Standard for Estimating Methane Emissions from Enteric Fermentation

California Livestock Methane Measurement, Mitigation and Thriving Environments Research Program (CLIM3ATE-RP)
Dr. Ermias Kebreab
University of California, Davis
Grant Agreement 22-1692-000-SG

This project and a final report were supported by the California Department of Food and Agriculture (CDFA), Office of Agricultural Resilience and Sustainability (OARS) through the CLIM3ATE research program, grant 23-0503-000-SG. The contents are solely the responsibility of the authors and do not necessarily represent the official views of the CDFA.

Contents

About the CLIM3ATE Research Program

The California Livestock Methane Measurement, Mitigation, and Thriving Environments Research Program (CLIM3ATE-RP) is a research funding initiative administered by the California Department of Food and Agriculture's Office of Agricultural Resilience and Sustainability (OARS).

CLIM3ATE-RP was launched with funds from the Budget Act of 2021 (SB 170, Chapter 240) to support applied research that advances California's climate goals and strengthens the long-term environmental and economic sustainability of the state's livestock sector.

Research Program Focus Areas

CLIM3ATE-RP funded research in three critical areas related to methane emissions and manure management in livestock operations. The three impact areas of the CLIM3ATE-RP are:

1. Verification of Methane Reduction Strategies,
2. Alternative Methane Reduction Strategies and
3. Manure Recycling and Innovative Product Development.

In the 2022 funding cycle, CDFA awarded six research projects totaling $4.7 million in funding.

How UCD's Project "Standard for Estimating Methane Emissions from Enteric Fermentation" Accomplishes CLIM3ATE-RP Goals

This project received $74,318 in funding to support Goal 1 of the CLIM3ATE-RP to verify methane reduction strategies by establishing a standard for estimating methane emissions from enteric fermentation.

Project Summary

Recent research findings have shown that Animal Feed Additives may reduce enteric methane emissions by as much as 50 percent. However, evaluation strategies for feed additives are inconsistent across studies, making it difficult to understand the additives' full potential in reducing enteric methane emissions from the livestock sector. These inconsistencies include but are not limited to enteric methane measurement techniques and equipment, treatment and control group sample size, study length, animal type, and animal production stage. This project aims to extend the work supported by CA Air Resources Board by developing a standard for estimating methane emissions from enteric fermentation. This can also be used as a standard for CDFA supported projects related to feed additives.

The following report will provide an overview of the findings of this project. Additionally, the research data published from this work can be found here: Systematic review for optimizing sample size in dairy cow methane emission studies: a comprehensive methodological approach.

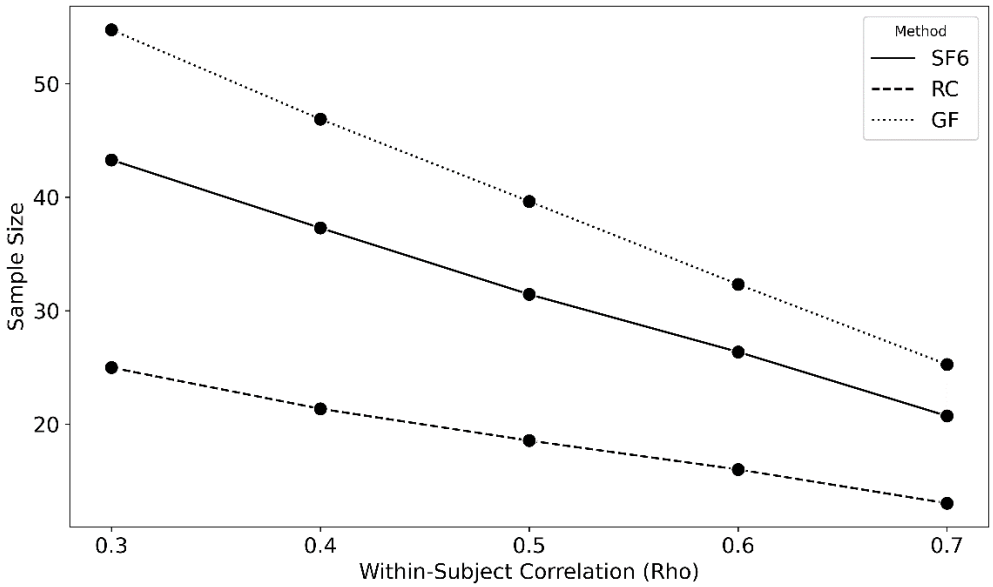
Project Details

PROJECT ACTIVITIES PERFORMED

Task 1: Literature Review
All the literature review was accomplished in previous reporting periods so none to add.

Task 2: Data Collection
All data collection was accomplished in previous reporting periods so none to add.

Task 3: Data Analysis
Based on feedback we got from reviewers, further data analysis was conducted. There was some assumption made on the data analysis and reviewers asked justification for choosing a parameter. In response, we presented the impact of within-subject correlation on the required average sample size for studies using a Latin square design:



Impact of within-subject correlation on the required average sample size for studies using a Latin square design to measure enteric CH_4 in dairy cows. This calculation assumes an expected CH_4 yield reduction of 10%, employing one of three measurement methods: RC (Open-circuit respirometry chambers), GF (the GreenFeed system), or SF6 (the sulfur hexafluoride tracer technique).

