Project No:	ON-00715 & ON-00555
Contract No:	AusVet PO4500012146 & PO4500012146 Uni of Tas PO4500013727
AWI Project Manager:	Bridget Peachey
Contractor Name:	AusVet Pty Ltd
Prepared By:	Dr Alison Hillman, Dr Rohan Sadler, Dr Ben Madin – AusVet Including a review from Dr Brian Horton Uni of Tas (Appendix C)
Publication Date:	September 2020

Explanatory modelling - Exploratory data analysis of two research flocks

Published by Australian Wool Innovation Limited, Level 6, 68 Harrington Street, THE ROCKS, NSW, 2000

Table of Contents

OVERVIEW OF THIS FINAL REPORT

This document includes:

- A Final Report by AusVet Pty Ltd. on the outcomes from additional analysis of existing data from the DPIRD and CSIRO breech flystrike resistant flock projects (EC940, WP468 and ON-00169) to investigate whether explanatory modelling has the potential to further explain variation between the breech flystrike susceptible and resistant sheep (Main body of report and Appendix 1).
- An explanation from AusVet on the use of predictive and explanatory statistical modelling as complementary tools when seeking to understand, influence and respond to disease occurrence (Appendix 2).
- Results from a comparison by Dr Brian Horton (University of Tasmania) between a FlyBoss modelling tool to estimate the risk of breech flystrike, developed using the predictive modelling data generated under the DPIRD and CSIRO breech flystrike resistant flock projects, and the explanatory modelling work undertaken by AusVet (Appendix 3).

Summary of Explanatory Modelling Work

1.1 Introduction

Australian Wool Innovation and its partners Department of Primary Industries and Resource Development (Western Australia) (DPIRD) and Commonwealth Scientific and Industrial Research Organisation (CSIRO) have undertaken extensive research on ovine breech flystrike in Merino sheep in Australia, particularly focusing on genetics and breeding strategies to reduce the incidence of ovine breech flystrike.

This study aimed to utilise existing data from these projects to clarify aspects of ovine breech flystrike causality in Australian winter rainfall and summer rainfall farming conditions. Specifically, the project objectives were to quantitatively evaluate:

- the association between ovine breech flystrike and dags,
- the association between ovine breech flystrike and breech wrinkle, and
- the association between ovine breech flystrike and breech cover, specifically considering whether the effect breech cover on the occurrence of breech flystrike varies depending on the degree of breech wrinkle or dags.

1.2 Methods

Three datasets were provided to Ausvet by Australian Wool Innovation and its partners DPIRD and CSIRO. Differences in environmental and management factors experienced by these flocks would have potential implications on the validity of results if these data were analysed together. To avoid this, analyses were undertaken separately on the data from winter rainfall environments (DPIRD data) and summer rainfall environments (CSIRO data).

A previously developed causal web of known and hypothesised risk factors for ovine breech flystrike in Australia (reported in Hillman and Madin, 2018) was used to construct directed acyclic graphs (DAGs) that specifically considered each study objective. The directed acyclic graphs were analysed to systematically identify covariates required to condition multivariable models investigating breech strike causality. Appropriately conditioned multivariable models allow unbiased measurement of the total causal effect of the exposures of interest (dags, breech wrinkle and breech cover) on ovine breech flystrike. The use of the graphs ensures postulated models remain parsimonious and account for potential confounding factors.
Statistical analysis was undertaken using Andersen and Gill survival models, and model results were averaged using an information theoretic approach (Burnham and Anderson, 2002).

1.3 Results

Andersen and Gill models provide quantitative output in the form of a hazard ratio. The hazard ratio compares breech flystrike occurrence between sheep with different dag/ breech wrinkle/ breech cover scores across the duration of the study. A hazard ratio of greater than one indicates that the hazard (breech flystrike) is more likely to occur in the presence of the putative risk factor; a hazard ratio less than one suggests that the presence of the risk factor is protective. Results are highly unlikely to have occurred by chance if the 95% confidence interval (CI) does not include 1, and the p-value is less than 0.05.

Dag was strongly associated with breech strike in both summer and winter rainfall environments, with the likelihood of breech flystrike occurrence increased ordinally with increasing dag score. Compared to sheep in the lowest categories of dag score, the relative likelihood of breech flystrike peaked at over 20 times higher for sheep with dag scores of 4 or above in winter rainfall environments (hazard ratio = 28.5; 95% CI 23.3 – 35.0; p<0.001) and just under 20 times higher for sheep with dag scores of 4 or above in summer rainfall environments (hazard ratio = 19.1; 95% CI 14.7 – 24.7; p<0.001).

Breech wrinkle was also strongly associated with breech flystrike occurrence, with likelihood of disease increasing ordinally in sheep in a summer rainfall environment: compared to sheep with a breech wrinkle score of 1, the relative likelihood of disease peaked at 8 times higher (95% CI 5.62 – 11.4; p<0.001) in sheep with a breech wrinkle score of 5. However, in winter rainfall environments the relative increase in likelihood of breech flystrike occurrence was similarly increased across sheep with breech wrinkle scores of 2 – 3.5, as compared to those with a breech wrinkle score of 1.5 or lower.

In summer rainfall environments the likelihood of occurrence of breech flystrike also increased ordinally with breech cover score, with relative likelihood peaking at peaking at 6.78 (95% CI 2.45 – 18.8; 95% CI <0.001) amongst sheep with a breech cover score of 5, compared to sheep with a breech cover score of 1 – 2. In contrast, results suggest that in winter rainfall environments, the causal effect of breech cover on the occurrence of breech flystrike was very limited.

1.4 Discussion

The quantitative results of this modelling are expected to be of practical value to wool producers in economically optimising production management strategies against the risk of breech flystrike. Results are also expected to be of value to the wool industry, in prioritising allocation of resources for research on intervention strategies that are likely to have the greatest impact in reducing the occurrence of breech flystrike.

The association between dag score and the likelihood of breech flystrike increased ordinally in both the winter rainfall and summer rainfall environments. This suggests that all successful efforts to reduce dag score entail benefit in reduced likelihood of disease.

The same ordinal pattern was seen in the likelihood of breech flystrike increasing with breech wrinkle scores in the summer rainfall environment, suggesting that all successful efforts to reduce breech wrinkle score entail benefit in reduced likelihood of breech flystrike occurrence. Meanwhile, the winter rainfall environment analyses found that the likelihood of breech flystrike was relatively increased in sheep with breech wrinkle scores greater than 1.5; but the magnitude of increased likelihood was similar across breech wrinkle score 2 to 3.5. This suggests that there may be no benefit in reducing breech wrinkle scores within the 2–3.5 range in winter rainfall environments, in terms of reducing the likelihood of breech flystrike.

