



Within-herd prevalence threshold for the detection of *Mycobacterium avium* ssp. *paratuberculosis* antibody–positive dairy herds using pooled milk samples: A field study

M. Krieger,^{1*} S. Eisenberg,² H. Köhler,³ F. Freise,¹ and A. Campe¹

¹Department of Biometry, Epidemiology and Information Processing, WHO Collaborating Centre for Research and Training for Health at the Human-Animal-Environment Interface, University of Veterinary Medicine, D-30559 Hanover, Germany

²Animal Diseases Fund of Lower Saxony, Brühlstraße 9, 30169 Hanover, Germany

³Institute of Molecular Pathogenesis, Friedrich-Loeffler-Institute, Federal Research Institute for Animal Health, Naumburger Straße 96 a, 07743 Jena, Germany

ABSTRACT

Herd-level diagnosis of paratuberculosis using a pool-milk ELISA (pool size: $n \leq 50$) is a novel, economical, and convenient method to identify blood serological *Mycobacterium avium* ssp. *paratuberculosis* (MAP) antibody–positive herds. To date, the diagnostic performance of the pool-milk ELISA has been described only under laboratory conditions where herd prevalence was simulated by the preparation of milk pools consisting of milk samples of cows with a known MAP status determined by fecal culture. In our observational study, test performance under field conditions was studied using pooled milk and individual blood samples. A total of 486 herds within the MAP prevalence reduction program of Lower Saxony, from which pooled milk and individual blood ELISA results were available, were assigned to this study. Data were analyzed for the period between January 1 and December 31, 2018, the first year after herd testing became obligatory in this federal state of Germany. To evaluate whether pooled milk samples reliably distinguish between herds with a MAP-apparent blood serological within-herd prevalence (MAP-Ab-WHP_{app}) $\geq 5\%$ and herds with a MAP-Ab-WHP_{app} $< 5\%$, the distribution of the MAP-Ab-WHP_{app} was compared between pool-positive and pool-negative herds. The MAP-Ab-WHP_{app} was 3.4% (median; 95% confidence interval = 0–11.4%) in pool-positive herds and 1.2% (median; 95% confidence interval = 0–6.4%) in pool-negative herds. Only 10.8% ($n = 12$) of the pool sample–negative herds had a MAP-Ab-WHP_{app} $\geq 5\%$ and were therefore false negatives, given the aims of the MAP prevalence reduction program. Hence, the pool-milk sampling strategy seems well suited to dis-

tinguish between herds with a MAP-Ab-WHP_{app} $\geq 5\%$ and herds with a MAP-Ab-WHP_{app} $< 5\%$ since only 10% of serum MAP-ELISA positive herds were missed. Employing a logistic regression model, we estimated that the minimum blood serological MAP-Ab-WHP_{app} to detect a pool-positive herd with a probability of 95% was 8%, which fits well with the aim of the MAP prevalence reduction program to focus on herds with a MAP-Ab-WHP_{app} of $\geq 5\%$. Despite the limitations of the control approach, which include milk pool sample collection and a low sensitivity of the ELISA used in milk pools and serum samples, the aims of the MAP prevalence reduction program can be achieved. The results of these field data support that pool-milk sample ELISA is a useful, economical, and low labor–intensive tool to identify herds seropositive for MAP in a MAP prevalence reduction program.

Key words: *Mycobacterium avium* ssp. *paratuberculosis*, pooled milk sample, herd-level diagnosis, ELISA

INTRODUCTION

Paratuberculosis, also known as Johne's disease, is a chronic enteritis in mainly domestic and wild ruminants (Chiodini et al., 1984) and is caused by *Mycobacterium avium* ssp. *paratuberculosis* (MAP; Thorel et al., 1990). Cattle are most susceptible to an infection within the first 6 mo of life (Windsor and Whittington, 2010). The major source of infection is the uptake of infectious fecal material (Doré et al., 2012). Clinical signs, which include chronic progressive weight loss and chronic or intermittent diarrhea, occur after a long incubation period of 2 to 10 yr (Whitlock and Buergelt, 1996; Clarke, 1997). Once animals are infected, there is no cure, and the disease eventually leads to the death of infected animals. Economic losses due to decreased milk production, premature culling, and reduced slaughter value are substantial (Benedictus et al., 1987; Ott et al., 1999).

Received March 4, 2021.

Accepted August 13, 2021.

*Corresponding author: Mette.Krieger@tiho-hannover.de

Paratuberculosis is a disease with a global occurrence and a variable prevalence (Whittington et al., 2019). In Germany, herd prevalence range differs substantially between federal states, for example, 8.9% based on fecal culture and fecal PCR samples that were collected with boot swab samples in Hesse up to 30.2% based on fecal PCR samples collected with boot swabs in Saxony and Thuringia (Noll et al., 2017; Pützschel et al., 2017). However, because no standardized nationwide prevalence study has been conducted to date, a comparison has to be drawn carefully. Similar variation can be found regarding the apparent within-herd prevalence (**WHP_{app}**) in different regions of the world, for example, Europe, where prevalences range from 0% to 24% between member states (Nielsen and Toft, 2009). For the United States, WHP_{app} values of 6.9 to 10% are described (Thorne and Hardin, 1997; Johnson-Ifeorlundu and Kaneene, 1999; Raizman et al., 2011) and in India WHP_{app} values of 13.4 to 42.6% are described in cattle between different states (Singh et al., 2008; Trangadia et al., 2012; Bhutediya et al., 2017).

To date, no nationwide MAP control program is in place in Germany. However, voluntary control programs for paratuberculosis currently exist in 7 federal states of Germany (Thüringer Tierseuchenkasse, 2020). In Lower Saxony, MAP monitoring has been mandatory since 2017. For direct MAP detection, bacterial culture and PCR can be used, whereas ELISA kits are available for indirect detection. Unfortunately, all available tests have a rather low sensitivity and high specificity, making MAP control challenging (Fecteau, 2018). Tests can be used at the herd level or at the animal level depending on the control strategy (Collins et al., 2006). Bacterial culture of individual fecal samples is time consuming and costly, but with the exception of necropsy, is still considered as a reference method for MAP detection with a higher sensitivity and specificity compared with ELISA tests (Collins et al., 2006). Therefore, bacterial culture or PCR is often used for herd-level diagnosis as a first step in paratuberculosis monitoring either on pooled fecal or environmental samples to reduce costs (Kohl et al., 2019). The ELISA tests can be used as a cost-efficient way to test individual serum or milk samples but lack sensitivity (Collins et al., 2006). For herd-level diagnosis, the collection of pooled milk samples seems practicable and time saving. The use of bulk tank milk samples for MAP diagnosis resulted in the detection of herd-level sensitivities of 33.3% and 50% referring to fecal culture (Bauman et al., 2019) and in an estimated median herd-sensitivity of <10% (Sergeant et al., 2019). A better option might be the use of pool-milk samples where the number of individual milk samples in that pool is standardized. In contrast to bulk tank milk samples, standardized