Task 4: Interpretation/Discussion
We were asked to clarify interpretation of results and further explanation of the tool. Therefore, we included the following phrases 'The interface actively guides users by providing alerts if their chosen parameters, such as those falling outside the permitted range, need adjustment. Specifically, these alerts are implemented to ensure that the experimental design conforms to the basic principles and requirements essential for the selected design type. For example, for Latin square design, users are prompted through alerts if the number of treatments does not match the number of periods, or if either is less than three, as both conditions are fundamental for the validity of this design type. The underlying database of our tool serves two primary purposes: it acts as a repository of reference values for median methane yields and associated variability, providing a robust benchmark, and also functions as a guideline for users either unfamiliar with these parameters or seeking to validate their sample size assumptions. Contrary to any notion of rigidly imposed "permitted ranges" our design philosophy emphasizes flexibility and user empowerment, allowing researchers to input their own values for methane yield and standard deviation. For instance, should a user wish to explore the effects of a diet with a very low methane yield, such as 5 g/kg of DMI,

this input is entirely permissible within the tool's framework. Similarly, if a user's methodology or prior experience suggests an exceptionally low variability in methane yield measurements such values can also be directly entered into the tool. These user-defined inputs are not overridden or constrained by the database averages; instead, they are welcomed as part of the tool's flexible input mechanism."

Task 5: Reporting

The work has now been accepted in the Journal of Dairy Science. It can be accessed here https://www.journalofdairyscience.org/article/S0022-0302(24)00915-9/fulltext

Results

Database Construction

The selected studies encompass various breeds, including Holstein, Jersey, Finnish Ayrshire, Italian Friesian, Danish Holstein, German Holstein, Brown Swiss, and Nordic Red. Some of these breeds are distinct, while others represent regional variations of the same breed (e.g., different Holstein variations). The predominance of Holstein and its regional variants in the database (85%) reflects the breed's widespread use in dairy production. Nevertheless, the inclusion of other breeds ensures our analysis captures a broader range of genetic diversity.

From the 150 studies selected, six focused on experiments with non-lactating cows. Four studies reported the variability of CH_4 yield as SD. It is important to note that for 25 of the studies, the tables included the largest SEM values. This suggests that the researchers were either pointing out the scenarios with the greatest variability or indicating the least precision in estimating the average of the measured variables. For simplicity, we adopted a conservative approach where these largest SEM are representative of the variability for all groups involved in these studies. Taking into account that some studies present data from two experiments or provide variability information for each group, our database ultimately comprises a total of 177 reports specifically detailing CH_4 yield and its associated variability.

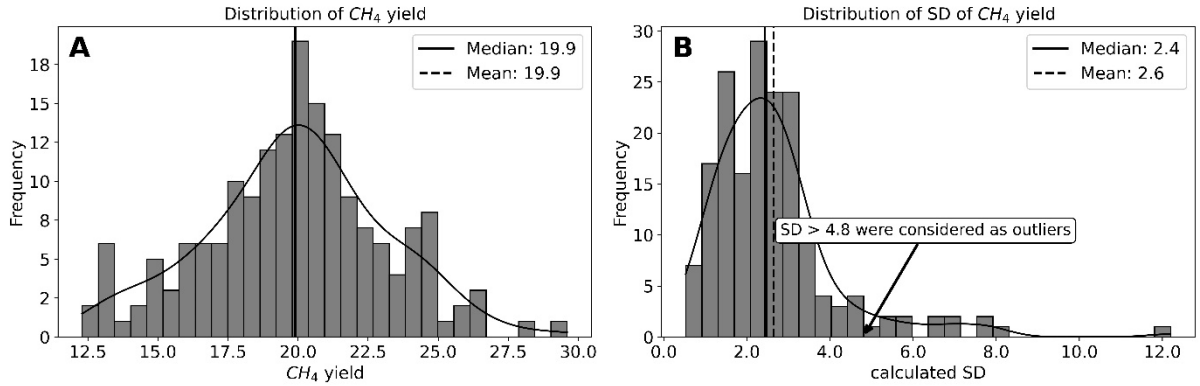


Figure 2 illustrates the distribution of CH_4 yield and its calculated SD in the selected studies. The transformation of SEM to SD is sensitive to the number of measurements (Equation 1), potentially leading to distortions in certain studies. To address this concern, we set a threshold at twice the median value of the SD distribution (i.e., the threshold was 4.8) to identify potentially inflated SD values. Thirteen reports, from 10 studies, with calculated SD values that exceeded this threshold (7.3%) were identified as outliers and excluded from subsequent analyses. This exclusion criterion was chosen to minimize the influence of abnormally high SD values on the sample size calculations. Due to one study contributing two reports, of which one report was identified as an outlier, the remaining database accepted for analysis consisted of a total of 141 studies.

Given the ambiguous use of the "±" sign in the four reports from one of the accepted studies, which leaves unclear whether the value that follows represents the SD or the SEM, we assumed that the authors were referring to SEM. Seven accepted studies, each with one report, reported the variability as standard error (**SE**) of CH_4 yield without specifying whether it was the SEM. Given the common practice of reporting SEM in animal studies and considering the context in which SE was used in these papers, we have inferred that the authors were referring to SEM. These seven studies had an average calculated SD of 2.1, aligning closely with the median SD of the CH_4 yield observed in the database (Figure 2).

The database, as made available on Zenodo (https://zenodo.org/records/10356506), represents the entirety of our assembled dataset, featuring 177 individual reports of CH_4 yield and its variability from 150 studies. This complete database includes both the studies included in our analysis and any that were initially considered but subsequently excluded, ensuring full transparency and accessibility for future research. It is crucial to distinguish between the total number of reports within our database and the subset of reports that were subject to detailed analyses.

Specifically, 13 reports were identified as outliers based on their SD values (i.e., SD greater than 4.8). Therefore, although the database comprises 177 reports, the analyses presented herein are based on a subset of 164 reports.

Database Analysis

A statistical summary of key variables related to CH_4 yield in dairy cows is given in Table 1. These variables include the number of observations (n), per experimental group, used in the statistical analysis, OM, CP, Ether Extract (EE), NDF, CH_4 yield, and calculated SD of CH_4 yield.

