013. An assessment of changes in seasonal and annual extreme rainfall in the UK between 1961 and 2009. Int. J. Climatol. 33, 1178–1194.

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

Kearney, J., 2010. Food consumption trends and drivers. Philos. Trans. R. Soc. B Biol. Sci. 365, 2793–2807.

Kemp, M.H., Nolan, A.M., Cripps, P.J., Fitzpatrick, J.L., 2008. Animal-based measurements of the severity of mastitis in dairy cows. Vet. Rec. 163, 175–179.

Kerby, M., 2009. The reproductive performance of 42 commercial dairy herds in the south-west of England in 2006/2007 - Observations, lessons learnt and thoughts for the future. Cattle Pract. 17, 26–32.

Kleijnen, J.P., Sargent, R.G., 2000. A methodology for fitting and validating metamodels in simulation. Eur. J. Oper. Res. 120, 14–29.

Kristensen, E., Dueholm, L., Vink, D., Andersen, J.E., Jakobsen, E.B., Illum-Nielsen, S., Petersen, F.A., Enevoldsen, C., 2006. Within- and Across-Person Uniformity of Body Condition Scoring in Danish Holstein Cattle. J. Dairy Sci. 89, 3721–3728.

Kristula, M.A., Reeves, M., Redlus, H., Uhlinger, C., 1995. A preliminary investigation of the association between the first postpartum milk fat test and first insemination pregnancy rates. Prev. Vet. Med. 23, 95–100.
Lamprou, A., Jia, G., Taflanidis, A.A., 2013. Life-cycle seismic loss estimation and global sensitivity analysis based on stochastic ground motion modeling. Eng. Struct. 54, 192–206.
Laney, D., 2001. 3D Data management: Controlling data volume, velocity and variety. META Group.
Lavon, Y., Ezra, E., Leitner, G., Wolfenson, D., 2011. Association of conception rate with pattern and level of somatic cell count elevation relative to time of insemination in dairy cows. J. Dairy Sci. 94, 4538–4545.
LeBlanc, S., 2007. Economics of improving reproductive performance in dairy herds. Adv. Dairy Technol. Vol 19, Advances in Dairy Technology 19, 201–214.
LeBlanc, S.J., Duffield, T.F., Leslie, K.E., Bateman, K.G., Keefe, G.P., Walton, J.S., Johnson, W.H., 2002. Defining and Diagnosing Postpartum Clinical Endometritis and its Impact on Reproductive Performance in Dairy Cows. J. Dairy Sci. 85, 2223–2236.
LeBlanc, S.J., Leslie, K.E., Duffield, T.F., 2005. Metabolic predictors of displaced abomasum in dairy cattle. J. Dairy Sci. 88, 159–170.
Leroy, J., Vanholder, T., Van Knegsel, A.T.M., Garcia-Ispierto, I., Bols, P.E.J., 2008. Nutrient prioritization in dairy cows early postpartum: mismatch between metabolism and fertility? Reprod. Domest. Anim. 43, 96–103.
Leroy, J.L.M.R., Vanholder, T., Mateusen, B., Christophe, A., Opsomer, G., de Kruif, A., Genicot, G., Van Soom, A., 2005. Non-esterified fatty acids in follicular fluid of dairy cows and their effect on developmental capacity of bovine oocytes in vitro. Reproduction 130, 485 –495.
Lloyd, S., 2000. Ultimate physical limits to computation. Nature 406, 1047–1054.
Loeffler, S.H., de Vries, M.J., Schukken, Y.H., 1999. The effects of time of disease occurrence, milk yield, and body condition on fertility in dairy cows. J. Dairy Sci. 82, 2589 – 2604.
Lopez, H., Satter, L.D., Wiltbank, M.C., 2004. Relationship between level of milk production and estrous behavior of lactating dairy cows. Anim. Reprod. Sci. 81, 209–223.
Lu, Z., Schukken, Y.H., Smith, R.L., Gröhn, Y.T., 2013. Using vaccination to prevent the invasion of Mycobacterium avium subsp. paratuberculosis in dairy herds: A stochastic simulation study. Prev. Vet. Med. 110, 335–345.
Lucy, M.C., 2001. Reproductive Loss in High-Producing Dairy Cattle: Where Will It End? J. Dairy Sci. 84, 1277–1293.
Lutz, W., Sanderson, W., Scherbov, S., 2001. The end of world population growth. Nature 412, 543–545.
Lynch, C., 2008. Big data: How do your data grow? Nature 455, 28–29.
MacGillivray, H.T., Dodd, R.J., 1982. Monte-Carlo simulations of galaxy systems. Astrophys. Space Sci. 86, 437–452.
Machado, V.S., Caixeta, L.S., McArt, J.A.A., Bicalho, R.C., 2010. The effect of claw horn disruption lesions and body condition score at dry-off on survivability, reproductive performance, and milk production in the subsequent lactation. J. Dairy Sci. 93, 4071–4078.
Madouasse, A., 2009. An Evaluation of Milk Recording, Somatic Cell Counts and Reproductive Performance in a Large Cohort of Dairy Herds in England and Wales. University of Nottingham.
Madouasse, A., Huxley, J.N., Browne, W.J., Bradley, A.J., Dryden, I.L., Green, M.J., 2010a. 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.