pooled milk samples can be generated under laboratory conditions, and the number and portion of individuals in pooled milk samples can be determined (Geue et al., 2007). In Germany, pooled milk samples ($n \leq 50$) are collected within the scope of mandatory BHV1 monitoring 3 times a year. Either pooled milk samples consisting of a maximum of 50 animals per pool are collected 3 times a year with a 4-mo interval or individual blood samples of all animals of the herd are collected annually (BGBl. 1, 2015). Existing pool samples seem to be an easy and economical method for herd-level diagnosis of MAP. However, results have shown that pooling decreases the detection probability of affected herds due to the dilution of positive samples by the larger number of negative samples (Wells et al., 2002). Therefore, cutoff values for the interpretation of available ELISA test kits were validated for use in pool-milk sample analysis (Köhler et al., 2017). The blood serological within-pool prevalence limit for the identification of paratuberculosis-infected dairy herds in pooled milk samples ($n = 50$) with a probability of 95% was estimated to be at least 8.9% under experimental conditions (Köhler et al., 2017).

Therefore, the primary objective of this study was to assess the use of milk pool testing for detection of MAP-positive herds under field conditions. Since most herds in Lower Saxony consist of more than 50 cows (i.e., several pool samples will be collected per farm), we hypothesized that under field conditions pooled milk sample analysis at one point in time would identify herds with a MAP-WHP_{app} even below 8.9%. Therefore, the minimum MAP-WHP_{app} needed to detect a MAP-positive herd by pool-milk ELISA with a 95% probability was determined. Furthermore, we hypothesized that the probability of detecting a herd with pool-milk sampling as positive increases with increasing herd size, because in larger herds more pool samples are collected.

MATERIALS AND METHODS

Study Herds

Enrolled dairy herds were located in Lower Saxony and Bremen and participated in the MAP prevalence reduction program of Lower Saxony. Data collected between January 1 and December 31, 2018 were included. In 2018, 10,749 herds participated in the program with 993,225 dairy cows older than 24 mo. Seven hundred seventy-four of these herds (7.2%) with 87,163 cows older than 24 mo took pooled milk and individual serum samples. Out of 10,749 herds, 490 farms (4.6%) with 72,903 cows older than 24 mo took at least 2 pooled milk samples per sampling event as well as in-

dividual serum samples. Taking at least 2 pooled milk samples per sampling was chosen to make sure that enrolled herds consisted of more than 50 lactating cows. Individual serum samples were taken from at least 50% of cows older than 24 mo. With regard to the available data, herd IDs, the binary result of the serum samples (positive/negative), the quantitative test result [sample to positive ratio in percent (S/P%)], and the binary result of the S/P% of the pooled milk samples, the number of individuals milked in each pool-milk sample and number of pooled milk samples at first test event were included in the analysis. If farms performed repeated pool-milk sampling in 2018, the results of all samplings were included in the analysis. Data management was conducted to check data quality. Therefore, testing in duplicate and repeated laboratory testing of the same pool sample was reviewed to ensure plausibility.

Positive and negative herds were determined by pool-milk ELISA results; farms with at least one positive milk pool were classified as milk MAP-antibody positive.

Sampling and Sample Testing

Pooled milk samples were defined as samples consisting of a maximum of 50 milk samples collected and pooled during test-day sampling. When conventional test-day sampling occurred, milk separated by an in-line electronic milk meter was pooled in a bucket for up to 50 cows and mixed before a 50-mL pool-milk sample was collected. Pooled milk samples as well as individual test-day milk samples collected by the milking robot were transported to the laboratory. Milk samples collected by milking robot were pooled in the laboratory.

Serum samples were taken out of the coccygeal vessel with a Vacutainer system (Becton, Dickinson and Co.) by the herd veterinarian. All samples were sent to one out of 4 different laboratories in Lower Saxony (Food and Veterinary Institute Oldenburg, Food and Veterinary Institute Braunschweig/Hannover, Institute for Animal Health of the Agricultural Investigation and Research Institute Nord-West (Oldenburg), or Georg-August-University Göttingen-Department for Microbiology and Animal Hygiene) and were tested using a commercial ELISA (Paratuberculosis Screening Ab Test; IDEXX Europa B.V.). Serum samples were analyzed according to the manufacturer's manual, and pooled milk samples were analyzed in duplicate. The results were expressed as S/P%, with cutoff values according to the manufacturer's manual [blood serum samples: negative ≤ 45 , suspect > 45 to < 55 , positive ≥ 55 (IDEXX, 2018)]. An adjusted cutoff value of 14.87 was used for pooled milk samples (Köhler et al., 2017).

Statistical Analyses

Statistical analyses were conducted using SAS software, version 9.4 (SAS Institute Inc.). Herd size was defined as the number of milk sampled cows on the day of sample collection. Therefore, the number of individuals milked in each pool-milk sample was summarized per farm. Because a farm can sample pooled milk samples more than once in a year, herd size and pool size were computed for each time a farm had sampled its animals. The herd size was determined to ensure that a herd consisted of more than 50 lactating cows. Pool size was determined to evaluate possible diagnostic misclassification.

The WHP_{app} per farm was calculated as the percentage of ELISA-positive serum results out of all tested serum samples per farm. Furthermore, the WHP_{app} was calculated as a prevalence at one point in time. Therefore, a herd was excluded from further analyses if it took blood serum samples more than once in a year because their WHP_{app} differed between test events. Because our statistical unit was the herd, a positive and a negative herd was determined according to the pool-milk result. Milk antibody-positive herds had at least one antibody-positive pooled milk sample, whereas negative herds had none. The WHP_{app} was dichotomized by pool-positive and pool-negative herds.

Descriptive analyses of the WHP_{app} distribution were performed by calculating measures of central tendency and dispersion to compare pool-positive and pool-negative herds. The Rogan-Gladen estimator was used to correct the WHP_{app} for the sensitivity and specificity of the ELISA used in blood serum. Herds that were classified correctly as positive or negative by their pool sample result were presented together with false-positive and false-negative herds. A herd was classified as correctly positive if it had at least one positive pool sample and a WHP_{app} above 0%. A herd was classified as correctly negative if it had no positive pool sample and a WHP_{app} of 0%. A herd was false positive if it had at least one positive pool sample result and a WHP_{app} of 0%, and false negative if it had no positive pool sample and a WHP_{app} above 0%. Spearman's rank correlation coefficient was calculated at the pool-milk sample level to analyze the relationship between the WHP_{app} of a farm ($n = 486$) and its quantitative test results of the pooled milk samples of one sampling event ($n = 1,629$). In pool sample-negative herds, the first sampling event was used, in pool sample-positive herds the last positive sampling event was used. In a second step, Spearman's rank correlation coefficient was calculated at the herd level. Therefore, the association between the median of a farm's quantitative pool test results ($n = 468$) and its WHP_{app} was calculated.