In sheep in summer rainfall environments, there was again an ordinal association observed between breech cover and the occurrence of breech flystrike, suggesting that all successful efforts to reduce breech cover score will entail benefit in the reduced likelihood of breech flystrike occurrence. In contrast, in sheep in winter rainfall environments the effect of breech cover on the occurrence of breech flystrike was very limited. Avoiding breech cover scores of 4 or higher may reduce the likelihood of breech flystrike occurrence in winter rainfall environments; but this study does not suggest a clear benefit is gained by further reducing flock breech cover scores, in terms of the likelihood of breech flystrike.

Potential limitations of study findings primarily relate to the use of pre-existing data to answer research questions that were not the specific goals of the data collections. There are many practical advantages to utilising existing data to answer additional research questions. However, as the data collection was not designed with these analyses in mind, some consequential insufficiencies in the availability of certain types of data may have influenced study findings.

1.5 Conclusions

This study finds that:

- In winter rainfall environments where dag is a management problem, management interventions that successfully reduce dag will have the greatest benefit in reducing likelihood of breech flystrike. Interventions aimed at reducing breech wrinkle scores will also yield protective benefit. Reducing breech cover scores is only of benefit if scores are excessive (4 or higher).
- In summer rainfall environments, management interventions that successfully reduce dag scores, breech wrinkle scores and breech cover scores will yield protective benefit against breech flystrike.
- Quantitative measures of the benefits to be gained are expected to be of value to economically optimising production management strategies and prioritising allocation of research funding in developing intervention strategies to reduce the occurrence of disease.

Introduction

The identification and application of strategies to minimise ovine breech flystrike in the absence of mulesing is important for the future of the Australian wool industry.

The occurrence of ovine breech flystrike is a classic example of a disease presentation with multifactorial causation. Elucidating the causal web of ovine breech flystrike (including interactions between various risk factors) was identified as a priority to inform further disease research strategy. With this in mind, a preliminary causal web was developed for ovine breech flystrike in Australia, in consultation with experts in parasitology, molecular biology, animal production and genetics (reported in Hillman and Madin, 2018). At this time, it was identified that further studies taking an explanatory approach to statistical modelling had the potential to clarify the causal role of various risk factors in ovine breech flystrike occurrence. Identifying causal risk factors would aid in identification and prioritisation of intervention strategies to yield the greatest economic benefit in reducing flystrike incidence across different farming systems in Australia (Hillman and Madin, 2018).

Explanatory modelling aims to estimate the causal effect of a particular risk factor of the occurrence of a given outcome. A key aspect of explanatory modelling is considering *a priori* the putative causal framework for the association between the risk factor and the outcome, and systematically identifying relevant covariates to control for confounding in a multivariable model. Inappropriate inclusion of covariates in a multivariable model can induce bias when measuring causal effects (Greenland et al., 1999; Shrier & Platt, 2008; Martin, 2014; Textor et al., 2016; Bello et al., 2018).

Australian Wool Innovation and its partners Department of Primary Industries and Resource Development (Western Australia) (DPIRD) and Commonwealth Scientific and Industrial Research Organisation (CSIRO) have undertaken extensive research on ovine breech flystrike in Merino sheep in Australia, particularly focusing on genetics and breeding strategies to reduce the incidence of ovine breech flystrike. Data accumulated through these projects is summarised in Hillman, Sadler and Madin (2019).

This study aimed to utilise these data to clarify aspects of ovine breech flystrike causality in Australian winter rainfall and summer rainfall farming conditions. Specifically, the project objectives were to quantitatively measure:

- the association between ovine breech flystrike and dags,
- the association between ovine breech flystrike and breech wrinkle, and
- the association between ovine breech flystrike and breech cover, specifically considering whether the effect breech cover on the occurrence of breech flystrike varies depending on the degree of breech wrinkle or dags.

Methods

This study took an explanatory approach to estimate the putative causal effects of dags, breech wrinkle and breech cover on breech flystrike. This incorporates the use of directed acyclic graphs to identify an appropriate set or sets of covariates on which to condition a multivariable model, to appropriately control for confounding whilst estimating the causal effect of a particular risk factor on breech flystrike. Systematic assessment of appropriate adjustment sets avoids inappropriate inclusion of covariates in a multivariable model, which may induce bias in estimates of causal effect (Greenland et al., 1999; Shrier and Platt, 2008).

Directed acyclic graphs and model adjustment set identification

The causal web developed for known and hypothesised risk factors for ovine breech flystrike in Australia (Hillman and Madin, 2018) was used to construct directed acyclic graphs that specifically considered each study objective, using the software DAGitty v2.3 (Textor et al., 2016) (

Figure 1,
Figure 2,
Figure 3). Ancestor variables that were not on potentially biasing pathways between the exposure-of-interest (dags, breech wrinkle or breech cover) and the outcome (breech flystrike) were generally not included in these graphs.

Application of directed acyclic graphs to investigation of causal effects:

- The causal, biasing and non-biasing paths identified in a DAG are relevant to selecting appropriate variables to control for confounding bias in a multivariable model, when estimating the magnitude of the causal effect of the exposure-of-interest (e.g. breech wrinkle) on the risk of breech flystrike.
- When assessing the association between a putative risk factor (e.g. breech wrinkle) and an outcome (e.g. breech flystrike), biasing paths that may influence statistical findings can blocked (statistically removed) by appropriately controlling for a variable along these paths.
    - Where multiple biasing paths exist, this may require blocking on multiple variables, depending on the structure of the causal and biasing paths.
- Software programs can analyse an entire network of biasing and causal pathways in a complex directed acyclic graph to systematically identify a set or multiple sets of appropriate variable(s) to include in a model, to block biasing pathways when assessing a relationship between a putative risk factor and an outcome (e.g. DAGitty, as used in this study).

When estimating causal effects, inappropriately including a variable in a model when it is not required can induce bias, and so is to be avoided by taking this systematic approach to covariate inclusion.

Figure legend for the directed acyclic graph nodes

exposure-of-interest (dags, breech wrinkle or breech cover)

outcome (breech flystrike)

ancestors of the exposure

ancestor variables of the outcome

	ancestor variables of both the exposure and outcome
	variables for which there were no data amongst the provided data set (or able to be obtained through external data sources)
Figure legend for the directed acyclic graph edges between the nodes	
	causal paths from the exposure-of-interest (dags, breech wrinkle or breech cover) to breech flystrike
	potentially biasing paths when considering the association between the exposure-of-interest (dags, breech wrinkle or breech cover) and breech flystrike
	non-biasing paths.

