Milk

AN APPRAISAL OF THE RELIABILITY OF DIFFERENT SYSTEMS OF MEASURING THE COMPOSITION OF MILK IN FARM BULK MILK TANKS IN ONTARIO

M. A. MacGregor and D. A. Head
School of Agricultural Economics
and Extension Education
Ontario Agricultural College
University of Guelph

September 1972

Publication AE/72/16

AN APPRAISAL OF THE RELIABILITY OF DIFFERENT SYSTEMS OF MEASURING THE COMPOSITION OF MILK IN FARM BULK MILK TANKS IN ONTARIO

M.A. MacGregor and D.A. Head

School of Agricultural Economics
and Extension Education
Ontario Agricultural College
University of Guelph

September 1972

Publication AE/72/16

ACKNOWLEDGEMENTS

The research on which this publication is based was conducted in part under the contract for research in Agricultural Economics between the University of Guelph and the Ontario Ministry of Agriculture and Food. It was also supported by a grant-in-aid from the Ontario Milk Marketing Board.

The authors are deeply indebted to many individuals from the Ontario Milk Marketing Board, the Milk Commission of Ontario, and the Royal Dairy in Guelph for their direct assistance and advice in obtaining the basic data for this study. In particular, the Central Milk Testing Laboratory in Guelph was most cooperative in carrying out the analysis of the milk samples.

All of the contributions of the many people involved in this study are gratefully acknowledged. However, the authors accept full responsibility for the results and interpretation of the results herein reported. Any errors or omissions are the complete responsibility of the authors.

TABLE OF CONTENTS

LIST OF TABLES

AN APPRAISAL OF THE RELIABILITY OF DIFFERENT SYSTEMS OF MEASURING THE COMPOSITION OF MILK IN FARM BULK MILK TANKS IN ONTARIO

M. A. MacGregor* and D. A. Head+

I. INTRODUCTION

This study was undertaken in order to provide information concerning the reliability of the present milk sampling system in terms of estimating the amount of butterfat, protein, and lactose shipped by each milk producer in Ontario.

The value of nonfat solids in milk relative to butterfat has been increasing in recent years in Canada, as well as in other countries. In the United States, for example, the wholesale price of butter was eleven percent higher in December, 1970, than it was in December, 1965. Comparative price changes between the same dates for nonfat dry milk and cheddar cheese were eighty-eight percent

---

*Professor of Agricultural Economics, University of Guelph.

+Economist, Economics Branch, Ontario Ministry of Agriculture and Food.

and thirty-six percent respectively.[1] One need only compare retail prices of whole milk with those of two-percent or skim milk, or Canadian cheddar cheese prices with the price of butter, over the past ten or twenty years to demonstrate that relative prices of fat and nonfat components in milk are changing in Canada in a manner similar to that in the United States and elsewhere.

Interest has been shown in determining whether a change in payment practices--from a butterfat-content basis of payment to a multiple-component content basis of payment--should be made in Ontario.[2]

Before any change can take place in basing payments to producers on multiple components in milk, one must know, or be able to estimate, the amounts of fat and nonfat solids in the milk which each producer is selling. Sampling the milk of each producer, and estimating the amount of fat and of nonfat solids in milk samples, is a practical procedure to follow if the criteria of accuracy and reliability are met.

---

[1] Stewart Johnson, "Protein Price Differentials for Milk (Part II)", Dairy Marketing The Co-operative Extension Service, College of Agriculture, The University of Connecticut, February, 1971.

[2] This interest is reflected, in part, by the commissioning of at least one study concerning pricing milk on a multiple component basis. See R.P. Story, and R.D. Aplin, A Report on Multiple Component Differentials in Pricing Milk to The Ontario Milk Marketing Board June, 1969.

The sampling system used in Ontario at the present time, it has been assumed, meets these criteria. However, the system has not been examined in its entirety to determine whether or not the criteria in fact have been met.

There have been numerous studies conducted, and experiments performed, with a view to determining the accuracy of the composite sample estimates. H.G. Webster studied the standard methods of procedure in sampling, preserving and testing milk by the Babcock method and the errors inherent within the test itself.[3] Campbell et al performed experiments to compare the accuracy of composite sample estimates of butterfat with fresh milk sample estimates.[4]

More recently, studies have been undertaken to determine the effects of milk sample environments on results of infra-red analysis for fat, protein and lactose[5], the sampling frequency

---

[3] H.G. Webster, Investigation and Study of Factors Involved in The Accuracy of Sampling, Preserving and Testing Milk for Butterfat Technical Bulletin, Ontario Concentrated Milk Producer's Association, Toronto, March 1945.

[4] H.C. Campbell, George Jaggard, Dewitt Morris, "Accuracy of The Composite Test" Milk Dealer Volume 21 No. 3 December, 1931.

[5] D.A. Biggs, "A Study of the Effects of Milk Sample Environments on Results of Infra-red Analysis for Fat, Protein, and Lactose." (unpublished)

requirements for accurate estimates[6] and the precision and accuracy of infra-red milk analysis.[7]

Because none of these studies, or others, has examined the composite sampling system in toto, the present study was undertaken.

The purposes of the study are twofold:

1. To determine how well or how variably, the present sampling system estimates the true composition of milk at the farm level in terms of butterfat, protein and lactose;

2. To determine, using another sampling system, whether one can reduce variability in estimating the total tank composition of milk at the farm level in terms of butterfat, protein and lactose.

The following chapters describe, in turn, the present sampling system, the approach and methodology used in the analyses, the results of the analyses, and the conclusions drawn from the analyses.

---

[6] D.A. Biggs and J. Denreyer, "Sampling Frequency Requirements for Accurate Estimates of Milk Composition at the Herd Level" Department of Food Science, University of Guelph (unpublished)

[7] D.A. Biggs, "Milk Analysis with the Infra-red Milk Analyser" Journal of Dairy Science, Vol. 50, No. 5 and "Precision and Accuracy of Infra-red Milk Analysis" Department of Food Science, University of Guelph.

