The effect of the number of observations used for Fourier transform infrared model calibration for bovine milk fat composition on the estimated genetic parameters of the predicted data

M. J. M. Rutten,1 H. Bovenhuis, and J. A. M. van Arendonk
Animal Breeding and Genomics Centre, Wageningen University, PO Box 338, 6700 AH Wageningen, the Netherlands

ABSTRACT

Fourier transform infrared spectroscopy is a suitable method to determine bovine milk fat composition. However, the determination of fat composition by gas chromatography, required for calibration of the infrared prediction model, is expensive and labor intensive. It has recently been shown that the number of calibration samples is strongly related to the model’s validation r² (i.e., accuracy of prediction). However, the effect of the number of calibration samples used, and therefore validation r², on the estimated genetic parameters of data predicted using the model needs to be established.

To this end, 235 calibration data subsets of different sizes were sampled: n = 100, n = 250, n = 500, and n = 1,000 calibration samples. Subsequently, these data subsets were used to calibrate fat composition prediction models for 2 specific fatty acids: C16:0 and C18u (where u = unsaturated). Next, genetic parameters were estimated on predicted fat composition data for these fatty acids. Strong relationships between the number of calibration samples and validation r², as well as strong genetic correlations were found. However, the use of n = 100 calibration samples resulted in a broad range of validation r² values and genetic correlations. Subsequent increases of the number of calibration samples resulted in narrowing patterns for validation r² as well as genetic correlations. The use of n = 1,000 calibration samples resulted in estimated genetic correlations varying within a range of 0.10 around the average, which seems acceptable.

Key words: milk, infrared, fatty acid, genetic parameter

INTRODUCTION

Fourier transform infrared (IR) spectroscopy is a suitable method to determine bovine milk fat composition (Soyeurt et al., 2006; Rutten et al., 2009). Infrared spectroscopy is already used for routine recording of the fat, protein, and lactose contents of milk and, therefore, an extension toward routinely measuring milk fat composition is relatively easy and inexpensive. In this way, fat composition data for a large number of animals could be generated. Data on IR-determined fat composition could be used at different levels in the dairy industry. First, individual cows or farms producing milk with a more desirable fat composition could be identified. Their milk could be collected separately; that is, creating separate milk streams, for the production of healthier products. An example of such a product is milk containing a higher amount of unsaturated fatty acids (Campina, 2007), achieved by feeding cows a diet enriched with linseed in this particular example. Second, several studies have reported that a cow’s milk fat composition, or the change in fat composition, has a relationship with her health status (Vlaeminck et al., 2006; Van Haelst et al., 2008). Hence, management information could be derived from IR-determined fat composition in the future. Third, IR-determined fat composition data could be used by dairy breeding organizations to estimate genetic parameters and breeding values. Genetic parameters, specifically genetic correlations between observations determined by GC and predicted by IR, are needed to reveal the potential genetic gain that can be achieved from selection on IR-predicted fat composition.

Received February 10, 2010.
Accepted June 25, 2010.
1Corresponding author: marc.rutten@wur.nl
of breeding values, would enable identification of genetically superior bulls with respect to fat composition. Estimation of breeding values also requires estimated genetic parameters. Even when selection response in fat composition is not targeted at this moment, genetic parameters of, and data on, IR-determined fat composition could be used to monitor changes in milk fat composition due to correlated responses of selection on current breeding goal or selection index traits. Rutten et al. (2009) showed that the number of observations used for calibration of an infrared prediction model for fat composition is strongly related to accuracy of prediction. When the accuracy of prediction is high, differences between predicted and observed fat composition are small. Consequently, estimation of genetic parameters of IR-predicted fat composition will yield roughly the same results as estimation of genetic parameters of fat composition determined by GC. However, when the accuracy of prediction is low, differences between predicted and observed fat composition are larger. Probably, differences between genetic parameters estimated on IR and GC data will get larger in this situation. This relationship, however, has not yet been established. Because GC is an expensive and time-consuming method, a guideline with respect to the number of observations required for calibration of an IR prediction model for fat composition and its effect on estimated genetic parameters is indispensable for animal breeding organizations.

