FACULTY OF VETERINARY MEDICINE
accredited by EAEVE

The prevalence of *Salmonella* Dublin in Alberta's dairy herds: A comparison between 2022 and 2024

PART 2: RESEARCH PAPER

Word count: 2994

Jessie van Heumen
Student number: 02113876

Supervisor: Prof. dr. Herman W. Barkema
Supervisor: Prof. dr. Sarne De Vliegher

A dissertation submitted to Ghent University in partial fulfilment of the requirements for the degree of Master of Veterinary Medicine

Academic year: 2024 – 2025

UNIVERSITEIT GENT

Preface

This thesis represents the final step in my studies and reflects not only academic work, but also personal development. I would not have been able to complete it without the support and guidance of several people, to whom I am sincerely grateful.

I would first like to thank Prof. Dr. Herman W. Barkema, Professor in Epidemiology of Infectious Diseases at the University of Calgary and NSERC Industrial Research Chair in Infectious Diseases of Dairy Cattle. From our first conversation at the coffee machine in Merelbeke till today, we have shared a very positive and supportive collaboration, for which I am truly grateful. Herman believed in me from the start and gave me the opportunity to experience research across different and meaningful international environments. His trust, encouragement and openness had a real impact on both my academic and personal development.

I also thank Dr. Waseem Shaukat, PhD candidate under Herman's supervision, for his clear guidance and quick, helpful responses throughout the writing process.

My thanks also go to Prof. Dr. Sarne De Vliegher for his flexibility within the Ghent University M-team, where international experiences and independent thinking are encouraged. This made it possible for me to take on this project abroad.

Thanks to Henry Grace for proofreading my English and for helping with statistics when things became difficult to untangle. I also thank Iris Steeg, who reviewed my work several times and helped me identify structural improvements when I had lost perspective.

Finally, I would like to thank Sofie Roelofs, a lifelong friend, for always being there. I also want to thank Lynn Bruins, for her steady support during the personal challenging moments. I have no doubts she will graduate soon as a compassionate veterinarian. A special thank you to my oma, who used to light a candle before every exam, her support is still with me.

Ghent, Belgium, May 2025

Jessie van Heumen

Abstract ENG

This Master thesis assessed changes in herd-level *Salmonella* Dublin status between 2022 and 2024 among 454 dairy farms in Alberta, Canada. Bulk tank milk samples were collected and tested using ELISA with two cut-offs: PP≥35% and PP≥15%, to determine herd positivity.

At cut-off PP≥35%, 85.2% of herds testing negative in 2022 remained negative in 2024, while 50% of positive herds remained positive. The lower cut-off increased sensitivity from 16.3% to 40.6%, identifying more positive herds and greater status changes from positive ↔ negative over time, at the cost of reducing specificity from 97.5% to 91.9%.

Univariable logistic regression analysis revealed strong, significant associations between overall herd status for *Salmonella* Dublin in 2022 and herd status outcomes for *Salmonella* Dublin in 2024 for both cut-offs (OR=13.3 at PP≥35%, OR=17.2 at PP≥15%), consistent across regions (north, central, south), herd types (Hutterite vs. Non-Hutterite), milking systems (conventional vs. automated), and farm sizes (small, medium, large).

McNemar tests confirmed overall stability in herd status, with some significant shifts in positive ↔ negative status, observed in specific regions and herd types, notably in Central Alberta (PP≥35%) and among Non-Hutterite farms (PP≥15%).

These findings suggest persistent herd status and highlight the importance of ongoing surveillance and biosecurity. Variability in test characteristics between cut-offs emphasizes the need for careful interpretation of herd status changes. Future research should focus on identifying herd-level risk factors and refining testing strategies to differentiate acute from chronic infections, improving surveillance and control of *Salmonella* Dublin in Alberta's dairy herds.

Abstract NL

Deze masterproef onderzocht veranderingen in de *Salmonella* Dublin status tussen 2022 en 2024 op 454 melkveebedrijven in Alberta, Canada. Bulk tank melk monsters werden verzameld en getest met ELISA, met twee afkapwaarden: PP≥35% en PP≥15%, om een positieve status van de boerderij te bepalen.

Bij de afkapwaarde van PP≥35% bleef 85,2% van de bedrijven die in 2022 negatief testten ook in 2024 negatief, terwijl 50% van de positief geteste bedrijven positief bleef. De lagere afkapwaarde (PP≥15%) verhoogde de gevoeligheid van 16,3% naar 40,6% en identificeerde meer positieve bedrijven en meer statusveranderingen over de tijd, maar verlaagde de specificiteit van 97,5% naar 91,9%.

Logistische regressie toonde sterke en significante associaties tussen de *Salmonella* Dublin status in 2022 en in 2024 voor beide afkapwaarden (OR=13,3 bij PP≥35%, OR=17,2 bij PP≥15%), consistent over regio's (noord, centraal, zuid), bedrijfstypes (Hutteriet vs. niet-Hutteriet), melksystemen (conventioneel vs. automatisch) en bedrijfsgroottes (klein, middelgroot, groot).

McNemar testen bevestigden een algemene stabiliteit in bedrijfsstatus, met enkele significante verschuivingen van positief ↔ negatief, vooral in Centraal-Alberta (PP≥35%) en bij niet-Hutterite bedrijven (PP≥15%).

Deze bevindingen wijzen op een aanhoudende infectiestatus en benadrukken het belang van voortdurende monitoring en bioveiligheid. De variatie in testkarakteristieken tussen afkapwaarden benadrukt de noodzaak van zorgvuldige interpretatie. Toekomstig onderzoek moet zich richten op het identificeren van risicofactoren op bedrijfsniveau en het verfijnen van teststrategieën om acute van chronische infecties te onderscheiden, ter ondersteuning van een gerichte aanpak van *Salmonella* Dublin in de melkveehouderij van Alberta.

Keywords

Salmonella Dublin – Dairy cattle – Alberta – Herd-level prevalence – Bulk tank milk – ELISA

Introduction

Background

Due to globalization and heightened awareness, biosecurity and the control of infectious diseases have become increasingly important in both global trade and on-farm practices. (Western) countries that rely on livestock exports face continued pressure to prevent the (re)introduction of diseases. This international pressure translates into national responsibilities. In Canada, maintaining emergency preparedness and preventing disease reintroduction remain essential. On farm level, infectious diseases threaten animal health and productivity, making biosecurity a priority. These concerns arise alongside trends such as declining dairy farm numbers, larger herd sizes, and increasing use of technology. Those developments create opportunities, such as using milk for disease diagnostics and monitoring. Since disease prevalence tends to rise with herd size, pathogens like *Salmonella,* capable of shedding in milk, pose a growing concern (Barkema et al., 2009; Barkema et al., 2015).

Salmonella enterica subspecies *enterica* serovar Dublin (*S.* Dublin) is a host-adapted bacterium in cattle that causes high calf mortality and reduced performance in adult cows (Velasquez-Munoz et al., 2024). Persistent carriers may shed bacteria intermittently, and environmental survival increases transmission risk (Perry et al., 2023). While *S.* Dublin predominantly affects cattle, it occasionally causes disease in humans through raw milk consumption (Fierer, 1983; Leedom, 2006; McDonough et al., 1999; Silva et al., 2014; Velasquez-Munoz et al., 2024). Most infections are self-limiting, but severe illnesses may occur in vulnerable groups (Helms et al., 2003). These risks, combined with growing concerns about antimicrobial resistance (AMR), make effective control efforts even more critical (Harvey et al., 2017; Velasquez-Munoz et al., 2024).

Although prevalence estimates of *S.* Dublin varies among countries and regions, it is an emerging pathogen on dairy farms in Canada (Perry et al., 2023). Bulk tank milk samples (BTMS) of 7.5% farms in Ontario (Nobrega et al., 2024) and 9.6% farms in Quebec (Um et al., 2022) tested positive for *S*. Dublin antibodies at a single timepoint. Additionally, 29.7% of farms in British Columbia (Boyd et al., 2025), and 15.6% in Alberta (Shaukat et al., 2024), tested positive at least once out of four quarterly tests. Therefore, Canada requires more data regarding the prevalence of *S.* Dublin.
As one of Canada's key dairy regions, Alberta plays a critical role in national milk production. It accounts for nearly 9% of the total, with 468 active producers. These numbers highlight Alberta's role in national milk production and need for herd health monitoring (Alberta Milk, 2022, 2024). This thesis follows the industry's regional division: South, Central, and North. (Alberta Milk, 2024; Shaukat et al., 2024).

Problem statement

The extent and impact of *Salmonella* Dublin in Alberta remains poorly understood. Given its importance, *S.* Dublin is a reportable disease in Alberta (Government, 2022). Alberta's Cattle Health Surveillance System (CHeSS) project identified 77 positive farms in 2021-2022 across 4 timepoints in Alberta, underscoring the need for closer monitoring (Shaukat et al., 2024). Hutterite farms, religious communal farms occupying ~4% of Alberta's farmland, tested positive less often than Non-Hutterite herds (Adandom et al., 2023; Shaukat et al., 2024). However, broader epidemiological trends in different stratifications over time have not been fully explored.

Objective

This thesis aims to compare the *Salmonella* Dublin status of dairy herds in Alberta between 2022 and 2024, stratified by region, herd type, milking system, and herd size.

The main research question was: *Are there significant trends in herd-level S. Dublin status between 2022 and 2024, and do these trends differ by region, herd type, milking system, or herd size?*

Comparing 2022 and 2024 herd-level status will reveal trends that can inform future research on within-herd prevalence in Alberta's evolving dairy industry (Kent et al., 2021; Nielsen, 2013b; Shaukat et al., 2024). This study contributes to a larger project investigating risk factors, economic impact, and antimicrobial resistance in *S.* Dublin in Alberta. The results aim to support both public and animal health (McDonough et al., 1999; Shaukat et al., 2024).

Material and methods

Study Design and sampling

This longitudinal study aimed to assess changes in herd-level status of *Salmonella* Dublin between 2022 and 2024 in Alberta's dairy herds. During routine milk collection in June 2024, bulk tank milk samples (40 mL) were obtained from all operational dairy farms in Alberta, following the standard protocols outlined by Alberta Milk (Alberta Milk, 2022). As the entire population was sampled, no sample size calculation was required (Shaukat et al., 2024). Samples were mixed and kept at 4°C throughout transport and storage at the University of Calgary's Faculty of Veterinary Medicine laboratory. The July 2022 *S.* Dublin status of these herds was obtained from a previously conducted study (Shaukat et al., 2024). This thesis includes only the 454 herds sampled in both July 2022 and June 2024. These farms were further categorised based on Region (North, Central, South); Colony type (Hutterite or non-Hutterite); Milking system; either Automatic (AMS), or Conventional (CMS) and Herd size, defined by daily milk delivery as Small (≤3,600 L/day), Medium (3,600–7,200 L/day), or Large (>7,200 L/day) (Shaukat et al., 2024).

Diagnostic Testing

Samples were tested using PrioCheck *S.* Dublin Ab kit (ThermoFisher Scientific, 2024), detecting antibodies against *S.* Dublin (Attachment 1). This gave a semi-quantitative concentration in Optical Density Coefficient in percentage (ODC%) (Veling et al., 2002). The corrected Optic Density (OD) allowed the calculation of the PP (Percentage Positivity), enabling differentiation between negative and positive results (Shaukat et al., 2024). Higher ODC% values indicates a stronger correlation with antibody concentration, suggesting a greater spread of infection within the herd (Veling et al., 2002). Any background interference (negative control OD) was subtracted from the sample OD to give a corrected measurement. The PP was calculated as followed (Shaukat et al., 2024):

$$PP\% = \left(\frac{Corrected\ OD\ samples}{Corrected\ OD\ positive\ controls} * 100\right) - 10$$

A PP%≥35 cut-off defined positives, as recommended by the manufacturer. At this threshold, the test has a sensitivity of 16.3% and a specificity of 97.5% (Um et al., 2022; ThermoFisher Scientific, 2024). To improve sensitivity, an additional analysis was conducted using a lower cut-off of PP%≥15, which increases sensitivity to 40.6%, but reduces specificity to 91.9% (Um et al., 2022).

Statistical Analysis

Data from 2022 and 2024 was entered into sheets in Microsoft Excel (Microsoft Corporation, 2021). Tables were created presenting both cutoffs (PP≥35% and PP≥15%), followed by a descriptive analysis of herd-level status between 2022 and 2024. Proportions of herds that maintained their status (positive or negative) or changed status (from positive to negative, and from negative to positive) were calculated separately for each cut-off (Attachment 2). Then, the dataset was exported to R4.5.0 for analysis (code in Attachment 3). Univariable logistic regression was performed to assess the association between herd status for *Salmonella* Dublin in 2022 and herd status outcomes for *Salmonella* Dublin in 2024, giving odds ratios (OR), confidence intervals (95%CI), and corresponding p-values. Then, stratified logistic regression analysis with 2024 status as an outcome and 2022 status as an explanatory variable were performed stratified by region, herd type, herd size, and milking system. A simple logistic regression estimated the overall association between test results in 2022 and 2024, assuming this association was constant across all farms sizes and regions. Including interaction terms

with farm size or region allowed assessment of whether the strength of this association varied between different farm sizes or regions.

The odds ratio (OR), calculated using logistic regression, estimated the association between herd-level *Salmonella* Dublin status in 2022 and 2024. OR compares the odds of being positive in 2024 for farms positive in 2022 versus negative in 2022 (Table 1). The formula is as followed:

$$OR = \frac{\left(\frac{a}{b}\right)}{\left(\frac{c}{d}\right)}$$

Table 1: Calculation of OR with a 2x2 table tabulating 2022 and 2024 results. A; are farms tested positive in 2022 and 2024, b; tested positive in 2022 and negative in 2024, c; tested negative in 2022 and positive in 2024, d; tested negative in 2022 and 2024.

		2024 Status
2022 Status	*Positive*	*Negative*
Positive	a	b
Negative	c	d

McNemar tests were used to assess the significance of herd status changes (positive↔negative) based on discordant pairs. Analyses were stratified by region, herd type, milking system, and herd size, separately for both cut-offs.

Results

This analysis compared the *Salmonella* Dublin status of 454 Alberta dairy herds tested in both July 2022 and June 2024. Table 2 summarizes changes in apparent prevalence, stratified by region, herd type, milking system, and herd size. Attachment 2 represents full estimates of farms changing status.

Table 2: Apparent herd-level prevalence of Salmonella Dublin in 2022 and 2024 by region, colony status, milking system, and farm size, using cut-off values of ≥35% and ≥15% Percent Positivity (PP). Odds Ratios (OR), 95% Confidence Intervals (CI), and P-values from logistic regression (LR) are presented to assess associations between 2022 and 2024 status. An OR >1 indicates higher odds of testing positive in 2024 for herds that were positive in 2022. P-values < 0.05 are considered statistically significant. McNemar test P-values are presented indicating statistical significance between herds changing status from positive to negative and from negative to positive.