II. THE PRESENT SAMPLING SYSTEM[1]

Producer Participation

All producers of fluid and industrial milk in Ontario are necessarily included in the sampling procedure, because payment for shipped milk is based on estimates of butterfat content resulting from infra-red analysis of composite milk samples.

Sampling Responsibility

Milk sampling is the responsibility of the Milk Commission of Ontario. Plant graders, Central Milk Testing Laboratory personnel, and others directly involved in milk sampling act as agents of, or are employed by the Milk Commission of Ontario. Transport drivers are agents of the Ontario Milk Marketing Board.

Payment to producers, based on the estimates provided via the sampling system, is the responsibility of The Ontario Milk Marketing Board.

Sampling Periods

There are twenty-four "test periods" annually, each one of approximately fifteen days duration. The samples which are analysed are composited over the test period, with samples taken

---

[1]This chapter deals specifically with sampling procedures for bulk-milk shipments. Industrial milk shipped in cans is sampled at each plant. Compositing and storage procedures are the same in both cases.

from each producer's milk at the time of milk collection, usually every second day.

Size of Milk Samples

Each individual sample is one ounce in volume. Composite samples should be seven ounces in volume for analysis at the end of the test period.

Sample Collection

Upon arrival at a farm, the milk transporter, who must hold a valid milk grader's certificate, records the volume of milk to be shipped and the temperature at which the milk is being held. He subjectively grades the milk by sight, smell and taste.

Prior to the samples being taken, the milk is agitated for a minimum of five minutes in order to thoroughly mix the milk. A sample of milk is taken from the tank with either a dipper or a straw.[2] One ounce of the sample milk is transferred to a sample vial which is placed in a sample case containing similar samples of milk from other farms on the tranporter's route. Each vial is labelled with the producer's licence number. The sample case must always contain ice to ensure that the milk samples are kept at a temperature between thirty-two degrees Fahrenheit and forty degrees Fahrenheit until the transfer of each sample to the official composite sample is completed. Samples, which must be delivered for compositing within forty-eight hours of

---

[2]The dipper and straw are to be replaced by a Pipette.

their having been collected, are generally delivered the same day.[3]

Sample Compositing and Storage

The individual bulk-tank samples are delivered by the milk transporter to a regulation sample depot, which is generally the plant to which the milk is shipped. At the depot, each one-ounce sample of milk is transferred to an eight-ounce Monjonnier bottle labelled with a producer's licence number which corresponds to the one on the sample vial. The transfer of the contents of the vial to the composite bottle (into which two Dichromate preservative tablets have been placed) is accomplished either by, or in co-operation with, the Plant Milk Grader. In transferring the sample, the milk grader must invert the vial slowly six times to mix the milk, transfer the entire contents to the composite bottle, mix all milk in the composite bottle by a rotary method, and check the temperature to ensure that the sample is not over forty degrees Fahrenheit.

The composite bottles and their contents are held in storage at the plant at a temperature between thirty-two and forty degrees Fahrenheit. At the end of the two-week sampling period the composite

---

[3] Reference Manual for Bulk Milk Graders and Plant Milk Graders, O.D.A.F., Toronto. It is expected that the maximum time limit for delivery will be reduced to twenty-four hours to ensure sample quality. See "Preliminary Report to the Milk Commission of Ontario Quality Committee" January 19, 1971.

milk samples are taken by refrigerated truck to the Central Milk Testing Laboratory for analysis.

Preparation for Infra-red Analysis

Upon arrival at the Central Milk Testing Laboratory, the composite samples are placed in cooler-storage until such time as the analysis is to be made. Immediately prior to analysis, each composite sample is heated to forty degrees Centigrade to allow for more complete dispersion of butterfat throughout the milk. The samples are then homogenized, and are analyzed within five minutes.

Infra-red Analysis

Because the operation of the infra-red milk analysers and the calibration methods employed have been explained more adequately elsewhere only a few general statements will be made here.[4]

Essentially, the infra-red milk analyser is "a conventional double beam spectrometer modified to provide rapid wavelength changes and to make signals linear with component concentrations".[5] The determination of the percentage content in the milk sample of butterfat, protein, and lactose is carried out by measuring the amount of infra-red energy absorbed, in turn, by each of the three components. Results are recorded electronically using a computer program designed for this purpose.

---

[4]For an excellent description of IRMA see D.A. Biggs, "Milk Analysis with the Infra-red Milk Analyser" op. cit.

[5]Ibid

Analysis of a producer's sample is repeated if the butterfat estimate differs by three tenths of one percentage point, or more, from the analysis of the previous sample from that producer. Requests for repeat analysis, which are indicated to the IRMA operator via the computer program, occur for about five percent of the samples analysed.

Calibration

The eight infra-red milk analysers at the Central Milk Testing Laboratory are calibrated to a standard chemical solution each day prior to the commencement of routine milk sample analyses. Calibration or "control" milks are analysed following the calibration to the standard solution. The "control" milks are analysed both chemically and on the infra-red analysers to ensure that the standard solution calibration truly represents the desired calibration for milk.

Results of Infra-red Analysis

The estimates of butterfat content in milk samples are rounded from two decimal places to one prior to their being reported to the Ontario Milk Marketing Board at the end of each test period.[6] In cases where the two semi-monthly estimates average

---

[6] Values of five and over in the second decimal place are rounded up, biasing the estimates to some extent. See Biggs, D.A. "A comparison of previous and present systems for rounding the semi-monthly fat test results for industrial milk shippers."

at five in the second decimal place, payment is made on that basis. At all other times payments for the month are based on the rounded (to one-tenth of one percent) butterfat estimate.

III. OBJECTIVES, METHODOLOGY AND EXPERIMENTAL PROCEDURES

Objectives

Although the implementation of the present sampling system has improved the reliability and accuracy of estimating the component content in milk, there are potential areas within the system from which significant errors could come.[1]

In order to achieve the first purpose stated in Chapter I, it was felt necessary to examine not only the estimates of component content in the samples in the present system, but also the estimates resulting from the analyses of samples prepared in a system similar to it. Thus, a comparison of the estimates from the two systems would provide some measure of the reliability-in terms of estimate variability-of the present system. This was done by means of replication (rejecting the procedure of collecting and analysing milk samples) and of duplication (repeating the infra-red analysis on each sample).