The aim of this study was to establish the relationship between the number of calibration samples used for calibration of an IR prediction model for fat composition and its effect on estimated genetic parameters. To this end, data subsets of different sizes were sampled for calibration of the IR prediction models, and genetic parameters were subsequently estimated on the predictions for fat composition. The estimated genetic parameters reveal the potential genetic gain that can be achieved when selection is on IR predictions instead of GC observations of fat composition.

**MATERIALS AND METHODS**

**Data**

In this study, 1,917 milk samples from individual heifers were available. Milk was collected at one morning milking only. All milk sampling took place in the period from February through March 2005. In this period, the animals were housed indoors and fed silage, and therefore these samples are referred to as “winter” samples. In total, 395 farms in the Netherlands were involved in this study and approximately 5 heifers per farm were sampled, all with at least 87.5% Holstein-Friesian genes. Of those heifers, 888 were sired by 5 proven bulls (each sire having between 102 and 206 daughters in the data set), 840 heifers were sired by 50 young bulls (each sire having between 10 and 24 daughters in the data set), and another 189 heifers were sired by 46 other proven bulls (each sire having between 1 and 32 daughters in the data set). The total pedigree consisted of 26,300 animals.

Milk samples were transported at 4°C and, subsequently, milk fat was extracted on the day of sampling from about 400 mL of milk according to International Organization for Standardisation (ISO) standard 15884 (ISO-IDF, 2002a). Fatty acid methyl esters were prepared from milk fat as described in ISO standard 15884 (ISO-IDF, 2002a) and were analyzed according to ISO standard 15885 (ISO-IDF, 2002b) on a trace GC Ultra chromatograph (Thermo Electron Corporation, Waltham, MA), using a Varian Fame Select column (100 m × 0.25 mm i.d.; Varian Inc., Palo Alto, CA). For more details on the GC procedure, see Rutten et al. (2009). The individual fatty acids that were included in this study were C4:0, C6:0, C8:0, C10:0, C12:0, C14:0, C16:0, C18:0, and C18:1 cis-9. Groups of fatty acids included C6–C12 containing C6:0, C8:0, C10:0, and C12:0; C14–16 containing C14:0 and C16:0; C18u (where u = unsaturated) containing 5 identified C18:1 isomers, C18:2 cis-9,12, C18:3 cis-9,12,15, and conjugated linoleic acid (C18:2 cis-9 trans-11); and the ratio of saturated to unsaturated fatty acids (SFA:UFA). Hereafter, the groups of fatty acids and SFA:UFA will be no longer referred to as groups but simply as fatty acids. Genetic parameters for GC-determined fatty acids were presented in an earlier study by Stoop et al. (2008).

Fourier transform infrared (FTIR) absorption spectra were recorded on 10-mL milk subsamples using MilkoScan FT 6000 equipment (Foss, Hillerød, Denmark) at the certified laboratory of the Milk Control Station (Zutphen, the Netherlands). The FTIR spectra consisted of 1,060 IR frequencies (wavenumbers) ranging from 925 to 5,008 cm⁻¹. For logistical reasons, IR spectra were available for 1,815 of the total 1,917 milk samples.

**IR Prediction Model for Fat Composition**

All fatty acids mentioned above were predicted using the IR prediction model developed by Rutten et al. (2009) in their scenario “AA”. In total 3,622 observations from both winter (n = 1,822) and summer (n = 1,804) were available for calibration of the IR model. The available data were randomly split in 2 parts with an equal number of observations: one part for model calibration and one part for model validation. The vari-
able selection strategy of Höskuldsson (2001) was applied to omit irrelevant independent variables. Independent variables (wavenumbers) were ordered according to their average correlation to the dependent variables (GC fatty acids) expressed on a fat basis (g/100 g). Subsequently, the 200 highest correlated wavenumbers were selected for modeling. Partial least squares was used and the number of latent variables was determined by means of cross-validation in the calibration data set. The use of the IR prediction model enabled the prediction of fat composition for all samples having IR spectra available \((n = 1,815)\). The predictive ability of the IR model was quantified by validation \(r^2\), which was defined as the squared correlation between predicted (IR) and observed (GC) fatty acids. Therefore, validation \(r^2\) was always based on 1,815 paired observations of GC data and IR predictions.

