Challenges of health data recording - an Australian perspective

J.E. Pryce[1], M. Blake[2], K. E. Davis[2] and D.S. Beggs[3,4]

[1]Department of Primary Industries, Agribio, 5 Ring Road,Bundoora, VIC 3083, Australia
[2]Dairy Australia, Level 5, IBM Centre, 60 City Road, Southbank, VIC, 3006, Australia
[3]Warrnambool Veterinary Clinic, 514 Raglan Parade, Warrnambool, VIC, 3280, Australia
[4]Faculty of Veterinary Science, University of Melbourne, 250 Princes Highway Werribee 3030, Australia

Abstract

Australian dairy farmers record health data for 3 main reasons:

1. farm management decisions at an individual cow and a herd level;
2. to comply with farm quality assurance schemes (e.g. drug usage and withholding periods); and,
3. genetic evaluations.

There are at least 10 different farm management software packages available for recording farm data. Standard data interchange formats (DIF) are used for transferring data from the farm software to the Australian Dairy Herd Improvement Scheme (ADHIS) which uses the data for estimating breeding values. Currently, the only data uploaded to ADHIS are from herds that are enrolled in milk-recording and participation in this is relatively low (capturing approximately 46% of cows). There are many reasons for this, including a long drought through the 2000s which resulted in many farmers cost-cutting by stopping herd-testing; another is the move to in-line recording systems where data is currently not captured centrally. However, not participating in milk-recording does not have to exclude Australian dairy farmers from having health or fertility events incorporated in genetic evaluation runs. The main challenge to integrating health data from non-milk-recording herds into the central database is unique animal identification and identifying ancestry. New tools, such as cheap parentage identification could help ensure that integrity of pedigree is maintained, although there is obviously still a cost involved. The Australian dairy industry has laid the foundations for a new central database that will store and integrate data from multiple sources; the expectation is that the number of users and data providers will be much broader than for genetic evaluations. It is hoped that the new central database will enable data capture on a greater variety of traits and from many more farms than previously. Currently somatic cell count and survival are the only aspects of herd health that breeding values are estimated for in Australia. Data capture is again an issue for diseases such as mastitis and lameness. Understanding their genetic control is the first task, as lameness and mastitis may have a different genetic basis in Australia compared to housed production systems in the Northern hemisphere. For example, lameness is often a result of pressure on the hoof resulting from walking long distances.

Keywords: Data recording, genomics, mastitis, lameness.

Introduction

There are around 1.6 million dairy cows in Australia of which 46% (731,082) participate in herd-testing (ADHIS, 2012). The average herd-size is 222 and the average milk yield is 6,930 litres (ADHIS, 2012). Seventy eight percent of dairy cows are Holstein, 12% are Jersey, 5% are Holstein-Jersey crosses and 4% are red breeds (Figure 1).