Madouasse, A., Huxley, J.N., Browne, W.J., Bradley, A.J., Green, M.J., 2010b. Somatic cell count dynamics in a large sample of dairy herds in England and Wales. Prev. Vet. Med. 96, 56–64.

Markowetz, A., Błaszkiewicz, K., Montag, C., Switala, C., Schlaepfer, T.E., 2014. Psycho-Informatics: Big Data shaping modern psychometrics. Med. Hypotheses 82, 405–411.

Mawhinney, I., Biggadike, H., 1998. A field study of the Intercept planned breeding routine for dairy cows involving GnRH and PGF2alpha. Cattle Pract. U. K.

McDougall, S., 2001. Effects of periparturient diseases and conditions on the reproductive performance of New Zealand dairy cows. N. Z. Vet. J. 49, 60–67.

McGowan, M.R., Kirkland, P.D., Rodwell, B.J., Kerr, D.R., Carroll, C.L., 1993. A field investigation of the effects of bovine viral diarrhea virus infection around the time of insemination on the reproductive performance of cattle. Theriogenology 39, 443–449.

McGuirk, B.J., Going, I., Gilmour, A.R., 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.

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

Mee, J.F., 2004. Temporal trends in reproductive performance in Irish dairy herds and associated risk factors. Ir. Vet. J. 57, 158–166.

Melendez, P., Bartolome, J., Archbald, L.F., Donovan, A., 2003. The association between lameness, ovarian cysts and fertility in lactating dairy cows. Theriogenology 59, 927–937.

Metropolis, N., 1987. The beginnings of the Monte Carlo method. Los Alamos Sci. 15, 125–130.

Metropolis, N., Ulam, S., 1949. The Monte Carlo Method. J. Am. Stat. Assoc. 44, 335–341.

Moore, D.A., Cullor, J.S., Bondurant, R.H., Sischo, W.M., 1991. Preliminary field evidence for the association of clinical mastitis with altered interestrus intervals in dairy cattle. Theriogenology 36, 257–265.

Moore, D.A., Overton, M.W., Chebel, R.C., Truscott, M.L., BonDurant, R.H., 2005. Evaluation of factors that affect embryonic loss in dairy cattle. J. Am. Vet. Med. Assoc. 226, 1112–1118.

Morton, J.M., 2003. The InCalf Project - Identifying risk factors for reproductive performance in Australian dairy herds. Cattle Pract. 11, 201–208.

National Milk Records, 2009. Annual Production Report. Chippenham, UK.

Nava-Trujillo, H., Soto-Belloso, E., Hoet, A.E., 2010. Effects of clinical mastitis from calving to first service on reproductive performance in dual-purpose cows. Anim. Reprod. Sci. 121, 12–16.

Nebel, R.L., McGilliard, M.L., 1993. Interactions of high milk-yield and reproductive performance in dairy-cows. J. Dairy Sci. 76, 3257–3268.

Nelson, S.T., Martin, A.D., Osteras, O., 2010. Risk factors associated with cystic ovarian disease in Norwegian dairy cattle. Acta Vet. Scand. 52, 60–70.

Neuvians, T.P., Schams, D., Berisha, B., Pfaffl, M.W., 2004. Involvement of pro-inflammatory cytokines, mediators of inflammation, and basic fibroblast growth factor in prostaglandin F-2 alpha-induced luteolysis in bovine corpus luteum. Biol. Reprod. 70, 473–480.

Nielsen, T.D., Kudahl, A.B., Østergaard, S., Nielsen, L.R., 2013. Gross margin losses due to Salmonella Dublin infection in Danish dairy cattle herds estimated by simulation modelling. Prev. Vet. Med. 111, 51–62.