**Genetic Analysis**

A series of bivariate genetic analyses was performed to estimate genetic parameters of GC determined and IR predicted fatty acids. For all fatty acids, the following model, adopted from Stoop et al. (2008), was assumed:

\[
y_{ijklmn} = \mu + \beta_1 \times \text{dim}_i + \beta_2 \times e^{-0.05 \times \text{dim}_i} + \beta_3 \times \text{af}_{ij} + \beta_4 \times \text{af}_{ij}^2 + \text{season}_k + \text{sirecode}_l + f_m + a_n + e_{ijklmn},
\]

where \(y_{ijklmn}\) is an observation of an IR-predicted or GC-determined fatty acid on animal \(n\) of farm \(m\), with sirecode \(l\), age at first calving \(j\) in season \(k\), and DIM \(i\); \(\mu\) is the general mean; \(\text{dim}_i\) is a covariate for DIM, modeled with a Wilmink curve (Wilmink, 1987); \(\text{af}_{ij}\) is a covariate for the effect of age at first calving; \(\text{season}_k\) is a fixed effect with 3 classes for season of calving, summer (June to August 2004), autumn (September to November 2004), and winter (December 2004 to February 2005); \(\text{sirecode}_l\) is a fixed effect accounting for possible differences in genetic level between the groups of proven bull daughters and young bull daughters [3 classes: proven bulls aimed to have at least 200 daughters in the data, test bulls aimed to have at least 20 daughters in the data, and other bulls who did not qualify for one of the previous groups]; \(f_m\) is a random effect of farm \(m\); \(a_n\) is a random additive genetic effect for animal \(n\); and \(e_{ijklmn}\) is a random residual effect. In matrix notation, the model can be written as

\[
y = Xb + Wf + Za + e,
\]

where \(y\) is a vector of observations on a specific fatty acid both predicted by IR and determined by GC; \(X\) is a design matrix for the estimation of \(b\), a vector of fixed effects and regression coefficients; \(f\) is a vector of farm effects; \(a\) is a vector of additive genetic animal effects; \(e\) is a vector of residual effects; and \(W\) and \(Z\) are incidence matrices. The following assumptions were adopted:

\[
E[y] = Xb,
\]

\[
E[f] = E[a] = E[e] = 0,
\]

and

\[
\begin{bmatrix}
\mathbf{f} \\
\mathbf{a}
\end{bmatrix} =
\begin{bmatrix}
\mathbf{I} \otimes \mathbf{F} & 0 & 0 \\
0 & \mathbf{A} \otimes \mathbf{G} & 0 \\
0 & 0 & \mathbf{I} \otimes \mathbf{R}
\end{bmatrix},
\]

where \(\mathbf{F}\) is a \(2 \times 2\) variance-covariance matrix of farm effects for IR and GC fatty acids; \(\mathbf{G}\) is a \(2 \times 2\) variance-covariance matrix of additive genetic animal effects for IR and GC fatty acids; \(\mathbf{R}\) is a \(2 \times 2\) variance-covariance matrix of residual effects for IR and GC fatty acids; \(\mathbf{A}\) is the numerator relationship matrix (Henderson, 1975); \(\mathbf{I}\) is an identity matrix; and \(\otimes\) denotes the Kronecker product operator. Parameters were estimated by REML implemented in ASReml release 2.0 (Gilmour et al., 2006). Intraherd heritabilities (Heringstad et al., 2006) were calculated as \(h_{2\text{H}}^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_r^2}\); that is, ignoring farm variance \(\sigma_r^2\). Hereafter, \(h_{2\text{H}}^2\) will be denoted simply as \(h^2\).

**Resampling Scheme**

To study the effect of the number of observations used for FTIR model calibration on the estimated genetic parameters of the predicted data, a resampling scheme was performed for 2 fatty acids, C16:0 and C18u. The fatty acids C16:0 and C18u represent 2 traits having low and high validation \(r^2\), respectively (Rutten et al., 2009). In this resampling scheme, random sets of calibration samples were selected and IR prediction models were calibrated as described earlier [i.e., using the same 200 wavenumbers as Rutten et al. (2009) selected in their scenario “AA”]. First, 150 subsets of \(n = 100\) calibration samples were sampled; then, 50 subsets of \(n = 250\) calibration samples were sampled; then, 25 subsets of \(n = 500\) calibration samples were sampled; and, finally 10 subsets of \(n = 1000\) calibration samples were sampled. Next, these prediction models were used to
predict C16:0 or C18u (g/100 g) for all available milk samples (n = 1,815). In this way, 235 values for both C16:0 and C18u were predicted for each milk sample. Because C16:0 and C18u were also determined by GC for all milk samples, all 235 IR-predicted values could be paired to the GC-determined values of C16:0 and C18u. All 235 IR prediction models were validated by means of validation \( r^2 \) on all data (n = 1,815). Finally, bivariate genetic analyses were performed on the GC-determined values of C16:0 and C18u and all 235 corresponding IR-predicted values. The focus was especially on the estimated genetic correlations between the GC-determined fatty acids and all IR-predicted fatty acids, for reasons explained in the next section.