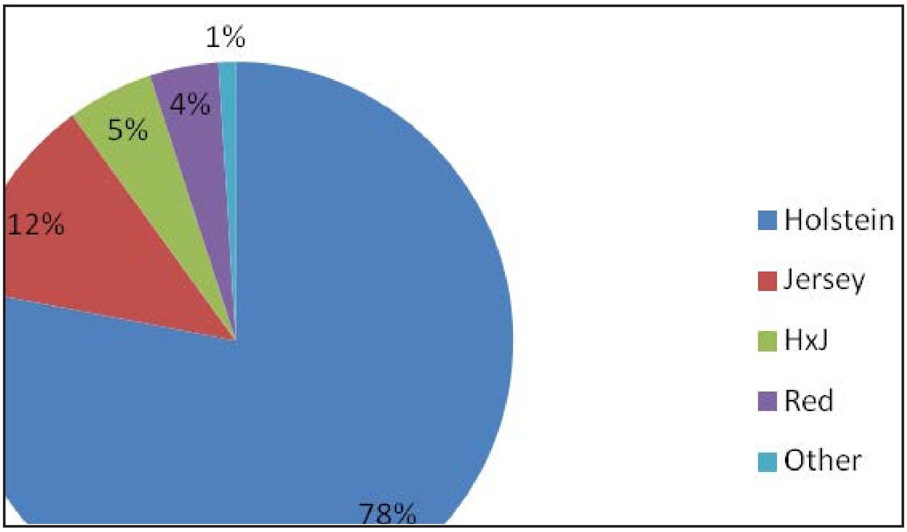


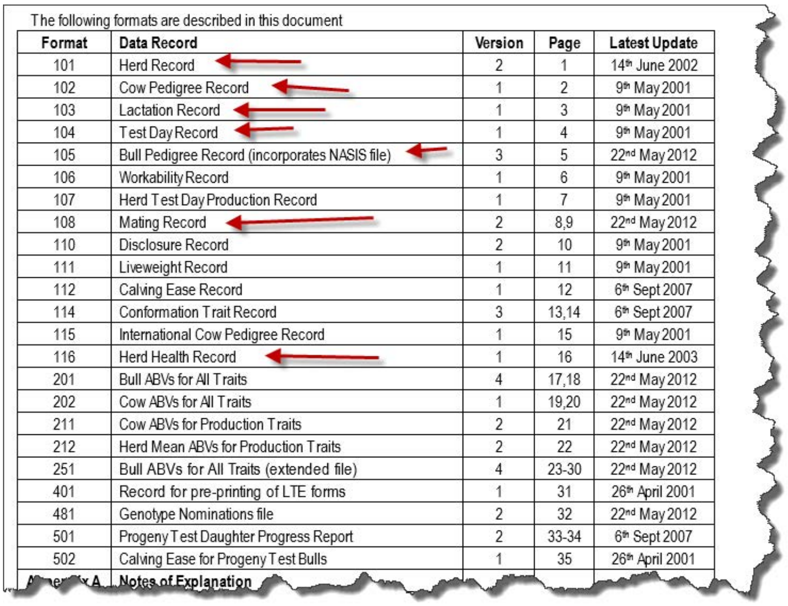
Figure 1. Breed composition of the Australian dairy population.

Recording system overview

Data Processing Centres (DPCs; that are generally part of milk recording organisations) provide data to the Australian Dairy Herd Improvement Scheme (ADHIS) for genetic evaluation purposes. There are 11 DPCs located across Australia. Farmers have an agreement with the DPC to allow their data to be used for research and genetic evaluations. Data is provided in standard data interchange format (DIF) as fixed length flat files. A snapshot of DIF format files is shown in Figure 2. For example, the DIF101 file includes information on the farm and owner, while the DIF102 includes cow information, such as animal identification, birth dates and sire identification. Approximately 1/3 of cows do not have their sire identification recorded which renders data from these cows useless for genetic evaluations. The health data DIF file (DIF116) has not been updated for around a decade, although there are plans to work on developing this file. DIF116 data are not currently used for genetic evaluations by ADHIS, as the proportion of herds recording health data is low. Currently the somatic cell count breeding value (a proxy for mastitis resistance) and longevity breeding value (for overall health and well-being) are the only "health" breeding values estimated by ADHIS.

As DIF files are transferred to ADHIS prior to a genetic evaluation run,this database can be considered to be static, but regularly updated.

The domestic farm data software providers generally comply with the DIF formats, meaning that data collected on most farms can be used for genetic evaluations. One of the main challenges for Australia is ensuring that data collected on farm actually reaches ADHIS. For example, there are logistical reasons why some fertility (mating information and pregnancy test results) does not reach ADHIS.



The following formats are described in this document

Format	Data Record	Version	Page	Latest Update
101	Herd Record	2	1	14th June 2002
102	Cow Pedigree Record	1	2	9th May 2001
103	Lactation Record	1	3	9th May 2001
104	Test Day Record	1	4	9th May 2001
105	Bull Pedigree Record (incorporates NASIS file)	3	5	22nd May 2012
106	Workability Record	1	6	9th May 2001
107	Herd Test Day Production Record	1	7	9th May 2001
108	Mating Record	2	8,9	22nd May 2012
110	Disclosure Record	2	10	9th May 2001
111	Liveweight Record	1	11	9th May 2001
112	Calving Ease Record	1	12	6th Sept 2007
114	Conformation Trait Record	3	13,14	6th Sept 2007
115	International Cow Pedigree Record	1	15	9th May 2001
116	Herd Health Record	1	16	14th June 2003
201	Bull ABVs for All Traits	4	17,18	22nd May 2012
202	Cow ABVs for All Traits	1	19,20	22nd May 2012
211	Cow ABVs for Production Traits	2	21	22nd May 2012
212	Herd Mean ABVs for Production Traits	2	22	22nd May 2012
251	Bull ABVs for All Traits (extended file)	4	23-30	22nd May 2012
401	Record for pre-printing of LTE forms	1	31	26th April 2001
481	Genotype Nominations file	2	32	22nd May 2012
501	Progeny Test Daughter Progress Report	2	33-34	6th Sept 2007
502	Calving Ease for Progeny Test Bulls	1	35	26th April 2001
Appendix A	Notes of Explanation			

Figure 2. A snap-shot of the data interchange format files document from an ADHIS manual. The DIF records with arrows next to them are required for genetic evaluation of health and fertility traits.