Table 1. Descriptive statistics of variables related to CH_4 yield in dairy cows

Statistic	n1	OM2	CP2	EE2,3	NDF2	CH_4 4	SD5
count6	164	132	148	121	146	164	164
Mean	13.6	92.5	16.3	3.8	36.0	19.9	2.3
SD7	23.1	1.5	1.9	1.2	6.6	3.1	0.9
min	3.0	85.1	9.1	1.4	22.9	12.3	0.5
25%	4.0	91.9	15.2	2.8	31.8	18.1	1.6
50%	8.0	92.9	16.2	3.6	34.9	19.9	2.3
75%	12.0	93.3	17.4	4.7	38.0	21.6	2.9
max	203.0	95.5	23.4	6.4	58.8	29.6	4.6

1n = number of observations, per experimental group, used in the statistical analysis.
2Chemical composition of the diets (% of DM).
3EE = Ether Extract.
4CH_4 = CH_4 yield (g/kg of DMI).
5SD = calculated standard deviation of CH_4 yield.
6Number of individual reports of CH_4 yield and its variability after elimination of outliers. Not all reports present data on all feed variables.
7Standard deviation of the variables in the table.

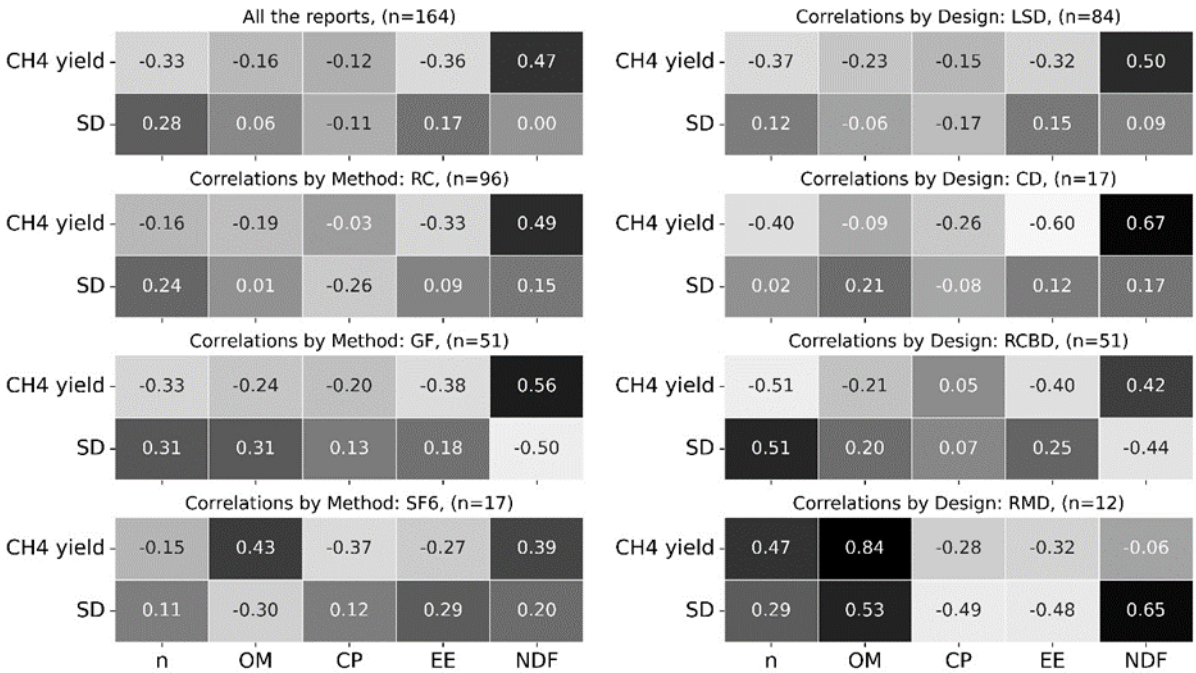


Figure 3 presents eight correlation heatmaps, each depicting two relationships: one between CH_4 yield, number of observations used in the statistical analysis (n), and diet chemical composition, and the other between the SD of CH_4 yield, number of observations, and diet composition. These correlations are shown for the entire database, broken down by measurement method and experimental design. Notably, the negative correlation between CH_4

yield and EE, and the positive correlation between CH_4 yield and NDF were stronger for the GF method and CD design. An inverse correlation between number of observations and CH_4 yield was observed in most cases except for RMD. There was a medium positive relationship between number of observations and SD in RCBD. A medium inverse correlation between NDF and SD was observed in the GF method and RCBD design.

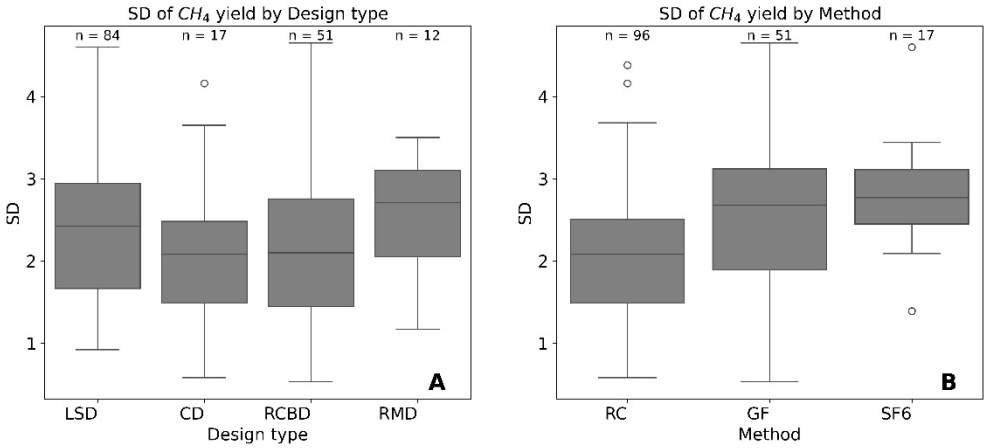


Figure 4. Boxplots of calculated standard deviation (SD) of CH4 yield by experimental design types (A) and measurement methods (B) in dairy cows. LSD = Latin Square Design, RCBD = Randomized Complete Block Design, CD = Crossover Design, RMD = Repeated Measures Design, RC = Open-circuit respirometry chambers, GF = the GreenFeed system, SF6 = the sulphur hexafluoride tracer technique.