---

[1] D.A. Biggs and J. Denreyer, op. cit.

Replications of both systems provided additional information of within-system variability of the estimates from which were drawn further comparisons between the two systems. It was felt that some measure could be made of the extent to which infra-red analysis at the Central Milk Testing Laboratory provided varying estimates of the component content in milk samples. This task was to be accomplished by duplicating the analyses to determine component content estimates in each sample.

To achieve the second purpose of the study, comparisons were made between the estimates derived from the present sampling system and those derived from a system in which non-composited fresh milk samples were analysed. The latter system was replicated, and duplicated. The estimates, rather than the milk samples, were composited over the period of the study. The comparisons of the estimates provided a measure of the variability within each system, between the replications of each system, and between the duplicated analyses. From each of those, comparisons were made between the two systems. Similar comparisons were made between the fresh milk sampling system and the system (described above) similar to the present composite sampling system.

Sampling Procedure

As was mentioned above, the study was concerned with comparing the estimates of butterfat, protein, and lactose resulting from duplicate infra-red analyses of samples collected in three different sampling systems, each one of which was replicated.

The time period chosen for the sampling was one test period, a fourteen-day span beginning

March 24, 1971, which coincided with the normal test period for those milk producers involved in the study.

Milk samples for each sampling system were collected from the bulk tanks of forty-eight dairy farms every two days during the test period. All of the farms involved were shipping milk to the Royal Dairy in Guelph.

The Present Sampling System Modified

The sampling procedure for the present system was followed, for the most part, as described in Chapter II. Constraints of time and availability of personnel necessitated some minor changes, however. In the opinion of the authors, the modifications to the present system, required for the study, did not detract from, or enhance, the results obtained.[2] Deviations from the normal procedure were as follows:

1. A fieldman rode in the truck with the transport driver,

2. A fieldman was responsible for determining the required amount of agitation prior to sample collection,

---

[2] In order to prove or disprove this point, one would be required to compare the estimates resulting from the analyses of samples collected and handled both under the present system and under the present system as modified here, using the same people at the same time, doing both jobs in the same manner. This task at best would be an extremely difficult one.

3. Two samples of one ounce each were collected by the driver,

4. In addition to the producer licence number, one vial was labelled "A", the other "AR". The "A" vial contained the official milk sample which was to be collected regardless of the study. The "AR" vial contained a sample of milk which was to be handled and analysed in a manner which replicated the present system. Each of the samples were composited in appropriately labelled composite bottles at the Royal Dairy,

5. Milk samples for the other sampling systems used in the study were carried in the truck,

6. Two estimates of component content in the official composite samples were acquired. The first estimate was deemed to be the official one, on which producer payments were to be based,

7. The infra-red analysis was performed off-line, on IRMA's #7 and 8.

The present system with its modifications as described here, and its replicate, were denoted during the study as System A and System AR respectively. This practice will be continued throughout the remaining chapters.

The Composite Sampling System

In System C milk samples were collected and composited in a manner similar to that in System A. There was, however, more control over the samples and their handling than there was over

those in System A. System C and its replicate CR were designed with greater control in order that the estimates derived from the C and CR samples, would represent the ideal results of the present sampling system. The following procedure was adopted for System C and its replicate.

An Ontario Milk Commission fieldman collected, with a dipper, a five-ounce milk sample, for each replicate of System C, at the time of each milk pick-up. Each sample was placed in a sample bottle labelled with the producer's licence number. The samples were placed in a sample case containing ice in order to ensure that the temperature of the samples would remain between thirty-two and forty degrees Fahrenheit. The two five-ounce samples from each farm were taken to the Central Milk Testing Laboratory. After inverting each sample bottle several times to ensure that the milk was thoroughly mixed, one ounce of the sample milk from each bottle was placed in an eight-ounce Monjonnier bottle containing two preserative tablets labelled with the appropriate licence number of the milk producer, and either "C" or "CR". The method of compositing followed by the fieldman was accomplished in the same manner as that outlined in the description of the present sampling system. The C and CR composite samples were stored within the required temperature range in the cooler facilities at the Central Milk Testing Laboratory. The composite samples in System C were analysed off-line at the end of the test period on IRMA's #7 and 8, following machine calibration, at the same time as System A samples were analysed.

In order to facilitate the collecting and handling of the samples in System C, the fieldman

rode with the milk transporter, and was given the responsibility of determining the length of time (greater than five minutes) required for bulk-tank agitation.

The Fresh Milk Sampling System

System B, and its replicate BR, were designed with a view to determining component estimates from non-composited, fresh milk samples.

In System B, a four-ounce milk sample from each farm was collected by a fieldman at the time of milk pick-up concurrently with the samples used in Systems A and C. It was noted above that a fieldman collected two five-ounce samples each milk pick-up, and from each of those samples, one ounce of milk was composited. The remaining four ounces of milk from each sample were placed in a sample bottle labelled "B" or "BR" along with the licence number of the producer from whose bulk-tank the milk was taken.[3] Because the infra-red analyses of the fresh milk samples could be arranged only for Mondays, Wednesdays and Fridays during the test period, each sample bottle used in the B System and its replicate had added to it one preservative tablet.

---

[3]The B and C samples were originally to have been collected separately from each bulk-tank. This was modified because it was felt to add unnecessary pressure to the time and personnel constraints already imposed on the sampling. Needless to say, this modification required employing the assumption of homogenity of the samples taken from adequately agitated milk in each bulk-tank.

For each bulk tank, two sets of seven analyses each were performed, one for each milk sample. The seven estimates in each set were added together, and a simple average of the estimates was determined. That is to say, instead of compositing the milk samples to determine an average estimate of component content over the test period (as was done in Systems A and C), the results of the seven analyses were composited. Thus, the data accrued in System B were comparable to those accrued in Systems A and C.

