***Herd enrollment and selection***

The source population for this study was the 145 farms that responded to a survey sent to all certified organic dairy farms producing cow milk in Vermont (n = 177) which aimed to quantify the frequency and diversity of winter housing and bedding types used by organic dairy farmers in the state, as previously described in Andrews et. al (Andrews et al. 2021). Dairy farms were eligible for enrollment in the current study if they: 1) responded to the initial survey in the winter of 2018-2019, 2) indicated they met the enrollment criteria of testing with the Dairy Herd Improvement Association (DHIA) at least monthly and milking between 35 and 120 cows, and 3) indicated they would be interested in further participation. Eligible farms were contacted from this source population if they responded that they were using 1 of the 4 categories of bedding/housing combinations for their winter housing system of interest to the current study: 1) a freestall system with bedded with sand, 2) a freestall system bedded with shavings/sawdust, 3) a tiestall system bedded with shavings/sawdust, or 4) a loose housing system deeply bedded with organic material (hereafter, “bedded pack”). The first three housing and bedding combinations were found to be the top three used by organic dairies in the state to house cows over the non-grazing season, and bedded packs were the primary housing style of interest for this project.

A list of eligible farms was made by housing/bedding combination, and were haphazardly/by convenience (?) contacted by phone or email provided in the previous survey in Spring 2019.

**OR**

A convenience sample was enrolled from a list of eligible farms (grouped by housing/bedding combination) using the phone number or email address provided in the previous survey in Spring 2019.

The aim was to complete the survey and sampling at 40 farms total, with 10 farms from each housing/bedding style: 1) freestall bedded with sand (FS), 2) freestall bedded with wood shavings or sawdust (FW), 3) tiestall bedded with wood shavings (TW), or 4) bedded pack system (BP). As this preliminary study design to select 10 farms of each type was outlined before getting the full results from the initial survey, it was anticipated that it would be possible to select enough organic dairies in Vermont using a bedded pack system as their primary winter housing system. However, out of the 17 farms from 2018-2019 survey that indicated at least some use of a bedded pack system, 1 farm was not interested in any further participation, 5 did not use DHIA testing at all, and 6 only used a bedded pack system as a secondary housing system in conjunction with a tiestall barn, or cows were only on the pack a few hours a day. As the number of farms using the bedding system of interest was significantly smaller than anticipated, the eligibility requirements were relaxed to include a farm where cows spend the majority (two-thirds) of their time in a bedded pack, with the remaining time in a tiestall with wood shavings. The survey was intended to study cows while they were in their winter housing system, so all herds visits were completed before any grazing had begun for the season.

Of the intended 40 herds to be recruited in the study (10 each from each of 4 housing/bedding combination categories: freestalls with sand, freestalls with wood shavings/sawdust, bedded packs, and tiestalls with wood), 21 herds that were contacted agreed and the survey and sampling were completed throughout April-May, 2019. All herds sampled during this period were housing their cows as they would in the winter months. Completion of the survey and sampling was suspended in mid-May, as farms began turning their cows out to pasture, with the intention of resuming in Spring 2020 to complete the remaining 19 herds. However, due to the COVID-19 pandemic, the decision was made to not resume the survey and sampling, and the final analysis included herds only completed in 2019 (n = 21). As there was only 1 farm sampled using a freestall facility bedded with sand, the initial plan to group farms by the four housing/bedding combinations specified was abandoned in favor of grouping farms by the three facility types used (tiestall, freestall, and bedded pack).

***Survey administration, sampling, and udder hygiene scoring***

At each farm visit, a survey was administered by the primary researcher at every visit which collected information which aimed to: 1) get a comprehensive understanding of factors potentially related to a cow’s mastitis risk on that particular farm, and 2) acquire a comprehensive understanding of housing and bedding management and related practices on the farm. The survey is included in its entirety in the provided in Supplemental Data (XXX). The survey was created and administered on a tablet using KoboCollect software, a free and open-source suite of tools for field data collection (http://www.kobotoolbox.org). The section about mastitis risk included producer concerns about bedding/mastitis risk; mastitis control, identification and record keeping; milking facilities, procedures, and hygiene practices; information about diet, vitamin and mineral supplementation, and water source; typical calving and periparturient practices; and fly control. The section about housing and bedding management included describing type of housing system used for both lactating and dry cows; classification and description of any bedding material used; and bedding management practices for each housing type used. The survey also collected some basic herd information (production numbers; number of lactating, dry, and youngstock; breeds; record-keeping systems). As the focus of this study was the use of bedded pack systems by organic dairy producers in Vermont, some additional questions were asked of these farms to gather more detailed information about management, monitoring, impressions comparing them to previously used systems, and initial construction of the pack. Completion of the survey took about 45 minutes on average, but ranged from roughly 30 minutes to 1.5 hours.

