***Text (or my notes) removed from Materials and Methods:***

Not sure how I’m handling culture results as of yet:

Because BBC (cfu/cm3) and BTM culture results (cfu/mL) were not normally distributed, results were transformed (log10) before further analysis.

***Model building notes:***

* 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, were discluded from further analysis (but listed in descriptive statistic tables)
* univariate analysis used to screen predictors, if unconditionally associated at a level of p<0.2 (using linear regression, single predictor for 6 UH outcomes which were numeric, continuous) were candidates for inclusion in multivariable model
* predictors that were completely correlated with one facility type (predictor of interest that will be forced into the model) were discluded from further analysis (but listed in descriptive statistic tables); binary categorical predictors with a category of less than 5 were unable to be combined and discluded from further analysis (but listed in descriptive statistic tables)
* pearson’s correlation coefficient was calculated for all numeric continuous predictors to check for high levels of correlation between predictor variables, and if found to be greater than 0.6 the predictor with a more highly significant relationship found in univariate analysis was eligible for inclusion in the multivariable model; chi-square, fisher’s exact tests (where appropriate) used to check for correlation between categorical variables (cut-off?); ANOVA used to check for correlation between numeric continuous variables and categorical variables (cut-off?)

***Comparison of Bulk Tank Udder Health Measures, Aerobic Culture Data, and Hygiene Scores by Facility Type***

* .. imported into R Statistical Programming Environment (R Core Team, 2022); R version 4.1.3 (One Push-Up) was released on 2022-03-10.
* R Core Team. 2022. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
* notes re: ANOVA/boxplots for 8 outcomes (SCC, wLS, unLS, %New, %chron, %any, avgHyg, 34Hyg)… Checking the assumptions for ANOVA: were checked for (1) outliers, (2) normality using a Shapiro-Wilk test of normality for each group being compared (with significance at p = 0.05) and (3) homogeneity of variances (using Levene’s test and also looking at the residuals vs. fitted values plot). The Tukey method was used for adjusting p-values for multiple comparisons using the “TukeyHSD” function of the “stats” package in R (R Core Team, 2022).

***How Patel organized her M and M:***

Herd enrollment and sampling

* Bedding sample collection
* BTM sample collection
* Udder hygiene scores
* Herd management practices
* Herd measures of UH

Laboratory analysis of bedding and bulk tank milk samples

* Bedding culture
* Bedding characteristics
* BTM culture

Data management and analysis

***Bedding and BTM M and M from patel’s paper:***

Frozen bedding and BTM samples were shipped on ice to the Laboratory for Udder Health (University of Minnesota Veterinary Diagnostic Laboratory, St. Paul) for analysis. After thawing at room temperature, 50 cm3 of packed bedding material was weighed and measured into a new Whirl-Pak bag Nasco, Fort Atkinson, WI), 250 mL of sterile water was added, and the contents were mixed and left to stand for 10 min. Serial 10-fold dilutions of the samples were made using sterile water (Becton Dickinson and Company, Franklin Lakes, NJ). Sample dilutions were plated onto MacConkey agar (gram-negative bacteria selection) and colistin naladixic acid agar (gram-positive bacteria selection, Becton Dickinson and Company) plates and incubated overnight at 37°C. For the MacConkey plates, lactose fermenting (pink) colonies were counted as coliform bacteria and all other colonies were counted as non-coliform gram-negative bacteria. Colonies with a confluent appearance on MacConkey agar were identified to the genus level using a MALDI Biotyper (Bruker Daltonics, Billerica, MA), and colonies identified as Klebsiella spp. were counted and reported as a percentage of total coliform count. For colistin naladixic acid plates, colony morphology in conjunction with catalase reaction and Gram stain were used to differentiate colonies of Staphylococcus spp., SSLO, and Bacillus spp. Total bacteria count (TBC) and counts of Bacillus spp., coliforms, Klebsiella spp., non-coliform gram-negatives, Staphylococcus spp., and SSLO were recorded as colony-forming units per cubic centimeter of wet bedding. The minimum limit of detection was 25 cfu/cm3 (max count of 6,250,000 cfu/mL).