Nordlund, K.V., Cook, N.B., 2004. Using herd records to monitor transition cow survival, productivity, and health. Vet. Clin. North Am. Food Anim. Pract. 20, 627–649.

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

Oakley, J.E., O'Hagan, A., 2004. Probabilistic sensitivity analysis of complex models: a Bayesian approach. J. R. Stat. Soc. Ser. B Stat. Methodol. 66, 751–769.

Ospina, P.A., Nydam, D.V., Stokol, T., Overton, T.R., 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.

Peake, K.A., Biggs, A.M., Argo, C.M., Smith, R.F., Christley, R.M., Routly, J.E., Dobson, H., 2011. Effects of lameness, subclinical mastitis and loss of body condition on the reproductive performance of dairy cows. Vet. Rec. 168, 301.

Pennington, J.A., Albright, J.L., Diekman, M.A., Callahan, C.J., 1985. Sexual Activity of Holstein Cows: Seasonal Effects1. J. Dairy Sci. 68, 3023–3030.

Perrin, L., Bostelmann, R.W., Sheldon, I.M., 2007. Reduced conception rates associated with bovine mastitis during a "window of opportunity". Vet. Rec. 161, 61–62.

Pinedo, P.J., Melendez, P., Villagomez-Cortes, J.A., Risco, C.A., 2009. Effect of high somatic cell counts on reproductive performance of Chilean dairy cattle. J. Dairy Sci. 92, 1575–1580.

Platz, S., Ahrens, F., Bendel, J., Meyer, H.H.D., Erhard, M.H., 2008. What happens with cow behavior when replacing concrete slatted floor by rubber coating: a case study. J. Dairy Sci. 91, 999–1004.

Plummer, M., Best, N., Cowles, K., Vines, K., 2006. Convergence diagnosis and output analysis for MCMC. R News 6, 7.

Podpečan, O., Mrkun, J., Zrimšek, P., 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.

R Core Development Team, 2010. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Raftery, A.E., Gneiting, T., Balabdaoui, F., Polakowski, M., 2005. Using Bayesian Model Averaging to Calibrate Forecast Ensembles. Mon. Weather Rev. 133, 1155–1174.

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

Rasbash, J., Steele, F., Browne, W.J., Goldstein, H., 2009. A User's Guide to MLwiN, v2.10. Centre for Multilevel Modelling, University of Bristol.

Ravagnolo, O., Misztal, I., 2002. Effect of Heat Stress on Nonreturn Rate in Holsteins: Fixed-Model Analyses. J. Dairy Sci. 85, 3101–3106.

Refsdal, A.O., 2007. Reproductive performance of Norwegian cattle from 1985 to 2005: trends and seasonality. Acta Vet. Scand. 49, 7.

Ripley, B.D., 2009. Stochastic Simulation. John Wiley & Sons.

Risco, C.A., Donovan, G.A., Hernandez, J., 1999. Clinical mastitis associated with abortion in dairy cows. J. Dairy Sci. 82, 1684–1689.

Roche, J.F., 2006. The effect of nutritional management of the dairy cow on reproductive efficiency. Anim. Reprod. Sci. 96, 282–296.

Roche, J.R., Friggens, N.C., Kay, J.K., Fisher, M.W., Stafford, K.J., Berry, D.P., 2009. Invited review: Body condition score and its association with dairy cow productivity, health, and welfare. J. Dairy Sci. 92, 5769–5801.

Roelofs, J., López-Gatius, F., Hunter, R.H.F., van Eerdenburg, F.J.C.M., Hanzen, C., 2010. When is a cow in estrus? Clinical and practical aspects. Theriogenology 74, 327–344.

Roelofs, J., Vaneerdenburg, F., Soede, N., Kemp, B., 2005. Pedometer readings for estrous detection and as predictor for time of ovulation in dairy cattle. Theriogenology 64, 1690–1703.

Roth, Z., Arav, A., Bor, A., Zeron, Y., Braw-Tal, R., Wolfenson, D., 2001. Improvement of quality of oocytes collected in the autumn by enhanced removal of impaired follicles from previously heat-stressed cows. Reproduction 122, 737–744.

Santos, J.E.P., Cerri, R.L.A., Ballou, M.A., Higginbotham, G.E., Kirk, J.H., 2004. Effect of timing of first clinical mastitis occurrence on lactational and reproductive performance of Holstein dairy cows. Anim. Reprod. Sci. 80, 31–45.