Table 1. Number of milk sampled cows in a herd in the first sampling in 2018, number of cows per pooled milk sample in the first sampling in 2018 and number of pooled milk samples in a herd in the first sampling in 2018

Item	N	Minimum	25% Quantile	Median	Interquartile range	75% Quantile	Maximum
Herd size ¹	486 ²	48	78	115	91	169	762
Number of cows per pooled milk sample	1,681 ³	2	37	44	13	50	53
Number of pooled milk samples at sampling day	486 ²	2	2	3	2	4	16

¹Defined as milk sampled cows at first pool sampling.

²Total number of farms.

³Total number of samples.

Apparent within-herd prevalence thresholds for the detection of herds seropositive for MAP using pooled milk samples were determined using logistic regression as outlined in the study of Donat et al. (2016). The predictor was the WHP_{app} of a herd, and the outcome of interest was the pool-milk sample result of a herd (all pooled milk samples negative/at least one positive pooled milk sample).

The WHP_{app} thresholds, along with corresponding asymptotic standard errors, were calculated for 2 different probabilities (90% and 95%) of a positive pool-milk sample result.

A second logistic regression for the association among the WHP_{app} of a herd, the number of pool samples at first test event (predictors), and the pool-milk sample result of a herd (outcome of interest) was performed as the number of pools in the first test event was representative of the number of pools in the other test events. The second logistic regression was performed to analyze whether the number of pool samples had an effect on the estimated WHP_{app} threshold. To visualize the effect of the number of pool samples at first test event on the required WHP_{app} threshold for a positive pool sample result, one WHP_{app} threshold ($WHP_{app,N}$) was calculated per farm. The $WHP_{app,N}$ threshold was calculated for a probability of detection of 95%. The formula used to calculate the $WHP_{app,N}$ threshold is as follows:

$$WHP_{app,N} = \frac{\ln\left(\frac{Pd}{1-Pd}\right) - a - b_2 \cdot N}{b_1},$$

where $WHP_{app,N}$ = the calculated apparent within-herd prevalence per farm; \ln = the natural logarithm; Pd = the desired probability of detection (i.e., 0.95); a = constant in the logistic regression (intercept); b_2 = regression coefficient of N ; N = number of pool samples at first test event per farm in 2018; b_1 = regression coefficient of the apparent within-herd prevalence.

Approximate confidence intervals were calculated for the estimated $WHP_{app,N}$ thresholds.

RESULTS

Data from 486 dairy farms were analyzed in this study, including 2,504 pooled milk samples and 70,456 serum samples from 72,451 cows older than 24 mo. The median herd size was 115 lactating cows per farm (minimum: 48, maximum: 762, Table 1). Out of the 486 analyzed herds, 485 herds had a herd size larger than 50 milk sampled cows. The median pool size was 44 cows (minimum: 2, maximum: 53) during the first sampling (Table 1). The pool size exceeded 50 cows ($n \leq 53$) in 5 of the pools during the first sampling. The rest of the pools ($n = 2,499$, 97.8%) contained a maximum of 50 milk samples. Of the 486 analyzed farms, 54.3% took pool-milk samples only once in the respective year ($n = 264$), 44.9% were sampled twice ($n = 218$), and 0.8% were sampled 3 times during 2018 ($n = 4$). Of the 490 farms, 0.8% ($n = 4$) took blood serum samples twice a year and were excluded from the analysis because their WHP_{app} differed between test events.

Difference Between Pool-Positive and Pool-Negative Herds

Pool-positive herds ($n = 375$) had a WHP_{app} of 3.4% (median; minimum: 0%, maximum: 33.3%, 95% CI = 0–11.4%), whereas pool-negative herds had a WHP_{app} of 1.2% (median; minimum: 0%, maximum: 13%, 95% CI = 0–6.4%, Figure 1). A total of 10.8% of the pool-negative herds ($n = 12$) had a WHP_{app} above 5%. Applying the Rogan-Gladen estimator results in a true WHP value of 4.1% (median, 95% CI = 2.1–6.1%) for pool-positive herds ($n = 375$), which had a median WHP_{app} value of 3.4%. A total of 79.6% of the farms ($n = 387$) were correctly classified as positive ($n = 356$) and negative ($n = 31$). A total of 16.5% ($n = 80$) of the herds were tested negative in all pool samples that had a WHP_{app} of >0%, and 3.9% ($n = 19$) of the herds were tested positive in at least one pool sample that had a WHP_{app} of 0% (Table 2). A weak but statistically significant correlation between the WHP_{app} of a farm and its quantitative test results of the pooled milk sam-

Table 2. Number of herds with a $WHP_{app} > 0\%$ and number of herds with a $WHP_{app} = 0\%$ classified by pool sample result on herd-level¹

Pool sample result on herd-level	$WHP_{app} > 0\%$	$WHP_{app} = 0\%$
Positive	356	19
Negative	80	31

¹ WHP_{app} = apparent within-herd prevalence.

ples was observed (P -value: < 0.01 ; Spearman's rank coefficient: 0.31). Analyzing the correlation between the WHP_{app} and the median of the farms quantitative test results of one sampling, we observed a higher and statistically significant correlation (P -value: < 0.01 ; Spearman's rank coefficient: 0.43).

Within-Herd Prevalence Thresholds for the Detection of Dairy Herds Seropositive for MAP Using Pooled Milk Samples in a Field Study

When the WHP_{app} increased by 1%, the odds of detecting at least one positive pooled milk sample increased by 40% [odds ratio (OR) = 1.40; 95% CI = 1.25–1.56; $P < 0.01$]. An increase of 5% WHP_{app} increased the probability for herd detection by 5.4 times ($OR^5 = 1.40^5 = 5.38$). The higher the WHP_{app} was, the higher the probability of detecting at least one positive milk pool per farm (Table 3). Herds with a WHP_{app} of at least 8% could be detected as positive by milk pools with a probability of 95% (Table 3). The number of pooled milk samples taken in the first sampling varied per farm between 2 and 16 (median: 3; Table 1). Results for $WHP_{app,N}$ thresholds are given in Figure 2. In the model including the number of pool samples at first test event, more pool samples reduced the threshold for the WHP_{app} with which a herd could be detected with a probability of at least 95% (Figure 2). When the WHP_{app} increased by 1%, the odds of detecting at least one positive pooled milk sample increased by 39% (OR = 1.39; 95% CI = 1.25–1.6; $P < 0.01$) including WHP_{app} and number of pool samples as predictors in the model. With each additional pool sample, the odds of identifying a farm as milk MAP-antibody positive increases by 13% (OR = 1.13, 95% CI = 0.99–1.29; $P = 0.06$).