		Apparent Prevalence at 35% PP				
Variable	**N (%)**	**2022**	**2024**	**OR (95% CI)**	**P-Value LG**	**P-Value Mcnemar**
Grand Total	**454 (100)**	**8,4%**	**10,6%**	**13,3 (6,38-28,2)**	**<0.001**	**0,194**
Regions						
Central	**179 (39,4)**	5,6%	11,2%	16,6 (4,27-71,9)	<0.001	0.034
North	**124 (27,3)**	8,1%	6,5%	18,3 (3,60-97,9)	<0.001	0.752
South	**151 (33,3)**	11,9%	13,2%	11,1 (3,67-34,6)	<0.001	0.823
Colonies						
Hutterite	**132 (29,1)**	6,1%	4,5%	24,2 (3,72-165,0)	<0.001	0.724
Non-Hutterite	**322 (70,1)**	9,3%	13,0%	11,7 (5,16-27,0)	<0.001	0.082
Systems						
AMS	**132 (29,1)**	6,1%	8,3%	71,4 (13,1-587,0)	<0.001	0.450
CMS	**322 (70,1)**	9,3%	11,5%	8,54 (3,67-19,7)	<0.001	0.349
Sizes						
Small	**213 (46,9)**	7,5%	9,4%	15,4 (4,92-49,5)	<0.001	0.502
Medium	**165 (36,3)**	7,9%	10,9%	22,72 (6,46-88,6)	<0.001	0.302
Large	**76 (16,7)**	11,8%	13,2%	4,3 (0,78-20,6)	0.073	1

		Apparent Prevalence at 15% PP				
Variable	**N (%)**	**2022**	**2024**	**OR (95% CI)**	**P-Value LG**	**P-Value McNemar**
Grand Total	**454 (100)**	**18,3%**	**20,5%**	**17,2 (9,89-30,7)**	**<0.001**	**0,268**
Regions						
Central	**179 (39,4)**	14,5%	19,0%	19,3 (7,48-54,0)	<0.001	0.153
North	**124 (27,3)**	16,1%	16,9%	79,2 (21,3-376)	<0.001	1
South	**151 (33,3)**	24,5%	25,2%	7,5 (3,31-17,5)	<0.001	1
Colonies						
Hutterite	**132 (29,1)**	17,4%	11,4%	16 (4,93-58,6)	<0.001	0.099
Non-Hutterite	**322 (70,1)**	18,6%	24,2%	20,8 (10,7-42,7)	<0.001	0.014
Systems						
AMS	**132 (29,1)**	12,9%	17,4%	22,7 (7,10-83,4)	<0.001	0.211
CMS	**322 (70,1)**	20,5%	21,7%	15,9 (8,44-30,8)	<0.001	0.671
Sizes						
Small	**213 (46,9)**	14,6%	17,8%	13,6 (5,82-33,2)	<0.001	0.281
Medium	**165 (36,3)**	18,2%	20,0%	23,9 (9,31-67,0)	<0.001	0.663
Large	**76 (16,7)**	28,9%	28,9%	14,4 (4,57-51,3)	<0.001	1

Overall results

Using cutoff PP≥35%, 85.2% of the 454 farms that tested negative in 2022 (n=387) remained negative in 2024, whereas 6.4% (n=29) became newly positive. Of the 38 farms that tested positive in 2022, 50% (n=19) remained positive, and 50% (n=19) became negative. Using cutoff PP≥15%, 89.8% of the 371 farms that tested negative in 2022 remained negative in 2024, while 10.2% became positive. Of the 83 farms that tested positive in 2022, 66.3% remained positive and 33.7% became negative. More status changes occurred at PP≥15% than at PP≥35%, While most herds remained negative, a minority changed status (Figure 1).

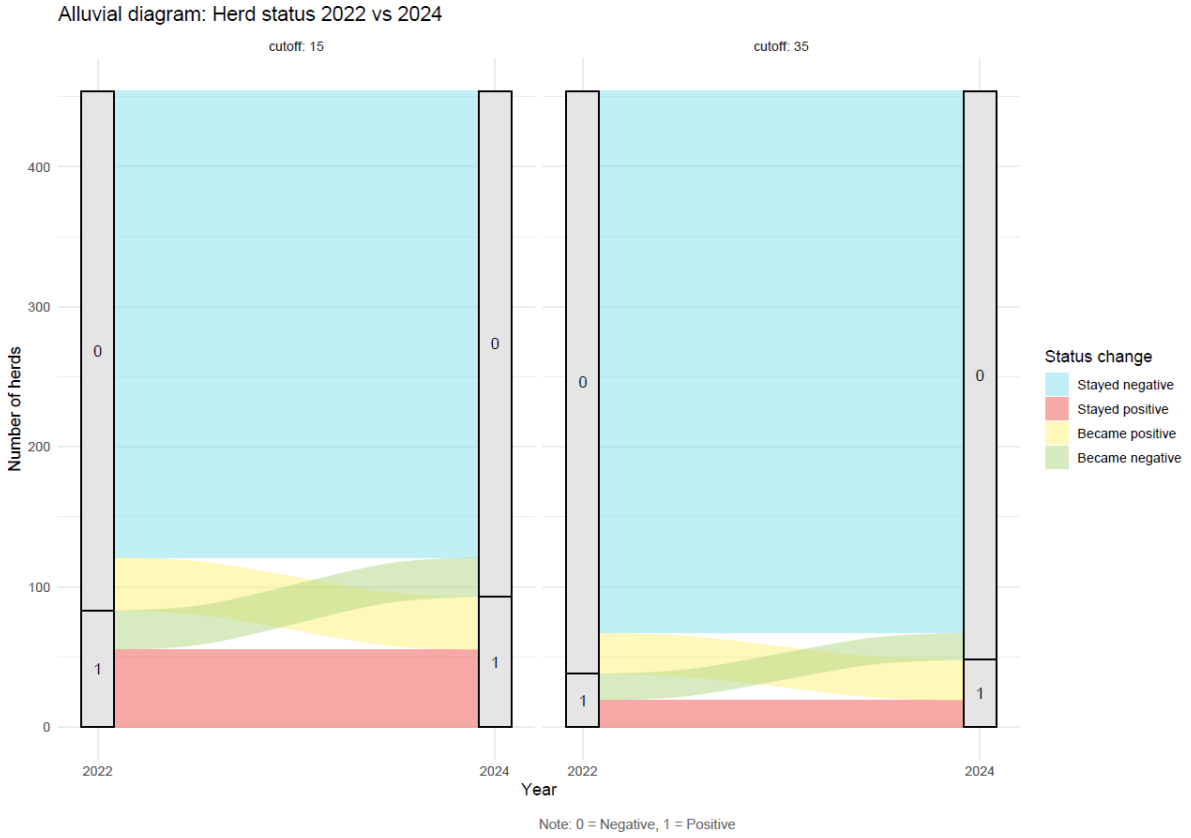

Figure 1: Change in herd status comparing 2022 to 2024 with a different cutoff: PP≥15% and PP≥35%. The alluvial diagram shows the number of herds that remained positive or negative, or changed status between the two time points.

Logistic regression showed a significant association between 2022 and 2024 status. At PP≥35%, positive herds in 2022 had 13.3 times higher odds of testing positive again in 2024. At PP≥15%, the association was even stronger (OR=17). Full estimates are presented in Table 2.

A McNemar test was conducted to examine the difference of the number of farms that changed in status between 2022 and 2024, separately for both cut-off values (PP≥15% and PP≥35%). No significant difference was found between the number of herds switching from positive to negative and vice versa (PP≥35%: p=0.19;PP≥15%: p=0.27).

Results stratified by region

Regional differences in herd status changes were observed for both cut-offs. At PP≥35%, 60% of positive herds in the Central region remained positive, compared to 40% in the North. At PP≥15%, the North showed the most stability, with 80% of herds remaining positive, while the South had the most newly positive herds (17 farms;25%).

Logistic regression analyses showed a significant association between 2022 and 2024 status across all regions. At PP≥35%, odds of remaining positive were highest in the North (OR=18.3), followed by Central (OR=16.6) and South (OR=11.1) region, as illustrated in figure 2. Similar results were found using PP≥15%, full estimates are presented in Table 2. Although the strength of association varied slightly between regions, interaction terms were not statistically significant, indicating that the relationship between 2022-2024 test status did not differ by region (p-values>0.05). Details can be found in R code, Attachment 3.

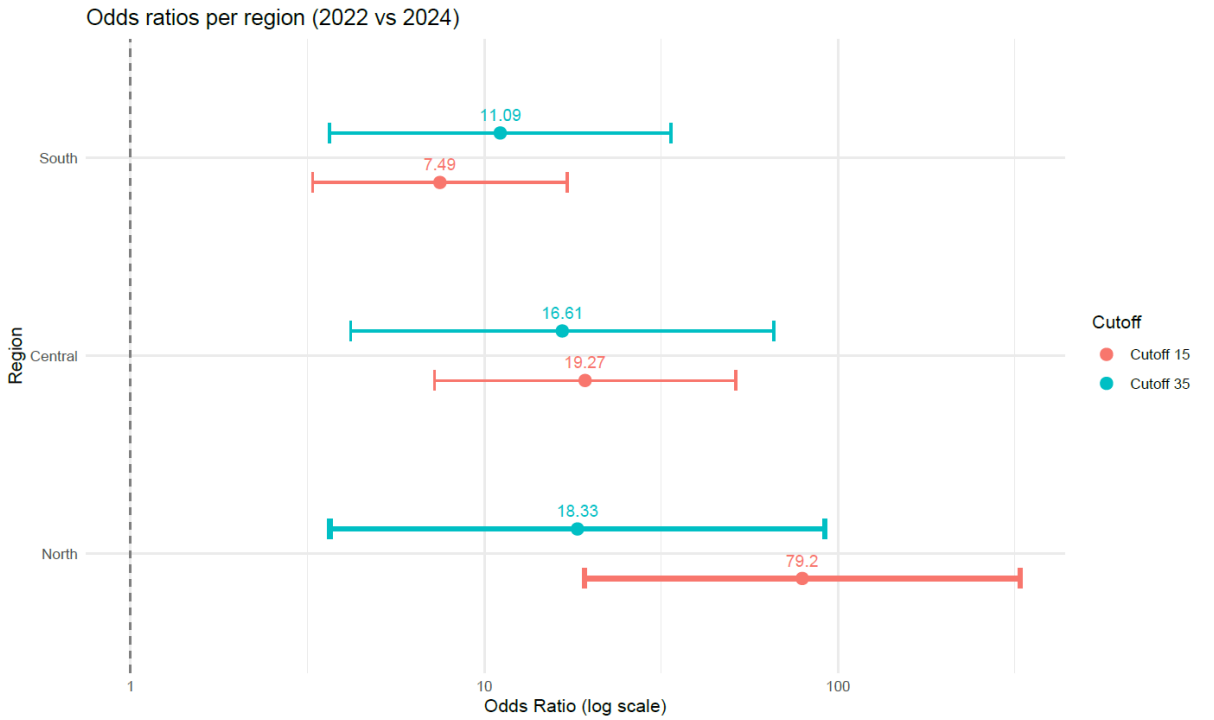

Figure 2: Forest plots. Odds ratios with 95% confidence intervals (CI) from logistic regression analyses comparing test results from 2022 and 2024, stratified by region (North, Central, South) separately for both cutoff values (PP≥35% and PP≥15%). An odds ratio above 1 indicates increased odds of testing positive in 2024 when already positive in 2022; the North region showed the strongest association at both Cut-offs. Smaller CI PP≥15% are results of more data information due to more positive tests.

At PP≥35%, McNemar showed significant status change over time in the Central region (p=0.03), indicating a meaningful shift despite high overall positivity. No significant shifts were found in the North or South, or at PP≥15%.

Results stratified by colony

Most farms remained negative at both cut-offs. Notable status changes were observed among Non-Hutterite farms showed more status changes: at PP≥35%, 26 became positive and 14 negative and at PP≥15%, 33 became positive and 15 negative. In contrast, Hutterite farms showed fewer and declining changes.

With logistic regression analysis in both groups and at both cut-offs, a positive test in 2022 was strongly associated with a positive result in 2024 (Table 2). Interestingly, the strength of this association reversed between cut-offs: at PP≥35%, odds were higher in Hutterite farms (OR=24.2, CI:3.72–165.0) than in Non-Hutterite farms (OR=11.7, CI:5.16–27.0), while at PP≥15%, the pattern was stronger in Non-Hutterite herds (OR=20.8, CI:10.7–42.7 vs. OR=16.0, CI:4.93–58.6).

McNemar tests showed a significant change in Non-Hutterite farms (PP≥15%;p=0.014), and a trend toward change at PP≥35% (p=0.082) (figure 3). In contrast, no significant changes were observed in Hutterite farms at either cut-off.

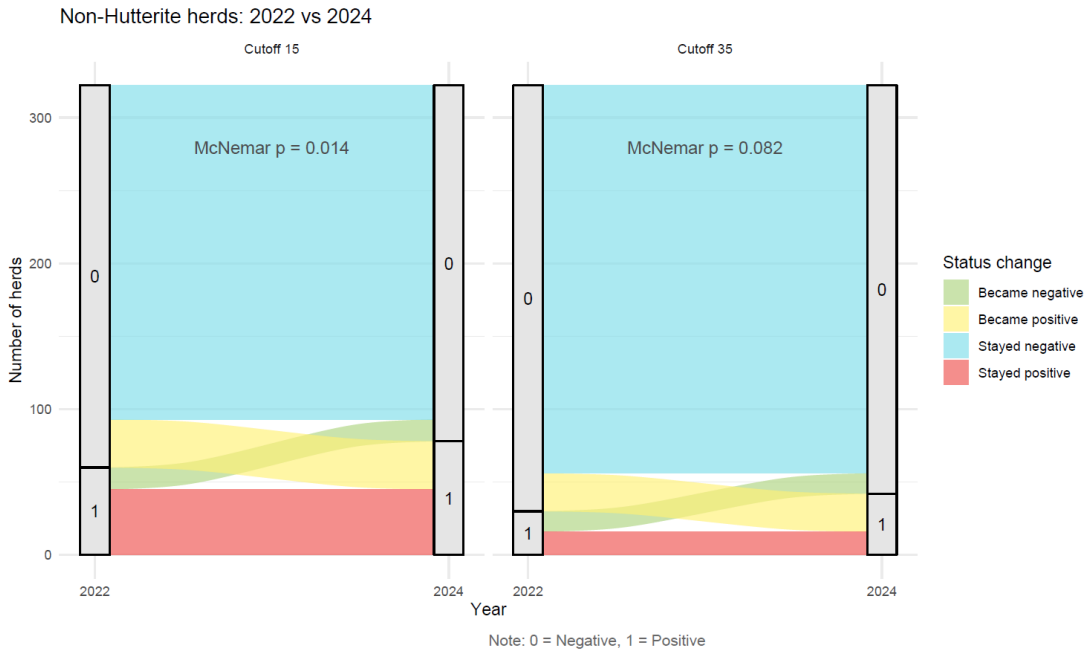

Figure 3: Change in herd status of Non-Hutterite farms, comparing 2022 to 2024 with a cutoff PP≥15% and PP≥35%. The alluvial diagram shows the number of herds that remained positive or negative, or changed status between the two time points. A significant shift in status at PP≥15% was observed among Non-Hutterite farms (McNemar p = 0.014), indicating a meaningful number of herds changed from positive to negative or vice versa. A trend towards significant was observed at PP≥35% (McNemar p=0.082). No significant change was found in Hutterite farms (p = 0.099).

Results stratified by milking system

At both cut-offs, most farms remained negative. CMS farms had more newly positive cases than AMS farms. At PP≥35%, 13 CMS farms remained positive and 24 became newly positive, compared to 5 new positives in AMS herds. At PP≥15%, 43 CMS farms stayed positive and 27 became positive, versus 12 and 11 in AMS herds, respectively. Overall, few farms became newly positive in either system. These pattens are shown in Figure 4.

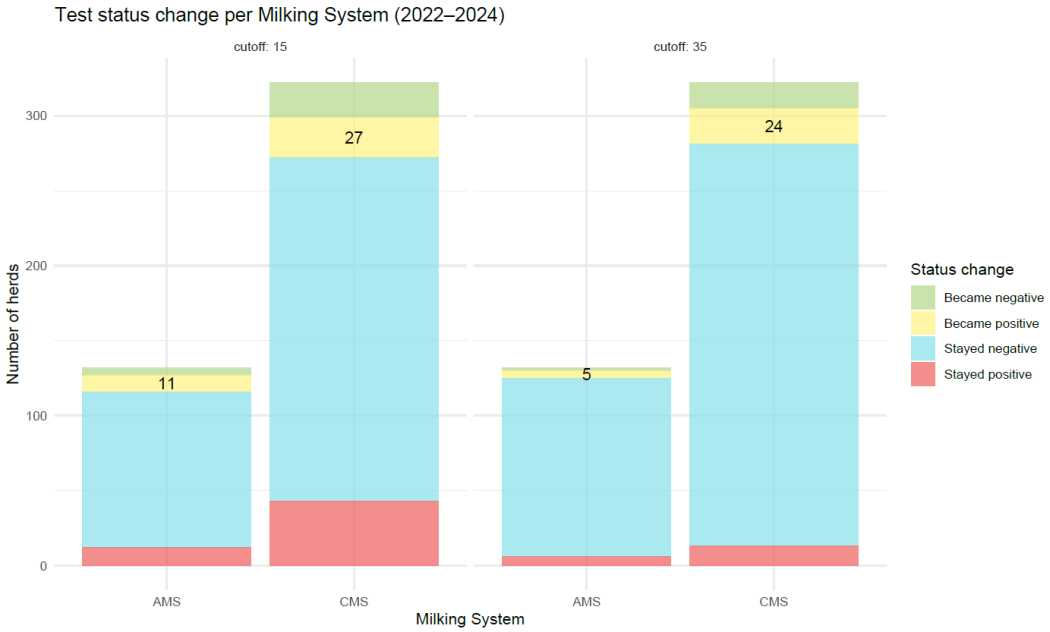

Figure 4: Distribution of herd-level test status changes between 2022 and 2024 by milking system (CMS vs. AMS) for cut-off values PP≥15% and PP≥35%. Bars represent the number of farms that remained negative, remained positive, became positive, or became negative, showing that CMS has highest number of new cases compared to AMS.