After thawing to room temperature, bulk tank milk and a 10-fold dilution of the 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 d at 37°C. Lactose fermenting (pink) colonies on MacConkey medium were counted and reported as coliform bacteria. All β-hemolytic colonies on Focus medium were counted and identified to the species level using a MALDI Biotyper, as these colonies were suspect for *Streptococcus agalactiae*. All colonies on Focus medium that were not identified as *Strep. agalactiae* were counted and recorded as SSLO. β-Hemolytic colonies on Factor medium were counted and identified to the species level using a MALDI Biotyper, and those with a confidence score ≥2.0 for *Staph. aureus* were counted and reported as such. Non-hemolytic colonies of Staphylococcus spp. (based on colony morphology, catalase reaction, or Gram stain) were counted and reported as NAS. For Mycoplasma spp., 0.1 mL of BTM was swabbed across the entire surface of a Mycoplasma agar plate and incubated for 7 d in a 7% CO2 incubator at 37°C. Plates were examined for Mycoplasma spp., and colonies were counted by a trained microbiology technician. For each BTM sample, total counts of coliforms, NAS, SSLO, *Staph. aureus*, *Strep. agalactiae*, and Mycoplasma spp. were recorded as colony-forming units per milliliter of milk. The minimum limit of detection for the BTM culture protocol was 5 cfu/mL (max count of 62,500 cfu/mL).

For the beddings, we use MacConkey and CNA agar.

Bulk Tanks use Factor (gram positive selective), Focus (Strep selective; FKA MKTK) and MacConkey

* Considerations for dealing with 40 herd bedding data
  + Cut points from culture data (e.g., less than 10 CFU; max count of 6,250,000)
    - Make them categorical?
  + Zeroes in bedding culture data
    - Make difficult to just log transform bedding bacteria counts

***Bedding Sample Collection.*** Wearing clean disposable gloves, the sampler collected unused bedding from the bedding storage area by collecting grab samples from the top 5 cm of bedding from 15 random locations in the pile. After mixing in a clean bucket, a composite sample was transferred to two 1-quart (946.4 mL) Ziploc (SC Johnson, Racine, WI) bags. The age of the unused bedding (days that it had been in storage) was recorded. Used bedding was collected as a grab sample from the top 5 cm of bedding in the back one-third of 15 randomly selected stalls or locations in the yard, representing up to 5 lactating pens, and then mixed well in a clean bucket before being transferred into two 1-quart Ziploc bags. Samplers avoided manure pats. If more than 5 lactating pens existed, then samples were collected from 5 pens housing early- or peak-lactation cows and heifers. The age of the used bedding sample was recorded as the days since fresh bedding was most recently added to the stall or resting area. All bedding samples were placed on ice at the farm and then frozen at −20°C until being transported to the laboratory for analysis.

***(Tucker’s draft) Bedding Sample Collection:*** For loose housing, used bedding subsamples were collected at 5 meters apart along a systematic transect with a random starting point within each barn of interest; barn area was calculated and one grab-sample was collected per 25 m2. Using fresh disposable gloves, subsamples were collected from the top 10 cm of the bedding surface using fingers to penetrate compacted material if necessary. For tiestall and freestall housing, used bedding subsamples were collected from the rear third of every other stall. Subsamples from each barn were pooled and homogenized in a disposable-plastic lined tote to form a composite sample for each barn. Homogenized bedding was collected in .25 liter aliquots into ziplock bags (if necessary with long fiber bedding) or whirlpak bags and stored on ice in a cooler for transport to the lab for storage at -80 C or further analysis. Temperature was measured at time of sampling using a >>>compost thermometer>>>.

***Text removed from Results:***

A one-way ANOVA was performed to compare the effect of [independent variable] on [dependent variable].

A one-way ANOVA revealed that there [was or was not] a statistically significant difference in [dependent variable] between at least two groups (F(between groups df, within groups df) = [F-value], p = [p-value]).

Tukey’s HSD Test for multiple comparisons found that the mean value of [dependent variable] was significantly different between [group name] and [group name] (p = [p-value], 95% C.I. = [lower, upper]).

There was no statistically significant difference between [group name] and [group name] (p=[p-value]).