Methodology and Statistical Design

The three sampling systems described above provided twelve estimates of each of butterfat, protein and lactose content in the milk shipped from each of forty-eight farms over a two-week period.

The analyses of the estimates for each component were separate throughout the study, and no attempt was made to relate one to another the estimates of the various components' content in the milk from each farm.

For each component, there were two groups of analyses performed. The first group of analyses was used to determine differences among estimates of component content in the milk shipped from each farm within each sampling system. Thus three analyses, one for each of Systems A, B, and C, were performed for each component.

The second group of analyses was used to determine differences among estimates of component content in the milk shipped from each farm across all systems. Thus one analysis was performed for

each component.

Both groups of analyses were employed within the context of a randomized block design. In the former, there were forty-eight treatments, one for each farm, and two blocks, one for each replication of a sampling system. In the latter, the replications of all systems were considered as blocks and farms again were treatments.

The data used in the analyses were stated, for butterfat, unadjusted protein, and lactose, to two decimal places. The correction factor for adjusting the protein estimates was built into the computer program.[4]

Analysis of variance with hierarchical classification was used in both groups of analyses to isolate the sources of variation, to quantify the extent to which each source of variation contributed to total variation, and, using F tests, to determine the statistical significance of the variability

---

[4]The protein correction formula is

$$P = (X - 3.60\ (.21) + Y,$$

where P is the corrected protein value,
X is the butterfat estimate for that sample as determined by IRMA,
Y is the uncorrected protein estimate as determined by IRMA,

Adjustment formula was developed for protein because of measurement interferences by other milk components. For a discussion of this aspect of IRMA see Biggs, D.A. "Milk Analysis with the Infra-red Milk Analyzer" Journal of Dairy Science Vol. 50, No. 5, p 799-803

associated with different souces of variation.[5]

In the within-system analyses, the potential sources of variation were identified as among-farm differences, differences due to replication variability and differences between duplicate infra-red analyses. The hierarchical classification in the within-system analyses was such that duplication variability was nested within replication variability, and both of these were nested within among-farm variability.

The analysis of the estimates for each component within each system proceeded in the following manner. For simplicity, the analysis of butterfat estimates in System A is discussed.

There were four estimates of fat content in the milk shipped from each farm in System A, one from each duplicate analysis of two replicates. A simple average value (mean estimate) was calculated for each of the forty-eight farms from which milk samples were taken. The average value of all farm means was determined by summing the forty-eight farm means, and dividing by forty-eight. The resulting value was subtracted from each farm mean and squared. The

---

[5] Hierarchical classification in analysis of variance is discussed in a number of textbooks on Statistics, See for example, Jerome C.R. Li, Statistical Inference I, Edwards Brothers Inc., Ann Arbour, Michigan, 1964 Pp 374ff. The Statistical formulae used in this study were presented in Walter T. Federer, Experimental Design Theory and Application MacMillan Co., New York, 1955 pg. 97.

squared values were summed, providing a measure of the variability which could be attributed to the differences of estimated butterfat content among all forty-eight farms. This measure, known as among-farm sums of squares, will be discussed below in the context of the results of the analyses.

There were forty-eight estimates of fat, one for each farm, resulting from the duplicate infrared analyses of the two composite samples collected in the replicated Sampling System A.

A measure of the variability of fat estimates which could be attributed to replication differences was determined by averaging the value of the butterfat estimates provided by duplicate analyses in each replication, summing these values over the forty-eight farms, dividing by forty-eight to give an average of the mean values of butterfat estimates in each replication, and adding together the summed squared values of the difference between each of the two means and an average of both means. This measure of variability will be referred to below as replication sums of squares.

Duplication, or within-IRMA, variability, as measured by duplication sums of squares, was determined as follows. An average estimate of butterfat for each of the four estimates for all forty-eight farms was computed. The squared differences between each of these four averages and the simple average of all four averages were added together.

For each of the three sources of variation, an average of variability was computed. This was accomplished by dividing the sums of squares for each

source of variation, discussed above, by their respective degrees of freedom.[6] This value is referred to below as the mean square or variance.

Because of the nesting of variability, described previously, it was necessary to remove duplication variability from replication variability, and both of these from among-farm variability, in order to determine whether the various sources of variation significantly contributed to the variation among all estimates.

This separation was accomplished by dividing the among-farm mean square by replication mean square and by dividing replication mean square by duplication mean square. It could be concluded that if the former ratio were larger than one, some variability among the butterfat estimates in System A could be attributed solely to the differences in butterfat content estimates among farms. Similarly, if the latter ratio were larger than one, some variability among the butterfat estimates in System A could be attributed solely to the differences in butterfat content estimates between the two replications. These conclusions follow logically from the fact that the presence of no variability among estimates of butterfat content in System A precludes the computing of sums of squares and the ensuing variances. This would occur only if all estimated

---

[6] For simplicity's sake, "degrees of freedom" may be defined as the least number of deviations which have to be known before the remaining ones can be calculated. For a discussion of degrees of freedom, and its relationship to X^2 distributions, refer to Jerome C.R. li, op. cit.

values were equal. (In that case, all forty-eight farms would be shipping milk containing exactly the same percentage of butterfat, and the sampling system was measuring this equality perfectly, or if differences existed, the sampling system was not able to detect them). If variability exists, and the system can estimate them, sums of squares and mean squares can be computed. Further, if the F ratio $\frac{\text{among-farm mean squares}}{\text{replication mean squares}}$ is equal to one, then the combination replication variability plus duplication variability accounts for all among-farm variability (in which replication and duplication variability are nested) and no among-farm variability per se exists. If the F ratio $\frac{\text{replication mean square}}{\text{duplication mean square}}$ is equal to one, then duplication variability and replication variability are the same.