The distribution of the CH_4 yield SD across the different experimental designs and measurement methods are given in Figure 4. All designs showed a similar range of SD variability (Figure 4A). The SF6 method exhibited a narrower range of SD, indicating more consistent reports, while RC and GF show a broader range, suggesting to higher variability (Figure 4B). The two-way ANOVA results (i.e., Type III Sums of Squares) revealed that the measurement method (P= 0.04), but not the experimental design type (P = 0.59), had a significant effect on calculated SD. The interaction between design type and measurement method was not statistically significant (P = 0.96). In the multiple comparisons within designs and within methods, only significant difference on calculated SD was detected between RC and GF (P= 0.01) and between RC and SF6 (P = 0.02).

Total Sample Size Calculations

The term "total sample size" is used consistently throughout this document to denote the total number of animals required for an experiment. However, its application differs between within-subject designs (i.e., LSD) and between- subject designs (i.e., RCBD). In LSD, the total number of animals undergo all treatments in a crossover manner, while in RCBD, the total is divided among the treatments, with each animal receiving only one treatment.

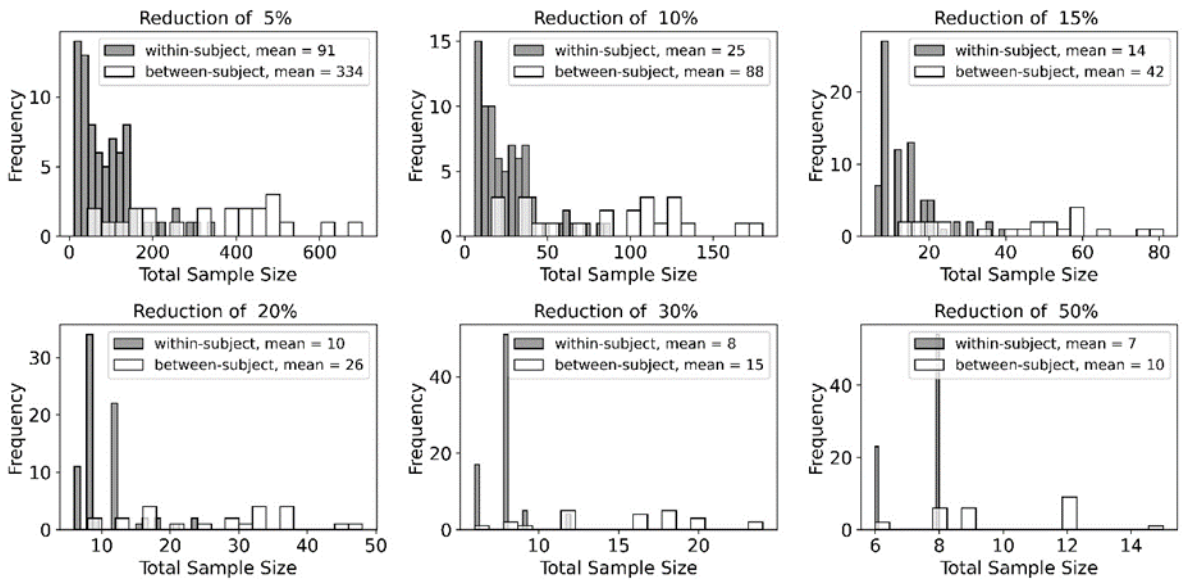


Figure 5. Comparative distribution of required total sample sizes for within-subject (n = 77) and between-subject (n = 25) designs, with three or four treatments, to detect varied CH_4 yield reduction levels (5, 10, 15, 20, 30, and 50%) in dairy cows.

The distributions of the total sample sizes required in balanced LSD (3x3 and 4x4), and RCBD designs with three or four treatments (i.e., 102 reports) to detect six different levels of CH_4 yield reduction, ranging from 5 to 50%, are given in Figure 5. This figure indicates that the mean total sample size is generally lower in within-subject studies compared to between-subject studies. Figure 5 serves as a general guideline for researchers planning three or four-treatment studies, without specifying the measurement method, based on the database developed in the current study.

Table 2. Averages of the total sample size (i.e., total number of animals) required in within- and between-subject designs to detect six levels of CH_4 yield reduction in dairy cows1

CH_4 yield (g/kg of DMI) reduction (%)							
Design type2	n3	5	10	15	20	30	50
Open-circuit respirometry chambers							
LSD	37	45	14	9	8	7	7
RCBD	14	324	85	41	26	15	10
The GreenFeed system							
LSD	14	133	36	19	12	8	7
RCBD	7	287	76	37	23	14	9
The sulphur hexafluoride (SF6) tracer technique							
LSD	12	82	23	13	10	8	8
RCBD	2	572	147	68	41	21	11

1Calculations were performed using experiments with three or four treatments. In LSD, the number of animals per treatment is equal to the total number of animals required in the experiment (i.e., total sample size).

2LSD = Latin Square Design; RCBD = Randomized Complete Block Design.
3n = number of experiments for each combination of method and design after outliers detection.

The analysis takes a step further by breaking down these sample size requirements, categorized by measurement method and experimental design (Table 2). For each combination, Table 2 presents the mean total sample sizes required to detect specific levels of CH_4 yield reduction. Sample size calculations from 16 reports (15.7%) were considered outliers and were not included in the table. For outlier detection, we calculated the first (Q1) and third quartiles (Q3) for each combination of method, design, and number of treatments. We then determined the Interquartile Range (IQR) by subtracting Q1 from Q3. The boundaries for outliers were set at 1.5 times the IQR below Q1 and above Q3, respectively. This threshold is widely accepted for outlier detection (Barbato et al., 2011). Data points falling above the upper bound were specifically considered outliers and subsequently removed. The decision to focus exclusively on upper bound outliers is driven by our objective to identify and exclude data points that might lead to an overestimation of the necessary sample size, ensuring a conservative approach to the efficient use of resources in future experiments.