Logistic regression showed a strong association between 2022 and 2024 test results for both CMS and AMS herds (all p<0.001), with a stronger but less precise effect in AMS herds. Similar results were found for both cut-offs, suggesting persistent herd-level positivity across systems. McNemar tests detected no significant changes over time, indicating stable prevalence and highlighting that, despite strong associations, within-herd status shifts were limited (Table 2)

Results stratified by farm size

Status changes by farm size were evaluated at both cut-offs. At PP≥15%, small farms showed the most transitions, mainly from negative to positive, while changes were fewer in medium and large farms. At PP≥35%, status remained largely stable across all sizes, with few new positives. Although small farms had the most status changes, this did not result in the largest increase in prevalence, reflecting outcomes rather than transitions in table 3.

Table 3: Descriptive analysis: Apparent herd-level prevalence of Salmonella Dublin in 2022 and 2024, stratified by farm size and using a cut-off of PP≥35%. Values represent the percentage of farms testing positive at each time point and the absolute percentage change over time. Medium sizes farms show a biggest increase of absolute change in PP. More details can be found in Attachment 2.

Apparent Prevalence (%) Stratified by Farm Size

	2022	2024	Abs % Chng
All Sizes	*8,4%*	*10,6%*	*2,2%*
Small	7,5%	9,4%	1,9%
Medium	7,9%	10,9%	3,0%
Large	11,8%	13,2%	1,3%

Notes: *Positive if results are PP>=35%*

Logistic regression showed that a positive test in 2022 strongly predicted a positive result in 2024. At cut-off PP≥35%, this was significant for small and medium farms, but not for large farms. This was possibly due to the smaller number of large farms (76/454). At cut-off PP≥15%, the association was significant for all farm sizes, with the strongest effect in medium farms. Interaction terms showed no significant differences between farm sizes. McNemar tests showed no significant shift for any farm size (both cut-offs p>0.28).

Discussion

Overall, a moderate increase in *Salmonella* Dublin prevalence was observed between 2022 and 2024, especially at the more sensitive cut-off of PP≥15%. This suggests that small fluctuations in antibody levels can be detected, though this comes with a higher risk of false-positives and fluctuating results. The stricter PP≥35% cut-off likely identified herds with more persistent or high-level antibody production. Most herds showed stable test results over time, especially at the higher cut-off. The observed increase may reflect factors beyond infection dynamics, such as changes in laboratory procedures, herd management, or biosecurity.

A strong association between 2022 and 2024 test results across all regions suggests persistent herd-level infection. The North showed the most stable positives at PP≥15%, indicating stable infection pressure. The South had the most new positives at PP≥15%, possibly due to higher transmission or introduction into previously negative farms. Similar numbers in the South and Central at PP≥35% suggest emerging persistent infections. The observed regional differences may be due to chance, variation in sample size, or local management practices, as no significant interaction was found between region and test status over time. Structural factors such as climate or regulation are unlikely explanations, given their relative uniformity across Alberta. However, the higher farm density in the Central region, as reported by Shaukat et al. (2024), may contribute to more frequent status changes, as reflected in the significant McNemar test result ($p=0.03$).

Differences in herd structure and management likely influence *Salmonella* Dublin dynamics. Non-Hutterite farms showed more status changes between 2022 and 2024, confirmed by a significant McNemar result at PP≥15% ($p=0.014$) and a trend at PP≥35% ($p=0.082$). These changes suggest ongoing transmission and clearance, potentially driven by greater external contact and variable biosecurity in more open systems. Logistic regression further indicated a higher likelihood of persistent positivity at PP≥15%, possibly indicating repeated exposure or fluctuating antibody levels around the cut-off.
In contrast, the isolated structure of Hutterite farms may explain their stable status over time. At PP≥35%, high odds for testing positive again in 2024 given their 2022 positive status, may indicate more persistent infections. PP≥15% identifies patterns but increases the risk of false positives, while PP≥35% offers better specificity for persistent infections. However, wider confidence intervals at PP≥35%, likely due to few positive farms, may reduce interpretative precision.

At both cut-offs, AMS and CMS herds showed a strong association between 2022 and 2024 for testing positive again ($p<0.001$), indicating persistent infection. The higher number of new positives in CMS farms may reflect greater exposure or transmission risk. In contrast, the higher odds of repeated positivity (2022-2024) in AMS farms points to challenges in clearing infection, although wide confidence intervals limit the certainty of this finding. A non-significant change in status positive ↔ negative found place according to the McNemar test, underscoring the possible trouble of clearing the infection. The overall stability in apparent prevalence across systems may reflect that *Salmonella* Dublin transmission is more strongly linked to animal contact than to milking system type.

Across all farm sizes, a positive test in 2022 strongly predicted a positive result in 2024, particularly at the lower cut-off (PP≥15%). This suggests persistent infection over time, with the strongest association seen in medium-sized farms at both cut-offs. The lack of significant association in large farms at PP≥35% is likely due to the smaller sample size rather than absence of effect. The stable prevalence observed may reflect more consistent management and biosecurity practices in larger, well-structured

operations. Although small farms had the most status changes, medium-sized farms showed the largest increase in apparent prevalence at PP≥35%, while overall prevalence remained stable. This highlights outcomes rather than the number of changes.
The McNemar test showed no significant status changes overall. More transitions in small farms may reflect their larger sample size (n=213) or greater animal movement, causing short-term fluctuations without long-term impact. Although some differences in association strength were observed between farm sizes, there were no significant interaction terms. In other words, although some farm sizes showed stronger associations than others, these differences could be due to random variation.

Beyond Statistics: Considerations

Variation in test performance may arise from lab or herd-level factors, though consistent methods reduced this risk (Nyman et al., 2013). The ELISA was more sensitive at PP≥15%, enabling earlier detection but increasing false positives. Cross-reactions, particularly with *S.* Typhimurium, and antibody persistence in older animals can cause false positive results as well (Biosystems, 2017; Nielsen & Ersboll, 2004; Nielsen et al., 2004). The indirect ELISA detects O-antigen factors, which may lead to cross-reactivity. A blocking ELISA could reduce this, but was not used (Hoorfar et al., 1994). Additionally, antibody levels vary with infection stage. Low IgG in early infection can result in negative OD values despite active infection (Nielsen, 2013). Other factors, such as age distribution and immune status, may also reduce herd-level test validity (Nielsen & Ersboll, 2004).

Future improvements could involve testing strategies that identify acute from chronic infections and identify carriers (Veling et al., 2000), increasing validity ELISA validity (Nielsen & Ersboll, 2004). Combining results with ongoing research support ELISA interpretation and (within)herd-level surveillance. Although quantitative statistical methods cannot always capture the full complexity of biological processes and disease dynamics (Lakshman et al., 2000), combining logistic regression and McNemar tests in this study provided insights into infection persistence and direction of the status changes over time.

Beyond statistical findings, herd-level changes in *S.* Dublin status may be influenced by unmeasured factors, such as changes in herd management or biosecurity practices. The introduction of new animals, differences in hygiene protocols, or regional disease awareness efforts could contribute to the persistence or clearance of infection. These were not assessed but could explain persistence or clearance patterns. Future research, including farmer surveys, could help identify key risk factors influencing the rise in seroprevalence.

Conclusions

This study found a general increase in *Salmonella* Dublin prevalence and strong persistence of herd status across Alberta. While no significant differences were found between subgroups, within-group changes were observed, particularly in Central Alberta and among non-Hutterite farms. These results highlight the need for continued monitoring and improved biosecurity. Future research should focus on herd-level risk factors, including within-herd transmission, with targeted testing strategies to distinguish acute from chronic infections and better identify carriers.

References

Adandom, H. C., Ofori-Dei, S. M., & Hallstrom, L. K. (2023). Health and well-being of Hutterite farmers in Alberta: Results from the Sustainable Farm Families Alberta program. *Can J Rural Med*, *28*(3), 123-130. https://doi.org/10.4103/cjrm.cjrm_96_22

Alberta Milk. (2022). *Alberta Milk producer handbook*. A. Milk. https://albertamilk.com/wp-content/uploads/2022/08/2022-Producer-Policy-Handbook-Final.pdf

Alberta Milk. (2024). Annual report Alberta Milk. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://albertamilk.com/wp-content/uploads/2024/11/Alberta-Milk-Annual-Report-2023-24.pdf

Andersen, H. J., Pedersen, L. H., Aarestrup, F. M., & Chriél, M. (2003). Evaluation of the Surveillance Program of *Streptococcus agalactiae* in Danish dairy herds. *Journal of Dairy Science*, *86*(4), 1233-1239. https://doi.org/10.3168/jds.s0022-0302(03)73707-2

Barkema, H. W., Green, M. J., Bradley, A. J., & Zadoks, R. N. (2009). Invited review: The role of contagious disease in udder health. *J Dairy Sci*, *92*(10), 4717-4729. https://doi.org/10.3168/jds.2009-2347

Barkema, H. W., Keyserlingk, M. A. G. v., Kastelic, J. P., Lam, T. J. G. M., Luby, C., Roy, J.-P., LeBlanc, S. J., Keefe, G. P., & Kelton, D. F. (2015). Invited review: Changes in the dairy industry affecting dairy cattle health and welfare. *J of Dairy Sci*, *98*(11), 7426-7445.

Boyd E, Cuthbert E, Dick J, Ghosh K, Leung D, Renaud DL, Himsworth C. Understanding Salmonella Dublin in British Columbia through bulk tank milk surveillance. J Dairy Sci. 2025 Mar;108(3):2749-2755. doi: 10.3168/jds.2024-25710. Epub 2024 Dec 6. PMID: 39647617

Biosystems, A. (2017). *PrioCHECK Salmonella antibody ELISA kits (bovine)*. Thermo Fisher Scientific Inc. Retrieved March 12, 2025 from https://www.thermofisher.com/order/catalog/product/7610640

Fierer, J. (1983). Invasive Salmonella dublin infections associated with drinking raw milk. *West J Med*, *138*(5), 665-669. https://www.ncbi.nlm.nih.gov/pubmed/6880181

Government. (2022). *Reportable and notifiable diseases regulation: Animal Health Act*. Edmonton, AB T5K 2P7: Alberta King's Printer Retrieved from https://www.alberta.ca/reportable-and-notifiable-animal-diseases

Harvey, R. R., Friedman, C. R., Crim, S. M., Judd, M., Barrett, K. A., Tolar, B., Folster, J. P., Griffin, P. M., & Brown, A. C. (2017). Epidemiology of *Salmonella enterica* Serotype Dublin Infections among Humans, United States, 1968–2013. *Emerg Infect Dis*, *23*(9), 1493-1501. https://doi.org/10.3201/eid2309.170136

Helms, M., Vastrup, P., Gerner-Smidt, P., & Molbak, K. (2003). Short and long term mortality associated with foodborne bacterial gastrointestinal infections: registry based study. *BMJ*, *326*(7385), 357. https://www.ncbi.nlm.nih.gov/pubmed/12586666

Hoorfar, J., Feld, N. C., Schirmer, A. L., Bitsch, V., & Lind, P. (1994). Serodiagnosis of Salmonella dublin infection in Danish dairy herds using O-antigen based enzyme-linked immunosorbent assay. *Can J Vet Res*, *58*(4), 268-274. https://www.ncbi.nlm.nih.gov/pubmed/7889458

Lakshman, M., Sinha, L., Biswas, M., Charles, M., & Arora, N. K. (2000). Quantitative vs qualitative research methods. *Indian J Pediatr 67*(5), 369-377. https://doi.org/10.1007/bf02820690

Leedom, J. M. (2006). Milk of nonhuman origin and infectious diseases in humans. *Clin Infect Dis*, *43*(5), 610-615. https://doi.org/10.1086/507035

McDonough, P. L., Fogelman, D., Shin, S. J., Brunner, M. A., & Lein, D. H. (1999). *Salmonella enterica* Serotype Dublin Infection: an emerging infectious disease for the Northeastern United States. *J Clin Microbiol*, *37*(8), 2418-2427. https://doi.org/10.1128/jcm.37.8.2418-2427.1999

Microsoft Corporation. (2021). *Microsoft Excel for Microsoft 365 MSO*. In Microsoft. https://www.microsoft.com

Nielsen, L. R. (2013). Review of pathogenesis and diagnostic methods of immediate relevance for epidemiology and control of Salmonella Dublin in cattle. *Vet Microbiol*, *162*(1), 1-9. https://doi.org/10.1016/j.vetmic.2012.08.003

Nielsen, L. R., & Ersboll, A. K. (2004). Age-stratified validation of an indirect Salmonella Dublin serum enzyme-linked immunosorbent assay for individual diagnosis in cattle. *J Vet Diagn Invest*, *16*(3), 212-218. https://doi.org/10.1177/104063870401600306

Nielsen, L. R., Toft, N., & Ersboll, A. K. (2004). Evaluation of an indirect serum ELISA and a bacteriological faecal culture test for diagnosis of Salmonella serotype Dublin in cattle using latent class models. *J Appl Microbiol*, *96*(2), 311-319. https://doi.org/10.1046/j.1365-2672.2004.02151.x

Nobrega, D. B., French, J. E., & Kelton, D. F. (2023). A scoping review of the testing of bulk milk to detect infectious diseases of dairy cattle: Diseases caused by bacteria. *J Dairy Sc*, *106*(3), 1986-2006. https://doi.org/10.3168/jds.2022-22395

Nobrega, D. B., Miltenburg, C., Séguin, G., & Kelton, D. F. (2024). Prevalence and spatial distribution of infectious diseases of dairy cattle in Ontario, Canada. *J Dairy Sc*, *107*(7), 5029-5040. https://doi.org/10.3168/jds.2023-24197

Nyman, A.-K. J., Ågren, E. C., Bergström, K., & Wahlström, H. (2013). Evaluation of the specificity of three enzyme-linked immunosorbent assays for detection of antibodies against Salmonella in bovine bulk milk. *Acta Vet Scandi*, *55*(1), 5. https://doi.org/10.1186/1751-0147-55-5

Perry, K. V., Kelton, D. F., Dufour, S., Miltenburg, C., Umana Sedo, S. G., & Renaud, D. L. (2023). Risk factors for Salmonella Dublin on dairy farms in Ontario, Canada. *J Dairy Sci*, *106*(12), 9426-9439. https://doi.org/10.3168/jds.2023-23517

Shaukat, W., Jong, E. d., McCubbin, K., Biesheuvel, M., Meer, F. v. d., Buck, J. d., Lhermie, C., Hall, D., Kalbfleisch, K., Kastelic, J., Orsel, K., & Barkema, H. (2024). Herd-level prevalence of bovine leukemia virus, Salmonella Dublin and Neospora caninum in Alberta dairy herds using ELISA on bulk tank milk samples. *J Dairy Sci* 107(10) :8313-8328. https://doi.org/10.3168/jks.2023-24511

Silva, C., Calva, E., & Maloy, S. (2014). One health and food-borne disease: Salmonella transmission between Humans, Animals, and Plants. *Microbiol Spectr*, *2*(1), OH-0020-2013. https://doi.org/10.1128/microbiolspec.OH-0020-2013

ThermoFisher Scientific. (2024). *PrioCHECKtm S. Dublin Ab Strip kit*. https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0013903_7610640_UG_en.pdf

Um, M. M., Castonguay, M. H., Arsenault, J., Bergeron, L., Cote, G., Fecteau, G., Francoz, D., Giguere, J., Amine, K. M., Morin, I., & Dufour, S. (2022). Estimation of the accuracy of an ELISA test applied to bulk tank milk for predicting herd-level status for Salmonella Dublin in dairy herds using Bayesian Latent Class Models. *Prev Vet Med*, *206*, 105699. https://doi.org/10.1016/j.prevetmed.2022.105699

Velasquez-Munoz, A., Castro-Vargas, R., Cullens-Nobis, F. M., Mani, R., & Abuelo, A. (2024). Review: Salmonella Dublin in dairy cattle. *Front Vet Sci*, *10*, 1331767. https://doi.org/10.3389/fvets.2023.1331767

Veling, J., Barkema, H. W., van der Schans, J., van Zijderveld, F., & Verhoeff, J. (2002). Herd-level diagnosis for Salmonella enterica subsp. enterica serovar Dublin infection in bovine dairy herds. *Prev Vet Med*, *53*(1-2), 31-42. https://doi.org/10.1016/s0167-5877(01)00276-8

Veling, J., Van Zijderveld, F. G., Van Zijderveld-Van Bemmel, A. M., Barkema, H. W., & Schukken, Y. H. (2000). Evaluation of Three Newly Developed Enzyme-Linked Immunosorbent Assays and Two Agglutination Tests for Detecting *Salmonella enterica* subsp. *enterica* serovar Dublin infections in dairy cattle. *J Clin Microbio*, *38*(12), 4402-4407. https://doi.org/10.1128/jcm.38.12.4402-4407.2000

Attachments

Attachment 1: ThermoFisher ELISA PrioCheck *S.* Dublin Ab kit

applied biosystems — INSTRUCTIONS FOR USE

PrioCHECK™ S. Dublin Ab Strip Kit

ELISA for *in vitro* detection of antibodies against *Salmonella* in milk of cattle

Catalog Number 7610640 — Pub. No. MAN0013903 Rev. B

WARNING! Read the Safety Data Sheets (SDSs) and follow the handling instructions. Wear appropriate protective eyewear, clothing, and gloves. Safety Data Sheets (SDSs) are available from **thermofisher.com/support.**

WARNING! POTENTIAL BIOHAZARD. Read the biological hazard safety information at this product's page at **thermofisher.com.** Wear appropriate protective eyewear, clothing, and gloves.