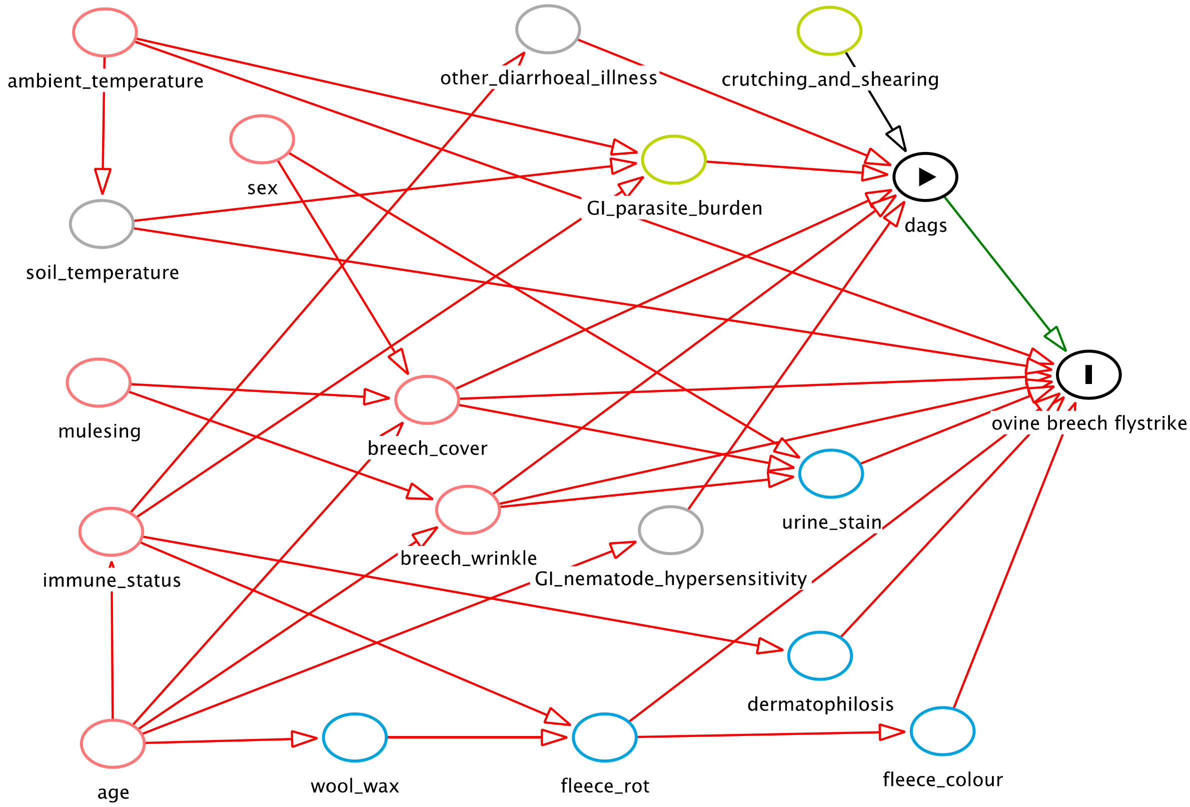


Figure 1 Directed acyclic graph considering the putative association between dags and breech flystrike in sheep.

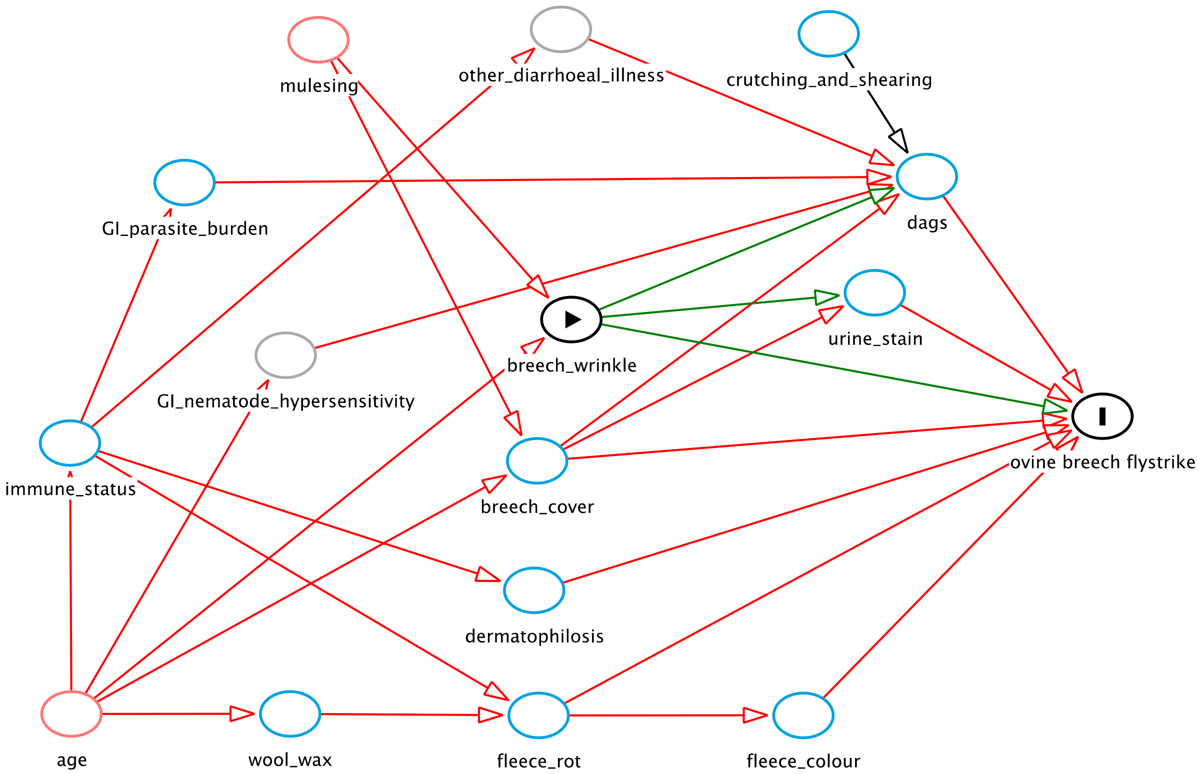


Figure 2 Directed acyclic graph considering the putative association between breech wrinkle and breech flystrike in sheep.