**NEW IMI**

Gloves

Depth bedding stalls

**CHRONIC IMI**

Clip/flame udders

Prop. Dirty cows

Avg. hygiene score (although correlated with 3+4 cows obvi)

**ANY IMI**

Depth bedding stalls

Prop. Dirty cows

Avg. hygiene score (although correlated with 3+4 cows obvi)

**BTM SCC**

Lying surface

Depth bedding stalls

**AVG LS (UNWEIGHTED)**

Vit E selenium

Dry-off product

Depth bedding stalls

Prop. Dirty cows

Avg. hygiene score (although correlated with 3+4 cows obvi)

**AVG LS (WEIGHTED)**

Depth bedding stalls

Prop. Dirty cows

**PROP 3+4**

Depth bedded pack

Lying surface

Depth bedding stalls

**AVG HYGIENE**

Depth bedded pack

Lying surface

Depth bedding stalls

**FAMD:**

Calculates the principal dimensions (linear combinations of the original variables which better account for the variance in the dataset)

Then, use the eigenvalues and percentage variance explained by each PD to provide insight into how informative the original variables are

Goal is identify principal components along which the variation in the data is maximal

***Stuff removed from Discussion/Conclusions (or my notes/outline)***

* ***Theme 1:* Deeper bedding, better hygiene;** Increased bedding depth measures also tended to be associated with lower mean udder hygiene scores.
  + **Lower mean udder hygiene scores:**
    - (1) deeper bedded pack -> lower avg hygiene score
    - (2) deep bedding vs. mattress/concrete lower avg hygiene score
    - (3) deeper bedding in freestalls/tiestalls -> lower avg hygiene score
* ***Theme 2:* Better udder hygiene, better milk quality;** Farms with lower mean udder hygiene scores tended towards having lower percent chronic IMI, lower percent any IMI, and lower weighted average linear score.
  + **Lower percent CHRONIC IMI**:
    - (1) Lower proportion of dirty udders -> fewer chronic IMI
    - (2) Lower avg hygiene -> fewer chronic IMI
    - (3) Clip/flame udders -> fewer chronic IMI
  + **Lower percent ANY IMI:**
    - (1) Lower avg hygiene -> fewer any IMI
    - (2) Lower proportion of dirty udders -> fewer any IMI
  + **Lower weighted avg LS:**
    - (1) Lower proportion of dirty udders -> lower weighted avg LS
* ***Theme 3:* Deeper bedding, better milk quality;** farms with deeper bedding showed a tendency toward a lower bulk tank somatic cell count, lower percent new IMI, lower percent any IMI, lower weighted average linear score
  + **Lower SCC**:
    - (1) deep bedding vs. mattress/concrete had lower SCC
    - (2) deeper bedding in freestalls/tiestalls -> lower SCC
  + **Lower percent NEW IMI:**
    - (1) deeper bedding in freestalls/tiestalls -> fewer new IMI
  + **Lower percent ANY IMI:**
    - (1) deeper bedding in freestalls/tiestalls -> fewer any IMI
  + **Lower weighted avg LS:**
    - (1) deeper bedding in freestalls/tiestalls -> lower avg weighted LS
* **Misc:**
  + **Lower percent NEW IMI:**
    - (1) Consistent glove use -> lower NEW IMI
* ***Theme 4:* Bedded packs can be considered a viable option for pasture-based herds looking for a more affordable loose-housing system (in the Northeast? In Vermont?)**
  + We didn’t see difference in hygiene between facility types, BUT Robles 2020 found cows in freestalls had Cows in freestall barns more often had dirty lower legs, and upper legs/flanks compared to tie-stall barns
  + Hygiene is hard to compare between system types: Cook 2002; different parts of cow will be dirtier for freestall, tiestall, loose-housing with no stalls. We looked just at udder hygiene
  + Can borrow language from Robles 2020 blurb

Farms with deeper bedding showed a tendency toward a lower bulk tank somatic cell count, lower percent new IMI, lower percent any IMI, lower weighted average linear score, and lower mean udder hygiene score. Farms with lower mean udder hygiene scores tended towards having lower percent chronic IMI, lower percent any IMI, and lower weighted average linear score. Increased bedding depth measures also tended to be associated with lower mean udder hygiene scores.

Overall, farms with more deeply-bedded cows had improved bulk tank milk quality (as measured by SCC), as well as better udder health metrics and udder hygiene scores. Unsurprisingly, better udder hygiene was associated with improved udder health measures. Additionally, bedded pack systems did not differ significantly in their milk quality, udder health, or hygiene measures when compared to the more commonly used winter housing systems for organic cows in the state. Bedded packs can therefore be considered as a viable option for pasture-based herds looking for a more affordable loose-housing system.

