Early prediction of lactational milk, fat and protein using daily milk data

Oded Nir (Markusfeld), Gil Katz, Liron Reuveni
The primer:

J. I. Weller and E. Ezra, Genetic and phenotypic analysis of daily Israeli Holstein milk, fat, and protein production as determined by a real-time milk analyzer, JDC, Vol. 99 No. 12, 2016

✓ Scope: >37,000 Holstein cows spanning over 2 years
✓ Finds agreement between Afimilk's inline milk lab real time analysis and between DHIA monthly tests.
✓ Selected for 'Editor's Choice' of JDSc

Slides are by the courtesy of Ephraim Ezra
Heritabilities, genetic and environmental correlations among 7,866 first parity 305 d lactations computed from the ICBA and AfiLab records.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Heritabilities</th>
<th>Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ICBA</td>
<td>AfiLab</td>
</tr>
<tr>
<td>Milk (kg)</td>
<td>0.33</td>
<td>0.35</td>
</tr>
<tr>
<td>Fat (kg)</td>
<td>0.23</td>
<td>0.31</td>
</tr>
<tr>
<td>Protein (kg)</td>
<td>0.27</td>
<td>0.32</td>
</tr>
<tr>
<td>% fat</td>
<td>0.48</td>
<td>0.57</td>
</tr>
<tr>
<td>% protein</td>
<td>0.55</td>
<td>0.46</td>
</tr>
</tbody>
</table>

Heritabilities were higher for the AfiLab records for all traits, except for % protein.
Phenotypic correlations among complete and extended 1st parity lactations the last ICBA test day and the last two weeks of AfiLab records.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Mean days in milk at truncation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30</td>
</tr>
<tr>
<td>ICBA</td>
<td>0.67</td>
</tr>
<tr>
<td>Afilab</td>
<td>0.77</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trait</th>
<th>Mean days in milk at truncation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30</td>
</tr>
<tr>
<td>ICBA</td>
<td>0.70</td>
</tr>
<tr>
<td>Afilab</td>
<td>0.72</td>
</tr>
</tbody>
</table>
Prediction of complete lactations in Afifarm

➢ Our objective: To adapt the large scale retrospective study’s method to a prospective prediction of complete (305_days) lactations in individual herds
✓ For selection
✓ For production planning (quota, summer/winter)
➢ The operational need: To enable farmers to get the decision as early as possible, but before breeding
## Waiting Periods

<table>
<thead>
<tr>
<th>Herds</th>
<th>Cows/herd</th>
<th>Voluntary waiting period (days)</th>
<th>Days to 1&lt;sup&gt;st&lt;/sup&gt; AI</th>
</tr>
</thead>
<tbody>
<tr>
<td>13,885</td>
<td>158.4 ± 325 SD</td>
<td>58.4 ± 5.6 SD</td>
<td>95.2 ± 26.9 SD</td>
</tr>
</tbody>
</table>

Ferguson J.D. & Skidmore A. (2013). JDS 96 (2) 1269 -1289

<table>
<thead>
<tr>
<th>Days to 1&lt;sup&gt;st&lt;/sup&gt; AI</th>
<th>50</th>
<th>51 - 80</th>
<th>81 - 110</th>
<th>111 - 150</th>
</tr>
</thead>
<tbody>
<tr>
<td>1&lt;sup&gt;st&lt;/sup&gt; lactation</td>
<td>0.4%</td>
<td>41.4%</td>
<td>45.2%</td>
<td>13.0%</td>
</tr>
<tr>
<td>2&lt;sup&gt;nd&lt;/sup&gt; lactation</td>
<td>9.7%</td>
<td>58.4%</td>
<td>26.9%</td>
<td>5.1%</td>
</tr>
</tbody>
</table>


**Our objective is to be able to make the decision at 60 DIM**
Material & methods

The population
✓ 14 herds
✓ 324 – 721 cows/herd
✓ 11,840 – 13,635 kg annual yield