**Comparison of Alternative Selection Indices**

In dairy cattle breeding programs, young bulls are usually tested based on the performance of approximately 100 daughters. In the context of selection index theory (Hazel, 1943), this source of information is weighted very heavily in the selection index (i.e., it will get a relatively large b-value) compared with other potential information sources. This is because the term \( n \) (i.e., the number of daughters) appears in the corresponding P-matrix element. In the comparison of selection indices with respect to dairy cattle breeding programs targeting the same breeding goal, it was therefore assumed that it was sufficient to compare indices consisting of elements corresponding only to the information of a group of \( n \) progeny.

Selection index theory states that genetic gain from selection can be calculated as \( \Delta G = i \cdot r_{II} \cdot \sigma_H \), where \( i \) is intensity of selection, \( r_{II} \) is the accuracy of selection, and \( \sigma_H \) is the standard deviation of the breeding goal. Because \( r_{II} = \sqrt{\sigma_I^2 / \sigma_H^2} \) and therefore \( \sigma_I = r_{II} \cdot \sigma_H \), we can alternatively write \( \Delta G = i \cdot \sigma_I \). The latter expression shows that it is actually sufficient to compare only the standard deviations of alternative selection indices to reveal the relative amount of genetic gain that can be expected from selection on the alternative selection index. In the current study, one of the aims was to evaluate the consequences of replacing observations on fat composition determined by GC by IR-predicted fat composition in terms of genetic gain. Derivation of the ratio of index standard deviations leads to the following expression (see Appendix):

\[
\frac{\sigma_{I_{IR}}}{\sigma_{I_{GC}}} = \frac{r_A h_{IR}}{h_{GC}} \times \sqrt{1 + \frac{1}{n} h_{IR}^2 (n-1)}
\]

where \( n \) is the number of daughters, \( r_A \) is the genetic correlation, and \( h_{IR}^2 \) and \( h_{GC}^2 \) are heritabilities of IR- and GC-obtained observations, respectively. Looking at the latter expression, it becomes clear that when \( n \to \infty \),

\[
\frac{\sigma_{I_{IR}}}{\sigma_{I_{GC}}} \to r_A.
\]

The genetic correlation is, therefore, largely determining the relative amount of selection response that can be expected from the use of IR-predicted fat composition instead of GC-determined fat composition in dairy cattle breeding schemes where progeny testing is practiced. Hence, in the next section, the focus is mainly on the genetic correlations between IR and GC obtained observations.

**RESULTS AND DISCUSSION**

**Validation \( r^2 \)**

The term validation \( r^2 \) is used throughout this paper to quantify the predictive ability of the IR prediction models in the current data set. The IR prediction model developed by Rutten et al. (2009) was used to predict fat composition of the milk samples used in this study’s full genetic analysis. The process of validation of the latter model was not carried out in this study and therefore the term “validation \( r^2 \)” in connection with the predictions using this model is strictly wrong and should simply be termed “\( r^2 \)”. However, because the models developed in the resampling scheme were actually validated in this study, the term “validation \( r^2 \)” was chosen in all cases to prevent confusion in terminology.