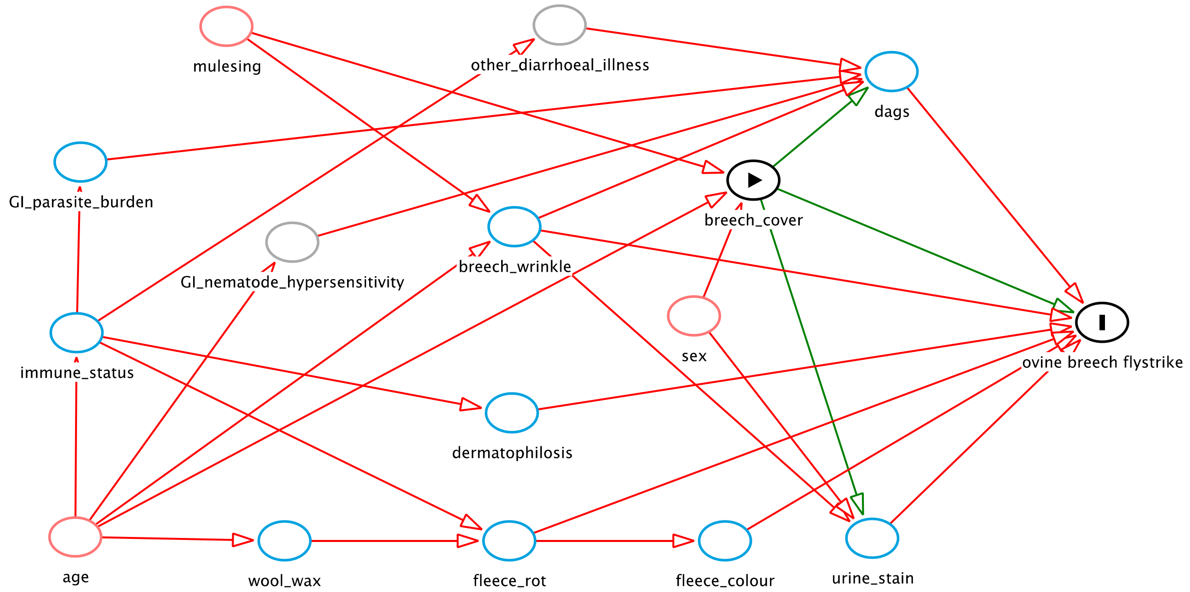


Figure 3 Directed acyclic graph considering the putative association between breech cover and breech flystrike in sheep.

Adjustment sets for multivariable analysis

The directed acyclic graphs were systematically analysed using the software DAGitty to identify sets of covariates required for inclusion in a multivariable model, to obtain an unbiased measure of the total causal effect of the relevant exposure on ovine breech flystrike (Table 1).

Table 1 Minimal sufficient adjustment set options for examining the association of the respective exposure to ovine breech flystrike, as per DAGitty output

Exposure-of-interest	Minimal sufficient adjustment set options to measure total effect
Dag	• gastrointestinal parasite burden, breech cover, breech wrinkle, immune status and age; or • gastrointestinal parasite burden, breech cover, breech wrinkle, immune status, fleece rot and sex; or • gastrointestinal parasite burden, breech cover, breech wrinkle, immune status, fleece rot and urine stain; or • gastrointestinal parasite burden, breech cover, breech wrinkle, immune status, urine stain and wool wax; or • gastrointestinal parasite burden, breech cover, breech wrinkle, immune status, sex and wool wax
Breech wrinkle	• age, breech cover and sex; or • age and mulesing
Breech cover	• age, breech wrinkle and sex; or • age, mulesing and sex

There are a few considerations regarding the application of this DAGitty output to the AWI breech flystrike data:

- 'immune status' was proxied by body condition score in these analyses;
- gastrointestinal parasite burden data were available regarding faecal egg count for strongyles, and faecal egg counts specifically for *Nematodirus* sp.; and
- for the winter rainfall environment data, all sheep entered the study from approximately the same age (marking); and mulesed animals were excluded from the breech cover and breech wrinkle analyses, as there were too few for a balanced data set. Age and mulesing were therefore controlled by design, with no requirement for adjustment as part of data analysis.

  - Mulesed animals were not excluded from the dag analyses, as mulesing status was not a potential confounder of the relationship between dags and breech flystrike.

In regard to the application of this DAGitty output to the CSIRO (summer rainfall environment) data:

- there was no data or insufficient data available regarding mulesing, gastrointestinal parasite burden, body condition score (to proxy immune status, urine stain and wool wax: the relevant models (Table 1) could not be adjusted for these variables.

Statistical analysis
Three datasets were provided to Ausvet by Australian Wool Innovation and its partners DPIRD and CSIRO. These were:

- Excel spreadsheet 'BSRworkfile 2005to2014 from birth to before post weaning for AWI.xlsx' ('DPIRD 2005–2014 data') from DPIRD;
- Excel spreadsheet 'BSRworkfile 2006to2014_including 2012 animals available in 2014 for AWI.xlsx' ('DPIRD 2006–2014 data') from DPIRD; and
- A MySQL dump of the CSIRO Armidale Breech Strike Flock database ('CSIRO data').

Analyses were undertaken separately for sheep in winter rainfall environments (DPIRD data) and summer rainfall environments (CSIRO data), given the differences in environmental and management factors of these flocks and the potential implications on validity of results, were these data sets to be combined (Hillman, Sadler and Madin, 2019).

Quality control of data from a winter rainfall environment
The two spreadsheets provided by DPIRD were uploaded into R v3.5.1 (R Core Team, 2018) for description and analysis.

Data cleaning was undertaken in keeping with standard approaches to data quality control (Rindler, et al. 2013). Corrections for data validity (e.g., normalise different reference dates for date fields in the provided excel spreadsheets; type conversion; and correcting errors in allowable values such as retyping "m" or "male" as "MALE" for the 'sex' variable) were made following extensive data profiling of available data fields.

More specifically, covariates (breech wrinkle score, dag score, breech cover score, fleece rot score, urine stain score, immune status (proxied by body condition score) and gastrointestinal parasite burden (represented by faecal egg counts for strongyles and *Nematodirus* sp.) were proxied to the timepoints of measurement of the presence or absence of breech strike. This was undertaken by matching in the nearest covariate measurement to the breech strike measurement.

Presumed data entry errors were corrected to the nearest half score (e.g. 1.53 corrected to 1.5); or if the error was exceptionally large, to the nearest plausible value (e.g. breech wrinkle score of 15 corrected to 1.5; presumed error in decimal placement). All score data were grouped into ordered categorical variables to improve the balance of the data, as very few sheep had some scores (groups were score 1 – 1.5, score 2 – 2.5, score 3 – 3.5 and score 4 – 5).

Gastrointestinal parasite burden was represented by faecal egg counts for strongyles (modelled as a scaled numeric variable) and *Nematodirus* sp. (modelled as a binary variable).

Sheep with missing data regarding important dates relevant to the study (e.g. time of marking), exposures, the outcome or required covariates were excluded from the analyses. Further, mulesed animals were excluded from the breech cover and breech wrinkle analyses, as there were too few mulesed animals to allow for a balanced data set when including this as a covariate.

Quality control of data from a summer rainfall environment
Data were extracted from the CSIRO database using SQL and R was used to describe and analyse these data, as per Hillman, Sadler and Madin (2019). Similar approaches to data cleaning were adopted as per the winter rainfall data. In general, the validity of the data received was greater for the summer rainfall data than for the

winter rainfall data. However, data sparsity (i.e., lack of completeness) for variables such as faecal egg count, condition score, urine stain, gastrointestinal parasite burden and wool wax were greater for the summer rainfall data, and hence these were excluded from the relevant analyses (Table 1). Variables with reasonable temporal coverage (e.g., at least 10,000 time-point records for 5,500 sheep) still had widespread missingness (i.e., they may be complete for some covariates, but missing for other covariates, at any point in time). The missing values in these instances were interpolated from the nearest non-null valued record in time.