Delving into an example, the database indicates that for a 4x4 LSD using RC, the average CH_4 yield is 20.8 with a SD of 2.2. From these values, we calculate an effect size of 0.4094 for an expected reduction of 10% in CH_4 yield in three of the four treatments. This leads to a required total sample size of 16 animals to ensure the study is adequately powered (0.95) assuming an ρ correlation of 0.5. Similarly, for RCBD experiments featuring four treatments and RC, the average CH_4 yield is 21.1, with a corresponding SD of 1.9. The resultant effect size of 0.4809 necessitates a more substantial total sample size of 80 animals, equating to 20 animals per treatment.

It is crucial to recognize that, when applying the information from Table 2, the number of reports varies considerably across different combinations of methods and designs. The results offer researchers a reference for the required sample size in both within- and between-subject experiments featuring three or four treatments, to achieve a statistical power of 0.95, assuming the expected CH_4 yield reduction is the same in the experimental groups. For more specific calculations, researchers may use the web-based tool developed in this work.

Sensitivity Analyses

Figure 6 presents plots illustrating the relationship between sample sizes calculations using the average CH_4 yield from all groups versus those derived using the highest group in each study. In all cases, the CCC values were high (all above 0.9), suggesting a strong agreement between the two methods across all expected CH_4 yield reduction percentages. These results are critical for ensuring the robustness of the methodology used in the study. By demonstrating that the sample sizes calculated using all groups are in strong agreement with those using only the highest group. 123 reports, from studies that reported a single CH_4 yield variability and with calculated SD lower than 4.8, were used in this analysis.

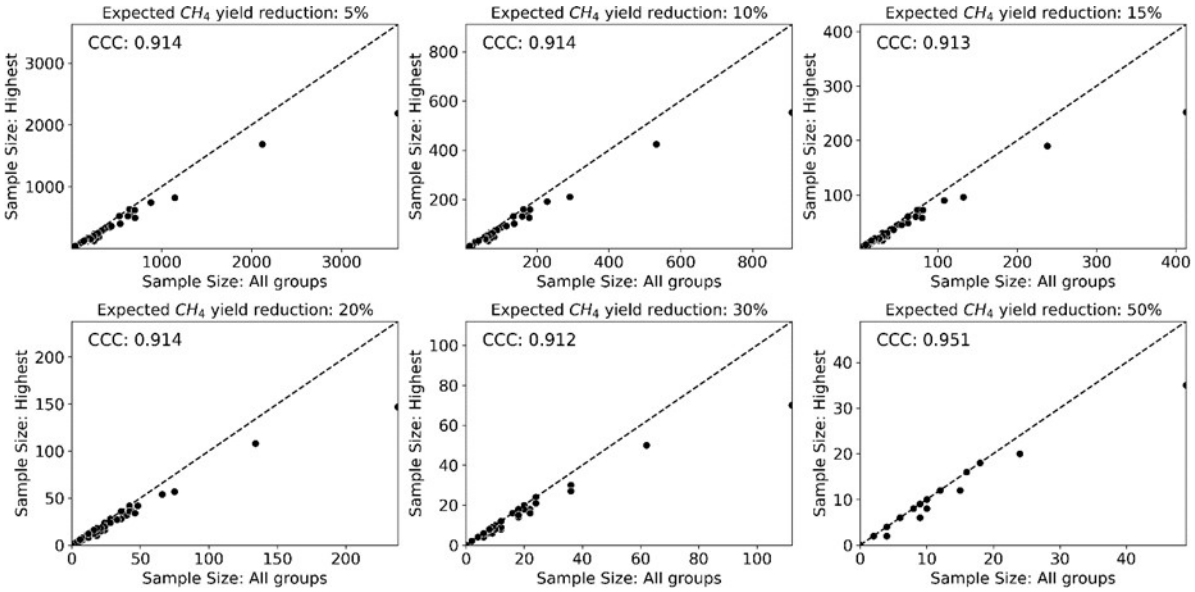


Figure 6. Comparative analysis of sample size calculation using the highest CH_4 yield versus the mean of all groups in dairy cows CH_4 experiments: Concordance Correlation Coefficient (CCC) evaluation across different expected CH_4 yield reduction percentages