Validation \( r^2 \) in the resampling scheme was based on the whole data set (i.e., including calibration samples, which is a potential source for upward bias of validation \( r^2 \)). Recalculation of validation \( r^2 \) excluding calibration samples revealed that the differences with validation \( r^2 \) including calibration samples varied between −0.035 and 0.039 for C18u. These differences can be considered small, which is because the number of validation samples in the remainder was still quite large after the exclusion of calibration samples from the validation data set. A benefit of inclusion of calibration samples in the validation data set, in this particular study, was that validation \( r^2 \) could be directly related to the estimated genetic parameters, which are also estimated from all samples. Furthermore, but probably of lesser importance, validation \( r^2 \) values were directly comparable between models based on different numbers of calibration samples.
Relationship of the Number of Calibration Samples Used and Genetic Parameters

In Figure 1, the results from the resampling scheme are plotted for the genetic correlations between C16:0 determined by GC and C16:0 predicted using the IR prediction model. When \( n = 100 \) calibration samples were used, the estimated genetic correlations varied from 0 to 0.75. Also in terms of validation \( r^2 \), the use of \( n = 100 \) calibration samples resulted in some variation: validation \( r^2 \) varied approximately from 0.05 to 0.30. These ranges narrowed in a stepwise manner when \( n = 250 \), \( n = 500 \), and finally when \( n = 1,000 \) calibration samples were used. For C16:0, only the use of \( n = 1,000 \) calibration samples resulted in reasonably invariable estimates of genetic correlations around \( r_A = 0.56 \). Furthermore, Figure 1 shows that there is a nonlinear relationship between the number of calibration samples and validation \( r^2 \) on the one hand and genetic correlation on the other hand. In this case, simple quadratic regressions were fitted without taking precautions for the regression lines exceeding a genetic correlation of 1.

The use of all observations (\( n = 1,815 \)) for calibration resulted in a relatively low validation \( r^2 \) of 0.53 and a corresponding genetic correlation of 0.86. In the case of C16:0, increasing the number of calibration samples translated directly into higher potential genetic gain because of the increase of the genetic correlation. However, there was an end to the increase of validation \( r^2 \) and therefore genetic correlation. Both Soyeurt et al. (2006) and Rutten et al. (2009) have shown that some fatty acids can be modeled more successfully by IR spectra than others. This suggests that after a certain threshold for the number of calibration samples is reached, there is no room for further improvement of validation \( r^2 \).

In Figure 2, the genetic correlations from the resampling scheme for C18u are presented. When \( n = 100 \) calibration samples were used, the estimated genetic correlations varied from 0.02 to 0.99. In terms of validation \( r^2 \), the use of \( n = 100 \) calibration samples resulted...
Both these ranges were larger for C18u than for C16:0. When increasing the number of calibration samples, the same narrowing pattern of genetic correlations could be observed. The increase of the number of calibration samples from \( n = 1,000 \) to \( n = 1,815 \) did increase validation \( r^2 \) but did not increase the genetic correlation. The case of C18u is therefore an example where there is not much room for improvement for either validation \( r^2 \) or genetic correlation after a certain threshold for the number of calibration samples is reached. In general, C16:0 and C18u represent 2 fatty acids that show moderate and excellent prospects, respectively, for modeling using IR spectra (Rutten et al., 2009). Validation \( r^2 \) values for the other fatty acids analyzed in this study lay roughly between those of C16:0 and C18u. Therefore, it is expected that similar plots of genetic correlations for the other fatty acids analyzed in this study would appear intermediate between results shown in Figures 1 and 2.

Both Figures 1 and 2 show that the use of \( n = 1,000 \) calibration samples resulted in genetic correlations that varied within a range of roughly 0.10. This means that deviations from the average “true” genetic correlation are roughly 0.05. Earlier we showed that the ratio of expected genetic gain from selection on observations of fat composition determined by GC or predicted based on IR spectra in dairy breeding programs with progeny testing can be directly quantified by the genetic correlation. Therefore, a maximum deviation of 0.05 for the genetic correlation translates directly to a maximum over- or underestimation of 5% in terms of genetic gain. With the use of \( n = 500 \) calibration samples, the ranges in which the genetic correlations varied increased, which seems to make the choice for \( n = 500 \) calibration samples less suitable (the range was 0.22 for C16:0 and 0.16 for C18u when \( n = 500 \)).

Figures 3 and 4 show plots of heritabilities for C16:0 and C18u obtained from the resampling schemes. Generally, the same narrowing patterns were found for heri-
tabilities as were found for genetic correlations when the number of calibration samples, and therefore validation $r^2$, increases. In contrast to the genetic correlations, heritabilities showed an almost linear relationship to validation $r^2$. These figures show once more that the IR prediction models require roughly $n = 1,000$ calibration samples to be able to also estimate heritabilities that lie within a limited range of values.