Age was approximated for the summer rainfall data set, using a spring lambing assumption (1 September) and the year of birth; age was then split into quartiles to produced ordered categories of age. As for the winter rainfall data, score data were grouped into ordered categorical variables where needed to improve the balance of the data: dag and breech wrinkle were grouped into score 1, score 2, score 3 and score 4–5; breech cover was grouped into score 1– 2, score 3, score 4 and score 5; and fleece rot was grouped into score 0, score 1, score 2, score 3 and scores 4 and above. As previously mentioned, given the lack of available data, the variables relating to mulesing, gastrointestinal parasite burden, immune status (body condition score proxy), urine stain and wool wax could not be used where required in the respective models (Table 1).

Analysis

Data were analysed in Andersen and Gill survival models, using R v3.5.3 and the package survival (Therneau, 2019). These models enable consideration of covariates that change over time (for example, breech traits and dag) in the analytical approach.

Analyses considered the time to first occurrence of breech flystrike: this was considered the most relevant outcome for this research to consider, given the nature of breech flystrike in a farm production setting and practical application of the findings of these analyses. Thus, any data from an animal subsequent to the first occurrence of breech flystrike was discarded.

The proportional hazards assumption was tested using the cox.zph function of the survival package (Therneau, 2019). As the data set was large, validation of the models focussed on assessing the plots of the scaled Schoenfeld residuals against transformed survival time — particularly considering the plots relating to the exposures of interest. The plots for the exposure-of-interest from the most supported model for each hypothesis (as measured by the information theoretic modelling, discussed below in Section 0) are presented in Appendix A. Where a substantial departure of the smoothed curve fitted to the plot deviated from a regression coefficient of 0, the model was stratified by problematic covariates, or fitted with interaction terms between problematic covariates and time.

Statistical hypothesis tests of a non-zero slope were also examined, but given lesser weight in interpretation, given the large size of the data set and thus relatively high risk of Type I error (In and Lee, 2019). In view of the relative subjectivity required in this approach to validation of the proportional hazards assumption, robust standard errors were used in the models, as an extra step to support the validity of model outputs (Lin and Wei, 1989).

If more than one model was considered valid for each hypothesis, model results were averaged by an information theoretic approach (Burnham and Anderson, 2002), using the MuMIn package in R (Barton, 2019).

For the breech cover models, the anova.coxph function in the survival package (Therneau, 2019) was used to test for interaction. This compared the most supported model for the association between breech cover and breech flystrike, to that model with an interaction term for breech wrinkle or dag (respectively). The most supported breech cover model did not contain dag as a covariate—as per the directed acyclic graph, it was not required (Table 1). However, to enable a valid test for interaction with dag score, dag was added to this model when comparing it to the model considering interaction by dag score.

Results

Descriptive findings

A detailed descriptive analysis of the DPIRD data was reported previously (Hillman, Sadler and Madin, 2019).

Descriptive findings: winter rainfall environment data (DPIRD)

The winter rainfall environment data comprised records from 11,212 sheep over the marking to hogget shearing period. After excluding incomplete data, the dag analyses included 6103 sheep, of which 1012 (16.6%) were struck on the breech at least once across the duration of the study. 3128 were ewes (51.3%), and 2975 were rams (48.7%); only 121 sheep (1.98%) were mulesed at marking.

Breech wrinkle, dag and breech cover scores varied for individuals across the different time points at which they were measured; and there was variation between when sheep were measured for particular traits. Considering the highest dag score recorded per sheep across the duration of that sheep's participation in the study, these highest scores were relatively evenly spread across the score categories. Meanwhile, the highest breech wrinkle score recorded for a sheep across the duration of the study was most commonly in the range of 1–1.5, and the highest breech cover score recorded was most commonly in the range of 3 to 5 (Table 2).

The median minimum faecal egg count (strongyles) recording per sheep was 100 (IQR 0 – 300, range 0 – 4950), and the median maximum faecal egg count (strongyle) per sheep was 400 (IQR 150 – 750, range 0 – 5350). At one point or more across the duration of the study, 1940 (31.8%) sheep had a positive *Nematodirus* sp. faecal egg count.

Note that in Table 2, the highest scores recorded per sheep are presented; variables such as breech cover varied during the course of the study, and the models are constructed to consider this variation in assessing the relationship with breech flystrike.

Data available for the breech wrinkle and breech cover analyses were similar to those available for the dag analyses, so for brevity are not specifically summarised here. There was some variation between models for each exposure-of-interest (dag, breech wrinkle or breech cover), because of missing data associated with covariates required to condition the model on, which varied between models for the different exposure type (Table 1).

Table 2 Highest scores recorded per sheep across the duration of the winter rainfall study, in the data available for the dag analyses

	Categories of score:				
	Score 1 – 1.5	Score 2 – 2.5	Score 3 – 3.5	Score 4 – 5	Total
Dag: No. sheep (%)	1111 (18.2%)	1572 (25.8%)	1805 (29.6%)	1615 (26.5%)	6103 (100%)
Breech wrinkle: No. sheep (%)	4934 (80.9%)	828 (13.6%)	280 (4.60%)	61 (1.00%)	6103 (100%)
Breech cover: No. sheep (%)	1 (0.02%)	876 (14.4%)	3145 (51.5%)	2081 (34.1%)	6103 (100%)
Urine stain: No. sheep (%)	5086 (83.3%)	786 (12.9%)	197 (3.23%)	34 (0.56%)	6103 (100%)
Fleece rot: No. sheep (%)	5583 (91.5%)	351 (5.75%)	101 (1.65%)	68 (1.11%)	6103 (100%)
Wool wax: No. sheep (%)	132 (2.16%)	3774 (61.8%)	1893 (31.0%)	304 (4.98%)	6103 (100%)
Body condition score† No. sheep (%)	9 (0.15%)	562 (9.21%)	3934 (64.5%)	1598 (26.2%)	6103 (100%)

†A proxy for immune status

Descriptive findings: summer rainfall environment data (CSIRO)
The summer rainfall environment data initially available for the breech strike analyses comprised observations from 4,805 sheep.

Data used for the breech wrinkle and breech cover analyses
After excluding incomplete data, records from 4662 sheep were available for the breech wrinkle analyses. 753 (16.2%) sheep experienced breech flystrike at least once. 2081 (44.6%) were wethers/rams and 2541 (54.5%) were ewes.

Breech wrinkle and breech cover scores (Table 3) and dag scores (**Error! Reference source not found.**) varied for individuals across the different time points at which they were measured; and there was variation between when sheep were measured for particular traits.