Cows & production variables
✓ AfiMilk MPC™ & AfiLab™
✓ Afifarm

Calibration models
✓ Complete records of milk, fat & protein in 305 days
✓ Calendar (12 month) year
✓ Stepwise multiple regression; Estimates of $p<0.01$ included in the final models

Prediction models
✓ Known yields
✓ Calved after the end of the calibration models

Validation
Criteria for Success

➢ $R^2$ = RSquare of the summary of fit

➢ $r$ = Correlations to actual production

➢ The gap (%) between the predicted & the actual production (APD)

➢ Rate of wrong positive cows (1 – specificity) for selection (cows culled by mistake)
Prediction after 54 DIM 14 herds

<table>
<thead>
<tr>
<th>(Kg) Variable</th>
<th>The gap (%) to actual production</th>
<th>cows culled by mistake (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>milk</td>
<td>-3.6%-7.0%</td>
<td>0.0%-18.5%</td>
</tr>
<tr>
<td>fat</td>
<td>-3.7%-7.8%</td>
<td>0.0%-23.8%</td>
</tr>
<tr>
<td>protein</td>
<td>-2.4%-6.5%</td>
<td>0.0%-13.2%</td>
</tr>
<tr>
<td>ECM</td>
<td>-1.8%-5.4%</td>
<td>0.0%-9.1%</td>
</tr>
</tbody>
</table>

✓ The gaps between actual & predicted production allow safe planning
✓ Selection for low production after 54 DIM could be unsatisfactory
Prediction improves with time (1) - Correlations mean r between predicted & actual yields (14 herds)

✓ Correlations improve with time in all production variables
✓ Estimates, are estimates, even those carried out at 294 DIM could be erroneous!
Prediction improves with time (2) – Production (APD)
Gaps between predicted & actual 305 d yields (14 herds)

✓ APD (production gaps between actual & estimated yields) were satisfactory for all production variables
✓ Production planning could be corrected with time
Prediction improves with time (3) - Selection for yield

Rates (%) of cows erroneously culled for selection/yield (14 herds)

<table>
<thead>
<tr>
<th>Days in milk to truncation</th>
<th>milk</th>
<th>fat</th>
<th>protein</th>
<th>ECM</th>
</tr>
</thead>
<tbody>
<tr>
<td>34</td>
<td>7.7%</td>
<td>5.2%</td>
<td>5.1%</td>
<td>4.8%</td>
</tr>
<tr>
<td>54</td>
<td>5.2%</td>
<td>5.4%</td>
<td>3.8%</td>
<td>3.7%</td>
</tr>
<tr>
<td>84</td>
<td>5.2%</td>
<td>5.2%</td>
<td>2.5%</td>
<td>3.8%</td>
</tr>
</tbody>
</table>

- ✓ 25% lowest yielders culled for selection
- ✓ Risk of erroneous culling is higher at 34 DIM
- ✓ Prediction improvement with time is not constant
## Selection, erroneous culling (%), Afimilk 34 DIM versus 1st milk test

A Sample Herd, lowest 25% culled

<table>
<thead>
<tr>
<th></th>
<th>1st test (4 to 41 DIM)</th>
<th>Afimilk (34 DIM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>milk</td>
<td>12.2%</td>
<td>24.4%</td>
</tr>
<tr>
<td>fat</td>
<td>26.8%</td>
<td>24.4%</td>
</tr>
<tr>
<td>protein</td>
<td>15.0%</td>
<td>9.8%</td>
</tr>
<tr>
<td>ECM</td>
<td>3.3%</td>
<td>3.3%</td>
</tr>
</tbody>
</table>

✓ Risk of erroneous culling after 1st test is higher than after 34 DIM in Afimilk
✓ Correlations with actual production are higher in Afimilk
How to improve prediction?