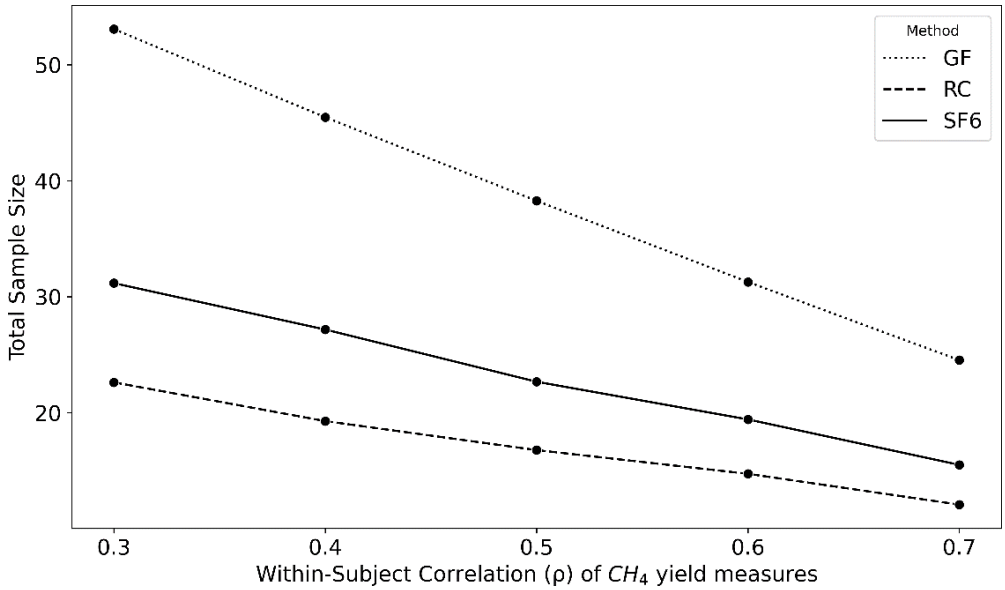


Figure 7. Impact of within-subject correlation of CH_4 yield measures on the required total sample size for studies using 3x3 or 4x4 Latin Square design to measure enteric CH_4 in dairy cows. This calculation assumes an expected CH_4 yield reduction of 10%, employing one of three measurement methods: RC (Open-circuit respirometry chambers), GF (the GreenFeed system), or SF6 (the sulfur hexafluoride tracer technique).

Figure 7 presents the influence of within-subject correlation of CH_4 yield measures on the requisite sample size for detecting a 10% reduction in CH_4 yield using 3x3 or 4x4 LSD. This relationship is represented for the three distinct measurement methodologies: RC, GF, and SF6. The graph portrays a descending trend line for each method, indicating that as the within-subject correlation increases from 0.3 to 0.7, the necessary sample size decreases.

Higher within-subject correlations often mean less variability due to individual differences, enhancing the power to detect treatment effects (Guo, 2013). 77 reports, from 3x3 or 4x4 LSD studies and calculated SD lower than 4.8, were used in this analysis.

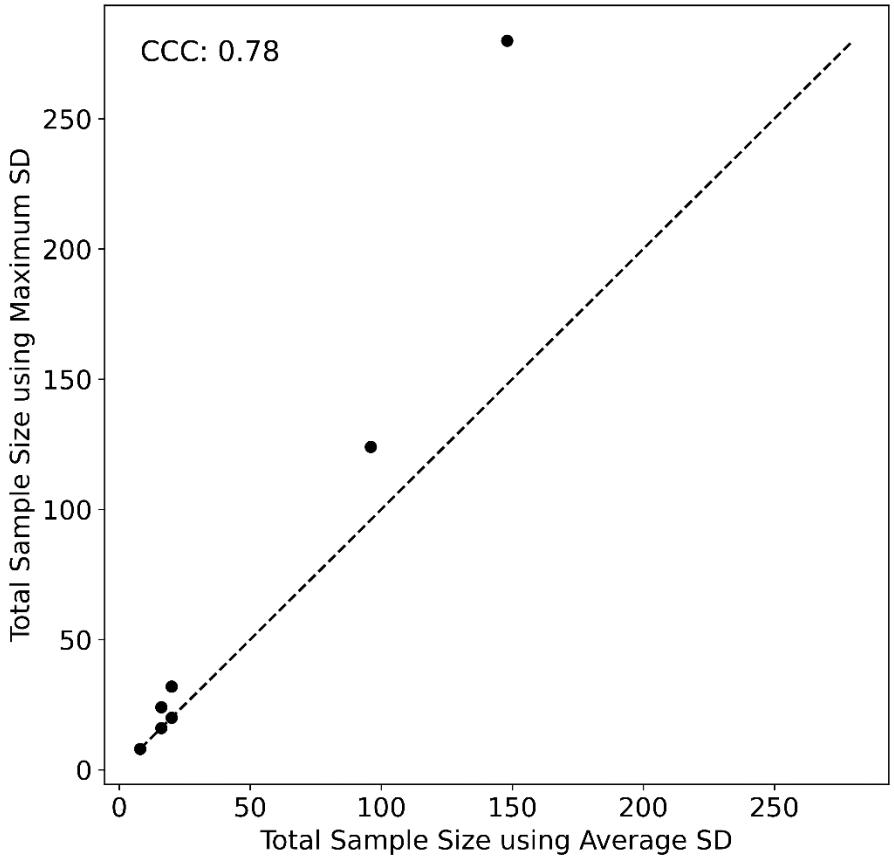


Figure 8. Correlation of total sample size calculations in a 4x4 Latin Square design to measure enteric CH_4 yield in dairy cows. The calculations were performed by comparing the average and the maximum SD of CH_4 yield from seven studies, anticipating a 10% reduction in emissions. Concordance Correlation Coefficient (CCC).

Figure 8 presents a scatter plot that compares sample size estimates derived from two distinct approaches: using the average SD versus using the maximum SD of CH_4 yield. The plot is annotated with a CCC of 0.78, indicative of a substantial positive agreement between the two sample size estimation methods. This value suggests that while there is a good level of concordance, some differences exist which merit consideration. The analysis is, however, limited by the fact that only seven of the studies provide the CH_4 yield variability for individual groups, thereby restricting the breadth and depth of this comparative evaluation.

Web-Based Sample Size Calculation Tool

Several statistical software packages are available for conducting power analysis, including G*Power (Faul et al., 2007), PASS (NCSS, Kaysville, Utah, USA), the Java applets for power and sample size (Lenth, 2001), and specialized functions within R (R Core Team, 2012) and SAS (SAS Institute Inc., Cary, NC). However, for the development of a standalone web-based sample size calculation tool, Python proved to be the more appropriate tool because of its simplicity, vast libraries for statistical and mathematical operations, and excellent frameworks for building web applications.