Table 3 Highest scores recorded per sheep across the duration of the summer rainfall study, in the data available for the breech wrinkle analyses

	Categories of score:					
	Score 1	Score 2	Score 3	Score 4	Score 5	Total
Breech wrinkle:	171	1304	1795	1031	321	4622
No. sheep (%)	(3.70%)	(28.2%)	(38.8%)	(22.3%)	(6.94%)	(100%)
Breech cover:	1	9	212	1507	2893	4622
No. sheep (%)	(0.02%)	(0.19%)	(4.59%)	(32.6%)	(62.6%)	(100%)

Considering the highest breech wrinkle score recorded per sheep across the duration of that sheep's participation in the study, the highest breech wrinkle score per sheep was most commonly in score 2 or 3. The highest breech cover score recorded was most commonly score 4 or 5 (Table 3). [Note that as for the winter rainfall environment data, Table 3 presents the highest scores recorded per sheep: variables such as breech cover varied during the course of the study, and the models are constructed to consider this variation in assessing the relationship with breech flystrike].

Data available for the breech cover analyses were similar in quantity and distribution to those available for the breech wrinkle analyses, and for brevity are not summarised here.

Data used for the dag analyses
After excluding incomplete data, records from 3629 sheep were available for the dag analyses. 687 (18.9%) sheep experienced breech flystrike at least once. 2041 (56.2%) were wethers/rams and 1588 (43.8%) were ewes.

Dag scores varied for individuals across the different time points at which they were measured; and there was variation between the times at which sheep were measured for particular traits. Considering the highest dag score recorded per sheep across the duration of that sheep's participation in the study, the highest dag score per sheep was most commonly score 1 or 2. The highest breech wrinkle score recorded per sheep was most commonly score 2 or 3; meanwhile, the highest breech cover score recorded per sheep was most commonly score 4 or 5 (Table 3). [Note that as for the winter rainfall environment data, Table 3 presents the highest scores recorded per sheep: variables such as breech cover varied during the course of the study, and the models are constructed to consider this variation in assessing the relationship with breech flystrike].

Table 4 Highest scores recorded per sheep across the duration of the summer rainfall study, in the data available for the dag analyses

	Categories of score:					
	Score 1	Score 2	Score 3	Score 4	Score 5	Total
Dag No. sheep (%)	1600 (44.1%)	1394 (38.4%)	425 (11.7%)	161 (4.4%)	49 (1.4%)	3629 (100%)
Breech wrinkle: No. sheep (%)	96 (2.65%)	1018 (28.1%)	1428 (39.3%)	811 (22.3%)	276 (7.6%)	3629 (100%)
Breech cover: No. sheep (%)	0 (0 %)	3 (0.1%)	132 (3.6%)	1110 (30.6%)	2384 (65.7%)	3629 (100%)

3629 sheep had scores available for fleece rot. 2997 (82.6%) sheep did not score higher than 0; 569 sheep (15.7%) had a maximum score of 1 or 2; 45 sheep (1.2%) had a maximum score of 3; and 18 sheep (0.5%) had a maximum score of 4 or higher.

Other variables relevant to the analysis (e.g. urine stain, body condition score and wool wax, Table 1), were too incomplete to consider as part of the analytical approach (Section 0), and so are not summarised here.

Explanatory modelling

Andersen and Gill models provide quantitative output in the form of a hazard ratio. The hazard ratio compares breech flystrike occurrence between sheep with different dag/ breech wrinkle/ breech cover scores across the duration of the study. For example, in Table 5, breech flystrike occurrence is compared between sheep with dag scores greater than 1.5, to the baseline group of sheep with dag scores of 1 – 1.5. Results are highly unlikely to have occurred by chance if the 95% confidence interval (CI) does not include 1, and the p-value is less than 0.05. The hazard ratio for sheep of dag score 2 to 2.5 was 2.32: this indicates that any point of the duration of the study, sheep with dag score 2 to 2.5 were 2.32 times more likely (95% CI 1.92–2.79) to be struck on the breech for the first time than sheep with dag score 1 to 1.5.

Dags as a risk factor in winter rainfall environments

For the period of marking to hogget shearing in sheep in a winter rainfall environment, dag score was associated with the occurrence of breech flystrike (Table 5). The likelihood of a sheep experiencing breech flystrike increased with dag score, up to 28.5 times (95% CI 23.3 – 35.0) in sheep with a dag score of 4 – 5, compared to sheep with a dag score of 1 – 1.5.

Table 5 Total effect of breech flystrike by dag score, in sheep in a winter rainfall environment from birth to hogget shearing

	Hazard ratio	95% CI	p-value
Dag score 1 – 1.5	1	—	—
Dag score 2 – 2.5	2.32	1.92 – 2.79	<0.001
Dag score 3 – 3.5	4.69	3.89 – 5.65	<0.001
Dag score 4 – 5	28.5	23.3 – 35.0	<0.001

Dags as a risk factor in summer rainfall environments

Dag score was associated with the occurrence of breech flystrike in summer rainfall environments (Table 6). Compared to sheep with a dag score of 1, the likelihood of a sheep experiencing breech flystrike was three times greater in sheep with a dag score of 2, over nine times greater in a sheep with a dag score of 3, and nineteen times greater in sheep with a dag score of 4 or higher.

Table 6 Total effect of dag score on breech flystrike, in sheep in a summer rainfall environment

	Hazard ratio	95% CI	p-value
Dag score 1	1	—	—
Dag score 2	3.08	2.53 – 3.75	<0.001
Dag score 3	9.55	7.66 – 11.9	<0.001
Dag score 4 – 5	19.1	14.7 – 24.7	<0.001

Breech wrinkle as a risk factor in winter rainfall environments

For the period of marking to hogget shearing in sheep in a winter rainfall environment, the degree of breech wrinkle was associated with the occurrence of breech flystrike (Table 5). Sheep with breech wrinkle scores of 2 – 2.5 had over two times the likelihood of breech flystrike occurrence throughout the duration of this study (95% CI 1.62 – 2.64), compared to sheep of breech wrinkle scores 1 – 1.5. Meanwhile, results for breech wrinkle scores 2 – 2.5 and 3 – 3.5 are similar, suggesting that reducing wrinkle from 3.5 to 2 may not provide any protective benefit against breech flystrike in sheep winter rainfall environments.

Table 7 Total effect of breech wrinkle score on breech flystrike, in sheep in a winter rainfall environment from birth to hogget shearing

	Hazard ratio	95% CI	p-value
Breech wrinkle score 1 – 1.5	1	—	—
Breech wrinkle score 2 – 2.5	2.14	1.69 – 2.71	<0.001
Breech wrinkle score 3 – 3.5	2.32	1.51 – 3.57	<0.001
Breech wrinkle score 4 – 5	1.60	0.51 – 5.05	0.42

Breech wrinkle as a risk factor in summer rainfall environments

Breech wrinkle was very strongly associated with breech flystrike occurrence in summer rainfall environments (p<0.001), with the likelihood of breech flystrike occurrence increasing ordinally with breech wrinkle score (Table 8).