While the producer and primary researcher completed the survey, a technician collected a bulk tank milk sample was collected directly from the top of the bulk tank after agitating the milk for at least 5 minutes, using a 250-mL sterile single-use vial (Blue Dippas™, Dynalon Products, England). Samples were kept on ice in a cooler until they could be frozen and stored at −20°C in the laboratory before being sent to a diagnostic lab for analysis. Also during this time, the technician completed an on-farm observation sheet, which collected information about the bulk tank information, cow identification, air quality, and any outdoor exercise area. Additionally, measurements of the housing facilities were recorded for freestalls and tiestalls (stall sizes, pen sizes, stocking density, trainer use), as well as observations about bedded packs when applicable (temperature, depth, sq. ft per animal). Hygiene scoring was completed by the same technician on the day of every visit, with a minimum of 30 randomly selected cows housed in the same pens from which used bedding samples were collected. A 4-point scoring system described by Schreiner and Ruegg was used, where 1 = free of dirt, 2 = slightly dirty (2–10% of surface area), 3 = moderately covered with dirt (10–30% of surface area), and 4 = covered with caked on-dirt (>30% of surface area) (Schreiner and Ruegg 2002). Materials were left with producers to record and collect milk samples of cows with clinical mastitis in the 30 days following the farm visit, but participation in this aspect of the study was too low to include in any analyses.

***Herd-level udder health measurements***

Herd-level DHIA test results for the test day closest in proximity to the farm visit (less than 30 days) were captured from the record processing center working with each herd (Lancaster DHIA, Manheim, PA; Dairy One Co-Op. Inc., Ithaca, NY). Information captured included test dates, number of lactating cows, standardized 150-day milk production, test-day average cow-level linear score (LS; unweighted), average test day milk yield (lbs/cow), and weighted average somatic cell-count (cow-level). The weighted average somatic cell count (cow-level) was used to calculate the weighted average linear score (cow-level) for a herd. The following udder health measures were also captured from DHIA records: proportion of cows with an intramammary infection on most recent test day (“any IMI”), where infection was defined as a linear score of ≥4.0; the proportion of cows with a new IMI (“new IMI”), where a new IMI was defined as a LS changing from <4.0 to ≥4.0 over the last 2 tests; and the proportion of cows with a chronic intramammary infection (“chronic IMI”), where a chronic IMI was defined as having a LS ≥4.0 on the last two tests.

***Bulk Tank Milk Culture***

Frozen bulk tank milk samples were shipped on ice to the Laboratory for Udder Health (University of Minnesota Veterinary Diagnostic Laboratory, St. Paul) for analysis. Methodology for bulk tank milk cultures at the Laboratory of Udder Health have been thoroughly described elsewhere (Patel et al. 2019). Briefly, thawed, room-temperature bulk tank milk and a 10-fold dilution of each bulk tank milk sample were plated onto MacConkey, Factor (gram-positive selective agar; University of Minnesota, 2016), and Focus (University of Minnesota, St. Paul, selective for SSLO bacteria) media plates and incubated for 2 days at 37°C. Any lactose-fermenting (pink-colored) colonies on MacConkey medium were counted and reported as coliform bacteria. Any β-hemolytic colonies on Focus medium were counted and identified to the species level using a MALDI Biotyper (suspect *Streptococcus agalactiae*), while β-Hemolytic colonies on Factor medium were counted and identified to the species level using a MALDI Biotyper (suspect *Staph. aureus*). Any β-Hemolytic colonies with a confidence score ≥2.0 for *Staph. aureus* were counted and reported as such. All remaining colonies on Focus medium that were not identified as *Strep. agalactiae* were counted and recorded as streptococci or strep-like organisms, and non-hemolytic colonies on Factor media of *Staphylococcus* spp. (based on colony morphology, catalase reaction, or Gram stain) were counted and reported as NAS. Bulk tank samples were also cultured for *Mycoplasma* spp. (0.1 mL milk was swabbed across a Mycoplasma agar plate, then placed in a 7% CO2 incubator at 37°C for 7 days, after which they were examined for *Mycoplasma* spp. by a trained microbiology technician). For each bulk tank milk sample, counts were generated for coliform organisms, non-*aureus* staphylococci (NAS), streptococci and strep-like organisms (SSLO), *Staph. aureus*, *Strep. agalactiae*, and *Mycoplasma* spp. as total colony-forming units per mL. The lower threshold of detection for this bulk tank milk culture protocol was 5 cfu/mL, and the upper threshold was 62,500 cfu/mL.