Examples of software used in Australia

There are at least 10 farm software programs that are used to record cow and health data. However, the 3 most popular are Mistro, EasyDairy and DairyData. Mistro and EasyDairy are farm software providers, while DairyData was designed and is sold by veterinarians. Therefore, in DairyData some extra attention is paid to recording of reproduction and health data.

Health data recording in Australia

Health data is voluntarily recorded in Australia. A record consists of a cow's health event and a date of occurrence. Most of the data recorded is for preventative measures such as vaccination, drenching, dry cow therapy etc. There are 340 codes that farmers can use to enter treatments and health events. Having a large number of codes may make it difficult for the farmer to decipher the correct code to use. Also, the incidences of a lot of the categories are low. For this reason, in estimating genetic parameters for health disorders, Haile-Mariam and Goddard (2010) summarised the data as:

1. Udder health including mastitis - clinical, teat injury, sore teats, black spots on teat.
2. Reproductive disorders including abortion -early (no new lactation), ovarian cysts, uterine prolapse, uterine infection / metritis, uterine irrigation, anoestrus-inactive ovaries, retained foetal membranes, deformed calf.
3. Leg problems including dislocated hip, downer cow, paralysis at calving, arthritis, lameness.
4. Data on all disorders including all recorded treatments (excluding dry cow therapy, vaccination and drenching) and disease events to investigate whether there is a genetic component to overall disease resistance. A summary of health events recorded in 2007 and 2008 is presented in table 1. Additionally heritability estimates are also presented, which are consistently low, but comparable to other studies of similar data estimated in other countries.

Table 1. Incidences of health disorders and heritabilities with standard errors in brackets (adapted from the study of Haile-Mariam and Goddard, 2010).

	Udder		**Reproductive**		**Leg**		**'All disorders'**	
Year	**2007**	**2008**	**2007**	**2008**	**2007**	**2008**	**2007**	**2008**
Herds	91	112	99	154	45	58	203	225
Records	18291	14928	20810	19874	9695	7192	40256	45356
Incidence	0.142	0.129	0.056	0.058	0.053	0.055	0.31	0.214
Heritability	0.035 (0.009)		0.006 (0.004)		0.013 (0.009)		0.016 (0.004)	

Capturing extra data

Currently, Dairy Futures CRC in conjunction with ADHIS and Warrnambool Vet Clinic are collaborating on a project to increase the capture of fertility data (mating information and pregnancy test results) that would not normally be uploaded to the ADHIS database. It is expected that data for other traits, such as health and disease, capture of data will also increase as a result of this project. With some effort it should be possible increase the amount of data available to estimate breeding values.

Currently, the proportion of herds contributing valid fertility data to genetic evaluations is fairly low. However, this varies by state. For herds where data was provided to ADHIS (via their Data Processing Centre; DPC), 49.8% of Victorian herds have some mating data used for fertility ABV calculation in 2008, 2009 or 2010. The percentage of herds with mating data used in ABV calculation in other States is lower. There are some logistical reasons why the proportion of data is low, for example incomplete data capture from some of the herd-management software.

The current status of software systems is as follows:

- MISTRO - data transfer with most DPC's is working well
- EASY DAIRY / DAIRY DATA / DAIRY ID - We have tested the manual transfer of fertility & mating data from Farm to DPC to confirm the process will work. Once the fertility & mating data reaches the DPC it is then routinely transferred to ADHIS. Testing has now been completed and automation of data transfer has recently started for Easy Dairy users. Early results have been most encouraging with up to 9 years of farm fertility data with over 9 000 records moved from a single farm to DPC with the click of the on-farm software button.
- VETERINARY HELD FERTILITY DATA - A partner in the project is Warrnambool Veterinary Clinic, who own the popular Dairy Data software used by many veterinary practices and some farms to closely monitor fertility levels. Pregnancy test data is important because it can be used to calculate pregnancy rate, which is one of the traits in the multi-trait model. We can split the vet clients data into those herds that are herd testing and those herds that are not.
- FULLY AUTOMATED DAIRY PLATFORMS - High tech platform systems will be also studied to see what current data can be transferred from these systems in a suitable format to the Herd Improvement DPC platforms. These high tech systems are collecting massive amounts of cow and herd data 365 days of the year so it becomes important to see how we can include their data and make full use of their data .

By comparing exactly the same herds over the same period of time, an extra 18% of cows had valid fertility data in March 2013 compared to August 2012.