***Notes on M and M for model-building:***

**SCC:** no “significant” model (overall p for f-test below 0.5); no predictors significant either; model with only herd\_size had lowest AIC; f-test between full and reduced model showed herd\_size model "NO better" than reduced model (p>0.05); facility\_type not significant predictor

**NEW IMI:** 20 different “significant” models (overall p for f-test below 0.5); model with bedding\_conditioner, air\_quality\_TA, gloves, mast\_record had lowest AIC; f-test between best and reduced model showed best model "better" than reduced model (p<0.05); facility\_type not significant predictor

**CHRONIC IMI:** 2 “significant” (overall p for f-test below 0.5); model with bedding\_conditioner, clip\_udder, sel\_feed, perc\_udder\_3and4 had lowest AIC (FULL model); f-test between full and reduced model showed full model "better" than reduced model (p<0.05); facility\_type not significant predictor

**ANY IMI:** no “significant” models (overall p for f-test below 0.5); no significant predictors, but full model had lowest AIC; f-test between full and reduced model showed full model "better" than reduced model (p<0.05); facility\_type not significant predictor

**UNWEIGHTED LS:** 4 “significant” models (overall p for f-test below 0.5); model with bedding\_conditioner, dry\_prod, sel\_supp, mean\_hygiene had lowest AIC (4 unforced predictors); f-test between "best" and reduced model showed "best" model "better" than reduced model (p<0.05); facility\_type not significant predictor

**WEIGHTED LS:** no “significant” models (overall p for f-test below 0.5); no significant predictors; model with air\_quality\_TA had lowest AIC (1 unforced predictors) but \*not\* signif. at p<0.05 (p was 0.058); f-test between "best" and reduced model showed "best" model "better" than reduced model (p<0.05)

…. *Why no ANOVA on BTM culture data? Somehow, all interesting ones (coliforms, non-ag streps, staph sp., aureus) violate assumptions in one way or another – even when log transformed*

**Coliforms:** all either 5 or 0, so can’t really log transform and certainly not normally distributed

**Streps:** there are 2 outliers in the bedded pack group (n = 5) when log10 and log2 transformed, violating assumption of ANOVA; 3 outliers when try raw data; normally distributed overall, but not normally distributed when break down by facility type (residuals for FS); variance is homogenous though

**Aureus:** no outliers when log transformed (with 0’s cheating); tried both log10 and log2 transformed, but not normally distributed (Shapiro-Wilke test is significant); variance is homogenous though

**Staph spp:** there is 1 outlier in BP (Choiniere) for both log10 and log2; when removed, still normally distributed, and when not removed, normally distributed; homogeneity of variance between groups same; when do ANOVA without CF, no difference in facility type

… *Trying to model culture data from BTM: Simon suggested treating as COUNT data, trying either Poisson, Negative Binomial, or Zero-inflated models of either. Sandra’s lab has used log transformed culture data.*

**Strep:** Simon suggested treating as COUNT data, trying either Poisson, Negative Binomial, or Zero-inflated models of either. Strep data is ~ normally distributed, but there is certainly an outlier. So, had initially tried modelling as “count” data… did a bunch of tests in R to check for overdispersion (when variance much bigger than the mean; “40herd\_model\_BTM\_outcomes.Rmd”). Got a ton of significant predictors in univariate screening, prob bc small sample size. When variance higher than mean (overdispersed), risk of false positives increases (number positive from linear regression univariate screening for binomials: 6/21 predictors for scc modeling, vs. poisson for strep data: 17/21 predictors) ... "relatively small violations of Poisson’s assumptions can dramatically inflate false positives rates." With a Negative Binomial: "Instead of requiring the mean equal the variance, the variance is a quadratic function of the mean and can differ from it. This allows it to more faithfully model data which is overdispersed – that is, where the variance is greater than the mean. While it is possible for data to become so overdispersed that they violate the assumptions of the Negative Binomial, there is more flexibility than Poisson accommodates." DHARMa nonparametric dispersion test via sd of residuals fitted vs. simulated … indicated NB ok, Poisson NOT.

**SA:** Many more zeroes is SA data (8/21). Regular linear regression on log-transformed SA data is NOT appropriate; it is NOT normally distributed (according to Shapiro-Wilk test). Additionally, had to make all 8 zeroes the log of 2.5 CFU (half the lower limit of detection for culture, like Sam and Kruthika did). Really had a lot of difficulty choosing between the count-based modelling methods that Simon suggested; Poisson, Negative Binomial, Quasipoisson, Zero-inflated Poisson, Zero-inflated Negative Binomial.