***Data management and analysis***

Survey data collected through KoboCollect software (KoboCollect 2019) was downloaded as an Excel worksheet (Microsoft Corp., Redmond, WA), which contained the information from the questionnaire covering herd information, description of housing, bedding, and bedding management, as well as milking hygiene and mastitis control practices. Udder hygiene scores for individual cows at each farm were used to calculate both a mean udder hygiene score for that farm, as well as the proportion of cows with dirty udders (udder hygiene score ≥3) for each farm. Bulk tank milk culture data from the U. Minnesota Veterinary Diagnostic lab, DHIA test results, and farm-level udder hygiene outcomes were entered into an Excel database, and the accompanying data for each farm from the questionnaire was then entered into this database to combine the outcomes and possible predictor variables for each of the 21 farms. This Excel database containing questionnaire data, udder health, hygiene, and bulk tank milk findings was then imported into the R Statistical Programming Environment (R Core Team, 2023) for data cleaning, checking, and statistical analysis. The distribution of outcome variables was visually assessed in R to check for normality, and descriptive statistics were calculated to evaluate the distribution and data integrity and to identify missing data (means, variances, percentiles for numeric continuous variables, frequencies tabulations and percentages for categorical variables). Descriptive statistics generated included description of general herd characteristics/farm traits, lactating cow housing/facilities, lactating cow bedding/bedding management practices, milking hygiene procedures, and mastitis control practices for all 21 herds included in the study. Additionally, descriptive statistics were also produced to describe udder hygiene, bulk tank milk culture, and DHIA udder health outcomes, both for all herds (n = 21) and stratified by facility type (freestall, bedded pack, tiestall). Distribution of the raw somatic cell count (SCC) data, log2 transformed SCC data, and log10 transformed SCC data was assessed, and all were found to be similarly close to being normally distributed; therefore, raw SCC data was chosen for ease of interpretation. Continuous variables underwent correlation analysis to identify variables that were highly associated (R2 ≥ 0.60), and unconditional associations among categorical variables were evaluated using a Pearson’s chi-squared or Fischer’s Exact test as appropriate (p ≤ 0.05). An ANOVA was used to check for correlation between numeric continuous variables and categorical variables (p ≤ 0.05). When a categorical variable had many categories with a small number of observations in each, categories were combined when biologically plausible/reasonable in an attempt to have all categories of predictor variables contain at least 5 observations. If any predictor had only 1 observation in a group and there was no way to combine groups in a logical way, it was discluded from further analysis (but listed in descriptive statistic tables).

Univariate linear regression was performed in R using the “lme4” package to investigate the unconditional relationship between the six (seven?) udder health (+/- and production, if use STD150 milk) (BTSCC, avg. weighted LS; average unweighted LS; new IMI, any IMI, chronic IMI) and two hygiene outcomes of interest (mean hygiene score and proportion of dirty udders for each farm) and the previously-described herd-level independent variables. Any explanatory variable that was unconditionally associated with 1 or more of the outcomes of interest at p < 0.20 was then offered into a multivariable model investigating the relationship between the udder health or hygiene outcome and the herd-level predictor variables. Facility type was forced into these multivariable models, as it was the primary explanatory predictor of interest. A backward stepwise variable selection process was used, with the least significant variables being removed one by one until all remaining predictors had P ≤ 0.10. … Due to over parametrization and lack of statistical significance due to limited sample size, the multivariable modelling approach to investigate the conditional relationship between facility type and outcome of interest was abandoned in favor of reporting the univariate analysis results for all farms combined together (n = 21). Unconditional comparisons of bulk tank udder health measures, aerobic culture data, and hygiene scores by facility type was then carried out using an ANOVA test. Data was first checked to ensure the assumptions for ANOVA were met, including: (1) presence of outliers, (2) normality using a Shapiro-Wilk test (p ≤ 0.05), and (3) homogeneity of variances using Levene’s test and visual assessment of a residuals vs. fitted values plot. The Tukey method was used for adjusting p-values for multiple comparisons (“TukeyHSD” function of the “stats” package in R).