In testing the statistical significance of the variability which could be attributed to different sources of variation, each of the two F ratios discussed above was compared with the F ratio, for the required degrees of freedom, which would result if the populations from which the samples were taken were normally distributed. A higher value of the latter F ratio indicates a greater statistical significance of the estimates. For example, if one wished to state that the variability accounted for by among-farm differences was statistically significant niney-five percent of the time, the acceptable value of the latter F ratio would be less than if one wanted to state that the same hypothesis held true ninety-nine percent of the time.

For the purpose of this study, a judgement was made that acceptance of the hypothesis would be correct nine hundred and ninety-five times out of one thousand. That is to say, the significance level was 0.5%. This level was chosen so as to minimize the probability of rejecting a true hypothesis.

The hypotheses are stated below, following a discussion of the second group of analyses used.

The between-systems analyses were carried out in the same manner as the within-system analyses described above, except for the fact that the estimates of (say) butterfat for all systems were included. Thus there were twelve rather than four estimates for each farm, and an additional sum of squares and mean square computation required to account for the inclusion of all three sampling systems' estimates in the analysis.

In this group of analyses, replication and duplication variability were nested within the between-systems variability as well as in the among-farm variability. Among-farm variability was not nested within the between-systems variability.

The hypotheses tested in the study were as follows:

(a) For within system analyses, for each of butterfat, protein and lactose

(1) The component-estimate variance among farms is different from that between replications. This hypothesis was deemed to be true 99.5% of the time if

$$\frac{\text{Among-farm mean Squares}}{\text{Replication mean Squares}} \geq 2.12,$$

(2) The component-estimate variance between replicates is different from that between duplicates. This hypothesis was deemed to be true 99.5% of the time if

$$\frac{\text{Replication Mean Squares}}{\text{Duplication Mean Squares}} \geq 1.87$$

(b) For between-systems analyses, for each of butterfat, protein and lactose

(1) The component-estimate variance between systems is different from that between replications. This hypothesis was deemed to be true 99.5% of the time if

$$\frac{\text{Between-systems Mean Squares}}{\text{Replication Mean Squares}} \geq 5.40$$

(2) The component-estimate variance among farms if different from that between replications. This hypothesis was deemed to be true 99.5% of the time if

$$\frac{\text{Among-farm Mean Squares}}{\text{Replication Mean Squares}} \geq 1.32$$

(3) The component-estimate variance between replications is different from that between duplicates. This hypothesis was deemed to be true 99.5% of the time if

$$\frac{\text{Replication Mean Squares}}{\text{Duplication Mean Squares}} \geq 1.00$$

The results of the analyses are presented in the following chapter.

IV. RESULTS OF ANALYSES

Introduction

The previous chapter indicated the methods used in data collection and how the data were analysed. The results of the analyses are presented below in the following manner.

Presented first are the mean estimates of butterfat, protein and lactose content in the milk shipped from the forty-eight dairy farms included in the study. These mean estimates, and the accompanying standard deviation for each, are classified according to the sampling system from which the estimates are derived.[1]

Secondly, the analyses of variance associated with the butterfat estimates are discussed.

---

[1] The standard deviation employed is the root value of the replication variances in each case, thereby indicating within-farm rather than among-farm variation. In the sections discussing the analyses associated with each component, the standard deviation of all the estimates is presented.

Thirdly, the analyses of variance associated with the protein estimates are discussed.

Fourthly, the analyses of variance computed from the lactose estimates are presented.

Tabular results relevant to each section are included in the discussions.

Level of Estimates

It can be seen from table 4-1 below that the mean estimate of butterfat content in the milk shipped from all forty-eight farms was greater in System B than that in Systems A and C. The estimated standard deviation from the mean estimate in System B was less than that in the other two systems. The former result was anticipated and was probably due, to some extent, to lipolysis (decomposition of fat) occuring more in the composite samples than in the fresh milk samples. Only within the range of 3.874 - 3.881 was there an overlap of butterfat estimates in the first standard deviation for all systems.

TABLE 4-1

PERCENTAGE CONTENT IN MILK OF BUTTERFAT, PROTEIN AND LACTOSE ESTIMATED USING THREE SAMPLING SYSTEMS

Component		System A	System B	System C
Butterfat	Mean	3.845	3.892	3.849
	SD*	0.043	0.018	0.032
Protein	Mean	3.392	3.385	3.301
	SD*	0.047	0.011	0.041
Lactose	Mean	4.775	4.942	4.778
	SD*	0.047	0.013	0.034

SD*refers to ($\pm$) standard deviation of the estimates within-farm.

The lactose mean estimates determined in Systems A and C were lower than that determined in System B. While the standard deviation from the mean in System C indicated that two-thirds of the estimate fell within the range in which fell two-thirds of the estimates in System A, there was no overlapping of all three systems within one standard deviation as was the case with the butterfat estimates.

It had been expected that the difference between mean estimates of each component in Systems A and C would be less than that between Systems C and B. Past studies have indicated that this can be attributed to lipolysis in stored milk. It was also expected that the standard deviation from the mean (the square root of the replication variance) would be lower in System B than in Systems A and C. The latter expectation was verified in all cases. The former expectation was verified for butterfat and lactose, but not for protein.

In the case of protein, the difference between values obtained in Systems A and B was less than that between Systems B and C and between Systems A and C.

Analysis Associated with Butterfat Estimates[2]

The mean of all butterfat estimates determined in System A was found to be 3.845, with a

---

[2]See Tables 4-1, 4-2, 4-3, 4-4 and 4-5. The standard deviations in this section were calculated across all farms.