Got a ton of significant predictors in univariate screening, prob bc small sample size. Calculated the variance of the SA data: 4885.357 for sample variance, 4652.721 for population variance. Mean is 43.57143. When variance higher than mean (overdispersed), risk of false positives increases (number positive from linear regression univariate screening for binomials: 6/21 predictors for scc modeling, vs. poisson for sa data: 15/21 predictors) ... "Relatively small violations of Poisson’s assumptions can dramatically inflate false positives rates.” Tried a bunch of tests for overdispersion, and certainly seems that Poisson model for SA data is over-dispersed.

library(AER) *# has a test for overdispersion*

dispersiontest(sa1, trafo=1) *### not sure how to interp. this one- says alpha is huge, but the p isn't signif; also, I can’t tell if I’m supposed to use 1 or 0 for this “trafo” value*

Overdispersion test

data: sa1

z = 1.4671, p-value = 0.07118

alternative hypothesis: true alpha is greater than 0

sample estimates:

alpha

71.79762

*library(AER)*

*dispersiontest(sa1, trafo=0)*

Overdispersion test

data: sa1

z = 0.5343, p-value = 0.2966

alternative hypothesis: true alpha is greater than 0

sample estimates:

alpha

480.1924

Tried to make a **Negative Binomial model;** but I think the model won’t converge – too many 0’s? It gives errors that I can’t make sense of when I give the command to make the model, but then also gives a summary when asked for the NB summary

Call:

glm.nb(formula = milk\_Staph\_aureus ~ Facility\_type, data = lu,

init.theta = 355991.0487, link = log)

Deviance Residuals:

Min 1Q Median 3Q Max

-11.225 -5.457 -1.700 1.521 22.933

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 2.1972 0.1491 14.739 <2e-16 \*\*\*

Facility\_typeFS 1.4917 0.1624 9.182 <2e-16 \*\*\*

Facility\_typeTS 1.9459 0.1543 12.611 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for Negative Binomial(355991) family taken to be 1)

Null deviance: 1544.5 on 20 degrees of freedom

Residual deviance: 1263.0 on 18 degrees of freedom

AIC: 1346.2

DHARMa nonparametric dispersion test via sd of residuals fitted vs. simulated; both Poisson and NB are over 100 for dispersion, with highly significant p-values

**Negative binomial?** *"Instead of requiring the mean equal the variance, the variance is a quadratic function of the mean and can differ from it. This allows it to more faithfully model data which is overdispersed – that is, where the variance is greater than the mean. While it is possible for data to become so overdispersed that they violate the assumptions of the Negative Binomial, there is more flexibility than Poisson accommodates"*

Tried also using “P\_\_disp: A function to calculate Pearson Chi2 and its dispersion,” which gave a dispersion number of 85 for Poisson (over 1 is concerning) and same for Quasipoisson. *“The Quasi-Poisson Regression is a generalization of the Poisson regression and is used when modeling an overdispersed count variable. The Poisson model assumes that the variance is equal to the mean, which is not always a fair assumption. When the variance is greater than the mean, a Quasi-Poisson model, which assumes that the variance is a linear function of the mean, is more appropriate.”* A quasi-Poisson uses scaled SE to deal with overdispersion.

Then did a Vuong test to see if Zero-inflated Poisson (ZIP) better than regular Poisson:

sa1<-glm(milk\_Staph\_aureus ~ Facility\_type, family="poisson", data=lu)

vuong(sa1, model\_zip) *# we can see that our test statistic is significant (p=0.005), indicating that the zero-inflated model is superior to the standard Poisson model*

*## ok, now we know ZIP better than poisson, so make neg bin inflated and test those two*

model\_zinb<-zeroinfl(milk\_Staph\_aureus ~ Facility\_type, dist="negbin", data=lu)

summary(model\_zinb) *# ok cool it made a real model this time, as there were no errors thrown in the first command and the summary actually looks fine*

Tried a Vuong test to compare ZIP and ZINB, it says either model is preferable depending on which you list first.

**I think where I ended up is that a ZINB is the way to go,** because it is LESS overdispersed than a ZIP (BUT still over-dispersed), the log theta term from the ZINB is significant in the model summary, and the AIC for the ZINB model is much lower than the AIC for the ZIP model.

However, I can’t use the automated univariate way to screen predictors that I could with Poisson, or regular linear regression. It seems like days/weeks of work that would be pretty unlikely to yield any good results anyhow at this point.

