Developing metrics to rank individual herds according to data quality

M Stephen, J Bryant, M Camara and D Meadows
Overview

• Why we need data quality metrics
• Key metrics, and how we developed them
• What we have learned
• Future applications
Why we need data quality metrics

• Monitoring data quality over time
• Assessing the impact of poor quality data
• Targeting specific herds for research (i.e. novel trait collection)
• Feedback to farmers
Key metrics, and how we developed them

• Sire assignment
• Calving and mating dates
• Herd testing
• Conformation scoring
• Calving assistance recording
• Herd exit recording
Fate Diversity Index: Formula

\[ 1 - \sum_{i}^{n} \text{PropReason}_i^2 \]

\text{PropReason}_i = \text{proportion of animals culled or sold for reason } i
Fate Diversity Index: Example

Herd A: Fate Diversity Index 0.8

Herd B: Fate Diversity Index 0.2
What we have learned

• Data quality is extremely varied across herds
• Herd performance across multiple metrics is often inconsistent
• Most metrics are trending downwards
Data quality is extremely varied across herds

**Sire Assignment 2016**

- Number of herds
- Percentage of two year olds

**Calving and Mating 2016**

- Number of herds
- Percentage of parturitions with an associated mating
Data quality is extremely variable across herds.

**Average Fate Diversity Index 2016**

- 0 to 0.2: 1847
- 0.2 to 0.4: 1176
- 0.4 to 0.6: 2916
- 0.6 to 0.8: 3141
- 0.8 to 1.0: 190

**Herd Testing 2016**

- 0: 2299
- 1: 712
- 2: 1213
- 3: 1693
- 4: 2769
- > 4: 584
Herd performance across multiple metrics is often inconsistent.
Most metrics are trending downwards

- Sire assignment and ‘Mating and Calving’ metric’s appear constant
Future applications

- Monitor data quality over time
- Assess the impact of poor quality data
- Targeting specific herds for research (i.e. novel trait collection)
- Feedback to farmers (via Herd recorders)
Acknowledgements

Funded by New Zealand Dairy farmers through DairyNZ