Table 3. Estimated WHP_{app} threshold values for the detection of at least one positive pooled milk sample (for 2 different probabilities of detection)¹

Pd (%)	Calculation	WHP_{app} threshold (%) \pm SE	95% CI
90	$\{\ln^5 [0.9/(1 - 0.9)] - 0.2869\}/0.3326$	5.7 (\pm 0.7)	4.4–7.1
95	$\{\ln^5 [0.95/(1 - 0.95)] - 0.2869\}/0.3326$	8 (\pm 1.06)	5.9–10.1

¹Pd = desired probability of detection. WHP_{app} = apparent within-herd prevalence. \ln = natural logarithm.

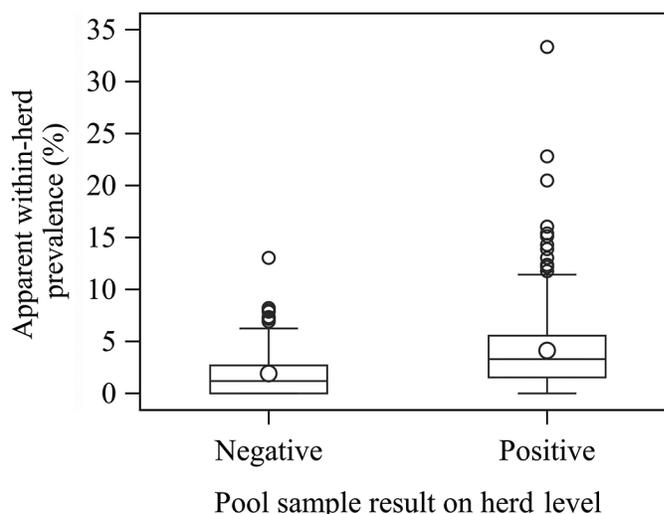


Figure 1. Apparent within-herd prevalence expressed as a percentage of positive serum samples per farm for the pool sample–negative and the pool sample–positive (sub)population. Lower whisker = minimum observation above 25% quantile – 1.5 \times interquartile range; lower line of the box = 25% quantile; horizontal line inside the box = median; circle inside the box = mean; upper line of the box = 75% quantile; upper whisker = maximum observation below 75% quantile + 1.5 \times interquartile range; circles outside the box = possible outliers (SAS Institute Inc., 2013).

DISCUSSION

This study illustrates an association between the within-herd prevalence of antibody-positive animals and the presence of positive pool milk samples in field samples.

Our analysis aimed to demonstrate that pooled milk samples are sufficient to detect serum MAP-antibody positive dairy herds with a $WHP_{app} < 8.9\%$ under field conditions, as a WHP_{app} value of 8.9% (± 1.62) for a probability of detection of 95% was estimated by Köhler et al. (2017). This hypothesis was confirmed, as the WHP_{app} threshold in our study was even lower (8%; ± 1.06 ; 95% probability of detection). Our results support the WHP_{app} value of 8.9% of Köhler et al. (2017) as the approximate confidence interval we determined under field conditions ranged between 5.9% and 10.1%. Furthermore, our confidence interval was even lower than the WHP_{app} confidence interval of up to 12.1% given by Köhler et al. (2017). As the analyzed herds

took individual serum samples only once in a year and serological testing was performed several weeks after pool-milk sampling, our results refer to one point in time.

We assumed that the probability of identifying a positive herd is higher in larger herds where more pool samples are taken at sampling day. The probability of finding at least one positive pool sample increased by 13% with each additional pool sample. Even though this finding is not significant, the number of pool-milk samples at sampling day showed a clear tendency to influence the likelihood of detecting at least one positive pooled milk sample. The number of pool samples at sampling day depends on the herd size. Therefore, larger herds are more likely to be detected, even if they have a WHP_{app} of $\leq 5\%$ and do not belong to the target group of the program.

Under experimental conditions, within-pool prevalence was identical to WHP, as only one pool per herd was tested (Köhler et al., 2017). In larger herds, the threshold might be even lower (Köhler et al., 2017). To evaluate this hypothesis, we computed $WHP_{app,N}$ thresholds according to the number of pool samples at sampling day. When applying the formula for $WHP_{app,N}$, only fixed parameters, except for N , are used, resulting in a linearity of the computed $WHP_{app,N}$ thresholds. As the number of pool samples and therefore the herd size increases, the WHP_{app} detection level decreases (Figure 2), so the use of pooled milk samples is a reliable tool to identify serum MAP-antibody positive herds in a

MAP prevalence reduction program where low MAP-prevalent farms are not in focus.

A possible reason for a lower WHP_{app} threshold in larger herds might be an accumulation of infected cows in one pool (Köhler et al., 2017). The target group of the voluntary MAP prevalence reduction program are farms with a within-herd prevalence of approximately 5% and higher. Since the aim of the Lower Saxonian MAP prevalence reduction program is prevalence reduction and not MAP eradication, pool-milk ELISA seems to be a feasible tool to identify program-relevant farms (Figure 1). Conclusively, the diagnostic test strategy of the pool-milk ELISA is fit for this purpose (Wright et al., 2006). Our purpose was to detect and confirm the (currently most) affected herds in the population. Hence, we wanted to use a method with a high specificity that is fit for disease confirmation for each tested herd as opposed to prevalence estimation for the whole population, which would have required a test fit for screening with a high sensitivity (OIE, 2021). Therefore, we consider the pool-milk ELISA fit for our purpose even if it has a considerably low sensitivity (Köhler et al., 2017). In the study presented here we estimated the minimum within-herd prevalence to detect a positive herd by applying the chosen test system to be 8%. This WHP_{app} threshold was determined by serum ELISA, but it applies to the pool-milk ELISA testing of herds. In the Lower Saxonian MAP prevalence reduction program it was intended to detect herds with a WHP_{app} threshold of at least 5%, because Donat et al. (2014) reported that from this WHP_{app} economic losses can be expected. Our results confirm that the use of pool-milk ELISA suffices to meet the intentions of the program as it identifies herds with at least 8% within-herd prevalence. As MAP- WHP_{app} has been shown to vary between 0% and 24% in Europe, but also reaches up to 42.6% (Singh et al., 2008; Nielsen and Toft, 2009) and prevalence estimations for Lower Saxony are still pending, we assume that the purpose to identify the (currently most) affected herds in the population was reached with the chosen threshold. Hence, we consider the difference between the intended and realized threshold (i.e., 5% and 8%, respectively) to be marginal. The threshold for the within-herd prevalence reflects the character of the infectious agent (Whittington et al., 2019) and indicates that, as opposed to other infectious agents, herds are already considerably affected with a comparably low WHP. Therefore, we conclude that the threshold in the reduction program was chosen in accordance with the requirements of the infectious agent and the current situation of the target population. Taking a detection probability of 95% as a basis, the aim of the MAP prevalence reduction program cannot be fulfilled completely