***\*\*\*\* BTM CULTURE RESULTS: DISCUSSION \*\*\*\*\*\*\*\****

**Staph spp**.: for tiestalls, numerically, the mean is much higher than the other 2 facility types; but not statistically significant. Also, more variation in tiestall farms – large outlier (665 CFU - JandL).

Run without giant tiestall outlier, means are more similar:

BP FS TS

"M (SD) = 53.00 (48.94)"; "M (SD) = 65.83 (47.69)"; "M (SD) = 75.56 (60.23) (range = 15-175)"

All our farms ranged 0-665, 0-175 without the huge outlier in TS; our staph. spp. mean 95.5 (36-155 95% CI).

BP’s ranged 0-130, mean 53 CFU/mL; with 95% CI 10-96

*Shane 2010:* the 6 BP ranged from 0-108. don’t give a mean

*Lobeck 2012:* Staph sp. mean (CI) 26.1 (CI: 2‐443) CFU/mL in winter for 6 BP; overall for their 18 farms 78.1 (CI: 4-1,544). *There were no differences in bacterial numbers among housing systems for Staphylococcus aureus, non‐ag Streptococcus, Staphylococcus species, and coliforms*

*Barberg 2007b:* no mention of staph spp.

**Streps:** – large outlier (1250 CFU - HoytHill).

BP FS TS

"M (SD) = 39.00 (25.35)"; "M (SD) = 89.17 (97.54)" "M (SD) = 145.00 (110.45)"

All our farms ranged 10-1250, mean=156.4 (95% CI 42-271); bedded packs ranged 10-80, mean = 39 (95% CI 17-61).

Bedded packs in Shane 2010 ranged from 98-48,400 CFU/mL in the winter. don’t give a mean. Our BP strep spp. in BTM much lower

Barberg 2007b: 6/12 BP farms had “high” levels of non-ag strep, but don’t give a number

Lobeck 2012: Strep sp. mean (CI) 911.2 (CI: 138-6011) CFU/mL in winter for 6 BP; overall for their 18 farms 445.4 (CI: 116-1704). Our BP strep spp. in BTM much lower; overall our farms had lower strep spp. too

**Aureus:** highest proportion of negative farms in BP category (3/5, vs. 3/10 TS and 2/6 FS); would wonder if is confounding: early adapters/producers open to innovative or newer ways of mgmt. are more progressive generally in their mgmt. style, so more likely to have adapted stricter SA controls?

Our BP 3/5 negative for SA. Overall our farms ranged 0-320, mean 43.6 (95% CI 14-73); BP ranged 0-30, mean 9 (95% CI 0-21)

Shane 2010 BP’s ranged 0-55. Shane 2010 3/6 BP farms in winter were negative

Barberg 2007b: only 1 farm had a “high” level of SA out of 12 BPs

Lobeck 2012: SA mean (CI) 6.2 (CI: 1.3-30.1) CFU/mL in winter for 6 BP; overall for their 18 farms 17.3 (CI: 3.3-91.2); *SA in BP also numerically lower for them, but not statistically different;* Our BP SA in BTM very similar;

**Coliforms:** all farms really low (0 or 5), not much interesting to say about this? Is there a larger population of farms to compare this to?

All our farms were pretty excellent farms already? All either 0 or 5. Our overall mean 1.2 (95% CI: 0.3-2.1); our BP mean 1, 95% CI 0-3

CFU/mL for bedded pack farms in Shane 2010 ranged 15-1,128 (CFU/mL) in winter (!methodology for culture different, but still same measurement of concentration!)

Barberg 2007b: 5/12 farms had “high” levels of coliforms – will need to check on seasonality

Lobeck 2012: Coliforms mean (CI) 63.7 (CI: 6-735) CFU/mL in winter for 6 BP; overall for their 18 farms 19.4 (CI: 1-252); Our BP coliforms in BTM much lower; overall our farms had lower coliforms too

***\*\*\*\* UNCONDITIONAL COMPARISONS: DISCUSSION \*\*\*\*\*\*\*\****

In comparison to previous work comparing different hygiene measures of dairy cows between facility types, only udder hygiene was measured in the current study. Other researchers studying this issue have taken a more comprehensive evaluation of dairy cow hygiene, including separate scores for different body regions (lower legs, upper legs, udder, flank, and belly). Cook (2002) has pointed out the challenges of comparing dairy cattle hygiene between different facility types; namely, cows in freestall barns tend to have dirtier lower legs from walking through alleyways, whereas tiestall cows are likely to have dirtier upper legs and flanks, and loose-housed cows may have a different pattern of manure deposition on their bodies depending on the cleanliness of their lying surface. Despite the limitation of only observing udder hygiene, we feel confident that udder hygiene of cows housed on bedded pack systems in the current study was similar to the more conventional housing systems used by dairies in the Northeast.