New central data system

In Australia, data relevant to farms is collected by multiple private businesses using many different systems and formats. There is a lot of duplication with the same information being entered in multiple systems and farmers are responsible for coordinating information from different sources. Information needed for day-to-day management and strategic decisions of the farm business is often not readily accessible. For this reason, a Central Data System (CDS) has been proposed for Australia. In contrast to the ADHIS database, the CDS is live rather than static. The overarching objective of the CDS for animal performance data is to provide a unified, authoritative dairy data system for supporting on-farm decision making, providing data for improved genetic evaluation, herd improvement and industry analysis. The key beneficiaries of the CDS will be Australian dairy farmers and organisations that innovate in response to the improved data-infrastructure. Its purpose is to enhance the adaptive capability of the dairy supply chain to improve farm margins and growth opportunities. The CDS is currently in the planning phase, with deployment expected in 2017.

Genomic reference population

One risk of replacing progeny-testing with breeding schemes that screen large numbers of young bulls and only select a small number of these for widespread use, is that fewer bulls will be added to the reference population on an annual basis than in the past. This would decrease the reliability of genomic prediction as the distance between the current dairy population and the majority of animals in the reference population increases (Lillehammer *et al.*,2010). Countries with small populations may be more affected by this issue than larger populations (McHugh *et al*., 2011). Considerable effort has gone into increasing the size of current reference populations and this effort must continue to ensure reference populations remain relevant to selection candidates. One of the strategies used to increase reference populations is to share genotypes with other countries.

Genotyping of cows is another way that the reference population can be grown. Genotyped females need to be incorporated cautiously, as there could be a risk that some of them are preferentially treated and therefore their phenotypes could be biased. Instead, directly targeting a group of randomly selected cows may be more beneficial. In Australia, the size of the male reference population is around 3000 Holstein males, so adding genotyped females to the reference population could improve the reliability of breeding values.

Recently, the Australian Dairy Futures Cooperative Research Centre's 10 000 Holstein Cow Genomes project and Jer-nomics project embarked on collecting DNA samples and genotyping 10 000 Holstein and 4 000 Jersey cows (from commercial herds). In April 2012, this information has become part of the Australian reference population. This has led to a 4-8% improvement in the reliability of breeding values depending on trait. Table 2 shows the increase in reliability of genomic breeding values for 437 young Holstein bulls achieved by adding close to 10 000 cows to the reference population.

Table 2. The reliability of genomic breeding values of 437 young bulls when bulls only were included in the reference population and when cows were also included

Trait	Bulls only	Bulls + cows	Change
Protein	54	61	7
Fat	54	61	7
Milk	54	61	7
Survival	30	36	6
Fertility	33	37	4
Somatic cell count	43	51	8
Milking speed	49	53	4
Temperament	49	53	4
Likability	49	53	4
Mammary system	39	44	5
Overall type	38	44	6
Udder depth	38	43	5
Udder texture	33	38	5

The 10 000 Holstein Cow Genomes project and Jer-nomics project were one-off genotyping events. We now intend to establish a more permanent genomic reference population that comprises around 100 herds.

Having the entire sequences of bulls may help to increase the accuracy further. The idea behind sequencing key ancestors of cattle breeds, is that we will have the causative mutations in the data set, i.e. we will be able to capture more of the genetic variation in a trait. The 1 000 bull genomes project has started with an aim to provide researchers with a large database for genomic prediction and genome wide association studies in all cattle breeds (*http://1000bullgenomes.com*).

Genomic reference populations may assist with difficult to measure traits, such as health, as efforts to record and evaluate these traits can happen in a small reference population and the benefits used by the entire population i.e. prediction equations are based on cows in the reference population that have phenotypes on a range of traits, possibly also including health traits.

Conclusions

Health data for dairy cows is sparsely recorded in Australia. However, through our endeavours to increase the amount of data collected for fertility, it is anticipated that more health data will also become available. A new central data system is planned which should help by connecting farmers with their data. Prediction of breeding values can be enhanced, especially for difficult to measure traits, such as health traits, in dedicated resource populations that are genotyped.

Haile-Mariam, M. & M.E. Goddard (2010) Preliminary genetic analyses of voluntarily supplied disease data in Australian dairy herds. Anim. Prod. Sci. 50: 186-192.

Lillehammer M, T.H.E. Meuwissen & A.K. Sonesson (2011) A comparison of dairy cattle breeding designs that use genomic selection. J. Dairy Sci. 94: 493-500.

Mc Hugh N, T.H.E. Meuwissen, A.R. Cromie & A.K. Sonesson (2011) Use of female information in dairy cattle breeding programs. J. Dairy Sci. 94: 4109-4118.