Schlageter-Tello, A., Bokkers, E.A.M., Koerkamp, P.W.G.G., Van Hertem, T., Viazzi, S., Romanini, C.E.B., Halachmi, I., Bahr, C., Berckmans, D., Lokhorst, K., 2014. Effect of merging levels of locomotion scores for dairy cows on intra- and interrater reliability and agreement. J. Dairy Sci. In Press.

Schrick, F.N., Hockett, M.E., Saxton, A.M., Lewis, M.J., Dowlen, H.H., Oliver, S.P., 2001. Influence of subclinical mastitis during early lactation on reproductive parameters. J. Dairy Sci. 84, 1407–1412.

Shalloo, L., Dillon, P., Rath, M., Wallace, M., 2004. Description and Validation of the Moorepark Dairy System Model. J. Dairy Sci. 87, 1945–1959.

Sheldon, I.M., Dobson, H., 2003. Reproductive challenges facing the cattle industry at the beginning of the 21st century. Reprod. Camb. Engl. Suppl. 61, 1–13.

Singer, J.D., Willett, J.B., 1993. It's about time: Using discrete-time survival analysis to study duration and the timing of events. J. Educ. Behav. Stat. 18, 155.

Sogstad, A.M., Østeraas, O., Fjeldaas, T., 2006. Bovine claw and limb disorders related to reproductive performance and production diseases. J. Dairy Sci. 89, 2519–2528.

Sørensen, J.T., Kristensen, E.S., Thysen, I., 1992. A stochastic model simulating the dairy herd on a PC. Agric. Syst. 39, 177–200.

Soto, P., Natzke, R.P., Hansen, P.J., 2003. Identification of possible mediators of embryonic mortality caused by mastitis: Actions of lipopolysaccharide, prostaglandin F-2 alpha and the nitric oxide generator, sodium nitroprusside dihydrate, on oocyte maturation and embryonic development in cattle. Am. J. Reprod. Immunol. 50, 263–272.

Spiegel, B.M.R., Targownik, L., Dulai, G.S., Gralnek, I.M., 2003. The cost-effectiveness of cyclooxygenase-2 selective inhibitors in the management of chronic arthritis. Ann. Intern. Med. 138, 795–806.

Spiegelhalter, D., Thomas, A., Best, N., Lunn, D., 2003. WinBUGS User Manual, version 1.4.

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

Statham, J.S., 2012. Fertility management of dairy cows. Cattle Pract. 20, 48–56.

Steele, F., 2003. A discrete-time multilevel mixture model for event history data with long-term survivors, with an application to an analysis of contraceptive sterilization in Bangladesh. Lifetime Data Anal. 9, 155–174.

Steeneveld, W., Schukken, Y.H., van Knegsel, A.T.M., Hogeveen, H., 2013. Effect of different dry period lengths on milk production and somatic cell count in subsequent lactations in commercial Dutch dairy herds. J. Dairy Sci. 96, 2988–3001.

Steinbach, R., Cairns, J., Grundy, C., Edwards, P., 2012. Cost benefit analysis of 20 mph zones in London. Inj. Prev. 19, 211–213.

Steinfeld, H., Gerber, P., Wassenaar, T., Castel, V., Rosales, M., de Haan, C., 2006. Livestock's long shadow.

Tekerli, M., Akinci, Z., Dogan, I., Akcan, A., 2000. Factors Affecting the Shape of Lactation Curves of Holstein Cows from the Balikesir Province of Turkey. J. Dairy Sci. 83, 1381–1386.

Tsekouras, G., Koutsoyiannis, D., 2014. Stochastic analysis and simulation of hydrometeorological processes associated with wind and solar energy. Renew. Energy 63, 624–633.

Van Rijmenam, M., 2013. How small data can also be big data. BigData Startup.