The current resampling scheme reveals the potential variation in validation $r^2$ in relation to the number of calibration samples used, but such an exercise is not common practice. In addition, standard errors of validation $r^2$, or any measure of uncertainty of validation $r^2$, are not commonly provided by statistical software packages. Therefore, the current study provides a guideline with respect to the number of calibration samples required when genetic analyses of predicted values of fat composition are targeted. In Figure 5, the fitted regression lines of Figure 1 (C16:0) and Figure 2 (C18u) are plotted together with a line representing their average. Because the 2 regression lines for C16:0 and C18u seem to show the same patterns, and C16:0 and C18u represent 2 fatty acids that show moderate and excellent prospects for modeling using IR spectra, their average pattern can be used as a general guideline for the relationship of genetic correlation and validation $r^2$ for the fatty acids used in this study.

**Potential Genetic Gain in Fat Composition**

In Table 1, $r^2$, intraherd heritabilities for observations on fat composition determined by GC and predicted based on IR spectra, and their genetic correlations are presented where all data was used. All data includes the $n = 1,917$ observations having observations on fat composition determined by GC, of which $n = 1,815$ also had fat composition predicted by IR available. The $r^2$ values varied from 0.43 to 0.77 and were lower than those reported by Rutten et al. (2009) because validation was performed on winter samples only in the current study. The same GC data as used in Stoop et al. (2008) were used and therefore our estimates of intraherd heritabilities based on the data determined by GC were equal to their estimates. The correspond-
ing estimates of intraherd heritabilities for predictions of fat composition based on IR spectra were generally roughly equal or lower. One exception was observed for C18:0, where the intraherd heritability based on IR predictions was larger than the intraherd heritability based on GC data (0.52 vs. 0.23), for no apparent reason.

Soyeurt et al. (2007) presented estimated heritabilities on IR-predicted fat composition data. Four specific fatty acids were defined equally in the latter and the current studies: C12:0, C14:0, C16:0, and C18:0. The heritabilities presented by Soyeurt et al. (2007) were smaller for C12:0 and C14:0 (0.29 and 0.31) and were roughly equal for C16:0 and C18:0 (0.38 and 0.30) compared with those in the current study. Because the data used by Soyeurt et al. (2007) were collected from several non-Holstein-Friesian breeds and because the genetic model was different between both studies, it is hard to place their presented heritabilities in perspective. Soyeurt et al. (2007) did not present any genetic correlations between the predicted and observed traits.

The last column of Table 1 presents genetic correlations, which ranged from 0.77 to 0.99. These genetic correlations indicate that the replacement of observations of fat composition by predictions based on IR spectra will yield, depending on the fatty acids of choice, at least 77% of the potential genetic gain that can be achieved by selection on GC-determined data in dairy breeding programs where progeny testing is practiced. For the fatty acids currently of interest to the dairy industry because of their association with human health (i.e., C14:0, C16:0, C18:0, and SFA:UFA; German and Dillard, 2006), the genetic correlations were 0.98, 0.86, 0.96, and 0.99, respectively. These results show that the use of IR-based prediction of milk fat composition does not compromise the achievable selection response from selection on GC values to a large extent. Considering that the routine recording of IR traits is cheap and relatively straightforward because the equipment and infrastructure are already present, we conclude that IR-based prediction of milk fat composition provides an excellent means for dairy breeding organizations to genetically improve the composition of milk fat.
CONCLUSIONS

A strong relationship between the number of calibration samples and validation $r^2$ on the one hand and genetic correlation on the other hand was found. The use of 1,000 calibration samples for the calibration of an IR prediction model for fat composition on a fat basis (g/100 g) gave the best results in subsequent genetic analyses: estimated genetic correlations varied within a range of 0.10. When observations on fat composition determined by GC were replaced by predictions based on IR spectra, the genetic gain of the human

![Figure 5](image)

**Figure 5.** Fitted regression lines of genetic correlations on validation $r^2$ for C16:0 and C18u (where u = unsaturated; dashed lines; see Figures 1 and 2) and their average (solid line).