TABLE 4-2

ANALYSIS OF VARIANCE OF BUTTERFAT ESTIMATES IN SYSTEM A

Source of Variation	Degrees of Freedom	Sums of Squares	Mean Squares	Standard Deviation	F
Among Farms	47	39.706	0.845	±0.919	447.1*
Replication	48	0.091	0.00189	±0.043	3.2*
Duplication	96	0.056	0.00059	±0.024	
Total	191	39.852			

* Significant at 0.5%

TABLE 4-3

ANALYSIS OF VARIANCE OF BUTTERFAT ESTIMATES IN SYSTEM C

Source of Variation	Degrees of Freedom	Sums of Squares	Mean Squares	Standard Deviation	F
Among Farms	47	38.234	0.813	±0.902	789.3*
Replication	48	0.049	0.00103	±0.032	2.1*
Duplication	96	0.048	0.00049	±0.022	
Total	191	38.331			

*Significant at 0.5%

TABLE 4-4

ANALYSIS OF VARIANCE OF BUTTERFAT ESTIMATES IN SYSTEM B

Source of Variation	Degrees of Freedom	Sums of Squares	Mean Squares	Standard Deviation	F
Among Farms	47	39.185	0.834	±0.914	2452.9*
Replication	48	0.017	0.00034	±0.018	4.9*
Duplication	96	0.007	0.00007	±0.008	
Total	191	39.209			

* Significant at 0.5%

TABLE 4-5

ANALYSIS OF VARIANCE OF BUTTERFAT ESTIMATES IN ALL SYSTEMS

Source of Variation	Degrees of Freedom	Sums of Squares	Mean Squares	Standard Deviation	F
Between Systems	2	0.259	0.12939	±0.36	118.7*
Among Farms	141	117.125	0.83067	±0.91	762.1*
Replication	144	0.157	0.00109	±0.033	2.9*
Duplication	288	0.111	0.00038	±0.020	
Total	575	117.651			

* Significant at 0.5%

standard deviation of ±0.457.

The mean of all butterfat estimates determined in System C was found to be 3.849, with a standard deviation of ±0.448.

The mean of all butterfat estimates determined in System B was found to be 3.892, with a standard deviation of ±0.453.

These mean values were higher than the one obtained for all Ontario producers in 1970, which was 3.630. This indicates a possible bias in the sample farms chosen. The standard deviation in the case of all farms for the study was representative, in terms of variability, of the whole population.[3]

Overall variability, measured by total sums of squares, was 39.852 in System A, 38.331 in System C and 39.209 in System B. Among-farm variability accounted for 99.6%, 99.7% and 99.9% of total sums of squares in Systems A, C and B respectively. Replication variability accounted for 0.2%, 0.1% and 0.4% of total sums of squares in Systems A, C and B respectively. Duplication (IRMA) variability accounted for 0.1%, 0.1% and 0.02% respectively. There was a highly significant amount of variation among farms in all three systems. Variation between replicates was statistically significant in Systems A, C and B. In all three systems the replicate and duplicate mean squares were extremely small.

---

[3] M.P. Csaba, The Ontario Milk Producers' Butterfat Tests (Preliminary Report) 1970. The Milk Commission of Ontario, ODAF.

F-tests were applied to determine whether replication mean squares between systems were significantly different from one another. It was concluded that the replication mean squares were larger in Systems A and C than those in System B. There was no significant difference between the replication mean squares in System A and the replication mean squares in System C, although the level of significance of System A's replication mean squares compared with System B's replication mean squares was higher than the replication mean squares of System C compared with the replication mean squares in System B.[4]

The grand mean (simple average of the means found in each of Systems A, C and B) for all systems was found to be 3.86176, with a standard deviation of $\pm$0.143. Overall variability, measured by the total sums of squares was 117.651, of which farm to farm variability was 117.125 or 99.6%. Between systems variability accounted for 0.259 or 0.2% of total variability. Replication and duplication variability accounted for approximately 0.1% each of total variability. Farm to farm variability was highly significant. Between-systems and replication variabilities were statistically significant. Replication and duplication mean squares were extremely small.

Because of the existence of significant variation estimates between systems, a further analysis was carried out in order to determine whether most of the difference occurred between

---

[4] See Table 4-14

Systems A and C, or between System B and the average of Systems A and C. The analysis indicated that essentially all of the variability could be attributed to the difference between System B and the average of Systems A and C.[5]

Analysis Associated with Protein Estimates[6]

The mean of all protein estimates in System A was found to be 3.392, with a standard deviation of ±0.256.

The mean of all protein estimates in System C was found to be 3.301, with a standard deviation of ±0.251.

The mean of all protein estimates in System B was found to be 3.385, with a standard deviation of ±0.259.

Overall variability, measured by total sums of squares, was 12.535 in System A, 11.952 in System C and 12.830 in System B. Among-farm variability accounted for 98.5%, 99.06% and 99.86% of total sums of squares in System A, C and B respectively. Between replicates variability accounted for 0.85%, 0.67% and 0.05% of total variation in Systems A, C and B respectively. In all three systems the replication and duplication mean squares were extremely small. F-tests of replication mean squares between

---

[5] See Table 4-15

[6] See Tables 4-6, 4-7, 4-8, 4-9 and footnote

systems resulted in the drawing of the following conclusions. The replication mean squares in Systems A and C were significantly higher than the replication mean square in System B. The level of significance of replication mean squares was higher between System A and System B than between System C and System B. There was no significant difference between the replication mean squares in System A and the replication mean squares in System C.[7]

The grand mean for all systems was 3.35930, with a standard deviation of .259. Overall variability, measured by total sums of squares, was 38,407, of which among-farm variability accounted for 96.6%. Between-systems variability, replication variability and duplication variability accounted for about 2.5%, 0.5% and 0.4% respectively of total sums of squares. Between systems variability and among-farm variability were statistically significant, the F value of the former being larger than the F value of the latter. Replication variability was significant statistically. Replication and duplication mean squares were extremely small.