The web service (samplesizecalculator.ucdavis.edu) offers an interface (Figure 9) designed for researchers and

practitioners interested in conducting experiments that necessitate sample size calculations, even if CH_4 yield is not the main variable of interest. At its core, the tool focuses on systematically determining the optimal sample size required for various experimental designs and CH_4 emission measurement methods. This ensures that experiments have sufficient statistical power to detect meaningful effects. The algorithms embedded within the tool can adjust to accommodate different experimental designs, including LSD, CD, RCBD, and RMD.

One of the standout features of the tool is its dynamic adaptation of input fields based on the user's selection of experimental designs and methods. By selecting a specific combination, the tool retrieves corresponding CH_4 yield and SD values from our database. Moreover, the interface assists users by issuing alerts when their selected parameters, like those exceeding or falling below allowed limits, require adjustments. These alerts aim to guarantee that the experimental setup adheres to the fundamental principles and criteria necessary for the chosen design type. For instance, in the case of LSD, users are prompted through alerts if the number of treatments does not match the number of periods, or if either is less than three, as both conditions are fundamental for the validity of this design type.

For sample size determination, the web tool initiates asynchronous requests to a server endpoint, processing user-provided parameters, including CH_4 yield, expected CH_4 yield reduction, SD of CH_4 yield, alpha level, power, correlation, number of treatments, and periods. Following the computation, the web tool presents a Cohen's value and offers comprehensive recommendations regarding the required sample size to ensure a balanced experimental design. To maintain a streamlined interface, any modifications to input or selection parameters clears previously calculated outputs. There are several ways in which the interface allows triggering of various computations and resets.

The underlying database of our tool serves two primary purposes: it acts as a repository of reference values for median CH_4 yields and associated variability, providing a robust benchmark, and also functions as a guideline for users either unfamiliar with these parameters or seeking to validate their sample size assumptions. Contrary to any notion of rigidly imposed "permitted ranges" our design philosophy emphasizes flexibility and user empowerment, allowing researchers to input their own values for CH_4 yield and SD. For instance, should a user wish to explore the effects of a diet with a very low CH_4 yield, such as 5 g/kg of DMI, this input is entirely permissible within the tool's framework. Similarly, if a user's methodology or prior experience suggests an exceptionally low variability in CH_4 yield measurements such values can also be directly entered into the tool. These user-defined inputs are not overridden or constrained by the database averages; instead, they are welcomed as part of the tool's flexible input mechanism.

Discussion

The main objective of this study was to support the planning of future experiments aimed at assessing the reducing of CH_4 emissions in dairy cows by providing a systematic framework for sample size calculation. Through a comprehensive literature review, extraction of key data, and subsequent database analysis, we have not only created a resourceful database but also demonstrated the utility of this resource in calculating sample sizes for various experimental designs and measurement methods of CH_4 emissions.

To develop the framework, we assume that sample size for future experiments on CH_4 emissions can be calculated by using the mean and variability on CH_4 yield, and the experimental design configuration (i.e., number of treatments and periods) from previous studies and by using Cohen's *d* or Cohen's *f* as the effect size metric in power analysis. To apply this approach in our framework, the first challenge was the lack of SD reports, as only four studies present the variability of CH_4 yield as SD in our database. This was a significant obstacle for calculating sample size since the SD is crucial for determining the effect size using the Cohen's approach. However, we overcome this challenge by using the mathematical relationship between SEM and SD (Barde and Barde, 2012).
In statistical analysis, the SD measures the spread of data points around the mean, while the SEM indicates the precision of the sample mean as an estimate of the population mean. SEM decreases with increasing sample size.

The SEM is often smaller than the SD, which may lead some readers or researchers to erroneously believe that the variability in the data is lower than it actually is. This distinction is critical in power analysis, where SD is the appropriate measure of variability (Altman and Bland, 2005; Nagele, 2003).
In alignment with best practices in statistical analysis, particularly for comparative group studies, it is recommended to calculate and report the variability for each group separately (Rowe, 2023). We found that only seven studies in our database reported the variability on this manner. Furthermore, in 16% of all studies, authors indicate that the SEM values they report are the maximum SEM, which indicates that SEM was computed for each group separately, and the highest SEM was selected for reporting. In contrast, the majority of studies reported a singular SEM value without clarification on its calculation. This leads us to posit that either (a) the SEM was calculated using combined data from all groups, or (b) the SEM was determined for each group independently and the values were sufficiently similar to report a single, averaged SEM for simplicity.

When translating SEM to SD, accurately identifying the method of SEM calculation is pivotal, since the resulting SD can differ substantially, becoming notably larger in studies with more animals. In our framework, we assumed that the reported SEM was calculated by adopting the approach of averaging group-specific SEM. This is based on the presumption that researchers aim to offer a balanced representation of the inherent variability within each group while maintaining simplicity in the results presentation. This approach aligns with established guidelines for scientific paper publication. For example, JDS (2023) specifies that SE should accompany any statistical value, such as a mean or the difference between two means, to facilitate future meta-analyses and provide readers with a measure of the experimental technique's efficiency. These guidelines further recommend that SE for individual means need only be presented separately if the means are based on differing numbers of observations or to highlight error variance heterogeneity.

Using the resulting SD and average CH_4 yield from each study, it was possible to calculate the samples sizes of six expected CH_4 yield reductions for different combinations of CH_4 measurement methods and experimental designs in dairy cows. In these calculations, we employed Cohen's *f*; however, it is noteworthy that for ANOVA, metrices such as Partial Eta-Squared (η^2_p) are frequently preferred. η^2_p provides a nuanced view of the variance explained by specific factors within an ANOVA framework, thus elucidating the proportion of variability accounted for by distinct factors, like treatments, in the presence of covariates (Cohen, 1973). However, similar to Omega squared (ω^2; Kirk, 2005), the calculation of η^2_p necessitates comprehensive statistical details, including sums of squares from an ANOVA output, which extends beyond simple means and SDs. Despite the depth η^2_p and ω^2 offer, our selection of Cohen's *f* as the effect size measure was guided by the information available in our database. This choice, while practical, required adjustment (Equation 4) to account for the complexity inherent in within-subject designs to the extent that η^2_p or ω^2 might achieve. We opted for Cohen's *f* for its practical compatibility with the

available data in our database, despite acknowledging the depth of analysis η^2_p or ω^2 might provide a more detailed analysis. These alternatives require more comprehensive statistical information than we had at our disposal.