Introduction

Salmonella infections in cattle can cause serious economical and welfare losses in the cattle industry. Infection can be transmitted to humans by the consumption of infected meat or dairy products and cause severe health problems or even death. Infections caused by *Salmonella* strains belonging to serotypes B, C1 and D are the most frequent occurring and serious infectious. *Salmonella* Dublin (serotype D) is adapted to cattle and unlike most other types of *Salmonella* bacteria, has the tendency to persist in herds for decades. Additionally *Salmonella* Dublin infections in humans are extremely invasive, and when compared to other *Salmonella* infection, mortality rate is high. In order to control the infection in infected herds it is necessary to cull the carriers and prevent production of new carriers. In Europe *Salmonella* programs to control infections in swine, poultry and eggs are already implemented or in the process of being implemented. Additionally some countries already implemented control programs for *Salmonella* in cattle. The Applied Biosystems™ PrioCHECK™ S. Dublin Ab Strip Kit originates from the Danish Veterinary Institute and has been successfully applied in the control program for *Salmonella* in Denmark since 2002.

The PrioCHECK™ S. Dublin Ab Strip Kit can be used to specifically detect infections caused by *Salmonella* Dublin, however cross reaction because of the O-antigen factors 1, 9 and 12 will occur. The test is suitable for large-scale screening of serum and (bulk) milk samples.

Test principle

The PrioCHECK™ S. Dublin Ab Strip Kit is an indirect ELISA for the detection of *Salmonella* antibodies in cattle directed against *Salmonella* Dublin and detects antibodies against *Salmonella* polysaccharide LPS O-antigens 1, 9 and 12. Plates are coated with the purified LPS isolated from *Salmonella* Dublin. The conjugate is goat- anti bovine IgG coupled to horse radish peroxidase. Test samples are placed in the wells of the test plate and incubated at room temperature (22±3°C). Subsequently plates are washed and the HRPO conjugate is added and incubated at room temperature (22±3°C). After the plates are washed the ready-to-use Chromogen (TMB) Substrate is dispensed to all wells of the test plate. After incubation at 22±3°C the color development is stopped and measured at 450 nm.

Kit components

5 plate kit for 450 samples. Store kit at 5±3°C until the expiry date. See kit label for actual expiry date.

The shelf life of diluted, opened or reconstituted components is noted below, where appropriate.

Component	Description
1: Test Plate	Five Test Plates.
2: Conjugate (30x)	30x concentrate, dilute before use. One vial containing 2.2 mL of Conjugate. Diluted Conjugate is not stable, prepare just before use.
3: Dilution Buffer (5x)	5x concentrate, dilute before use. One vial containing 60 mL of Dilution Buffer. Shelf life of dilution buffer working solution: 2 weeks at 22±3°C.
4: Washing Fluid (200x)	200x concentrate, dilute before use. One vial containing 60 mL of Washing Fluid. Shelf life of washing solution: 1 week at 22±3°C.
5: Negative Control	One vial containing 0.5 mL of Negative Control.
6: Validation Control	One vial containing 0.5 mL of Validation Control.
7: Positive Control	One vial containing 0.5 mL of Positive Control.
8: Chromogen (TMB) Substrate	Ready-to-use. One vial containing 60 mL of Chromogen (TMB) Substrate.
9: Stop Solution	Ready-to-use. One vial containing 60 mL of Stop Solution.
Additional kit contents	• Package insert • 15 plate sealers

Additional material required

Unless otherwise indicated, all materials are available through **thermofisher.com.**

Use	Description
General	Laboratory equipment according to national safety regulations.
Analysis of results	Plate Reader. The reader has to have an appropriate filter set to read the plates at 450 nm.
Optional	Plate washer.

Test procedure

Precautions

- National Safety Regulations must be strictly followed.
- The PrioCHECK™ S. Dublin Ab Strip Kit must be performed in laboratories suited for this purpose.
- Samples should be considered as potentially infectious and all items which contact the samples as potentially contaminated.

Notes

To achieve optimal results with the PrioCHECK™ S. Dublin Ab Strip Kit, the following aspects must be considered:

- **The Test Procedure protocol must be strictly followed.**
- All reagents of the kit must be equilibrated to room temperature (22±3°C) before use.
- Pipette tips have to be changed for every pipetting step.
- Separate solution reservoirs must be used for each reagent.
- Kit components must not be used after their expiry date or if changes in their appearance are observed.
- Kit components of different kit lot numbers must not be used together.
- Demineralized or water of equal quality must be used for the test.

Solutions to be made in advance

Dilution buffer working solution

The Dilution Buffer (Component 3) must be diluted 5 times in demineralized or distilled water. To perform a test with one plate, prepare 45 mL (add 9 mL Dilution Buffer (5x) to 36 mL demineralized or distilled water).
Shelf life of dilution buffer working solution: 2 weeks at 22±3°C.

Conjugate dilution

Prepare dilution of the Conjugate (Component 2) in dilution buffer working solution. To perform a test with one plate, prepare 12 mL (add 0.4 mL concentrated conjugate to 11.6 mL of dilution buffer working solution).
Note: The diluted conjugate must be prepared just before use.

Washing solution

The Washing Fluid (Component 4) must be diluted (200x) in demineralized water and is sufficient for a final volume of 12 liters. To perform a test with one plate, prepare 500 mL (add 2.5 mL Washing Solution (200x) to 497.5 mL demineralized or distilled water).
Stability of washing solution: 1 week stored at 22±3°C.

Incubation of control and test milk samples

1. Label each strip of the Test Plate (Component 1) with a marker pen.
2. Dispense 100 µL of the test milk samples to wells G1–H12 of the Test Plate.
3. Dispense 90 µL of the dilution buffer working solution to the wells A1–F1 of the Test Plate.
4. Dispense 10 µL of the Negative Control (Component 5) to wells A1 and B1.
5. Dispense 10 µL of the Validation Control (Component 6) to wells C1 and D1.
6. Dispense 10 µL of the Positive Control (Component 7) to wells E1 and F1.
7. Seal the test plate(s) with a plate sealer(s).
8. Shake the plate(s) during 1 minute, level 700 (for example SLT micro plate shaker EAS 2/4, rpm 1/min level 700, SLT lab instruments).
9. Incubate the Test Plate(s) for 60±5 minutes at room temperature (22±3°C).

Incubation with Conjugate

1. Empty the Test Plate and wash the plate 6 times with 200 to 300 µL diluted washing fluid. Tap the plate firmly after the last wash cycle.
2. Dispense 100 µL of the working solution of the conjugate to all wells.
3. Seal the test plate with a plate sealer.
4. Incubate the plate(s) for 60±5 minutes at room temperature (22±3°C).

Incubation with Chromogen (TMB) Substrate

1. Empty the Test Plate and wash the plate 6 times with 200 to 300 µL diluted washing fluid. Tap the plate firmly after the last wash cycle.
2. Dispense 100 µL of the Chromogen (TMB) Substrate (Component 8) to all wells.
3. Incubate the plate(s) 15 minutes at room temperature (22±3°C).
4. Add 100 µL of the Stop Solution (Component 9) to all wells.
5. Mix the content of the wells of the plate(s).

Note: Start the addition of stop solution 15 minutes after the first well was filled with the Chromogen (TMB) Substrate. Add the Stop Solution in the same order and at the same pace as the Chromogen (TMB) Substrate was dispensed.

Reading of the test and calculating the results

1. Measure the optical density (**OD**) of the wells at 450 nm, preferably within 15 minutes after color development has been stopped.
2. Calculate the mean OD_{450} value of the Negative Control (wells A1 and B1).
3. Calculate the mean OD_{450} value of the **Positive Control** (wells E1 and F1).
4. Calculate the **corrected** OD_{450} value of the **Positive Control**, Validation Control and all **samples** by subtracting the mean OD_{450} of the Negative Control (wells A1 and B1).
5. Calculate the percent positivity (**PP**) of all controls and of the **test samples** according to the formula below.

The OD_{450} of all **samples** is expressed as percent positivity (**PP**) of the OD_{450} of **Positive Control** (PC) (wells E1 and F1) **corrected** with the mean OD_{450} of the Negative Control (NC) (wells A1 and B1).

$$PP = (\text{corrected } OD_{450\ \text{test sample}} / \text{corrected } OD_{450\ \text{Positive Control}} \times 100) - 10$$

Result interpretation

Validation criteria

1. The mean OD_{450} of the Negative Control (wells A1 and B1) must be <0.4.
2. The OD_{450} of the **Positive Control** (not corrected) should be >1.000.
3. The percent positivity of the Validation Control must be ≥30.

Not meeting these criteria is reason to discard the results of that specific test plate.
Note: If the OD_{450} of the **Positive Control** (not corrected) is below 1.000 possibly the Chromogen (TMB) Substrate is too cold. In that case pre-warm the solution to 22±3°C or incubate up to 30 minutes.

Interpretation of the percent positivity

PP = <35%	Negative	*Salmonella*-specific antibodies are absent in the test sample.
PP = ≥35%	Positive	*Salmonella*-specific antibodies are present in the test sample.

In well-advanced *Salmonella* control programs the test can be used with a different cut-off. It remains the responsibility of the respective authorities/users to implement such cut-offs.

Customer and technical support

Technical support: visit **thermofisher.com/askaquestion**
Visit **thermofisher.com/support** for the latest in services and support, including:

- Worldwide contact telephone numbers
- Order and web support
- User guides, manuals, and protocols
- Certificates of Analysis
- Safety Data Sheets (SDSs; also known as MSDSs)
  NOTE: For SDSs for reagents and chemicals from other manufacturers, contact the manufacturer.

Limited product warranty

Life Technologies Corporation and/or its affiliate(s) warrant their products as set forth in the Life Technologies' General Terms and Conditions of Sale found on Life Technologies' website at **www.thermofisher.com/us/en/home/global/terms-and-conditions.html**. If you have any questions, please contact Life Technologies at **thermofisher.com/support**.

Prionics Lelystad B.V. | Platinastraat 33 | 8211 AR Lelystad | The Netherlands

Revision history of Pub. No. MAN0013903 (English)

Rev.	Date	Description
B	10 February 2025	A note was removed (see Incubation of control and test milk samples).
A.0	10 September 2019	• New document. Converted the legacy document (PrioCHECK *Salmonella* Ab Bovine Dublin milk 7610640 v1.2_e.doc) to the current document template, with associated updates to the publication number, limited license information, warranty, trademarks, and logos. • The product name was changed from PrioCHECK® Salmonella Ab bovine Dublin to PrioCHECK™ S. Dublin Ab Strip Kit.

Attachment 2: Tables descriptive analysis (Microsoft Corporation, 2021)

Apparent Prevalence at 35% PP

Variable	2022	2024	Abs Chng
Grand Total	**8,4%**	**10,6%**	**2,2%**

Notes: *Positive if results are PP>=35%*

Apparent Prevalence at 15% PP

Variable	2022	2024	Abs Chng
Grand Total	**18,3%**	**20,5%**	**2,2%**

Notes: *Positive if results are PP>=15%*

Apparent Prevalence (%) Stratified by Region

	2022	2024	Abs % Chng
All Regions	*8,4%*	*10,6%*	*2,2%*
Central	5,6%	11,2%	5,6%
North	8,1%	6,5%	-1,6%
South	11,9%	13,2%	1,3%

Notes: *Positive if results are PP>=35%*

Apparent Prevalence (%) Stratified by Region

	2022	2024	Abs % Chng
All Regions	*18,3%*	*20,5%*	*2,2%*
Central	14,5%	19,0%	4,5%
North	16,1%	16,9%	0,8%
South	24,5%	25,2%	0,7%

Notes: *Positive if results are PP>=15%*

Apparent Prevalence (%) Stratified by Colony

	2022	2024	Abs % Chng
All Colonies	*8,4%*	*10,6%*	*2,2%*
Hutterite	6,1%	4,5%	-1,5%
Non-Hutterite	9,3%	13,0%	3,7%

Notes: *Positive if results are PP>=35%*

Apparent Prevalence (%) Stratified by Colony

	2022	2024	Abs % Chng
All Colonies	*18,3%*	*20,5%*	*2,2%*
Hutterite	17,4%	11,4%	-6,1%
Non-Hutterite	18,6%	24,2%	5,6%

Notes: *Positive if results are PP>=15%*

Apparent Prevalence (%) Stratified by Milking System

	2022	2024	Abs % Chng
All Systems	*8,4%*	*10,6%*	*2,2%*
AMS	6,1%	8,3%	2,3%
CMS	9,3%	11,5%	2,2%

Notes: *Positive if results are PP>=35%*

Apparent Prevalence (%) Stratified by Milking System

	2022	2024	Abs % Chng
All Systems	*18,3%*	*20,5%*	*2,2%*
AMS	12,9%	17,4%	4,5%
CMS	20,5%	21,7%	1,2%

Notes: *Positive if results are PP>=15%*

Apparent Prevalence (%) Stratified by Farm Size

	2022	2024	Abs % Chng
All Sizes	*8,4%*	*10,6%*	*2,2%*
Small	7,5%	9,4%	1,9%
Medium	7,9%	10,9%	3,0%
Large	11,8%	13,2%	1,3%

Notes: *Positive if results are PP>=35%*

Apparent Prevalence (%) Stratified by Farm Size

	2022	2024	Abs % Chng
All Sizes	*18,3%*	*20,5%*	*2,2%*
Small	14,6%	17,8%	3,3%
Medium	18,2%	20,0%	1,8%
Large	28,9%	28,9%	0,0%