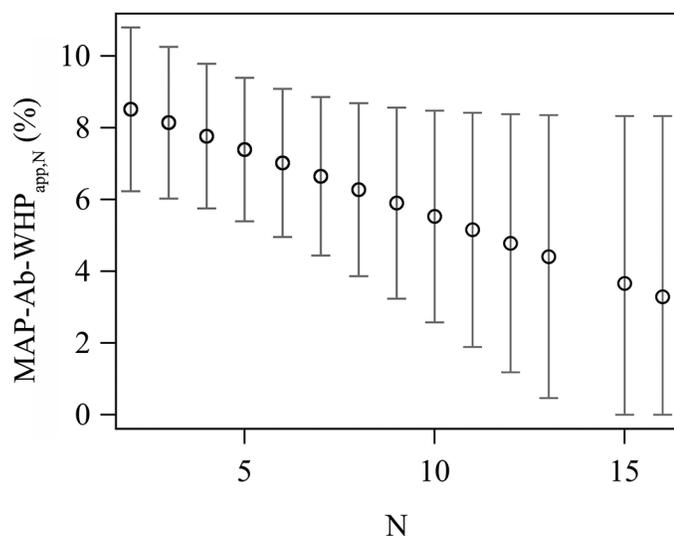


Figure 2. Association between MAP-Ab- $WHP_{app,N}$ threshold and number of pooled milk samples at sampling day per farm in 2018 (for a probability of detection of 95%) with approximate CI (95% CI; error bars); MAP-Ab- $WHP_{app,N}$ = apparent blood serological within-herd prevalence threshold per farm; N = number of pools at first test event.

by applying the pool-milk ELISA, but the majority of program-relevant farms can be identified. Because farms are tested repeatedly, the number of identified farms will increase over time. Hence, the magnitude of the realized threshold is sufficiently near the intended minimum detection limit of affected herds. Therefore, we conclude that this test is a cost-efficient way to confirm MAP-antibody positive herds with a WHP_{app} above the intended threshold. However, it must be considered that the threshold of 8% only applies to the analyzed study herds of 2018. Data from other years in the same program or other regions might produce other WHP_{app} thresholds. As mentioned in the introduction, the WHP_{app} differs among other regions of the world, with values ranging from 0% to approximately 40%. The pool-milk ELISA might perform only in those regions with a WHP_{app} of at least 8%. In regions with a $WHP_{app} > 8\%$, the ELISA might perform even better. It must be considered that WHP_{app} values can vary by the test used, the sample medium used, and the tested population (Nielsen and Toft, 2009).

With regard to sample collection, it must be considered that pool-milk samples collected in the milking parlor contain variable amounts of milk per cow proportional to their daily milk yield. This differs from the validation study where pools were composed in the laboratory (Köhler et al., 2017). This could have biased the probability of correctly classifying milk pools. On the other hand, one cow might be represented more than once within a pooled milk sample when the milk was taken by a milking robot. Consequently, the investigated number of pooled milk samples per farm and the apparent total number of lactating cows could be overrepresented in this analysis. As the size of the study population is large, it can be assumed that both reasons for possible overrepresentation of individual cows (i.e., regarding the volume of sample and the absolute number of samples per pool) did not affect the study results.

In general, a test is categorized as useful when it distinguishes between infected and noninfected herds (Christensen and Gardner, 2000). However, since all available MAP diagnostic tests are known for their low sensitivity, knowledge of these test characteristics must be considered when implementing their use in a control program. Under field conditions, low cost, ease of implementation, and time consumption are also test characteristics that should be assessed. Therefore, herd-level testing is often preferred to individual testing in national disease control programs, although often a prevalence threshold has to be taken into account. For example, in MAP control, environmental sampling methods have been used for MAP surveillance, where

a threshold WHP_{app} of 4% (90% probability of detection) was identified when boot swab and liquid manure samples were tested by fecal culture and direct fecal PCR (Donat et al., 2016). Pooling samples when the WHP is rather low always leads to a dilution of the antibody concentration in the pool and therefore to a loss of sensitivity (Wells et al., 2002; Tavoranpanich et al., 2004). The larger the pool is, the more difficult target detection will become (Wells et al., 2002). Hence, testing smaller milk pools was previously evaluated as an economical and convenient alternative to identify MAP-positive herds (Köhler et al., 2017) and is currently used in the MAP prevalence reduction program of Lower Saxony. The use of bulk milk samples for MAP diagnosis seemed inappropriate because of the low sensitivity, variable amounts of milk per cow, and different numbers of cows in such a sample (Geue et al., 2007). Despite these facts, which also apply to pooled milk samples analyzed in this study, the pool-milk ELISA seems sufficient to fulfill the aim of the Lower Saxonian MAP prevalence reduction program and to confirm the WHP_{app} threshold (e.g., 8.9%; 95% probability of detection) for herd-level detection by the method of Köhler et al. (2017) in the field.

To assess whether pooling affects the detection of herds with a WHP_{app} of $\geq 5\%$, we compared the WHP_{app} between pool-negative and pool-positive herds. As Figure 1 shows, distributions of WHP_{app} in pool-positive and pool-negative herds overlap. This is also represented by the confidence intervals of the 2 groups, which both include a WHP_{app} range of 0 to 6.4%. However, the pool-milk ELISA seems suitable to identify herds with a WHP_{app} of $\geq 5\%$ as intended by the program, which is also reflected by the WHP_{app} confidence interval of pool-negative herds as this is only up to 6.4%. A small proportion of pool-negative herds had a WHP_{app} above 5%. Because of their negative pool-milk result, these herds would not be considered for a prevalence reduction in the MAP prevalence reduction program. As a consequence, pool-negative herds have to be retested, as is stipulated in the MAP prevalence reduction program. Because the serum ELISA has a limited sensitivity of 58.2% and specificity of 99% (FLI, 2015), the WHP_{app} values presented here may be underestimated. With applying the Rogan-Gladen estimator for pool-positive herds ($n = 375$), we do not assume that the WHP_{app} was strongly biased by imperfect serum ELISA.