Table 8 Total effect of breech wrinkle score on breech flystrike, in sheep in a summer rainfall environment

	Hazard ratio	95% CI	p-value
Breech wrinkle score 1	1	—	—
Breech wrinkle score 2	2.08	1.54 – 2.81	<0.001
Breech wrinkle score 3	3.13	2.33 – 4.21	<0.001
Breech wrinkle score 4	4.87	3.59 – 6.62	<0.001
Breech wrinkle score 5	8.00	5.62 – 11.4	<0.001

Breech cover as a risk factor in winter rainfall environments

For the period of marking to hogget shearing in sheep in a winter rainfall environment, the causal effect of breech cover on the occurrence of breech flystrike was very limited (Table 9); and these very limited effects of breech cover on breech strike varied by dag score (test for statistical interaction p<0.001). There was good evidence (p<0.05) that, amongst sheep with dag score 1 – 1.5, a breech cover score of 4 – 5 was associated with 4.2 times the likelihood of breech flystrike during the study.
There was no variation in effect of breech cover on breech strike occurrence by degree of breech wrinkle (p=0.31).

Table 9 **Total effect of breech cover on breech flystrike, in sheep in a winter rainfall environment from birth to hogget shearing**

		Hazard ratio	95% CI	p-value
Dag score 1 – 1.5	Breech cover score 1 – 1.5	1	—	—
	Breech cover score 2 – 2.5	0.71	0.22 – 2.25	0.56
	Breech cover score 3 – 3.5	1.42	0.46 – 4.42	0.55
	Breech cover score 4 – 5	4.20	1.34 – 13.2	0.01
Dag score 2 – 2.5	Breech cover score 1 – 1.5	1	—	—
	Breech cover score 2 – 2.5	1.12	0.24 – 5.25	0.89
	Breech cover score 3 – 3.5	1.04	0.23 – 4.69	0.96
	Breech cover score 4 – 5	0.79	0.17 – 3.61	0.76
Dag score 3 – 3.5	Breech cover score 1 – 1.5	1	—	—
	Breech cover score 2 – 2.5	1.75	0.51 – 6.01	0.37
	Breech cover score 3 – 3.5	1.00	0.30 – 3.39	1.00
	Breech cover score 4 – 5	0.72	0.20 – 2.51	0.60
Dag score 4 – 5	Breech cover score 1 – 1.5	1	—	—
	Breech cover score 2 – 2.5	3.18	0.81 – 12.5	0.10
	Breech cover score 3 – 3.5	1.32	0.34 – 5.14	0.69
	Breech cover score 4 – 5	0.44	0.09 – 2.15	0.31

Breech cover as a risk factor in summer rainfall environments
In summer rainfall environments, sheep with breech cover scores of 3 had over three times the likelihood of breech flystrike occurrence (95% CI 1.08 – 8.45), compared to sheep with breech cover scores of 1 – 2; likelihood of breech flystrike occurrence increased further for breech cover scores 4 and 5 (
Table 10).
Table 10 Total effect of breech cover on breech flystrike, in sheep in a summer rainfall environment There was no evidence that the effect of breech cover on the likelihood of breech flystrike varied by dag score (p= 0.29) or by breech wrinkle score (p=0.20).

Table 10 **Total effect of breech cover on breech flystrike, in sheep in a summer rainfall environment**

	Hazard ratio	95% CI	p-value
Breech cover score 1 – 2	1	—	—
Breech cover score 3	3.02	1.08 – 8.45	0.036
Breech cover score 4	4.08	1.47 – 11.3	0.007
Breech cover score 5	6.78	2.45 – 18.8	<0.001

Discussion

The quantitative results of this modelling are expected to be of practical value to wool producers in economically optimising production strategies against the risk of breech flystrike. Results are also expected to be of value to the wool industry, in prioritising allocation of resources for research on intervention strategies that are likely to have the greatest impact in reducing risk of breech flystrike.

Association between dags and breech flystrike
The association between dags score and the likelihood of breech flystrike increased ordinally in both the winter rainfall environment and in the summer rainfall environment, suggesting that all successful efforts to reduce dag score entail benefit in reduced likelihood of the disease.

Association between breech wrinkle and breech flystrike
Similar to the dag analyses, the association between breech wrinkle score and breech flystrike increased ordinally in the summer rainfall environment, suggesting that all successful efforts to reduce breech wrinkle score will entail benefit in reduced likelihood of the disease.

However, the winter rainfall environment analyses found that while the likelihood of breech flystrike was relatively increased in sheep with breech wrinkle scores greater than 1.5, the magnitude of increased likelihood was similar across breech wrinkle score 2 to 3.5. This suggests that reducing breech wrinkle scores within this range (from (say) 3.5 to 2.5) may not be beneficial in terms of reducing the likelihood of breech flystrike.

While there was insufficient evidence of a relatively increased likelihood of breech flystrike amongst sheep of breech wrinkle score 4 – 5 in the winter rainfall environment data, this is considered likely to be attributable to lack of statistical power in this category, given that relatively few sheep had such extreme breech wrinkle scores in the provided data set.

Type I error: is finding statistical evidence against the null hypothesis (e.g. $p<0.10$), when the null hypothesis is actually true. For example, when comparing the likelihood of disease between two groups, concluding that the likelihood is different when actually it is not. This is more likely to occur in studies with very large sample sizes.

Type II error: is not finding statistical evidence against the null hypothesis (e.g. $p>0.10$), when the null hypothesis is actually false. For example, when comparing the likelihood of disease between two groups, concluding that the likelihood is not different when actually it is. This is more likely to occur in studies with small overall sample sizes, or where the size of one or more of the groups being compared is very low.

Statistical power: Statistical power has an inverse relationship with Type II error: a statistical comparison of low power has relatively high likelihood of Type II error.

Association between breech cover and breech flystrike

Results suggest that in winter rainfall environments, the total effect of breech cover on the occurrence of breech flystrike was very limited, and the effect did not vary between animals of different breech wrinkle scores. However, there was evidence that the effect of breech cover on breech flystrike varied by dag score, with the likelihood of breech flystrike increasing amongst sheep with breech cover scores 4 or higher, but only amongst sheep with a low dag score. As there is not a clear, biologically plausible rationale for these findings, they should be interpreted cautiously and subject to further investigation: it possible that this association is spurious (Type I error), given the relatively large sample size used in these analyses. This isolated breech cover association contrasts to the ordinal relationships identified for breech wrinkle and dag (where the likelihood of breech flystrike increased with the degree of the risk factor), which provide epidemiologically sound evidence of a true causal association. Overall, it can be concluded that breech cover is not associated with the likelihood of breech flystrike occurrence at scores 3.5 or below in winter rainfall environments; further investigation of whether the likelihood of flystrike occurrence truly increases when breech cover scores are higher, with or without interaction by dag, is warranted.