A further analysis indicated that Systems A and C versus System B accounted for slightly more than 80% of the total sums of squares, and hence most of the variability.[8]

---

[7] See Table 4-14

[8] See Table 4-15

TABLE 4-6

ANALYSIS OF VARIANCE OF PROTEIN ESTIMATES IN SYSTEM A

Source of Variation	Degrees of Freedom	Sums of Squares	Mean Squares	Standard Deviation	F
Among Farms	47	12.348	0.263	±0.513	119.0*
Replication	48	0.106	0.00221	±0.047	2.6*
Duplication	96	0.081	0.00085	±0.029	
Total	191	12.535			

* Significant at 0.5%

TABLE 4-7

ANALYSIS OF VARIANCE OF PROTEIN ESTIMATES IN SYSTEM C

Source Variation	Degrees of Freedom	Sums of Squares	Mean Squares	Standard Deviation	F
Among Farms	47	11.952	0.254	±0.505	153.0*
Replication	48	0.080	0.00166	±0.041	4.7*
Duplication	96	0.033	0.00035	±0.019	
Total	191	12.065			

* Significant at 0.5%

TABLE 4-8

ANALYSIS OF VARIANCE OF PROTEIN ESTIMATES IN SYSTEM B

Source of Variation	Degrees of Freedom	Sums of Squares	Mean Squares	Standard Deviation	F
Among Farms	47	12.813	0.273	±0.523	210.0*
Replication	48	0.006	0.00013	±0.011	1.1**
Duplication	96	0.011	0.00012	±0.011	
Total	191	12.830			

* Significant at 0.5%

TABLE 4-9

ANALYSIS OF VARIANCE OF PROTEIN ESTIMATES IN ALL SYSTEMS

Source of Variation	Degrees of Freedom	Sums of Squares	Mean Squares	Standard Deviation	F
Between Systems	2	0.976	0.48821	±0.221	367.1*
Among Farms	141	37.113	0.26321	±0.513	197.9*
Replication	144	0.192	0.00133	±0.036	3.0*
Duplication	288	0.125	0.00044	±0.021	
Total	575	38.407			

* Significant at 0.5%

Analyses Associated with Lactose Estimates[9]

The mean of all lactose estimates determined in System A was 4.775, with a standard deviation of ±0.16.

The mean of all lactose estimates determined in System C was 4.778, with a standard deviation of ±0.15.

The mean of all lactose estimates determined in System B was 4.942, with a standard deviation of ±0.12.

Overall variability, measured by total sums of squares was 4.336 in System A, 4.223 in System C, and 2.929 in System B. Among-farm variability accounted for 96.0%, 97.7% and 99.3% of total sums of squares in Systems A, C and B respectively. Between-replicate variability accounted for 1.8%, 1.3% and 0.3% of total variability in Systems A, C and B respectively. Duplication (IRMA) variability accounted for 2.1%, 1.0% and 0.4% of total variability in Systems A, C and B respectively. There was a highly significant amount of variability among farms in all three systems. Variation between replicates, as measured by mean squares, was statistically significant in System C, but not in Systems B and A. In all three systems replication and duplication mean squares were extremely small. F-tests of replication mean squares between systems resulted in the conclusions that there was no significant difference between Systems A and C in

---

[9] See Tables 4-1, 4-10, 4-11, 4-12 and 4-13 and footnote 2(p26).

terms of replication mean squares; that replication mean squares in Systems A and C were significantly higher than replication mean squares in System B; the level of significance was higher for replication mean squares in System A compared to System B, than that in System C compared to System B.[10]

The grand mean for all systems was 4.83174, with a standard deviation of ±0.161. Total sums of squares was 14.98, of which 11.199 or 74.8% was accounted for by differences among farms. Between-systems variability accounted for slightly less than 1% each of total variability. Between-systems variability was statistically significant, as was among-farm variability. Replication variability was not significant. Replication and duplication mean squares were extremely small. Further analysis indicated that essentially all of the variation between systems could be attributed to differences between the average of Systems A and C and System B.[11]

V. CONCLUSIONS

The purposes of the study stated in Chapter I were as follows:

1. To determine how well, or how variably, the sampling system estimates the true composition

---

[10] See Table 4-14

[11] See Table 4-15

TABLE 4-10

ANALYSIS OF VARIANCE OF LACTOSE ESTIMATES IN SYSTEM A

Source of Variation	Degrees of Freedom	Sums of Squares	Mean Squares	Standard Deviation	F
Among Farms	47	4.167	0.089	±0.299	54.3*
Replication	48	0.079	0.00164	±0.047	1.7*
Duplication	96	0.091	0.00095	±0.029	
Total	191	4.336			

* Significant at 0.5%

TABLE 4-11

ANALYSIS OF VARIANCE OF LACTOSE ESTIMATES IN SYSTEM C

Source of Variation	Degrees of Freedom	Sums of Squares	Mean Squares	Standard Deviation	F
Among Farms	47	4.124	0.088	±0.296	75.2*
Replication	48	0.056	0.00117	±0.034	2.7*
Duplication	96	0.042	0.00044	±0.021	
Total	191	4.223			

* Significant at 0.5%

TABLE 4-12

ANALYSIS OF VARIANCE OF LACTOSE ESTIMATES IN SYSTEM B

Source of Variation	Degrees of Freedom	Sums of Squares	Mean Squares	Standard Deviation	F
Among Farms	47	2.908	0.062	±0.25	344.0*
Replication	48	0.009	0.00018	±0.013	1.4*
Duplication	96	0.013	0.00013	±0.012	
Total	191	2.929			

* Significant at 0.5%

TABLE 4-13

ANALYSIS OF VARIANCE OF LACTOSE ESTIMATES IN ALL SYSTEMS

Source of Variation	Degrees of Freedom	Sums of Squares	Mean Squares	Standard Deviation	F
Between Systems	2	3.491	1.74565	±1.32	1763.3*
Among Farms	141	11.199	0.07942	±0.28	80.2*
Replication	144	0.143	0.00099	±0.031	1.9*
Duplication	288	0.146	0.00051	±0.023	
Total	575	14.980			

* Significant at 0.5%

TABLE 4-14

TESTS OF HOMOGENEITY OF REPLICATION

MEAN SQUARES

F Ratio of Replication Mean Squares determined by	Butterfat	Protein	Lactose
System A / System B	5.6*	17.0*	9.1*
System A / System C	1.8	1.3	1.4
System C / System B	3.0*	12.8*	6.5*

* Significant (F .005 = 2.13)

TABLE 4-15

SOURCES OF VARIATION AMONG SYSTEMS

Component	Source of Variation	Degrees of Freedom	Sums of Squares	Mean Squares	F
Butterfat	A vs C	1	0	0	0
	AC vs B	1	.259	.259	237.6
	Total	2	.259		
Protein	A vs C	1	.1897	.1897	142.6
	AC vs B	1	.795	.795	597.7
	Total	2	.976		
Lactose	A vs C	1	0	0	0
	AC vs B	1	3.5	3.5	3535.4

of milk at the farm level in terms of butterfat, protein, and lactose,

2. To determine, using another sampling system, whether one can reduce the variability in estimating the total tank composition of milk at the farm level in terms of butterfat, protein and lactose.