Notes: *Positive if results are PP>=15%*

	Apparent Prevalence at 35% PP			
	Became Positive	**Became Negative**	**Stayed Positive**	**Stayed Negative**
Grand Total	29 (6.4%)	19 (4.2%)	19 (4.2%)	387 (85.2%)
Regions				
Central	14 (7.8%)	4 (2.2%)	6 (3.4%)	155 (86.6%)
North	4 (3.2%)	6 (4.8%)	4 (3.2%)	110 (88.7%)
South	11 (7.3%)	9 (6.%)	9 (6.%)	122 (80.8%)
Colonies				
Hutterite	3 (2.3%)	5 (3.8%)	3 (2.3%)	121 (91.7%)
Non-Hutterite	26 (8.1%)	14 (4.3%)	16 (5.%)	266 (82.6%)
Systems				
AMS	5 (3.8%)	2 (1.5%)	6 (4.5%)	119 (90.2%)
CMS	24 (7.5%)	17 (5.3%)	13 (4.%)	268 (83.2%)
Sizes				
Small	12 (5.6%)	8 (3.8%)	8 (3.8%)	185 (86.9%)
Medium	10 (6.1%)	5 (3.%)	8 (4.8%)	142 (86.1%)
Large	7 (9.2%)	6 (7.9%)	3 (3.9%)	60 (78.9%)

Notes: *The percentage is based on the total number of farms within that stratum.*

	Apparent Prevalence at 15% PP			
	Became Positive	**Became Negative**	**Stayed Positive**	**Stayed Negative**
Grand Total	38 (8.4%)	28 (6.2%)	55 (12.1%)	333 (73.3%)
Regions				
Central	16 (8.9%)	8 (4.5%)	18 (10.1%)	137 (76.5%)
North	5 (4.%)	4 (3.2%)	16 (12.9%)	99 (79.8%)
South	17 (11.3%)	16 (10.6%)	21 (13.9%)	97 (64.2%)
Colonies				
Hutterite	5 (3.8%)	13 (9.8%)	10 (7.6%)	104 (78.8%)
Non-Hutterite	33 (10.2%)	15 (4.7%)	45 (14.%)	229 (71.1%)
Systems				
AMS	11 (8.3%)	5 (3.8%)	12 (9.1%)	104 (78.8%)
CMS	27 (8.4%)	23 (7.1%)	43 (13.4%)	229 (71.1%)
Sizes				
Small	19 (8.9%)	12 (5.6%)	19 (8.9%)	163 (76.5%)
Medium	12 (7.3%)	9 (5.5%)	21 (12.7%)	123 (74.5%)
Large	7 (9.2%)	7 (9.2%)	15 (19.7%)	47 (61.8%)

Notes: *The percentage is based on the total number of farms within that stratum.*