➢ Add variables
  ✓ Must have accurate individual data
  ✓ Must have an effect on the outcome

➢ Improve models
  ✓ Individual lactation numbers models vs. Common (all lactations combined models (allowing for lactation number))
**Improve Prediction – Add Variables – Genomics (PTA), 1**

Parameters of APD after 34 DIM, Herd 15 (175 cows)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Herd M (small), APD 34 DIM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>no PTA</td>
</tr>
<tr>
<td>milk</td>
<td>-5.5%</td>
</tr>
<tr>
<td>fat</td>
<td>-5.9%</td>
</tr>
<tr>
<td>protein</td>
<td>-3.0%</td>
</tr>
<tr>
<td>ECM</td>
<td>-4.0%</td>
</tr>
</tbody>
</table>

✓ **APD (Predicted – Actual yields (%))**

✓ **APDs are reduced in all production variables after adding PTA to the models**
Improve Prediction – Add Variables – Genomics (PTA) 2

Rates (%) of cows erroneously culled after 34 DIM, Herd 15, 175 cows

➢ Rate cows culled by mistake = (1 – specificity)

➢ Erroneously culling is reduced in all production variables after adding PTA to the models
Individual vs. Common models (2)—Add PTA to the Models

1st lactation cows erroneously culled by 34 DIM, Herd 10, n=721 cows

- Rates of wrong positive cows (1 – specificity) for selection (cows culled by mistake)
- Improve when Individual replace the common (combined) models
- Adding PTA to the individual cows models takes it even further
## Planning: Annual predicted production (2018) (NOA)

### Selected topics

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>n calved cows</td>
<td>62</td>
<td>65</td>
<td>63</td>
<td>76</td>
<td>100</td>
<td>67</td>
</tr>
<tr>
<td>6</td>
<td>n calved heifers</td>
<td>31</td>
<td>35</td>
<td>38</td>
<td>44</td>
<td>43</td>
<td>26</td>
</tr>
<tr>
<td>12</td>
<td>mean cows in milk</td>
<td>830</td>
<td>852</td>
<td>899.5</td>
<td>810</td>
<td>842</td>
<td>854.5</td>
</tr>
<tr>
<td>13</td>
<td>mean cows</td>
<td>965.5</td>
<td>970.5</td>
<td>979</td>
<td>992</td>
<td>1001.5</td>
<td>1002</td>
</tr>
<tr>
<td>14</td>
<td>mean dry cows</td>
<td>135.5</td>
<td>118.5</td>
<td>79.5</td>
<td>182</td>
<td>159.5</td>
<td>147.5</td>
</tr>
<tr>
<td>17</td>
<td>daily yield cows in milk (l)</td>
<td>32.95</td>
<td>33.91</td>
<td>34.4</td>
<td>30.04</td>
<td>31.01</td>
<td>32.95</td>
</tr>
<tr>
<td>18</td>
<td>daily yield all cows (l)</td>
<td>28.33</td>
<td>29.77</td>
<td>31.61</td>
<td>24.53</td>
<td>26.07</td>
<td>28.1</td>
</tr>
<tr>
<td>24</td>
<td>month marketing (l)</td>
<td>847,819</td>
<td>837,839</td>
<td>959,233</td>
<td>754,292</td>
<td>783,300</td>
<td>872,836</td>
</tr>
<tr>
<td>25</td>
<td>month quota (l)</td>
<td>828,891</td>
<td>757,367</td>
<td>869,061</td>
<td>788,720</td>
<td>783,821</td>
<td>842,608</td>
</tr>
<tr>
<td>28</td>
<td>above quota production (%)</td>
<td>2.3</td>
<td>10.6</td>
<td>10.4</td>
<td>-4.4</td>
<td>-0.1</td>
<td>3.6</td>
</tr>
</tbody>
</table>
Prediction is an ongoing process. The prospective prediction of all production variables improves with time from calving.

Accuracies of models based on AfiMilk MPC™ were higher than those established after the routine monthly milk tests.