Van Vliet, J.H., Van Eerdenburg, F.J.C.M., 1996. Sexual activities and oestrus detection in lactating Holstein cows. Appl. Anim. Behav. Sci. 50, 57–69.
Vanroose, G., Nauwynck, H., Van Soom, A., Vanopdenbosch, E., de Kruif, A., 1997. Susceptibility of zona-intact and zona-free in vitro-produced bovine embryos at different stages of development to infection with bovine herpesvirus-1. Theriogenology 47, 1389–1402.
Vanroose, G., Nauwynck, H., Van Soom, A., Vanopdenbosch, E., de Kruif, A., 1998. Replication of cytopathic and noncytopathic bovine viral diarrhea virus in zona-free and zona-intact in vitro-produced bovine embryos and the effect on embryo quality. Biol. Reprod. 58, 857 –866.
Veerkamp, R.F., Koenen, E.P.C., De Jong, G., 2001. Genetic Correlations Among Body Condition Score, Yield, and Fertility in First-Parity Cows Estimated by Random Regression Models. J. Dairy Sci. 84, 2327–2335.
Villa-Godoy, A., Hughes, T.L., Emery, R.S., Stanisiewski, E.P., Fogwell, R.L., 1990. Influence of Energy Balance and Body Condition on Estrus and Estrous Cycles in Holstein Heifers1. J. Dairy Sci. 73, 2759–2765.
Vonk Noordegraaf, A., Nielen, M., Franken, P., Dijkhuizen, A.A., 2002. Simulation modelling of BHV1-control programme at national level, with special attention to sensitivity analysis. Livest. Prod. Sci. 76, 153–170.
Vonk Noordegraaf, A., Nielen, M., Kleijnen, J.P., 2003. Sensitivity analysis by experimental design and metamodelling: Case study on simulation in national animal disease control. Eur. J. Oper. Res. 146, 433–443.
Wall, E., Brotherstone, S., Woolliams, J.A., Banos, G., Coffey, M.P., 2003. Genetic Evaluation of Fertility Using Direct and Correlated Traits. J. Dairy Sci. 86, 4093–4102.
Walsh, R.B., Walton, J.S., Kelton, D.F., LeBlanc, S.J., Leslie, K.E., Duffield, T.F., 2007. The Effect of Subclinical Ketosis in Early Lactation on Reproductive Performance of Postpartum Dairy Cows. J. Dairy Sci. 90, 2788–2796.
Walster, C., 2012. Adding value for farmers through herd health plans. Vet. Rec. 170, 524.
Wang, Y., Li, S., 2008. Worldwide trends in dairy production and consumption and calcium intake: is promoting consumption of dairy products a sustainable solution for inadequate calcium intake? Food Nutr. Bull. 29, 172–185.
Ward, J.S., Barker, A., 2013. Undefined By Data: A Survey of Big Data Definitions. University of St Andrews.
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.
Whitaker, D.A., Macrae, A.I., Burrough, E., 2004. Disposal and disease rates in British dairy herds between April 1998 and March 2002. Vet. Rec. 155, 43–47.
Whitlow, L.W., Hagler, W.M., 1999. An association of mycotoxins with production, health and reproduction in dairy cattle and guidelines for prevention and treatment. In: Biotechnology in the Feed Industry, Proceedings of Alltech's 15th Annual Symposium (TP Lyons and KA Jacques, Eds). Nottingham University Press, UK. p. 401.
WHO, 1996. WHO | Food Security [WWW Document]. URL http://www.who.int/trade/glossary/story028/en/ (accessed 3.12.14).
Wickham, H., 2007. Reshaping data with the reshape package. J. Stat. Softw. 21, 1–20.
Wickham, H., 2011. The split-apply-combine strategy for data analysis. J. Stat. Softw. 40.
Wilkinson, D.J., 2009. Stochastic modelling for quantitative description of heterogeneous biological systems. Nat. Rev. Genet. 10, 122–133.
Williams, E.J., Sibley, K., Miller, A.N., Lane, E.A., Fishwick, J., Nash, D.M., Herath, S., England, G.C.W., Dobson, H., Sheldon, I.M., 2008. The effect of Escherichia coli

lipopolysaccharide and tumour necrosis factor alpha on ovarian function. Am. J. Reprod. Immunol. 60, 462–473.
Windig, J.J., Calus, M.P.L., Veerkamp, R.F., 2005. High milk yields and the risk of mastitis in different herd environments. In: Hogeveen, H. (Ed.), Mastitis in Dairy Production - Current Knowledge and Future Ideas. Presented at the 4th IDF International mastitis conference, Wageningen, pp. 254–259.
Xu, Z.Z., McKnight, D.J., Vishwanath, R., Pitt, C.J., Burton, L.J., 1998. Estrus Detection Using Radiotelemetry or Visual Observation and Tail Painting for Dairy Cows on Pasture. J. Dairy Sci. 81, 2890–2896.
Yang, M., Goldstein, H., 2003. Modelling survival data in MLwiN 1.20. Centre for Multilevel Modelling, University of Bristol.
Yoshida, C., Yusuf, M., Nakao, T., 2009. Duration of estrus induced after GnRH-PGF2Î± protocol in dairy heifer. Anim. Sci. J. 80, 649–654.
Zhang, J., Yu, K.F., 1998. What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. J. Am. Med. Assoc. 280, 1690–1691.