Attachment 3: Statistical analysis R code

```
R version 4.5.0 (2025-04-11 ucrt) -- "How About a Twenty-Six"
Copyright (C) 2025 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> install.packages("readxl")  # indien nog niet geïnstalleerd
Installing package into 'C:/Users/jessi/AppData/Local/R/win-library/4.5'
(as 'lib' is unspecified)
--- Please select a CRAN mirror for use in this session ---
trying                                                                  URL
'https://ftp.belnet.be/mirror/CRAN/bin/windows/contrib/4.5/readxl_1.4.5.zip
'
Content type 'application/zip' length 765609 bytes (747 KB)
downloaded 747 KB

package 'readxl' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\jessi\AppData\Local\Temp\RtmpMH2Zyl\downloaded_packages
> library(readxl)
>                                 data                                    <-
read_excel("C:/Users/jessi/Documents/Data/Diergeneeskunde/Masterproef/MASTE
RPROEFdata for R.xlsx")

> str(data)
tibble [454 × 9] (S3: tbl_df/tbl/data.frame)
 $ farm          : num [1:454] 2 2 2 3 4 4 4 5 7 8 ...
 $ region        : chr [1:454] "Central" "Central" "South" "Central" ...
 $ colony        : chr [1:454] "Hutterite" "Non-Hutterite" "Hutterite" "Non-
Hutterite" ...
 $ 2022 cutoff 35: num [1:454] 0 0 0 0 0 0 1 0 0 0 ...
 $ 2024 cutoff 35: num [1:454] 0 0 0 0 0 0 1 0 0 0 ...
 $ 2022 cutoff 15: num [1:454] 0 0 1 0 0 0 1 0 0 0 ...
 $ 2024 cutoff 15: num [1:454] 0 0 1 0 0 0 1 0 0 0 ...
 $ milking_system: chr [1:454] "CMS" "AMS" "CMS" "CMS" ...
 $ farm size     : chr [1:454] "Small" "Small" "Medium" "Large" ...
> names(data)
[1] "farm"           "region"         "colony"         "2022 cutoff 35"
[5] "2024 cutoff 35" "2022 cutoff 15" "2024 cutoff 15" "milking_system"
[9] "farm size"
> head(data)
# A tibble: 6 × 9
   farm region  colony       `2022 cutoff 35` `2024 cutoff 35` `2022 cutoff
15`
  <dbl> <chr>   <chr>                   <dbl>           <dbl>           <dbl>
1     2 Central Hutterite                   0               0               0
2     2 Central Non-Hutterite               0               0               0
3     2 South   Hutterite                   0               0               1
```

```
4     3 Central Non-Hutterite               0               0               0
5     4 North   Non-Hutterite               0               0               0
6     4 South   Non-Hutterite               0               0               0
# ℹ 3 more variables: `2024 cutoff 15` <dbl>, milking_system <chr>,
#   `farm size` <chr>
> colnames(data) <- c("farm", "region", "colony",
+                     "cut35_2022", "cut35_2024",
+                     "cut15_2022", "cut15_2024",
+                     "milking_system", "farm_size")
> data$cut15_2022 <- as.factor(data$cut15_2022)
> data$cut15_2024 <- as.factor(data$cut15_2024)
> data$cut35_2022 <- as.factor(data$cut35_2022)
> data$cut35_2024 <- as.factor(data$cut35_2024)
> model_cut35 <- glm(cut35_2024 ~ cut35_2022, data = data, family = binomial)
> summary(model_cut35)

Call:
glm(formula = cut35_2024 ~ cut35_2022, family = binomial, data = data)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.5911     0.1925 -13.459  < 2e-16 ***
cut35_20221   2.5911     0.3773   6.868  6.5e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 306.44  on 453  degrees of freedom
Residual deviance: 263.09  on 452  degrees of freedom
AIC: 267.09

Number of Fisher Scoring iterations: 5

> exp(cbind(OR = coef(model_cut35), confint(model_cut35)))
Waiting for profiling to be done...
                    OR      2.5 %    97.5 %
(Intercept)  0.0749354 0.05026856  0.107219
cut35_20221 13.3448276 6.37932839 28.208345
> install.packages("broom")
Installing package into ‘C:/Users/jessi/AppData/Local/R/win-library/4.5’
(as ‘lib’ is unspecified)
trying                                                                  URL
'https://ftp.belnet.be/mirror/CRAN/bin/windows/contrib/4.5/broom_1.0.8.zip'
Content type 'application/zip' length 1946774 bytes (1.9 MB)
downloaded 1.9 MB

package ‘broom’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\jessi\AppData\Local\Temp\RtmpMH2Zyl\downloaded_packages
> library(broom)
>
> tidy(model_cut35, exponentiate = TRUE, conf.int = TRUE)
# A tibble: 2 × 7
  term        estimate std.error statistic  p.value conf.low conf.high
  <chr>          <dbl>     <dbl>     <dbl>    <dbl>    <dbl>     <dbl>
1 (Intercept)   0.0749     0.193    -13.5  2.74e-41   0.0503     0.107
2 cut35_20221  13.3        0.377      6.87 6.50e-12   6.38      28.2
> model_cut15 <- glm(cut15_2024 ~ cut15_2022, data = data, family = binomial)
> summary(model_cut15)
```

```
Call:
glm(formula = cut15_2024 ~ cut15_2022, family = binomial, data = data)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.1706     0.1712 -12.677   <2e-16 ***
cut15_20221   2.8457     0.2885   9.865   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 460.40  on 453  degrees of freedom
Residual deviance: 351.26  on 452  degrees of freedom
AIC: 355.26

Number of Fisher Scoring iterations: 4

> exp(cbind(OR = coef(model_cut15), confint(model_cut15)))
Waiting for profiling to be done...
                     OR      2.5 %     97.5 %
(Intercept)  0.1141141 0.08029274  0.1573988
cut15_20221 17.2133457 9.89059039 30.7200787
> library(broom)
> tidy(model_cut15, exponentiate = TRUE, conf.int = TRUE)
# A tibble: 2 × 7
  term        estimate std.error statistic  p.value conf.low conf.high
  <chr>          <dbl>     <dbl>     <dbl>    <dbl>    <dbl>     <dbl>
1 (Intercept)    0.114     0.171    -12.7  7.90e-37   0.0803     0.157
2 cut15_20221   17.2       0.288      9.86 5.90e-23   9.89      30.7
> install.packages("ggalluvial")
Installing package into ‘C:/Users/jessi/AppData/Local/R/win-library/4.5’
(as ‘lib’ is unspecified)
trying                                                                  URL
'https://ftp.belnet.be/mirror/CRAN/bin/windows/contrib/4.5/ggalluvial_0.12.
5.zip'
Content type 'application/zip' length 1671325 bytes (1.6 MB)
downloaded 1.6 MB

package ‘ggalluvial’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\jessi\AppData\Local\Temp\RtmpMH2Zyl\downloaded_packages
> install.packages("ggplot2")  # als nog niet geïnstalleerd
Installing package into ‘C:/Users/jessi/AppData/Local/R/win-library/4.5’
(as ‘lib’ is unspecified)
trying                                                                  URL
'https://ftp.belnet.be/mirror/CRAN/bin/windows/contrib/4.5/ggplot2_3.5.2.zi
p'
Content type 'application/zip' length 5016265 bytes (4.8 MB)
downloaded 4.8 MB

package ‘ggplot2’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\jessi\AppData\Local\Temp\RtmpMH2Zyl\downloaded_packages
> library(ggalluvial)
Loading required package: ggplot2
Need help? Try Stackoverflow: https://stackoverflow.com/tags/ggplot2
> library(ggplot2)
```

```
> library(dplyr)

Attaching package: ‘dplyr’

The following objects are masked from ‘package:stats’:

    filter, lag

The following objects are masked from ‘package:base’:

    intersect, setdiff, setequal, union

>
> # Voor cutoff 15
> df15 <- data %>%
+   count(cut15_2022, cut15_2024) %>%
+   rename(Status2022 = cut15_2022, Status2024 = cut15_2024, Freq = n) %>%
+   mutate(Cutoff = "15")
>
> # Voor cutoff 35
> df35 <- data %>%
+   count(cut35_2022, cut35_2024) %>%
+   rename(Status2022 = cut35_2022, Status2024 = cut35_2024, Freq = n) %>%
+   mutate(Cutoff = "35")
>
> # Combineer de datasets
> alluvial_combined <- bind_rows(df15, df35)
> library(ggalluvial)
> library(ggplot2)
>
> ggplot(alluvial_combined,
+        aes(axis1 = Status2022, axis2 = Status2024, y = Freq)) +
+   geom_alluvium(aes(fill = Status2022), width = 1/12) +
+   geom_stratum(width = 1/12, fill = "grey80", color = "black") +
+   geom_text(stat = "stratum", aes(label = after_stat(stratum))) +
+   scale_x_discrete(limits = c("2022", "2024"), expand = c(.05, .05)) +
+   labs(title = "Verandering in Salmonella Dublin-status",
+        x = "Jaar", y = "Aantal bedrijven") +
+   facet_wrap(~Cutoff, labeller = label_both) +
+   theme_minimal()
Warning messages:
1: In to_lodes_form(data = data, axes = axis_ind, discern = params$discern)
:
  Some strata appear at multiple axes.
2: In to_lodes_form(data = data, axes = axis_ind, discern = params$discern)
:
  Some strata appear at multiple axes.
3: In to_lodes_form(data = data, axes = axis_ind, discern = params$discern)
:
  Some strata appear at multiple axes.
> # Maak de plotobject apart (aanrader)
> p <- ggplot(alluvial_combined,
+        aes(axis1 = Status2022, axis2 = Status2024, y = Freq)) +
+   geom_alluvium(aes(fill = Status2022), width = 1/12) +
+   geom_stratum(width = 1/12, fill = "grey80", color = "black") +
+   geom_text(stat = "stratum", aes(label = after_stat(stratum))) +
+   scale_x_discrete(limits = c("2022", "2024"), expand = c(.05, .05)) +
+   labs(title = "Verandering in Salmonella Dublin-status",
+        x = "Jaar", y = "Aantal bedrijven") +
+   facet_wrap(~Cutoff, labeller = label_both) +
+   theme_minimal()
```

```
>
> # Sla op als PNG (pas pad aan als nodig)
> ggsave("alluvial_cutoffs15_35.png", plot = p, width = 8, height = 5, dpi =
300)
Warning messages:
1: In to_lodes_form(data = data, axes = axis_ind, discern = params$discern)
:
  Some strata appear at multiple axes.
2: In to_lodes_form(data = data, axes = axis_ind, discern = params$discern)
:
  Some strata appear at multiple axes.
3: In to_lodes_form(data = data, axes = axis_ind, discern = params$discern)
:
  Some strata appear at multiple axes.
> table35 <- table(data$cut35_2022, data$cut35_2024)
> table35

     0   1
  0 387  29
  1  19  19
> mcnemar.test(table35)

        McNemar's Chi-squared test with continuity correction

data:  table35
McNemar's chi-squared = 1.6875, df = 1, p-value = 0.1939

> table15 <- table(data$cut15_2022, data$cut15_2024)
> table15

     0   1
  0 333  38
  1  28  55
>
> mcnemar.test(table15)

        McNemar's Chi-squared test with continuity correction

data:  table15
McNemar's chi-squared = 1.2273, df = 1, p-value = 0.2679

> data$region <- as.factor(data$region)
> north_data   <- filter(data, region == "North")
> central_data <- filter(data, region == "Central")
> south_data   <- filter(data, region == "South")
> # North
> model_north <- glm(cut35_2024 ~ cut35_2022, data = north_data, family =
binomial)
>
> # Central
> model_central <- glm(cut35_2024 ~ cut35_2022, data = central_data, family
= binomial)
>
> # South
> model_south <- glm(cut35_2024 ~ cut35_2022, data = south_data, family =
binomial)
> # Odds ratios en 95% CI
> exp(cbind(OR = coef(model_north),   confint(model_north)))
Waiting for profiling to be done...
                    OR      2.5 %      97.5 %
(Intercept)  0.03636364 0.01114766  0.08651833
```

```
cut35_20221 18.33333333 3.59766142 97.92962720
> exp(cbind(OR = coef(model_central), confint(model_central)))
Waiting for profiling to be done...
                     OR     2.5 %     97.5 %
(Intercept)  0.09032258 0.0499119  0.1502987
cut35_20221 16.60714286 4.2652839 71.9388703
> exp(cbind(OR = coef(model_south),   confint(model_south)))
Waiting for profiling to be done...
                     OR     2.5 %     97.5 %
(Intercept)  0.09016393 0.0458359  0.1593651
cut35_20221 11.09090909 3.6662519 34.6282549
> library(broom)
>
> tidy(model_north,   exponentiate = TRUE, conf.int = TRUE)
# A tibble: 2 × 7
  term        estimate std.error statistic  p.value conf.low conf.high
  <chr>          <dbl>     <dbl>     <dbl>    <dbl>    <dbl>     <dbl>
1 (Intercept)   0.0364     0.509     -6.51 7.46e-11   0.0111    0.0865
2 cut35_20221  18.3        0.822      3.54 4.03e- 4   3.60     97.9
> tidy(model_central, exponentiate = TRUE, conf.int = TRUE)
# A tibble: 2 × 7
  term        estimate std.error statistic  p.value conf.low conf.high
  <chr>          <dbl>     <dbl>     <dbl>    <dbl>    <dbl>     <dbl>
1 (Intercept)   0.0903     0.279     -8.62 6.96e-18   0.0499     0.150
2 cut35_20221  16.6        0.703      4.00 6.45e- 5   4.27      71.9
> tidy(model_south,   exponentiate = TRUE, conf.int = TRUE)
# A tibble: 2 × 7
  term        estimate std.error statistic  p.value conf.low conf.high
  <chr>          <dbl>     <dbl>     <dbl>    <dbl>    <dbl>     <dbl>
1 (Intercept)   0.0902     0.315     -7.64 2.12e-14   0.0458     0.159
2 cut35_20221  11.1        0.567      4.24 2.19e- 5   3.67      34.6
> data$cut15_2022 <- as.factor(data$cut15_2022)
> data$cut15_2024 <- as.factor(data$cut15_2024)
> north_data   <- filter(data, region == "North")
> central_data <- filter(data, region == "Central")
> south_data   <- filter(data, region == "South")
> # North
> model15_north <- glm(cut15_2024 ~ cut15_2022, data = north_data, family =
binomial)
>
> # Central
> model15_central <- glm(cut15_2024 ~ cut15_2022, data = central_data, family
= binomial)
>
> # South
> model15_south <- glm(cut15_2024 ~ cut15_2022, data = south_data, family =
binomial)
> exp(cbind(OR = coef(model15_north),   confint(model15_north)))
Waiting for profiling to be done...
                     OR       2.5 %      97.5 %
(Intercept)  0.05050505  0.01782226   0.1117099
cut15_20221 79.19999913 21.28951224 376.3726872
> exp(cbind(OR = coef(model15_central), confint(model15_central)))
Waiting for profiling to be done...
                     OR     2.5 %     97.5 %
(Intercept)  0.1167883 0.0669158  0.1898213
cut15_20221 19.2656249 7.4761080 54.0462316
> exp(cbind(OR = coef(model15_south),   confint(model15_south)))
Waiting for profiling to be done...
                    OR     2.5 %     97.5 %
(Intercept) 0.1752577 0.1011196  0.2853302
```

```
cut15_20221 7.4889706 3.3121190 17.5479949
> library(broom)
>
> tidy(model15_north,   exponentiate = TRUE, conf.int = TRUE)
# A tibble: 2 × 7
  term        estimate std.error statistic  p.value conf.low conf.high
  <chr>          <dbl>     <dbl>     <dbl>    <dbl>    <dbl>     <dbl>
1 (Intercept)   0.0505     0.458     -6.51 7.31e-11   0.0178     0.112
2 cut15_20221  79.2        0.723      6.05 1.47e- 9  21.3      376.
> tidy(model15_central, exponentiate = TRUE, conf.int = TRUE)
# A tibble: 2 × 7
  term        estimate std.error statistic  p.value conf.low conf.high
  <chr>          <dbl>     <dbl>     <dbl>    <dbl>    <dbl>     <dbl>
1 (Intercept)    0.117     0.264     -8.13 4.35e-16   0.0669     0.190
2 cut15_20221   19.3       0.500      5.91 3.37e- 9   7.48      54.0
> tidy(model15_south,   exponentiate = TRUE, conf.int = TRUE)
# A tibble: 2 × 7
  term        estimate std.error statistic  p.value conf.low conf.high
  <chr>          <dbl>     <dbl>     <dbl>    <dbl>    <dbl>     <dbl>
1 (Intercept)    0.175     0.263     -6.62 3.51e-11    0.101     0.285
2 cut15_20221    7.49      0.423      4.76 1.98e- 6    3.31     17.5
> library(broom)
> library(dplyr)
>
> # Extract results per model
> res_15 <- bind_rows(
+       tidy(model15_north,  exponentiate  =  TRUE,  conf.int  =  TRUE)  %>%
mutate(region = "North", cutoff = "15"),
+      tidy(model15_central,  exponentiate  =  TRUE,  conf.int  =  TRUE)  %>%
mutate(region = "Central", cutoff = "15"),
+       tidy(model15_south,  exponentiate  =  TRUE,  conf.int  =  TRUE)  %>%
mutate(region = "South", cutoff = "15")
+ )
>
> res_35 <- bind_rows(
+   tidy(model_north, exponentiate = TRUE, conf.int = TRUE) %>% mutate(region
= "North", cutoff = "35"),
+       tidy(model_central,  exponentiate  =  TRUE,  conf.int  =  TRUE)  %>%
mutate(region = "Central", cutoff = "35"),
+   tidy(model_south, exponentiate = TRUE, conf.int = TRUE) %>% mutate(region
= "South", cutoff = "35")
+ )
>
> # Combineer resultaten
> results <- bind_rows(res_15, res_35) %>%
+   filter(term == "cut15_2022" | term == "cut35_2022") %>%
+   mutate(cutoff = factor(cutoff, levels = c("15", "35")))
> library(ggplot2)
>
> ggplot(results, aes(x = region, y = estimate, ymin = conf.low, ymax =
conf.high, color = cutoff)) +
+   geom_pointrange(position = position_dodge(width = 0.5)) +
+   geom_hline(yintercept = 1, linetype = "dashed") +
+   coord_flip() +
+   labs(
+     title = "Odds ratios per regio voor cutoff 15 en 35",
+     x = "Regio",
+     y = "Odds ratio (95% CI)",
+     color = "Cutoff"
+   ) +
+   theme_minimal()
```

```
Error in UseMethod("depth") :
  no applicable method for 'depth' applied to an object of class "NULL"
Error in UseMethod("depth") :
  no applicable method for 'depth' applied to an object of class "NULL"
> > > str(results)
tibble [0 × 9] (S3: tbl_df/tbl/data.frame)
 $ term     : chr(0)
 $ estimate : num(0)
 $ std.error: num(0)
 $ statistic: num(0)
 $ p.value  : num(0)
 $ conf.low : num(0)
 $ conf.high: num(0)
 $ region   : chr(0)
 $ cutoff   : Factor w/ 2 levels "15","35":
> tidy(model15_north)
# A tibble: 2 × 5
  term        estimate std.error statistic  p.value
  <chr>          <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)    -2.99     0.458     -6.51 7.31e-11
2 cut15_20221     4.37     0.723      6.05 1.47e- 9
> results <- bind_rows(res_15, res_35) %>%
+   filter(grepl("1$", term)) %>%  # pakt termen die eindigen op "1"
+   mutate(cutoff = factor(cutoff, levels = c("15", "35")))
> ggplot(results, aes(x = region, y = estimate, ymin = conf.low, ymax =
conf.high, color = cutoff)) +
+   geom_pointrange(position = position_dodge(width = 0.5)) +
+   geom_hline(yintercept = 1, linetype = "dashed") +
+   coord_flip() +
+   labs(
+     title = "Odds ratios per regio voor cutoff 15 en 35",
+     x = "Regio",
+     y = "Odds ratio (95% CI)",
+     color = "Cutoff"
+   ) +
+   theme_minimal()
> ggsave("forest_plot_cutoff15_35.png", width = 8, height = 5, dpi = 300)
> data$cut35_2022 <- as.factor(data$cut35_2022)
> data$cut35_2024 <- as.factor(data$cut35_2024)
> data$region <- as.factor(data$region)
> model_interactie_35 <- glm(cut35_2024 ~ cut35_2022 * region, data = data,
family = binomial)
> summary(model_interactie_35)

Call:
glm(formula = cut35_2024 ~ cut35_2022 * region, family = binomial,
    data = data)

Coefficients:
                         Estimate Std. Error z value Pr(>|z|)
(Intercept)             -2.404368   0.279070  -8.616  < 2e-16 ***
cut35_20221              2.809833   0.703240   3.996 6.45e-05 ***
regionNorth             -0.909818   0.580492  -1.567    0.117
regionSouth             -0.001758   0.420697  -0.004    0.997
cut35_20221:regionNorth  0.098888   1.081806   0.091    0.927
cut35_20221:regionSouth -0.403707   0.903258  -0.447    0.655
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)
```

```
    Null deviance: 306.44  on 453  degrees of freedom
Residual deviance: 258.98  on 448  degrees of freedom
AIC: 270.98

Number of Fisher Scoring iterations: 6

> exp(cbind(OR = coef(model_interactie_35), confint(model_interactie_35)))
Waiting for profiling to be done...
                               OR     2.5 %     97.5 %
(Intercept)              0.09032258 0.0499119  0.1502987
cut35_20221             16.60714286 4.2652839 71.9388703
regionNorth              0.40259740 0.1118106  1.1571075
regionSouth              0.99824356 0.4286554  2.2714109
cut35_20221:regionNorth  1.10394265 0.1278041  9.3961261
cut35_20221:regionSouth  0.66783969 0.1084800  3.8980795
> data$cut15_2022 <- as.factor(data$cut15_2022)
> data$cut15_2024 <- as.factor(data$cut15_2024)
> data$region <- as.factor(data$region)
> model_interactie_15 <- glm(cut15_2024 ~ cut15_2022 * region, data = data,
family = binomial)
> summary(model_interactie_15)

Call:
glm(formula = cut15_2024 ~ cut15_2022 * region, family = binomial,
    data = data)

Coefficients:
                        Estimate Std. Error z value Pr(>|z|)
(Intercept)              -2.1474     0.2642  -8.128 4.36e-16 ***
cut15_20221               2.9583     0.5004   5.912 3.37e-09 ***
regionNorth              -0.8383     0.5290  -1.585    0.113
regionSouth               0.4059     0.3727   1.089    0.276
cut15_20221:regionNorth   1.4137     0.8792   1.608    0.108
cut15_20221:regionSouth  -0.9449     0.6554  -1.442    0.149
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 460.40  on 453  degrees of freedom
Residual deviance: 341.38  on 448  degrees of freedom
AIC: 353.38

Number of Fisher Scoring iterations: 5

> exp(cbind(OR = coef(model_interactie_15), confint(model_interactie_15)))
Waiting for profiling to be done...
                               OR     2.5 %     97.5 %
(Intercept)              0.1167883 0.0669158  0.1898213
cut15_20221             19.2656250 7.4761079 54.0462282
regionNorth              0.4324495 0.1377609  1.1454348
regionSouth              1.5006443 0.7204930  3.1401219
cut15_20221:regionNorth  4.1109489 0.7718093 25.1679873
cut15_20221:regionSouth  0.3887219 0.1050679  1.3888584
> library(broom)
>
> tidy(model_interactie_15, exponentiate = TRUE, conf.int = TRUE)
# A tibble: 6 × 7
  term                     estimate std.error statistic  p.value conf.low
conf.high
```

```
  <chr>                       <dbl>      <dbl>      <dbl>     <dbl>     <dbl>
<dbl>
1 (Intercept)                 0.117      0.264      -8.13  4.36e-16    0.0669
0.190
2 cut15_20221            19.3       0.500      5.91 3.37e- 9   7.48      54.0
3 regionNorth             0.432     0.529     -1.58 1.13e- 1   0.138      1.15
4 regionSouth             1.50      0.373      1.09 2.76e- 1   0.720      3.14
5 cut15_20221:regionNo…   4.11      0.879      1.61 1.08e- 1   0.772     25.2
6 cut15_20221:regionSo…   0.389     0.655     -1.44 1.49e- 1   0.105      1.39
> library(dplyr)
>
> # North
>   table_north_35   <-   table(filter(data,   region   ==   "North")   %>%
select(cut35_2022, cut35_2024))
> cat("McNemar test - North (cutoff 35):\n")
McNemar test - North (cutoff 35):
> print(table_north_35)
          cut35_2024
cut35_2022   0   1
         0 110   4
         1   6   4
> mcnemar.test(table_north_35)

        McNemar's Chi-squared test with continuity correction

data:  table_north_35
McNemar's chi-squared = 0.1, df = 1, p-value = 0.7518

>
> # Central
>   table_central_35   <-   table(filter(data,   region   ==   "Central")   %>%
select(cut35_2022, cut35_2024))
> cat("\nMcNemar test - Central (cutoff 35):\n")

McNemar test - Central (cutoff 35):
> print(table_central_35)
          cut35_2024
cut35_2022   0   1
         0 155  14
         1   4   6
> mcnemar.test(table_central_35)

        McNemar's Chi-squared test with continuity correction

data:  table_central_35
McNemar's chi-squared = 4.5, df = 1, p-value = 0.03389

>
> # South
>   table_south_35   <-   table(filter(data,   region   ==   "South")   %>%
select(cut35_2022, cut35_2024))
> cat("\nMcNemar test - South (cutoff 35):\n")

McNemar test - South (cutoff 35):
> print(table_south_35)
          cut35_2024
cut35_2022   0   1
         0 122  11
         1   9   9
> mcnemar.test(table_south_35)
```

```
        McNemar's Chi-squared test with continuity correction

data:  table_south_35
McNemar's chi-squared = 0.05, df = 1, p-value = 0.8231

> # North
>   table_north_15   <-   table(filter(data,   region   ==   "North")   %>%
select(cut15_2022, cut15_2024))
> cat("\nMcNemar test - North (cutoff 15):\n")

McNemar test - North (cutoff 15):
> print(table_north_15)
          cut15_2024
cut15_2022  0  1
         0 99  5
         1  4 16
> mcnemar.test(table_north_15)