The results of analyses performed on estimates from different sampling systems have enabled these purposes to be achieved. Conclusions drawn from the results as they relate to the above purposes are stated below.

Butterfat Estimates in the Present System

The variance of butterfat estimates at the herd level, while statistically significant, was very small, amounting to less than two one-thousandths of one percentage point. The butterfat estimate variability attained in System C, while numerically less than that attained in the present system was not statistically different from it. Therefore, in terms of variability, it may be concluded that the present sampling system is performing adequately the task of estimating butterfat at the herd level.

Protein Estimates in the Present System

At the herd level, the variance in estimates of protein content in milk was small, amounting to less than two one-thousandths of one percentage point. That this variance was statistically significant, results to a large extent, from the fact that IRMA variability was very slight. Replication variability in System C was less than, but not significantly different from, the replication

variability in System A. Thus, in terms of variability, it can be concluded that the present sampling system can adequately estimate protein content in milk.

The different values for protein between Systems A and C could be accounted for by the calibration error in the two infra-red milk analysers that were used for the analysis. All samples in System A were analysed on the same machine; all samples in System C were analysed on another machine. Consequently the calibration error is built into the estimates of the means rather than into the error variance. This could have been avoided by simply randomizing the A and C samples to the two machines for the analysis.

Lactose Estimates in the Present System

As was the case with butterfat and protein, the variability in estimates in each of Systems A and C was very small and the differences between the two were not significant. It may be concluded, therefore, that the present sampling system is adequate in estimating lactose content in milk.

The Composite Sampling System Versus a Fresh Milk Sampling System

The results of the analysis indicated that the fresh milk sampling system used in the study provided estimates of all components with lower variability than did the composite sampling system. Although the greater variability of estimates in the composite system was statistically significant, it may be that non-statistical factors are of greater importance. Thus, while it may be concluded that the fresh milk sampling system is better than the

composite system in terms of estimate variability, the complexity and costs associated with establishing such a system may be greater than the benefits which might accrue because of its establishment. Such an evaluation, if felt to be necessary, could form the basis of further study.

There may be other potential uses for the fresh milk sampling system used in this study. For example, the fact that the estimate variance for each of three components was very small may be an indication of its ability to detect very small changes in component content in milk resulting from changes in (say) feed inputs or management practices.

The fresh milk sampling system provides less variable estimates of butterfat, protein and lactose content than does the composite sampling system. However, the smaller variability can be attributed to the fresh milk itself, and not to the system, because of less lipolysis.

There was no indication that the present system is not highly accurate. The present sampling system as it was observed in this study gives highly reliable estimates of butterfat, protein and lactose content in milk.

BIBLIOGRAPHY

Biggs, D.A. "A comparison of previous and present systems for rounding the semi-monthly fat test results for industrial milk shippers."

_____ "A comparison of Simple and Weighted Average Monthly Fat Tests as a Basis for Payments to Milk Producers"

_____ "A Study of the Effects of Milk Sample Environments on Results of Infrared Analysis for Fat, Protein, and Lactose", Parts I and II.

_____ "Computer Control of Irma Calibration"

_____ "Optimum Conditions for the Handling, Storage and Preparations of Samples for Infrared Analysis of Milk"

_____ "Suggested Procedures for Assessment of IRMA Results by Comparison with Results of Standard Milk Analysis"

_____ "Milk Analysis with the Infrared Milk Analyzer" Reprinted from the Journal of Dairy Science, Vol. 50, No. 5, Pages 799-803.

_____ "Precision and Accuracy of Infrared Milk Analysis" Department of Food Science, University of Guelph

Biggs, D.A. and Denreyer, J. "Sampling Frequency Requirements for Accurate Estimates of Milk Composition at the Herd Level", Department of Food Science, University of Guelph.

Campbell, H.C., George Jaggard and Dewitt Morris "Accuracy of the Composite Test", Milk Dealer, Vol. 21, No. 3, Dec. 1931.

Csaba, M., The Ontario Milk Producer's Butterfat and Protein Tests (Preliminary Report) 1970, The Milk Commission of Ontario, Ontario Department of Agriculture and Food.

Dairy Marketing, The Co-operative Extension Service, College of Agriculture, The University of Connecticut, January, 1960, February, 1968, February 1971, October, 1971.

Federer, Walter T. Experimental Design Theory and Application, MacMillan Co., New York, 1955

Graf, Truman, Trends in Component Pricing of Milk and Implications to The Dairy Industry, Ag. Econ. 40, University of Wisconsin, College of Agriculture, December, 1964.

Li, Jerome C.R., Statistical Inference I, Edwards Brothers Inc. Ann Arbour, Michigan, 1964.

Reference Manual For Bulk Milk Graders and Plant Milk Graders, Ontario Department of Agriculture and Food, Toronto

Story, R.P. and R.D. Aplin, A Report on Multiple Component Differentials in Pricing Milk To The Ontario Milk Marketing Board, June, 1969.

Szijarto, Leslie, Training Manual For The Supervisors and Technicians at the Central Milk Testing Laboratory, Guelph, Ontario. Ontario Department of Agriculture and Food, Toronto.

Webster, H.G. Investigating and Study of Factors Involved in The Accuracy of Sampling, Preserving and Testing Milk for Butterfat Ontario Concentrated Milk Producers' Association, Technical Bulletin, March 1945.