Predictions improved by adding cow variables such as lengths of gestational & dry periods, twins & stillbirth, metritis, conductivity, & fat to protein ratio.

Rate of erroneous culling (1-specificity) is the most efficient criteria for an early selection.
Summary & Conclusions (2)

✓ Extreme managemental changes might impair the quality of prediction, screening filters will minimize erroneous culling

✓ Separate calibration models applied to the various parities (1, 2, 3+) are superior to common ones

✓ Addition of genomic information may improve prediction

✓ Predicting in small herds is possible

✓ Accurate data is essential!

✓ The present results allow planning of production, selection & culling in 54 DIM, and even earlier in lactation
From Prediction to Genomics

\[ \text{Phenotype}(t) = \text{Genotype} + \text{Environment}(t, x, y, z \ldots) \]

- **genotype** - the complete heritable genetic identity
- **phenotype** - description of the actual physical characteristics
- **Environment** - the circumstances, objects, or conditions of the surroundings (management, \ldots, \ldots, \ldots)

a) the dependent variables were production traits and diseases;
b) the examined factors were the PTA and the relevant managemental factors
Can we make sure environmental factors are not mistaken for genomics in the selection process?

- Phenotype(t) = Genotype + Environment(t, x, y, z ....)

Is there function F that for a given P_t0, P_t1 can predict P_tn?

\[ F(P_{t0}, P_{t1}, P_{t2}, x_{t0}, x_{t1}, x_{t2}, y, z...) = P_{tn} \]
Selection for low Production is not Genomic Selection!

Herd 15, culled for low phenotypic yields with high PTA. 34 DIM

- Milk: 9.1%
- Fat: 20.0%
- Protein: 10.9%
- ECM: 12.7%

✓ Low phenotypic (actual) production = Lowest 25%
✓ High PTA = Highest 75%
✓ Don’t confuse!
Herd #1. Selection for high milk protein percentage. Mistaking the phenotype for the genotype.

- **protein lower than average first three milk tests:** 14.0%* red, 2.4% blue
- **fat/protein > 1.4 second milk test:** 28.6%† red, 7.7% blue
- **pta % protein:** 1.3% red, 1.1% blue

- **selected n=13**
- **others n=99**
Herd #2. Genetic & Environmental contributions to low milk persistency of first lactation cows.

**potential persistency** 91.1%

- High BCS at calving, 0.3%
- negative energy balance, 0.9%
- young" at calving", 0.5%
- high SCC, 0.4%

**Actual persistency**, 88.5%

PTA for persistency <0.4%

Environment 2.1%

BCS = body condition score; SCC = somatic cell count;
Herd #3. Genetic & Environmental contributions to lameness (up to 180 DIM) in first lactation cows.

actual lameness 26.0%

Potential rate, 14.2%

PTA for F&LC <median, 6.3%

SARA, 0.9%

summer calvings, 5.5%

F&LC = Feet & Leg Composite; SARA = subacute rumenal acidosis
Conclusions (2):

• The key to select the secondary property is the interaction the genomic*the economic importance of the property*the options of improving the phenotypic trait through managerial improvement.

• Improvement in availability and reliability of individual cow data (manual or automatic such as Afilab) will improve quantification of managerial factors to the phenotypic traits.

• In clinical trials, genomics should be taken into account.
Phenotypes, Genotype, and Environment

Where do we go from here??

- **Data**
  - Raw Data
  - Data integrity?

- **Information**
  - Processed Data
  - Analytical on-line Reports

- **Knowledge**
  - Descriptive Modeling
  - What happened?

- **Intelligence**
  - Predictive Modeling
  - Why did it happen?
  - What is going to happen?
  - Optimization

- **Prescriptive modeling**
  - What is the best that could happen?

1980 → 2020

- Where do we go from here??
- Phenotypes, Genotype, and Environment

Vital Know-how in every drop
Thank you!
Q&A