        McNemar's Chi-squared test with continuity correction

data:  table_north_15
McNemar's chi-squared = 0, df = 1, p-value = 1

>
> # Central
>  table_central_15   <-   table(filter(data,   region   ==   "Central")   %>%
select(cut15_2022, cut15_2024))
> cat("\nMcNemar test - Central (cutoff 15):\n")

McNemar test - Central (cutoff 15):
> print(table_central_15)
          cut15_2024
cut15_2022   0   1
         0 137  16
         1   8  18
> mcnemar.test(table_central_15)

        McNemar's Chi-squared test with continuity correction

data:  table_central_15
McNemar's chi-squared = 2.0417, df = 1, p-value = 0.153

>
> # South
>   table_south_15   <-   table(filter(data,   region   ==   "South")   %>%
select(cut15_2022, cut15_2024))
> cat("\nMcNemar test - South (cutoff 15):\n")

McNemar test - South (cutoff 15):
> print(table_south_15)
          cut15_2024
cut15_2022  0  1
         0 97 17
         1 16 21
> mcnemar.test(table_south_15)

        McNemar's Chi-squared test with continuity correction

data:  table_south_15
McNemar's chi-squared = 0, df = 1, p-value = 1
```

```
> data$cut35_2022 <- as.factor(data$cut35_2022)
> data$cut35_2024 <- as.factor(data$cut35_2024)
> data$cut15_2022 <- as.factor(data$cut15_2022)
> data$cut15_2024 <- as.factor(data$cut15_2024)
> data$colony <- as.factor(data$colony)  # Hutterite / Non-Hutterite
> data_hutterite <- filter(data, colony == "Hutterite")
> data_nonhutterite <- filter(data, colony == "Non-Hutterite")
> # Hutterite
> model_35_hutterite <- glm(cut35_2024 ~ cut35_2022, data = data_hutterite,
family = binomial)
>
> # Non-Hutterite
>   model_35_nonhutterite   <-   glm(cut35_2024   ~   cut35_2022,   data   =
data_nonhutterite, family = binomial)
>
> # Samenvatten
> summary(model_35_hutterite)

Call:
glm(formula   =   cut35_2024   ~   cut35_2022,   family   =   binomial,   data   =
data_hutterite)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -3.6972     0.5845  -6.326 2.52e-10 ***
cut35_20221   3.1864     0.9354   3.407 0.000658 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 48.816  on 131  degrees of freedom
Residual deviance: 38.842  on 130  degrees of freedom
AIC: 42.842

Number of Fisher Scoring iterations: 6

> summary(model_35_nonhutterite)

Call:
glm(formula   =   cut35_2024   ~   cut35_2022,   family   =   binomial,   data   =
data_nonhutterite)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.3254     0.2055 -11.317  < 2e-16 ***
cut35_20221   2.4589     0.4197   5.859 4.66e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 249.36  on 321  degrees of freedom
Residual deviance: 216.84  on 320  degrees of freedom
AIC: 220.84

Number of Fisher Scoring iterations: 5

>
> # Odds ratios en CI
> exp(cbind(OR = coef(model_35_hutterite), confint(model_35_hutterite)))
```

```
Waiting for profiling to be done...
                    OR      2.5 %       97.5 %
(Intercept)  0.02479339 0.00610893   0.06555077
cut35_20221 24.19999999 3.71935845 165.47235308
>           exp(cbind(OR           =           coef(model_35_nonhutterite),
confint(model_35_nonhutterite)))
Waiting for profiling to be done...
                    OR      2.5 %     97.5 %
(Intercept)  0.09774436 0.06379281  0.1432288
cut35_20221 11.69230769 5.15748096 27.0212600
> # Hutterite
> model_15_hutterite <- glm(cut15_2024 ~ cut15_2022, data = data_hutterite,
family = binomial)
>
> # Non-Hutterite
>   model_15_nonhutterite   <-   glm(cut15_2024   ~   cut15_2022,   data   =
data_nonhutterite, family = binomial)
>
> # Samenvatten
> summary(model_15_hutterite)

Call:
glm(formula   =   cut15_2024   ~   cut15_2022,   family   =   binomial,   data   =
data_hutterite)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -3.0350     0.4578   -6.63 3.37e-11 ***
cut15_20221   2.7726     0.6217    4.46 8.20e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 93.470  on 131  degrees of freedom
Residual deviance: 72.079  on 130  degrees of freedom
AIC: 76.079

Number of Fisher Scoring iterations: 5

> summary(model_15_nonhutterite)

Call:
glm(formula   =   cut15_2024   ~   cut15_2022,   family   =   binomial,   data   =
data_nonhutterite)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.9372     0.1862 -10.404   <2e-16 ***
cut15_20221   3.0358     0.3515   8.637   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 356.55  on 321  degrees of freedom
Residual deviance: 265.88  on 320  degrees of freedom
AIC: 269.88

Number of Fisher Scoring iterations: 4
```

```
>
> # Odds ratios en CI
> exp(cbind(OR = coef(model_15_hutterite), confint(model_15_hutterite)))
Waiting for profiling to be done...
                     OR      2.5 %     97.5 %
(Intercept)  0.04807692 0.01697822  0.1061934
cut15_20221 15.99999964 4.93116905 58.6409161
>           exp(cbind(OR           =           coef(model_15_nonhutterite),
confint(model_15_nonhutterite)))
Waiting for profiling to be done...
                    OR       2.5 %     97.5 %
(Intercept)  0.1441048  0.09827192  0.2043888
cut15_20221 20.8181818 10.68875055 42.6548696
> # Hutterite
>  table_hutterite_35  <-  table(filter(data,  colony  ==  "Hutterite")  %>%
select(cut35_2022, cut35_2024))
> cat("McNemar test - Hutterite (cutoff 35):\n")
McNemar test - Hutterite (cutoff 35):
> print(table_hutterite_35)
          cut35_2024
cut35_2022   0   1
         0 121   3
         1   5   3
> mcnemar.test(table_hutterite_35)

        McNemar's Chi-squared test with continuity correction

data:  table_hutterite_35
McNemar's chi-squared = 0.125, df = 1, p-value = 0.7237

>
> # Non-Hutterite
> table_nonhutterite_35 <- table(filter(data, colony == "Non-Hutterite") %>%
select(cut35_2022, cut35_2024))
> cat("\nMcNemar test - Non-Hutterite (cutoff 35):\n")

McNemar test - Non-Hutterite (cutoff 35):
> print(table_nonhutterite_35)
          cut35_2024
cut35_2022   0   1
         0 266  26
         1  14  16
> mcnemar.test(table_nonhutterite_35)

        McNemar's Chi-squared test with continuity correction

data:  table_nonhutterite_35
McNemar's chi-squared = 3.025, df = 1, p-value = 0.08199

> # Hutterite
>  table_hutterite_15  <-  table(filter(data,  colony  ==  "Hutterite")  %>%
select(cut15_2022, cut15_2024))
> cat("\nMcNemar test - Hutterite (cutoff 15):\n")

McNemar test - Hutterite (cutoff 15):
> print(table_hutterite_15)
          cut15_2024
cut15_2022   0   1
         0 104   5
         1  13  10
> mcnemar.test(table_hutterite_15)
```

```
        McNemar's Chi-squared test with continuity correction

data:  table_hutterite_15
McNemar's chi-squared = 2.7222, df = 1, p-value = 0.09896

>
> # Non-Hutterite
> table_nonhutterite_15 <- table(filter(data, colony == "Non-Hutterite") %>%
select(cut15_2022, cut15_2024))
> cat("\nMcNemar test - Non-Hutterite (cutoff 15):\n")

McNemar test - Non-Hutterite (cutoff 15):
> print(table_nonhutterite_15)
          cut15_2024
cut15_2022   0   1
         0 229  33
         1  15  45
> mcnemar.test(table_nonhutterite_15)

        McNemar's Chi-squared test with continuity correction

data:  table_nonhutterite_15
McNemar's chi-squared = 6.0208, df = 1, p-value = 0.01414

> library(dplyr)
>
> data_change_15 <- data %>%
+   mutate(change = case_when(
+     cut15_2022 == 0 & cut15_2024 == 0 ~ "Stayed Negative",
+     cut15_2022 == 0 & cut15_2024 == 1 ~ "Became Positive",
+     cut15_2022 == 1 & cut15_2024 == 0 ~ "Became Negative",
+     cut15_2022 == 1 & cut15_2024 == 1 ~ "Stayed Positive"
+   ))
> change_summary <- data_change_15 %>%
+   group_by(colony, change) %>%
+   summarise(count = n(), .groups = "drop")
> library(ggplot2)
>
> ggplot(change_summary, aes(x = change, y = count, fill = colony)) +
+   geom_bar(stat = "identity", position = position_dodge()) +
+   labs(
+     title = "Change in Test Results (2022 → 2024, Cutoff 15)",
+     x = "Change in Test Status",
+     y = "Number of Herds",
+     fill = "Colony Type"
+   ) +
+   theme_minimal()
> library(dplyr)
>
> # Bereken % positief in 2022 en 2024 voor cutoff 15
> positives_15 <- data %>%
+   select(colony, cut15_2022, cut15_2024) %>%
+   pivot_longer(cols = c(cut15_2022, cut15_2024),
+                names_to = "year",
+                names_prefix = "cut15_",
+                values_to = "status") %>%
+   group_by(colony, year) %>%
+   sum
Error in pivot_longer(., cols = c(cut15_2022, cut15_2024), names_to = "year",
:
```

```
  could not find function "pivot_longer"
> library(dplyr)
>
> # Bereken % positief in 2022 en 2024 voor cutoff 15
> positives_15 <- data %>%
+   select(colony, cut15_2022, cut15_2024) %>%
+   pivot_longer(cols = c(cut15_2022, cut15_2024),
+                names_to = "year",
+                names_prefix = "cut15_",
+                values_to = "status") %>%
+   group_by(colony, year) %>%
+    summarise(perc_positive = mean(as.numeric(status)) * 100, .groups =
"drop")
Error in pivot_longer(., cols = c(cut15_2022, cut15_2024), names_to = "year",
:
  could not find function "pivot_longer"
> library(tidyr)
> install.packages("tidyr")
Warning: package 'tidyr' is in use and will not be installed
> ggplot(positives_15, aes(x = colony, y = perc_positive, fill = year)) +
+   geom_bar(stat = "identity", position = position_dodge()) +
+   labs(
+    title = "Percentage Positive Test Results per Colony Type (Cutoff 15)",
+     x = "Colony Type",
+     y = "Percentage Positive",
+     fill = "Year"
+   ) +
+   theme_minimal() +
+   ylim(0, 100)
Error: object 'positives_15' not found
> positives_15 <- data %>%
+   select(colony, cut15_2022, cut15_2024) %>%
+   pivot_longer(cols = c(cut15_2022, cut15_2024),
+                names_to = "year",
+                names_prefix = "cut15_",
+                values_to = "status") %>%
+   group_by(colony, year) %>%
+    summarise(perc_positive = mean(as.numeric(status)) * 100, .groups =
"drop")
> print(positives_15)
# A tibble: 4 × 3
  colony        year  perc_positive
  <fct>         <chr>         <dbl>
1 Hutterite     2022           117.
2 Hutterite     2024           111.
3 Non-Hutterite 2022           119.
4 Non-Hutterite 2024           124.
> ggplot(positives_15, aes(x = colony, y = perc_positive, fill = year)) +
+   geom_bar(stat = "identity", position = position_dodge()) +
+   labs(
+    title = "Percentage Positive Test Results per Colony Type (Cutoff 15)",
+     x = "Colony Type",
+     y = "Percentage Positive",
+     fill = "Year"
+   ) +
+   theme_minimal() +
+   ylim(0, 100)
Warning message:
Removed 4 rows containing missing values or values outside the scale range
(`geom_bar()`).
Error in UseMethod("depth") :
```

```
  no applicable method for 'depth' applied to an object of class "NULL"
> > data$cut15_2022 <- as.numeric(as.character(data$cut15_2022))
> data$cut15_2024 <- as.numeric(as.character(data$cut15_2024))
> table(data$cut15_2022)

  0   1
371  83
> table(data$cut15_2024)

  0   1
361  93
> library(tidyr)
> library(dplyr)
>
> positives_15 <- data %>%
+   select(colony, cut15_2022, cut15_2024) %>%
+   pivot_longer(cols = c(cut15_2022, cut15_2024),
+                names_to = "year",
+                names_prefix = "cut15_",
+                values_to = "status") %>%
+   group_by(colony, year) %>%
+    summarise(perc_positive = mean(as.numeric(status)) * 100, .groups =
"drop")
> ggplot(positives_15, aes(x = colony, y = perc_positive, fill = year)) +
+   geom_bar(stat = "identity", position = position_dodge()) +
+   labs(
+    title = "Percentage Positive Test Results per Colony Type (Cutoff 15)",
+     x = "Colony Type",
+     y = "Percentage Positive",
+     fill = "Year"
+   ) +
+   theme_minimal() +
+   ylim(0, 100)
> p <- ggplot(positives_15, aes(x = colony, y = perc_positive, fill = year))
+
+   geom_bar(stat = "identity", position = position_dodge()) +
+   labs(
+    title = "Percentage Positive Test Results per Colony Type (Cutoff 15)",
+     x = "Colony Type",
+     y = "Percentage Positive",
+     fill = "Year"
+   ) +
+   theme_minimal() +
+   ylim(0, 100)
> ggsave("positive_percentages_cutoff15.png", plot = p, width = 8, height =
5, dpi = 300)
> data$cut35_2022 <- as.factor(data$cut35_2022)
> data$cut35_2024 <- as.factor(data$cut35_2024)
> data$milking_system <- as.factor(data$milking_system)
> data_AMS <- filter(data, milking_system == "AMS")
> data_CMS <- filter(data, milking_system == "CMS")
> # AMS
> model_ams <- glm(cut35_2024 ~ cut35_2022, data = data_AMS, family =
binomial)
>
> # CMS
> model_cms <- glm(cut35_2024 ~ cut35_2022, data = data_CMS, family =
binomial)
> summary(model_ams)

Call:
```

```
glm(formula = cut35_2024 ~ cut35_2022, family = binomial, data = data_AMS)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -3.1697     0.4565  -6.943 3.83e-12 ***
cut35_20221   4.2683     0.9355   4.563 5.05e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 75.725  on 131  degrees of freedom
Residual deviance: 50.901  on 130  degrees of freedom
AIC: 54.901

Number of Fisher Scoring iterations: 6

> summary(model_cms)

Call:
glm(formula = cut35_2024 ~ cut35_2022, family = binomial, data = data_CMS)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.4129     0.2131 -11.325  < 2e-16 ***
cut35_20221   2.1447     0.4256   5.039 4.68e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 229.68  on 321  degrees of freedom
Residual deviance: 206.96  on 320  degrees of freedom
AIC: 210.96

Number of Fisher Scoring iterations: 5

>
> exp(cbind(OR = coef(model_ams), confint(model_ams)))
Waiting for profiling to be done...
                     OR       2.5 %       97.5 %
(Intercept)  0.04201681  0.01486614   0.09249019
cut35_20221 71.40000000 13.06468329 586.65761653
> exp(cbind(OR = coef(model_cms), confint(model_cms)))
Waiting for profiling to be done...
                    OR     2.5 %     97.5 %
(Intercept) 0.08955224 0.0574512  0.1329405
cut35_20221 8.53921569 3.6749515 19.7255119
> data$cut15_2022 <- as.factor(data$cut15_2022)
> data$cut15_2024 <- as.factor(data$cut15_2024)
> data$milking_system <- as.factor(data$milking_system)
> data_AMS <- filter(data, milking_system == "AMS")
> data_CMS <- filter(data, milking_system == "CMS")
> # AMS
> model_ams_15 <- glm(cut15_2024 ~ cut15_2022, data = data_AMS, family =
binomial)
>
> # CMS
> model_cms_15 <- glm(cut15_2024 ~ cut15_2022, data = data_CMS, family =
binomial)
> summary(model_ams_15)
```

```
Call:
glm(formula = cut15_2024 ~ cut15_2022, family = binomial, data = data_AMS)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.2465     0.3171  -7.085 1.39e-12 ***
cut15_20221   3.1220     0.6196   5.039 4.68e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 122.113  on 131  degrees of freedom
Residual deviance:  93.145  on 130  degrees of freedom
AIC: 97.145

Number of Fisher Scoring iterations: 5

> summary(model_cms_15)

Call:
glm(formula = cut15_2024 ~ cut15_2022, family = binomial, data = data_CMS)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.1379     0.2035 -10.507   <2e-16 ***
cut15_20221   2.7636     0.3288   8.404   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 337.19  on 321  degrees of freedom
Residual deviance: 257.85  on 320  degrees of freedom
AIC: 261.85

Number of Fisher Scoring iterations: 4

>
> exp(cbind(OR = coef(model_ams_15), confint(model_ams_15)))
Waiting for profiling to be done...
                    OR     2.5 %     97.5 %
(Intercept)  0.1057692 0.0535783  0.1879545
cut15_20221 22.6909091 7.1048222 83.4219866
> exp(cbind(OR = coef(model_cms_15), confint(model_cms_15)))
Waiting for profiling to be done...
                    OR      2.5 %     97.5 %
(Intercept)  0.1179039 0.07735877  0.1723077
cut15_20221 15.8566827 8.44290939 30.7514009
> library(dplyr)
>
> # AMS
>  table_ams_35  <-  table(filter(data,  milking_system  ==  "AMS")  %>%
select(cut35_2022, cut35_2024))
> cat("McNemar test - AMS (cutoff 35):\n")
McNemar test - AMS (cutoff 35):
> print(table_ams_35)
          cut35_2024
cut35_2022   0   1
         0 119   5
```

```
         1   2   6
> mcnemar.test(table_ams_35)

        McNemar's Chi-squared test with continuity correction

data:  table_ams_35
McNemar's chi-squared = 0.57143, df = 1, p-value = 0.4497

>
> # CMS
>  table_cms_35  <-  table(filter(data,  milking_system  ==  "CMS")  %>%
select(cut35_2022, cut35_2024))
> cat("\nMcNemar test - CMS (cutoff 35):\n")

McNemar test - CMS (cutoff 35):
> print(table_cms_35)
          cut35_2024
cut35_2022   0   1
         0 268  24
         1  17  13
> mcnemar.test(table_cms_35)

        McNemar's Chi-squared test with continuity correction

data:  table_cms_35
McNemar's chi-squared = 0.87805, df = 1, p-value = 0.3487

> # AMS
>  table_ams_15  <-  table(filter(data,  milking_system  ==  "AMS")  %>%
select(cut15_2022, cut15_2024))
> cat("\nMcNemar test - AMS (cutoff 15):\n")

McNemar test - AMS (cutoff 15):
> print(table_ams_15)
          cut15_2024
cut15_2022   0   1
         0 104  11
         1   5  12
> mcnemar.test(table_ams_15)

        McNemar's Chi-squared test with continuity correction

data:  table_ams_15
McNemar's chi-squared = 1.5625, df = 1, p-value = 0.2113

>
> # CMS
>  table_cms_15  <-  table(filter(data,  milking_system  ==  "CMS")  %>%
select(cut15_2022, cut15_2024))
> cat("\nMcNemar test - CMS (cutoff 15):\n")

McNemar test - CMS (cutoff 15):
> print(table_cms_15)
          cut15_2024
cut15_2022   0   1
         0 229  27
         1  23  43
> mcnemar.test(table_cms_15)

        McNemar's Chi-squared test with continuity correction
```

```
data:  table_cms_15
McNemar's chi-squared = 0.18, df = 1, p-value = 0.6714

> library(dplyr)
>
> # Functie om statusverandering toe te voegen
> add_change_status <- function(df, cut_2022, cut_2024) {
+   df %>%
+     mutate(change = case_when(
+       !!sym(cut_2022) == 0 & !!sym(cut_2024) == 0 ~ "Remained Negative",
+       !!sym(cut_2022) == 0 & !!sym(cut_2024) == 1 ~ "Became Positive",
+       !!sym(cut_2022) == 1 & !!sym(cut_2024) == 0 ~ "Became Negative",
+       !!sym(cut_2022) == 1 & !!sym(cut_2024) == 1 ~ "Remained Positive"
+     ))
+ }
>
>  data_15  <-  add_change_status(data,  "cut15_2022",  "cut15_2024")  %>%
mutate(cutoff = "15")
>  data_35  <-  add_change_status(data,  "cut35_2022",  "cut35_2024")  %>%
mutate(cutoff = "35")
>
> # Combineer
> data_combined <- bind_rows(data_15, data_35)
> summary_data <- data_combined %>%
+   group_by(milking_system, cutoff, change) %>%
+   summarise(count = n(), .groups = "drop")
> library(ggplot2)
>
> ggplot(summary_data, aes(x = milking_system, y = count, fill = change)) +
+   geom_bar(stat = "identity") +
+   facet_wrap(~ cutoff) +
+   labs(
+     title = "Change in Salmonella Dublin Test Status by Milking System",
+     x = "Milking System",
+     y = "Number of Farms",
+     fill = "Status Change"
+   ) +
+   theme_minimal()
> p <- ggplot(summary_data, aes(x = milking_system, y = count, fill = change))
+
+   geom_bar(stat = "identity") +
+   facet_wrap(~ cutoff) +
+   labs(
+     title = "Change in Salmonella Dublin Test Status by Milking System",
+     x = "Milking System",
+     y = "Number of Farms",
+     fill = "Status Change"
+   ) +
+   theme_minimal()
> ggsave("status_change_milking_system.png", plot = p, width = 10, height =
6, dpi = 300)
> data$cut35_2022 <- as.factor(data$cut35_2022)
> data$cut35_2024 <- as.factor(data$cut35_2024)
> data$`farm size` <- as.factor(data$`farm size`)
Error in `$<-`:
! Assigned data `as.factor(data$\`farm size\`)` must be compatible with
existing data.
✖ Existing data has 454 rows.
✖ Assigned data has 0 rows.
ℹ Only vectors of size 1 are recycled.
Caused by error in `vectbl_recycle_rhs_rows()`:
```

```
! Can't recycle input of size 0 to size 454.
Run `rlang::last_trace()` to see where the error occurred.
Warning message:
Unknown or uninitialised column: `farm size`.
> colnames(data)
[1] "farm"              "region"           "colony"           "cut35_2022"
"cut35_2024"     "cut15_2022"
[7] "cut15_2024"     "milking_system" "farm_size"
> data$farm_size <- as.factor(data$farm_size)
> library(dplyr)
>
> data_small <- filter(data, farm_size == "Small")
> data_medium <- filter(data, farm_size == "Medium")
> data_large <- filter(data, farm_size == "Large")
>
> model_small <- glm(cut35_2024 ~ cut35_2022, data = data_small, family =
binomial)
> model_medium <- glm(cut35_2024 ~ cut35_2022, data = data_medium, family =
binomial)
> model_large <- glm(cut35_2024 ~ cut35_2022, data = data_large, family =
binomial)
>
> summary(model_small)

Call:
glm(formula = cut35_2024 ~ cut35_2022, family = binomial, data = data_small)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.7354     0.2979  -9.183  < 2e-16 ***
cut35_20221   2.7354     0.5820   4.700  2.6e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 132.68  on 212  degrees of freedom
Residual deviance: 112.59  on 211  degrees of freedom
AIC: 116.59

Number of Fisher Scoring iterations: 5

> summary(model_medium)

Call:
glm(formula = cut35_2024 ~ cut35_2022, family = binomial, data = data_medium)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.6532     0.3272  -8.110 5.08e-16 ***
cut35_20221   3.1232     0.6573   4.752 2.02e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 113.721  on 164  degrees of freedom
Residual deviance:  91.076  on 163  degrees of freedom
AIC: 95.076

Number of Fisher Scoring iterations: 5
```

```
> summary(model_large)

Call:
glm(formula = cut35_2024 ~ cut35_2022, family = binomial, data = data_large)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.1484     0.3994  -5.379 7.47e-08 ***
cut35_20221   1.4553     0.8121   1.792   0.0731 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 59.185  on 75  degrees of freedom
Residual deviance: 56.322  on 74  degrees of freedom
AIC: 60.322

Number of Fisher Scoring iterations: 4

>
> exp(cbind(OR = coef(model_small), confint(model_small)))
Waiting for profiling to be done...
                     OR      2.5 %     97.5 %
(Intercept)  0.06486486 0.03422213  0.1111666
cut35_20221 15.41666666 4.91813185 49.5413123
> exp(cbind(OR = coef(model_medium), confint(model_medium)))
Waiting for profiling to be done...
                     OR      2.5 %     97.5 %
(Intercept)  0.07042254 0.03468204  0.1267845
cut35_20221 22.72000000 6.46171167 88.5652092
> exp(cbind(OR = coef(model_large), confint(model_large)))
Waiting for profiling to be done...
                   OR      2.5 %     97.5 %
(Intercept) 0.1166667 0.04854011  0.2379875
cut35_20221 4.2857143 0.77549930 20.6334366
> # Zorg dat variabelen factor zijn
> data$cut15_2022 <- as.factor(data$cut15_2022)
> data$cut15_2024 <- as.factor(data$cut15_2024)
> data$farm_size <- as.factor(data$farm_size)
>
> # Data splitsen per farm size
> data_small <- filter(data, farm_size == "Small")
> data_medium <- filter(data, farm_size == "Medium")
> data_large <- filter(data, farm_size == "Large")
>
> # Logistische regressie per farm size
> model_small_15 <- glm(cut15_2024 ~ cut15_2022, data = data_small, family =
binomial)
> model_medium_15 <- glm(cut15_2024 ~ cut15_2022, data = data_medium, family
= binomial)
> model_large_15 <- glm(cut15_2024 ~ cut15_2022, data = data_large, family =
binomial)
>
> # Samenvatten en odds ratios berekenen
> summary(model_small_15)

Call:
glm(formula = cut15_2024 ~ cut15_2022, family = binomial, data = data_small)
```