<table>
<thead>
<tr>
<th>Trait</th>
<th>$r^2$</th>
<th>$h_{GC}^2$</th>
<th>$h_{IR}^2$</th>
<th>$r_A$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual fatty acids</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C4:0</td>
<td>0.62</td>
<td>0.42 ± 0.09</td>
<td>0.42 ± 0.09</td>
<td>0.94 ± 0.03</td>
</tr>
<tr>
<td>C6:0</td>
<td>0.69</td>
<td>0.46 ± 0.10</td>
<td>0.35 ± 0.09</td>
<td>0.97 ± 0.02</td>
</tr>
<tr>
<td>C8:0</td>
<td>0.70</td>
<td>0.59 ± 0.11</td>
<td>0.38 ± 0.09</td>
<td>0.99 ± 0.01</td>
</tr>
<tr>
<td>C10:0</td>
<td>0.71</td>
<td>0.69 ± 0.12</td>
<td>0.46 ± 0.10</td>
<td>0.98 ± 0.01</td>
</tr>
<tr>
<td>C12:0</td>
<td>0.60</td>
<td>0.65 ± 0.11</td>
<td>0.54 ± 0.11</td>
<td>0.97 ± 0.02</td>
</tr>
<tr>
<td>C14:0</td>
<td>0.66</td>
<td>0.66 ± 0.11</td>
<td>0.50 ± 0.10</td>
<td>0.99 ± 0.01</td>
</tr>
<tr>
<td>C16:0</td>
<td>0.53</td>
<td>0.42 ± 0.11</td>
<td>0.30 ± 0.09</td>
<td>0.86 ± 0.07</td>
</tr>
<tr>
<td>C18:0</td>
<td>0.43</td>
<td>0.24 ± 0.07</td>
<td>0.52 ± 0.10</td>
<td>0.82 ± 0.08</td>
</tr>
<tr>
<td>C18:1 cis-9</td>
<td>0.72</td>
<td>0.22 ± 0.08</td>
<td>0.25 ± 0.08</td>
<td>0.93 ± 0.05</td>
</tr>
<tr>
<td>Groups$^3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C6–12</td>
<td>0.77</td>
<td>0.67 ± 0.11</td>
<td>0.50 ± 0.10</td>
<td>0.99 ± 0.01</td>
</tr>
<tr>
<td>C14–16</td>
<td>0.57</td>
<td>0.14 ± 0.07</td>
<td>0.19 ± 0.07</td>
<td>0.77 ± 0.14</td>
</tr>
<tr>
<td>C18u</td>
<td>0.74</td>
<td>0.23 ± 0.08</td>
<td>0.24 ± 0.08</td>
<td>0.96 ± 0.03</td>
</tr>
<tr>
<td>SFA:UFA</td>
<td>0.77</td>
<td>0.28 ± 0.01</td>
<td>0.26 ± 0.01</td>
<td>0.99 ± 0.01</td>
</tr>
</tbody>
</table>

$^1$Parameter estimate ± standard error. All fatty acids are expressed on a fat basis (g/100 g).

$^2$Validation $r^2$ was calculated as the squared correlation between observations (n = 1,815) of fatty acids determined by GC and predicted by IR. Genetic parameters for observations of fatty acids determined by GC and predicted by IR were estimated simultaneously using a bivariate genetic model.

$^3$Groups: C18u = unsaturated C18 fatty acids; SFA:UFA = ratio of saturated to unsaturated fatty acids.
health-related fatty acids C14:0, C16:0, C18u, and SFA:UFA was only reduced to 98, 86, 96, and 99%, respectively, in dairy breeding schemes where progeny testing is practiced. Given that the routine recording of fat composition predicted based on IR spectra is cheap and relatively straightforward, we conclude that this methodology provides an excellent means for the dairy industry to genetically improve milk fat composition.

ACKNOWLEDGMENTS

This study was part of the Milk Genomics Initiative and the project “Melk op maat,” funded by Wageningen University (Wageningen, the Netherlands), the Dutch Dairy Association (NZO, Zoetermeer, the Netherlands), CRV (cooperative cattle improvement organization, Arnhem, the Netherlands), the Dutch Technology Foundation (STW, Utrecht, the Netherlands), and the provinces of Gelderland (Arnhem, the Netherlands) and Overijssel (Zwolle, the Netherlands). The authors acknowledge the participating herd owners, the Dutch Milk Control Station (Qlip, Zutphen, the Netherlands) and P. Bijma (Animal Breeding and Genomics Centre, Wageningen University, the Netherlands) for valuable suggestions concerning this study.