***\*\*\*\* DISCUSSION: aureus results different between BP, TS, FS \*\*\*\*\*\*\*\****

Although not statistically significant, the mean *Staph. aureus* cfu count in BTM for bedded pack farms was numerically smaller than that for tiestalls and freestalls. Bedded pack herds had the highest proportion of farms with BTM negative for *Staph. aureus* (60% for BP, vs. 33.3% for both TS and FS). As *Staph. aureus* is regarded mainly as a contagious mastitis pathogen, on-farm prevalence is more likely a function of milking time hygiene and other management practices associated with limiting cow-to-cow transmission vs. environmental factors (i.e., bedding material type and hygiene; Jayarao and Wolfgang, 2003; although recent studies have described environmental reservoirs of transient sources: Leuenberger et al., 2019). The lower prevalence of *Staph. aureus* on bedded pack farms in the current study may be a function of confounding. Producers using this less-traditional housing type (bedded packs) may inherently be more open to newer technologies. A Dutch study found more progressive dairy farmers (younger in age, had children with higher education levels, more likely to invest in the future for their farm) were more likely to have a lower bulk tank milk somatic cell count (Barkema et al., 1998). We speculate that an extension of this relationship is that progressive dairy farmers already using a bedded pack may be more likely to have implemented stricter control programs for contagious mastitis pathogens such as *Staph. aureus*.

**Author Contributions**

Caitlin: Planning and execution of field study, creation of survey tool and on-farm observation sheet, survey administration and sample collection, sample processing, data organization & analysis, manuscript writing

XXXXX

John: Study design, creation of survey tool, sample processing, data analysis, manuscript preparation

**Conflict of Interest**

The authors declare no conflicts of interest.

**Survey of Management Practices, Bulk Tank Milk Bacteriology, and Udder Health and Hygiene Metrics on Organic Dairy Farms in Vermont**

**Bulk Tank Milk Bacteriology, and Udder Health and Hygiene Metrics on Organic Dairy Farms in Vermont by Facility Type**

**Observational Study of Udder Health, Hygiene, and Milk Quality on Vermont Organic Dairy Farms by Facility Type**

**Observational Study of Udder Health, Hygiene, and Milk Quality on Vermont Organic Dairy Farms Using Different Facility Types**

**Observational Study of Udder Health, Hygiene, and Milk Quality on Vermont Organic Dairy Farms**

For all models, variables that were unconditionally associated with an outcome but correlated with another independent variable offered into the model were excluded in order to prevent multicollinearity. In this instance, the variable with a lower P-value for an unconditional association with an outcome was chosen for inclusion.

***Things need to include with supplemental data:***

1. Full survey i gave to farmers
2. Observation sheet
3. Supplemental table 1: farm mgmt. practices, housing characteristics
4. Supplemental table 2: bedding material, bedding mgmt. practices
5. Supplemental table 3: milking procedures
6. Supplemental table 4: mastitis control practices

***\*\*\*\* BIGGER THINGS TO RECONCILE AFTER READING SANDRA’S COMMENTS \*\*\*\*\*\*\****

* ~~Clarify “winter” housing system; spring vs. winter vs. “non-pasture season”~~
* The two models for bedding depth: categorical (deep vs. not deep) vs. numeric predictor (depth of the bedding, measured)
* “Primary” vs. “Secondary” objectives; needs to be reorganized throughout manuscript
* All things where I say “tended,” “*state you found no statistical difference, but saw numeric differences (report mean, 95% CI) that may be biologically important, but which failed to reach threshold for declaring significance, possibly due to small sample size*”
* Add limitations sections:
  + BP measured vs. stalls estimated by producers