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.1493     0.2424  -8.867  < 2e-16 ***
cut15_20221   2.6088     0.4413   5.912 3.38e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 199.78  on 212  degrees of freedom
Residual deviance: 163.19  on 211  degrees of freedom
AIC: 167.19

Number of Fisher Scoring iterations: 4

> summary(model_medium_15)

Call:
glm(formula = cut15_2024 ~ cut15_2022, family = binomial, data = data_medium)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.3273     0.3024  -7.695 1.41e-14 ***
cut15_20221   3.1746     0.5002   6.347 2.20e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 165.13  on 164  degrees of freedom
Residual deviance: 117.64  on 163  degrees of freedom
AIC: 121.64

Number of Fisher Scoring iterations: 5

> summary(model_large_15)

Call:
glm(formula = cut15_2024 ~ cut15_2022, family = binomial, data = data_large)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.9042     0.4051  -4.700 2.60e-06 ***
cut15_20221   2.6664     0.6113   4.362 1.29e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 91.455  on 75  degrees of freedom
Residual deviance: 69.175  on 74  degrees of freedom
AIC: 73.175

Number of Fisher Scoring iterations: 4

>
> exp(cbind(OR = coef(model_small_15), confint(model_small_15)))
Waiting for profiling to be done...
                    OR      2.5 %     97.5 %
(Intercept)  0.1165644 0.07015079  0.1824239
cut15_20221 13.5833332 5.82130433 33.1735329
```

```
> exp(cbind(OR = coef(model_medium_15), confint(model_medium_15)))
Waiting for profiling to be done...
                     OR    2.5 %     97.5 %
(Intercept)  0.09756098 0.051100  0.1690094
cut15_20221 23.91666667 9.312399 67.0275940
> exp(cbind(OR = coef(model_large_15), confint(model_large_15)))
Waiting for profiling to be done...
                    OR      2.5 %     97.5 %
(Intercept)  0.1489362 0.06143769  0.3082368
cut15_20221 14.3877551 4.57073176 51.2663921
> # Zorg dat variabelen factor zijn
> data$cut35_2022 <- as.factor(data$cut35_2022)
> data$cut35_2024 <- as.factor(data$cut35_2024)
> data$farm_size <- as.factor(data$farm_size)
>
> # Model met interactie tussen cut35_2022 en farm_size
> model_interactie_farmsize_35 <- glm(cut35_2024 ~ cut35_2022 * farm_size,
data = data, family = binomial)
>
> # Samenvatten
> summary(model_interactie_farmsize_35)

Call:
glm(formula = cut35_2024 ~ cut35_2022 * farm_size, family = binomial,
    data = data)

Coefficients:
                            Estimate Std. Error z value Pr(>|z|)
(Intercept)                  -2.1484     0.3994  -5.379 7.49e-08 ***
cut35_20221                   1.4553     0.8121   1.792   0.0731 .
farm_sizeMedium              -0.5048     0.5163  -0.978   0.3282
farm_sizeSmall               -0.5870     0.4983  -1.178   0.2387
cut35_20221:farm_sizeMedium   1.6680     1.0448   1.596   0.1104
cut35_20221:farm_sizeSmall    1.2802     0.9991   1.281   0.2001
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 306.44  on 453  degrees of freedom
Residual deviance: 259.99  on 448  degrees of freedom
AIC: 271.99

Number of Fisher Scoring iterations: 5

>
> # Odds ratios en betrouwbaarheidsintervallen
>         exp(cbind(OR         =         coef(model_interactie_farmsize_35),
confint(model_interactie_farmsize_35)))
Waiting for profiling to be done...
                                    OR      2.5 %     97.5 %
(Intercept)                 0.1166667 0.04854009  0.2379875
cut35_20221                 4.2857143 0.77549935 20.6334365
farm_sizeMedium             0.6036217 0.22126250  1.7320141
farm_sizeSmall              0.5559846 0.21351935  1.5529997
cut35_20221:farm_sizeMedium 5.3013333 0.70642160 44.8941056
cut35_20221:farm_sizeSmall  3.5972222 0.51752008 27.6043220
> # Zorg dat variabelen factor zijn
> data$cut15_2022 <- as.factor(data$cut15_2022)
> data$cut15_2024 <- as.factor(data$cut15_2024)
> data$farm_size <- as.factor(data$farm_size)
```

```
>
> # Model met interactie
> model_interactie_farmsize_15 <- glm(cut15_2024 ~ cut15_2022 * farm_size,
data = data, family = binomial)
>
> # Samenvatten
> summary(model_interactie_farmsize_15)

Call:
glm(formula = cut15_2024 ~ cut15_2022 * farm_size, family = binomial,
    data = data)

Coefficients:
                            Estimate Std. Error z value Pr(>|z|)
(Intercept)                -1.90424    0.40513  -4.700 2.60e-06 ***
cut15_20221                 2.66638    0.61128   4.362 1.29e-05 ***
farm_sizeMedium            -0.42304    0.50557  -0.837    0.403
farm_sizeSmall             -0.24507    0.47212  -0.519    0.604
cut15_20221:farm_sizeMedium  0.50820    0.78984   0.643    0.520
cut15_20221:farm_sizeSmall  -0.05753    0.75392  -0.076    0.939
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 460.4  on 453  degrees of freedom
Residual deviance: 350.0  on 448  degrees of freedom
AIC: 362

Number of Fisher Scoring iterations: 5

>
> # Odds ratios en betrouwbaarheidsintervallen
>         exp(cbind(OR         =         coef(model_interactie_farmsize_15),
confint(model_interactie_farmsize_15)))
Waiting for profiling to be done...
                                  OR      2.5 %     97.5 %
(Intercept)                0.1489362 0.06143769  0.3082368
cut15_20221               14.3877551 4.57073175 51.2663919
farm_sizeMedium            0.6550523 0.24779728  1.8527758
farm_sizeSmall             0.7826468 0.32214330  2.1049207
cut15_20221:farm_sizeMedium  1.6622931 0.34306720  7.7402150
cut15_20221:farm_sizeSmall           0.9440898   0.20652811       4.0381300

> library(dplyr)
>
> sizes <- unique(data$farm_size)
>
> for (size in sizes) {
+   cat("McNemar test -", size, "(cutoff 35):\n")
+    tbl <- table(filter(data, farm_size == size) %>% select(cut35_2022,
cut35_2024))
+   print(tbl)
+   print(mcnemar.test(tbl))
+   cat("\n")
+ }
McNemar test - Small (cutoff 35):
          cut35_2024
cut35_2022   0   1
         0 185  12
         1   8   8
```

```
        McNemar's Chi-squared test with continuity correction

data:  tbl
McNemar's chi-squared = 0.45, df = 1, p-value = 0.5023


McNemar test - Medium (cutoff 35):
          cut35_2024
cut35_2022   0   1
         0 142  10
         1   5   8

        McNemar's Chi-squared test with continuity correction

data:  tbl
McNemar's chi-squared = 1.0667, df = 1, p-value = 0.3017


McNemar test - Large (cutoff 35):
          cut35_2024
cut35_2022  0  1
         0 60  7
         1  6  3

        McNemar's Chi-squared test with continuity correction

data:  tbl
McNemar's chi-squared = 0, df = 1, p-value = 1


> for (size in sizes) {
+   cat("McNemar test -", size, "(cutoff 15):\n")
+    tbl <- table(filter(data, farm_size == size) %>% select(cut15_2022,
cut15_2024))
+   print(tbl)
+   print(mcnemar.test(tbl))
+   cat("\n")
+ }
McNemar test - Small (cutoff 15):
          cut15_2024
cut15_2022   0   1
         0 163  19
         1  12  19

        McNemar's Chi-squared test with continuity correction

data:  tbl
McNemar's chi-squared = 1.1613, df = 1, p-value = 0.2812


McNemar test - Medium (cutoff 15):
          cut15_2024
cut15_2022   0   1
         0 123  12
         1   9  21

        McNemar's Chi-squared test with continuity correction

data:  tbl
McNemar's chi-squared = 0.19048, df = 1, p-value = 0.6625
```

```
McNemar test - Large (cutoff 15):
          cut15_2024
cut15_2022  0  1
         0 47  7
         1  7 15

        McNemar's Chi-squared test

data:  tbl
McNemar's chi-squared = 0, df = 1, p-value = 1
```

Additional alluvial plot:

```
p_alluvial <- ggplot(alluvial_data,
       aes(axis1 = jaar2022, axis2 = jaar2024, y = Freq)) +
  geom_alluvium(aes(fill = status_change), width = 1/12, alpha = 0.66) +
  geom_stratum(width = 1/12, fill = "grey90", color = "black") +
  geom_text(stat = "stratum", aes(label = after_stat(stratum))) +
  geom_text(data = p_labels,
            aes(x = 1.5, y = max(alluvial_data$Freq) * 1.05, label = label),
            inherit.aes = FALSE, size = 4, color = "gray30") +
  scale_x_discrete(limits = c("2022", "2024"), expand = c(.05, .05)) +
  scale_fill_manual(
    name = "Status change",
    values = c(
      "Stayed negative" = "#80DEEA",
      "Stayed positive" = "#EF5350",
      "Became positive" = "#FFF176",
      "Became negative" = "#AED581"
    )
  ) +
  facet_wrap(~cutoff) +
  labs(
    title = "Non-Hutterite herds: 2022 vs 2024",
    x = "Year",
    y = "Number of herds"
  ) +
  theme_minimal()

library(patchwork)

note <- ggplot() +
  annotate("text", x = 0, y = 0, label = "Note: 0 = Negative, 1 = Positive",
           hjust = 0, size = 3.5, color = "gray40") +
  theme_void()

final_plot <- p_alluvial / note + plot_layout(heights = c(1, 0.05))
final_plot

ggsave(
  filename                                                                =
"C:/Users/jessi/Documents/alluvial_nonhutterite_cutoffs_2022_2024.pdf",
  plot = final_plot,
  width = 10,
  height = 6,
  dpi = 300,
  device = cairo_pd
```