REFERENCES


Vlaeminck, B.; V. Fievez; S. Tammimga; R. J. Dewhurst; A. van Vuuren; D. De Brabander; and D. Demeyer. 2006. Milk odd- and branched-chain fatty acids in relation to the rumen fermentation pattern. J. Dairy Sci. 89:3954–3964.


APPENDIX

Calculations

Selection index weight (Hazel, 1943) for information on fat composition determined by GC from GC is the phenotypic variance, and $\sigma_2$ is the genetic variance. The index variance is then

$$ \sigma^2_{\text{I}_{GC}} = b_{GC}^T P_{GC} b_{GC} = \left( P_{GC}^{-1} G_{GC} \right)' P_{GC} b_{GC} $$

$$ = G_{GC}' b_{GC} = \frac{1}{2} \sigma^2_{A_{GC}} \left( \frac{1}{2} \frac{nh^2_{GC}}{h^2_{GC} (n-1)} \right) $$

where $\sigma^2_{A_{GC}}$ is the genetic variance. Selection index weight for information on fat composition predicted by infrared from $n$ half-sib daughters,
where the breeding goal is fat composition determined by GC:

\[ b_{IR} = P_{IR}^{-1} G_{IR} \]

\[ = \left( \frac{n}{\sigma_{p_{IR}}^2 + \frac{1}{4} h_{IR}^2 (n-1) \sigma_{p_{IR}}^2} + \frac{1}{4} h_{IR}^2 (n-1) \sigma_{p_{IR}}^2 \right) \times \frac{\sqrt{2}}{2} \sigma_{\gamma_{IR}}^2 h_{GC} \sigma_{\delta_{GC}}. \]  

\[ [A3] \]

The index variance is then

\[ \sigma_{IR}^2 = b_{IR}^T P_{IR} b_{IR} = \left( P_{IR}^{-1} G_{IR} \right)^T P_{IR} b_{IR} = G_{IR}^T b_{IR} \]

\[ = \frac{\sqrt{2}}{2} \sigma_{\gamma_{IR}}^2 h_{GC} \sigma_{\delta_{GC}} \left( \frac{\frac{1}{2} n r_A h_{IR}^2 h_{GC} \sigma_{\delta_{GC}}}{\sigma_{p_{IR}}^2 + \frac{1}{4} h_{IR}^2 (n-1) \sigma_{p_{IR}}^2} \right) \]

\[ = \frac{\sqrt{2}}{2} \sigma_{\gamma_{IR}}^2 \left( \frac{\frac{1}{2} n r_A^2 h_{IR}^2}{1 + \frac{1}{4} h_{IR}^2 (n-1)} \right). \]  

\[ [A4] \]

Using [A2] and [A4], we can write the ratio of index variances as

\[ \frac{\sigma_{IR}^2}{\sigma_{GC}^2} = \frac{\frac{1}{2} n r_A^2 h_{IR}^2}{1 + \frac{1}{4} h_{IR}^2 (n-1)} \]

\[ = \frac{\sqrt{2}}{2} \sigma_{\gamma_{IR}}^2 \left( \frac{\frac{1}{2} n r_A^2 h_{IR}^2}{1 + \frac{1}{4} h_{IR}^2 (n-1)} \right) \]

\[ = \frac{\left( r_A \right)^2 h_{IR}^2 (n-1)}{h_{GC}^2 + \frac{1}{4} h_{IR}^2 h_{GC} (n-1)}. \]  

\[ [A5] \]

and, by taking the square root, we arrive at the ratio of index standard deviations:

\[ \frac{\sigma_{IR}}{\sigma_{GC}} = \frac{r_A h_{IR}}{h_{GC}} \sqrt{1 + \frac{1}{4} h_{IR}^2 (n-1) \sigma_{GC}}. \]

Note that if \( n = 1 \), \( \frac{\sigma_{IR}}{\sigma_{GC}} = \frac{r_A h_{IR}}{h_{GC}} \), and if \( n \to \infty \),

\[ \frac{\sigma_{IR}}{\sigma_{GC}} \to r_A. \]