The fact that serological testing was performed several weeks after pool-milk sampling and therefore MAP-antibody positive cows might have been removed from the herd in the meantime might explain why a few herds (3.9%; $n = 19$) tested positive in at least one

pool sample that had a WHP_{app} of 0% (Table 3). In addition, it has to be assumed that false-positive results have occurred as well.

The Spearman's rank correlation coefficient for the association between the WHP_{app} of a farm and its quantitative pool test results was weak but statistically significant. This result may be explained if infected cows were unequally distributed in the cow cohorts that contributed to the milk pools per farm (Köhler et al., 2017). This may be because MAP infection events occur in age cohorts rather than in individual animals only (Zare et al., 2013). In addition, cows have a social structure, and it might be expected that age cohorts will still form loose clusters after introduction to the milking herd (Gygax et al., 2010; Boyland et al., 2016; de Freslon et al., 2019). As cattle become infected during calthood (Windsor and Whittington, 2010), the probability of becoming infected is more likely for calves of the same birth cohort compared with a random calf (Zare et al., 2013). Hence, the antibody concentration of the pool, in which cows of the MAP-infected birth cohort are milked, might be higher than in other pools of the same herd. Therefore, the probability that a farm will be identified as milk MAP-antibody positive might be higher in a milk pool sample collected in the field compared with a milk pool sample prepared randomly in a laboratory. As we analyzed the association between WHP_{app} and pool-milk results at the pool-milk sample level and not at the herd level, this birth cohort effect might have influenced the analytical results. We cannot further substantiate this hypothesis within the underlying data, as we lack individual identification of cows. Nevertheless, we assessed the same association at the herd level. There, the association was considerably better, underlining our assumption.

The pool-milk ELISA used in this study was characterized with a sensitivity of 44% and a specificity of 99% when a median of 50 cows per pool were tested (Köhler et al., 2017). In our study, the pool size (Table 1) was similar to the validation study of Köhler et al. (2017). Hence, we expected a similar test performance under the field conditions given in the MAP prevalence reduction program of Lower Saxony, Germany. Nevertheless, diagnostic misclassification due to different pool sizes cannot be excluded (Wells et al., 2002). The pooling of up to 50 individual milk samples has been shown to be a reliable tool to distinguish between herds considerably affected by MAP and herds with low or zero MAP infection rate. Thus, it is a convenient and useful alternative to individual or environmental sampling for herds with a WHP_{app} of $\geq 5\%$, especially when samples are already available due to other monitoring programs. As other parts of the world show similar WHP levels, the use of milk pool serology might be a

useful tool in other countries or under other conditions as well. As the study population corresponds to the target population and as the participation in the MAP prevalence reduction program is mandatory since 2017, the results of the study population reflect the actuality in Lower Saxony.

CONCLUSIONS

Based on the primary aim of the Lower Saxonian MAP prevalence reduction program to reduce the inner-herd prevalence in high MAP-antibody-prevalent herds, we investigated whether milk pool serology operates sufficiently reliable. The results of this field study support the use of pool-milk samples to detect high MAP-antibody-prevalent herds within this population, since most herds with a MAP-antibody prevalence $>5\%$ were identified using this testing strategy. Furthermore, we estimated the minimum WHP_{app} for a pool sample-positive herd to be 8% (95% detection probability), which corresponds well with the aim of the MAP prevalence reduction program to focus on herds with at least 5% WHP_{app} . As these data are generated by the analysis of a census in Lower Saxony and the WHP of MAP is similar compared with other dairy populations the use of milk pool serology seems a useful tool in other dairy populations with similar MAP prevalence. Therefore, pool sampling is suitable for the purpose of the MAP prevalence reduction program evaluated here and may be used in similar programs, as testing pool samples is an economic and convenient detection method.

ACKNOWLEDGMENTS

The authors thank the staff of the Lower Saxonian milk control association (Milchkontrollverband Elbe-Weser e.V., Landeskontrollverband Weser-Ems e.V., Germany) and the laboratories [Food and Veterinary Institute Oldenburg (Germany), Food and Veterinary Institute Braunschweig/Hannover (Germany), Institute for Animal Health of the Agricultural Investigation and Research Institute Nord-West (Oldenburg, Germany), and Department for Microbiology and Animal Hygiene, Georg-August-University Göttingen (Germany)] for excellent communication. The analysis of the data from the Lower Saxonian MAP prevalence reduction program was supported by Landesvereinigung der Milchwirtschaft Niedersachsen e.V. (Germany) and by the Lower Saxonian Animal Diseases Fund (Germany). The authors have not stated any conflicts of interest. M. Krieger: conceptualization, analysis and interpretation, writing of the original draft; S. Eisenberg: provision of data, conceptualization, review and editing of

the article; F. Freise: statistical analysis, review and editing of the article; H. Köhler: review and editing of the article; A. Campe: supervision and project administration, review and editing of the article.