***Clarify with Sandra:***

* aerobic culture data: i did check normality of both raw and log10 transformed; results on note paper; did Krusal-Wallis on all (ok if not normal). Should I just be reporting median (range) for this data if I wanted to leave it as the raw data (and not log10 transform)?
  + *Both in results section, discussion section, and* ***Table 1***
  + *Ask her to peek at updated table 1; remove 95% CI?*
* generally, for something not normally distributed, most appropriate to report median (range) NOT any means or 95% CI?
* Looked at distribution of BTSCC for log2, log10, raw; none *appeared* to be not normally dist. but still inappropriate to do stats as if they were?
* How best make tables- linear regressions for all univariate outcomes of interest (***Table 4***)
* Table of terms/group sizes offered to models (now ***TABLE 3***); where should this go (M and M; Results)? Is this helpful? Group size of each model will vary bc DHIA data availability changes so it’s HUGE (and kind of needs to stay that way)

Beginning of discussion- what’s the hook? Strengths … the first study that’s’… right at beginning.

* Compare to other studies
* Limitations-

All predictors, level of predictor, and group sizes offered to multivariable models for each of the eight different outcomes of interest are summarized in Table 3.

The main predictor of interest, facility type (FS, TS, or BP), was forced into the model for each of the eight outcomes.

Overall, this model was significant (*P* = 0.01), and shown to be better than the reduced model (ANOVA, *P* = 0.003)

other (non-facility) management-related risk factors associated with bulk tank milk quality, udder health, udder hygiene, and milk production in organic VT dairy herds.

Univariate analysis of farm management factors associated with bulk tank milk quality, udder health and udder hygiene measures

John 10.30.2023

* Objective 2 results

STRENGTHS

* Direct comparison BP to TS and FS (for all outcomes; hygiene, milk quality, milk production)
* First description ALL organic herds
* FIRST to describe BP in the northeast

LIMITATIONS

* SMALLLLLLLLLLLLL sample sizes
  + Future studies could be larger
* Only 1 BTM sample (addressed this); also, hard to pinpoint where on the farm pathogens coming from by only looking at BTM (and not quarter samples)
* Observational study; can’t make any statements about causality, it’s really “just an association”
* Missing data: some farms missing DHIA data
  + % new, chronic, elev: n = 19; 2 BP missing (SP and Donegan)
  + Avg SCS: n=20; 1 BP missing (Donegan)
  + STD150 milk: n=18; 1 bp (Donegan), 2 TS missing (BJ and JandL)
  + BTSCC, 2 hygiene: everyone, n=21
* Practices recommended for years for mastitis control (teat dipping) widely adapted by herds, so had limited power to identify (or was completely unable to analyze) associations between certain practices and BTM quality or hygiene (many, many other studies shown desirable and consistent associations with udder health for these practices; so these practices shouldn’t be ruled out of mastitis control programs just based on this single, small study)
* Bias
  + How to refute this? In spite of volunteers, avg herd size was similar to ... production, BTSCC, how does compare to general population
  + **selection bias:** The herds who participated in the study were a convenience sample of all Vermont organic dairy farms; likely share a lot of qualities with the population in general of VT organic farms (target population), but also may differ in general herd characteristics, management practices and bulk tank milk quality. This convenience sample would be an example of selection bias; may affect estimates of different BTM outcomes or hygiene outcomes (“but not associations between risk factors and BTM/hygiene outcomes?”). Possible that people willing to participate were in some way systematically (more progressive?) different than the target population.
    - What is typical herd size, BTSCC, and production for organic dairies generally in the US? Compare to this?
  + **residual confounding (or just confounding in the univariate models period);** maybe this is only applicable to models that Simon was making for his thesis papers; we don’t really claim to present associations with risk factors where anything is “controlled” for… multivariable models were problematic and tried to focus on univariate analysis
    - Maybe address that by focusing on UNIVARIATE analysis, we certainly are likely (could be likely) to be missing confounding… univariate models presented don’t control for other factors besides the predictor of interest and the outcome of interest.. are NOT adjusting for confounders in any of the models (BUT try to be up front about that).
    - Unknown, unmeasured confounders; OR imperfectly measured confounders
  + **misclassification bias:** aerobic culture of bulk tank milk not a “perfect” test, variable sensitivity and specificity for identifying bacteria growing in the milk; BUT would assume these measurement errors likely occur randomly, so bias would be nondifferential (towards the null)
  + **what kind of bias would it be people self-reporting answers?** Is this “covariate misclassification? Exposure misclassification?” likely also nondifferential, and would bias towards the null

FUTURE DIRECTIONS

* Looking into depth of bedding in stalls/BP and hygiene or milk quality more?