In experiments examining effectiveness using within-subject designs, the power analysis of a repeated-measures ANOVA is influenced by several elements. These include the sample size, the number of repeated measurements (k), the average correlation within subjects across these observations (ρ), and the degree to which the sphericity assumption is satisfied (Hertzog, 2008). The terms k and ρ in Equation 4 are important because they account the interdependence of measurements taken from the same animal. The multiplication by k in the numerator serves to amplify the effect size in direct proportion to the number of repeated measures, underpinning the principle that more repeated measurements typically reduce within-subject variability and increase statistical power (Vickers, 2003). However, this increase is tempered by the term $(1-\rho)$ in the denominator, which compensates for the correlation between within-subject measurements. This is logical because with strongly correlated repeated measures, each additional measurement contributes less novel information.

In between-subject designs, the choice of the sample size calculation method hinges only on the number of groups being compared. For two-group comparisons, "t tests: Means: Difference between two independent means" is appropriate. For more than two groups, "ANOVA: Fixed effects, omnibus, one-way" is the method of choice. In the context of an ANOVA, an omnibus test checks if there are any differences among the group means, without specifying where those differences are (DeJarnette and Mamidala, 2023).

The correlation heatmaps (Figure 3) present several intriguing relationships between variables. The observed negative correlation between CH_4 yield and the number of observations (n) used in the statistical analysis, except for RMD, may be that larger sample sizes offer a wider representation of the population, potentially incorporating individuals or conditions that exhibit lower CH_4 emissions, thereby reducing the average CH_4 yield observed in these studies. Additionally, this correlation may not signify a causal link but could instead indicate publication biases; larger studies demonstrating significant CH_4 reduction are more likely to be published. Interestingly, a significant positive correlation between OM and CH_4 yield is observed in RMD and SF6, aligning with the known dynamics of the rumen fermentation process. However, this relationship appears to be slightly inverse for the others designs, a phenomenon that is challenging to reconcile with existing theories, leaving the underpinning reasons for these observations unexplained within the context of the present study. While a positive correlation between OM and CH_4 yield was observed in SF6 and RMD, aligning with the anticipated dynamics of the rumen fermentation process, an inverse relationship was noted in the other designs and methods. A plausible explanation for this negative association could be the specific characteristics of diets. Feeds associated with lower CH_4 yields, such as many concentrate ingredients and corn silage, typically have lower ash contents compared to those associated with higher CH_4 yields, like grass silage (Gastelen et al., 2019).

Differences across measurement methods and designs may be due to various factors, including the sensitivity of the measurement methods (Hristov et al., 2018), the efficiency of the designs, and the inherent variability in the data. The results on sample size calculation indicate that the choice of CH_4 measurement method and experimental design can significantly influence the required sample size (Table 2). Regardless of the type of method, LSD experiments require less animals than RCBD, because the number of animals per treatment is equal to the total number of animals required in the experiment (i.e., total sample size) in LSD.

Table 2 presents an unexpected outcome within the RCBD design when utilizing RC and the GF system for studies with three and four treatments. The conventional expectation, based on the higher accuracy and reduced calculated SD of RC (Figure 4B), would be a smaller required number of animals to detect specific levels of CH_4 yield reduction. While this holds true in the context of LSD, it does not in RCBD. Our analysis supports the fact that RC has a smaller calculated SD than GF, with a P-value of 0.96 indicating no interaction between design type and measurement method on calculated SD. However, across all levels of CH_4 yield reduction presented in Table 2, GF exhibited larger average effect sizes than RC, with an overall average effect size of 1.46 for GF compared to 1.14 for RC. This indicates that for a given percentage reduction, the effect size was higher for GF, which account for the reduced number of animals required to detect the same level of reduction with GF, despite the larger calculated

SD. Similarly, Table 2 provides a contrast within the LSD design, where the SF6 technique requires fewer animals than GF as the latter has an overall smaller average effect size of 0.647 compared to 0.791 for SF6. These findings may appear contradictory at first glance but highlight the complexity of determining sample size, where effect size and SD both play pivotal roles. The larger effect sizes associated with GF, in RCBD, and SF6, in LSD, may provide sufficient power to detect changes with fewer animals, counterbalancing the effects of a larger calculated SD.

All these results have important implications for experimental planning, as it suggests that the choice of measurement method and design can significantly impact the resources required for an experiment. Furthermore, the results highlight the importance of considering the variability within each group when estimating the sample size, underscoring the need for extensive preliminary research to estimate this variability as accurately as possible. By taking these factors into account, researchers can design more efficient and effective experiments, minimizing both Type II errors and unnecessary resource usage.

CHALLENGES AND ACCOMPLISHMENTS

1. *Describe any challenges that occurred during the project term and the corrective actions and/or changes to the project as a result.*
   No challenges except that we started later than anticipated as recruiting a post-doc took longer than expected.

2. *Describe any positive developments that have occurred outside of the project's original intent that you experienced during the reporting period and any project changes as a result.*
   We are now in discussion with Global Methane Hub to extend what we have done in this project to beef as well. We have been asked to give talks and trainings on the tool as it is badly needed in the field.

FUTURE EXPECTATIONS

Describe activities regarding practice implementation and data collection you plan to continue after the project term ends.

The paper is now accepted and available here: Systematic review for optimizing sample size in dairy cow methane emission studies in temperate regions: A comprehensive methodological approach - Journal of Dairy Science

PROJECT SITE PICTURES

This is computer-based study so no pictures to attach.