REFERENCES

- Bauman, C. A., A. Jones-Bitton, J. Jansen, D. Kelton, and P. Menzies. 2019. Evaluation of bulk tank milk PCR and bulk tank milk modified ELISA tests for the detection of paratuberculosis at the herd level in goat and sheep dairies in Ontario, Canada. *J. Dairy Sci.* 102:511–520. <https://doi.org/10.3168/jds.2018-15020>.
- Benedictus, G., A. Dijkhuizen, and J. Stelwagen. 1987. Economic losses due to paratuberculosis in dairy cattle. *Vet. Rec.* 121:142–146. <https://doi.org/10.1136/vr.121.7.142>.
- BGBl. 1 (Bundesgesetzblatt 1). 2015. BHV1-Verordnung in the version of May 19, 2015. Pages 767–779. Bundesanzeiger Verlag.
- Bhutediya, J. M., P. Dandapat, A. Chakrabarty, R. Das, P. K. Nanda, S. Bandyopadhyay, and T. K. Biswas. 2017. Prevalence of paratuberculosis in organized and unorganized dairy cattle herds in West Bengal, India. *Vet. World* 10:574–579. <https://doi.org/10.14202/vetworld.2017.574-579>.
- Boylard, N. K., D. T. Mlynski, R. James, L. J. N. Brent, and D. P. Croft. 2016. The social network structure of a dynamic group of dairy cows: From individual to group level patterns. *Appl. Anim. Behav. Sci.* 174:1–10. <https://doi.org/10.1016/j.applanim.2015.11.016>.
- Chiodini, R. J., H. J. Van Kruiningen, and R. S. Merkal. 1984. Ruminant paratuberculosis (Johne's disease): The current status and future prospects. *Cornell Vet.* 74:218–262.
- Christensen, J., and I. A. Gardner. 2000. Herd-level interpretation of test results for epidemiologic studies of animal diseases. *Prev. Vet. Med.* 45:83–106. [https://doi.org/10.1016/S0167-5877\(00\)00118-5](https://doi.org/10.1016/S0167-5877(00)00118-5).
- Clarke, C. J. 1997. The pathology and pathogenesis of paratuberculosis in ruminants and other species. *J. Comp. Pathol.* 116:217–261. [https://doi.org/10.1016/S0021-9975\(97\)80001-1](https://doi.org/10.1016/S0021-9975(97)80001-1).
- Collins, M. T., I. A. Gardner, F. B. Garry, A. J. Roussel, and S. J. Wells. 2006. Consensus recommendations on diagnostic testing for the detection of paratuberculosis in cattle in the United States. *J. Am. Vet. Med. Assoc.* 229:1912–1919. <https://doi.org/10.2460/javma.229.12.1912>.
- de Freslon, I., B. Martinez-Lopez, J. Belkhiria, A. Strappini, and G. Monti. 2019. Use of social network analysis to improve the understanding of social behaviour in dairy cattle and its impact on disease transmission. *Appl. Anim. Behav. Sci.* 213:47–54. <https://doi.org/10.1016/j.applanim.2019.01.006>.
- Donat, K., N. Hahn, T. Eisenberg, K. Schlez, H. Köhler, W. Wolter, M. Rohde, R. Pützschel, U. Rösler, K. Failing, and P. M. Zschöck. 2016. Within-herd prevalence thresholds for the detection of *Mycobacterium avium* subspecies *paratuberculosis*-positive dairy herds using boot swabs and liquid manure samples. *Epidemiol. Infect.* 144:413–424. <https://doi.org/10.1017/S0950268815000977>.
- Donat, K., A. Soschinka, G. Erhardt, and H. R. Brandt. 2014. Paratuberculosis: Decrease in milk production of German Holstein dairy cows shedding *Mycobacterium avium* ssp. *paratuberculosis* depends on within-herd prevalence. *Animal* 8:852–858. <https://doi.org/10.1017/S1751731114000305>.
- Doré, E., J. Paré, G. Côté, S. Buczinski, O. Labrecque, J. P. Roy, and G. Fecteau. 2012. Risk factors associated with transmission of *Mycobacterium avium* ssp. *paratuberculosis* to calves within dairy herd: A systematic review. *J. Vet. Intern. Med.* 26:32–45. <https://doi.org/10.1111/j.1939-1676.2011.00854.x>.
- Fecteau, M. E. 2018. Paratuberculosis in cattle. *Vet. Clin. North Am. Food Anim. Pract.* 34:209–222. <https://doi.org/10.1016/j.cvfa.2017.10.011>.
- FLI. 2015. Test characteristics of the ELISA assays which are licensed in Germany for the detection of antibodies against *Mycobacterium avium* ssp. *paratuberculosis* in cattle, determined using the reference panels for serum and milk of the NRL for paratuberculosis, Accessed May 5, 2021. https://www.openagrar.de/receive/openagrar_mods_00061681.
- Geue, L., H. Köhler, W. Klawonn, K. Dräger, R. G. Hess, and F. J. Conraths. 2007. Investigations on suitability of ELISA for the detection of antibodies against *Mycobacterium avium* ssp. *paratuberculosis* in bulk milk samples from Rhineland-Palatinate. *Berl. Munch. Tierarztl. Wochenschr.* 120:67–78.
- Gygax, L., G. Neisen, and B. Wechsler. 2010. Socio-spatial relationships in dairy cows. *Ethology* 116:10–23. <https://doi.org/10.1111/j.1439-0310.2009.01708.x>.
- IDEXX. 2018. *Mycobacterium paratuberculosis* Antibody Test Kit. Accessed Oct. 17, 2019. <https://www.idexx.de/de/livestock/resources-support/product-inserts/>.
- Johnson-Ifearulundu, Y., and J. B. Kaneene. 1999. Distribution and environmental risk factors for paratuberculosis in dairy cattle herds in Michigan. *Am. J. Vet. Res.* 60:589–596.
- Kohl, J. L., S. Eisenberg, I. Noll, M. Zschöck, T. Eisenberg, and K. Donat. 2019. Two-stage control of paratuberculosis: Herd-status surveillance as the basis for operational measures to reduce the prevalence. Experiences from Lower Saxony, Hesse, Thuringia and Tyrol. *Tierarztl. Prax. Ausg. G Grosstiere Nutztiere* 47:171–183.
- Köhler, H., M. Ziller, F. Gierke, and K. Donat. 2017. Within-pool prevalence limits for the identification of paratuberculosis infected herds using antibody detection in pooled milk samples. *Berl. Munch. Tierarztl. Wochenschr.* 130:34–41.
- Nielsen, S. S., and N. Toft. 2009. A review of prevalences of paratuberculosis in farmed animals in Europe. *Prev. Vet. Med.* 88:1–14. <https://doi.org/10.1016/j.prevetmed.2008.07.003>.
- Noll, I., T. Eisenberg, K. Failing, M. Rohde, K. Schlez, W. Wolter, A. Fawzy, and M. Zschöck. 2017. Herd prevalence of *Mycobacterium avium* ssp. *paratuberculosis* (MAP) in Hessian dairy herds. *Berl. Munch. Tierarztl. Wochenschr.* 130:13–20.
- OIE. 2021. Principles and methods of validation of diagnostic assays for infectious diseases. Accessed Aug. 3, 2021. https://www.oie.int/fileadmin/Home/eng/Health_standards/tahm/1.01.06_VALIDATION.pdf.
- Ott, S. L., S. J. Wells, and B. A. Wagner. 1999. Herd-level economic losses associated with Johne's disease on US dairy operations. *Prev. Vet. Med.* 40:179–192. [https://doi.org/10.1016/S0167-5877\(99\)00037-9](https://doi.org/10.1016/S0167-5877(99)00037-9).
- Pützschel, R., E. Einax, V. Zoche-Golob, and K. Donat. 2017. Spread of infection with *Mycobacterium avium* ssp. *paratuberculosis* (MAP) in cattle herds in Saxony and Thuringia on herd level. *Berl. Munch. Tierarztl. Wochenschr.* 130:21–28.
- Raizman, E. A., S. J. Wells, C. A. Muñoz-Zanzi, and S. Tavornpanich. 2011. Estimated within-herd prevalence (WHP) of *Mycobacterium avium* ssp. *paratuberculosis* in a sample of Minnesota dairy herds using bacterial culture of pooled fecal samples. *Can. J. Vet. Res.* 75:112–116.
- SAS Institute Inc. 2013. The BOXPLOT Procedure. Page 1123 in *SAS/STAT® 13.1 User's Guide*. SAS Institute Inc.
- Sergeant, E. S. G., C. G. McAloon, J. A. Tratalos, L. R. Citer, D. A. Graham, and S. J. More. 2019. Evaluation of national surveillance methods for detection of Irish dairy herds infected with *Mycobacterium avium* ssp. *paratuberculosis*. *J. Dairy Sci.* 102:2525–2538. <https://doi.org/10.3168/jds.2018-15696>.
- Singh, S. V., A. V. Singh, R. Singh, S. Sharma, N. Shukla, S. Misra, P. K. Singh, J. S. Sohal, H. Kumar, P. K. Patil, P. Misra, and K. S. Sandhu. 2008. Sero-prevalence of bovine Johne's disease in buffaloes and cattle population of North India using indigenous ELISA kit based on native *Mycobacterium avium* subspecies *paratuberculosis* 'Bison type' genotype of goat origin. *Comp. Immunol. Microbiol. Infect. Dis.* 31:419–433. <https://doi.org/10.1016/j.cimid.2007.06.002>.
- Tavornpanich, S., I. A. Gardner, R. J. Anderson, S. Shin, R. H. Whitlock, T. Fyock, J. M. Adaska, R. L. Walker, and S. K. Hietala. 2004. Evaluation of microbial culture of pooled fecal samples for detection of *Mycobacterium avium* subsp. *paratuberculosis* in large dairy herds. *Am. J. Vet. Res.* 65:1061–1070. <https://doi.org/10.2460/ajvr.2004.65.1061>.

- Thorel, M. F., M. Krichevsky, and V. V. Lévy-Frébault. 1990. Numerical taxonomy of mycobactin-dependent mycobacteria, emended description of *Mycobacterium avium*, and description of *Mycobacterium avium* ssp. *avium* ssp. nov., *Mycobacterium avium* ssp. *paratuberculosis* ssp. nov., and *Mycobacterium avium* ssp. *silvaticum* ssp. nov. *Int. J. Syst. Bacteriol.* 40:254–260. <https://doi.org/10.1099/00207713-40-3-254>.
- Thorne, J. G., and L. E. Hardin. 1997. Estimated prevalence of paratuberculosis in Missouri, USA cattle. *Prev. Vet. Med.* 31:51–57. [https://doi.org/10.1016/S0167-5877\(96\)01138-5](https://doi.org/10.1016/S0167-5877(96)01138-5).
- Thüringer Tierseuchenkasse. 2020. Paratuberkulose bei Rindern. Accessed Jun. 3, 2020. <https://www.paratbc.de/>.
- Trangadia, B. J., S. K. Rana, K. Nagmani, and V. A. Srinivasan. 2012. Serological investigation of bovine brucellosis, Johne's disease and infectious bovine rhinotracheitis in two states of India. *J. Adv. Vet. Res.* 2:38–41.
- Wells, S. J., R. H. Whitlock, C. J. Lindeman, and T. Fyock. 2002. Evaluation of bacteriologic culture of pooled fecal samples for detection of *Mycobacterium paratuberculosis*. *Am. J. Vet. Res.* 63:1207–1211. <https://doi.org/10.2460/ajvr.2002.63.1207>.
- Whitlock, R. H., and C. Buergelt. 1996. Preclinical and clinical manifestations of paratuberculosis (including pathology). *Vet. Clin. North Am. Food Anim. Pract.* 12:345–356. [https://doi.org/10.1016/S0749-0720\(15\)30410-2](https://doi.org/10.1016/S0749-0720(15)30410-2).
- Whittington, R., K. Donat, M. F. Weber, D. Kelton, S. S. Nielsen, S. Eisenberg, N. Arrigoni, R. Juste, J. L. Sáez, N. Dhand, A. Santi, A. Michel, H. Barkema, P. Kralik, P. Kostoulas, L. Citer, F. Grifin, R. Barwell, M. A. S. Moreira, I. Slana, H. Koehler, S. V. Singh, H. S. Yoo, G. Chávez-Gris, A. Goodridge, M. Ocepek, J. Garrido, K. Stevenson, M. Collins, B. Alonso, K. Cirone, F. Paolicchi, L. Gavey, M. T. Rahman, E. de Marchin, W. Van Praet, C. Bauman, G. Fecteau, S. McKenna, M. Salgado, J. Fernández-Silva, R. Dzedzinska, G. Echeverría, J. Seppänen, V. Thibault, V. Fridriksdottir, A. Derakhshandeh, M. Haghkhal, L. Ruocco, S. Kawaji, E. Momotani, C. Heuer, S. Norton, S. Cadmus, A. Agdestein, A. Kampen, J. Sztejn, J. Frössling, E. Schwan, G. Caldow, S. Strain, M. Carter, S. Wells, M. Munyeme, R. Wolf, R. Gurung, C. Verdugo, C. Fourichon, T. Yamamoto, S. Thapaliya, E. Di Labio, M. Ekgat, A. Gil, A. N. Alesandre, J. Piaggio, A. Suanes, and J. H. de Waard. 2019. Control of paratuberculosis: Who, why and how. A review of 48 countries. *BMC Vet. Res.* 15:198. <https://doi.org/10.1186/s12917-019-1943-4>.
- Windsor, P. A., and R. J. Whittington. 2010. Evidence for age susceptibility of cattle to Johne's disease. *Vet. J.* 184:37–44. <https://doi.org/10.1016/j.tvjl.2009.01.007>.
- Wright, P., S. Edwards, A. Diallo, and R. Jacobson. 2006. Development of a framework for international certification by the OIE of diagnostic tests validated as fit for purpose. Pages 27–35 in First International Conference of the OIE Reference Laboratories and Collaborating Centres. Florianopolis, Brazil, D. B. (Basel) 128, Karger.
- Zare, Y., G. E. Shook, M. T. Collins, and B. W. Kirkpatrick. 2013. Evidence of birth seasonality and clustering of *Mycobacterium avium* subspecies *paratuberculosis* infection in US dairy herds. *Prev. Vet. Med.* 112:276–284. <https://doi.org/10.1016/j.prevetmed.2013.07.016>.

ORCID

- M. Krieger  <https://orcid.org/0000-0003-3536-0236>
 S. Eisenberg  <https://orcid.org/0000-0003-0079-3015>
 H. Köhler  <https://orcid.org/0000-0002-0422-4349>
 F. Freise  <https://orcid.org/0000-0002-8493-8359>
 A. Campe  <https://orcid.org/0000-0001-9390-759X>