In contrast, study results suggest that in summer rainfall environments, breech cover was associated with breech flystrike occurrence. Again, the ordinal relationship displayed in this association provides epidemiologically sound evidence of a true causal relationship and suggests that successful efforts to reduce breech cover score will entail benefit in the reduced likelihood of disease occurrence. In summer rainfall environments, the association between breech cover and breech flystrike did not vary depending on dag score or breech wrinkle score.

Measurement of direct and indirect effects

This study measured the total effects of dags, breech wrinkle and breech cover on breech flystrike occurrence, given the practical applications of these results are particularly relevant to producers.
More specifically, measurement of indirect and direct effects is possible.

- Measuring the total effect of an exposure variable (e.g. breech wrinkle) on the outcome (breech strike) includes measurement of direct effects and indirect effects (which are effects mediated through other variables). For example, breech wrinkle is hypothesised to have a direct effect on breech strike (e.g. by creating a protective environment favourable for larval development) and indirect effects mediated through urine stain and dags (

- Figure 2).
- Measuring the direct effect of an exposure variable (e.g. breech wrinkle) on the outcome (e.g. breech strike) only measures the effect that is not mediated by other variables.
    - In consideration of the effect of dags on breech strike, it is hypothesised that dags only influence the risk of flystrike directly (
    - Figure 2) and so the direct effects are also the total effects in this case.
- Measuring indirect effects only involves measuring the effects of the exposure variable that are mediated through other variables.

However, further analyses to measure direct and indirect effects are not recommended at this stage, as practical applications of such results have not been identified.

Specific commentary on this study, including limitations
The validity of these model results depends on the validity of the underlying causal assumptions, as represented in the directed acyclic graphs (Section 0). These graphs were constructed based on a causal web created by a group of experts on ovine breech flystrike (reported in Hillman and Madin, 2018).

There are many practical advantages to utilising existing data to answer research questions. However, given that these data have not been collected with these specific research questions in mind, the potential limitations must also be considered. For example, data on covariates required to address some aspects of this study's objectives were either not available or very limited, particularly from the summer rainfall environment. Where these covariates were not able to be included in the model, the model findings are at risk of confounding bias. In this study, the model most at risk of confounding (in terms of missing covariate data) was the dag model from the summer rainfall environment. However, for this model, analyses were repeated to include the limited condition score, urine stain and fleece rot data (thus with the model restricted to relatively few animals, with all bar one of them being ewes) and the results were similar (results not shown).

Further to the above, where data for certain covariates were collected (e.g. breech wrinkle scores and faecal egg counts), the need to infer values at the time of strike occurrence may entail inaccuracy, which may result in residual confounding influencing results. This is more likely to have affected the summer rainfall environment results than the winter rainfall results.

The data used in these analyses originate from two research stations, and so may not represent the variability amongst the broader sheep farming populations for the respective environments, which may affect the relevance of the results presented herein. However, it must also be noted that these data were collected over a number of years, in the same local environment and under consistent approaches to flock management, which limits the likelihood of several key potential sources of bias in investigating these hypotheses across many sites (e.g. differences in management strategies and environmental conditions between farms in a multi-farm study). There was some variation in approaches to management within these research stations—for example, three years of delayed crutching at the DPIRD station. However, given that crutching was on the causal pathway of the relationship between dags and breech flystrike (

Figure 1), and not confounding the relationship between breech wrinkle/breech cover and breech flystrike (
Figure 2,
Figure 3), this didn't require adjusting for in the explanatory models.

This data analysis considered the relative likelihood of the first occurrence of breech strike between animals of varying traits. This decision was made in view of practical applications of the results. Additionally, there were some concerns about the applicability of the repeat-event data to a farm setting— for example, where another strike is recorded several days after treatment of a strike that was noted very early on (due to research flock monitoring practices that are unlikely to reflect farming practices), this may have constituted the same strike event in a farming situation. Additionally, the role of a strike wound attracting more flies would need to be considered in relation to the analyses. Further research may be aimed at exploring the relationships with repeat strike events.

Finally, as data in this study originated from Merino sheep, the results may not apply to other breeds of sheep.

Conclusions

This study finds that:

- In winter rainfall environments where dag is a management problem, management interventions that successfully reduce dag will have the greatest benefit in reducing likelihood of breech flystrike. Interventions aimed at reducing breech wrinkle scores will also yield protective benefit. Reducing breech cover scores is only of benefit if scores are excessive (4 or higher).
- In summer rainfall environments, management interventions that successfully reduce dag scores, breech wrinkle scores and breech cover scores will yield protective benefit against breech flystrike.
- Quantitative measures of the benefits to be gained are expected to be of value to economically optimising production management strategies and prioritising allocation of research funding in developing intervention strategies to reduce the occurrence of disease.

References

Barton K, 2019. *MuMIn: Multi-Model Inference*. R package version 1.43.15. Available online: https://cran.r-project.org/web/packages/MuMIn/index.html

Bello NM, Ferreira VC, Gianola D, Rosa GJ, 2018. Conceptual framework for investigating causal effects from observational data in livestock. *Journal of Animal Science*, 96, 4045-62.

Burnham KP, Anderson DR, 2002. Model selection and multimodel inference: a practical information-theoretic approach. Springer-Verlag, New York.

Greenland S, Pearl J, Robins JM, 1999. Causal diagrams for epidemiologic research. *Epidemiology*, 10, 37-48.

Hillman A, Madin B, 2018. *Understanding risk factors for ovine breech flystrike (ON–00510)*. Report to Australian Wool Innovation, Sydney.

Hillman A, Sadler R, Madin B, 2019. Ovine breech flystrike in Australia: exploratory data analysis of two research flocks (ON–00555). Report to Australian Wool Innovation, Sydney.

In J, Lee, DK, 2019. Survival Analysis: part II—applied clinical data analysis. *Korean Journal of Anaesthesiology*, 72, 441 – 457.

Lin DY, Wei LJ, 1989. The robust inference for the cox proportional hazards model. *Journal of the American Statistical Association*, 84, 1074 – 1078.

Martin W, 2014. Making valid causal inferences from observational data. *Preventative Veterinary Medicine,* 113, 281-297.

R Core Team, 2019. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rindler A, McClowry S, Hillard R, Mueller S, 2013. Information development using mike2.0. BookBaby.
Shrier I, Platt RW, 2008. Reducing bias through directed acyclic graphs. *BMC medical research* methodology, 8, 70.

Textor J, van der Zander B, Gilthorpe MK, Liskiewicz M, Ellison GTH, 2016. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. *International Journal of Epidemiology*, 45(6), pp.1887–1894

Therneau TM, 2019. *Survival analysis*. R package version 3.1–8. Available online: https://cran.r-project.org/web/packages/survival/survival.pdf

Appendix A Testing the proportional hazards assumption

Dag score models

For the winter rainfall environment model, the smooth curve of the plot of the scaled Schoenfeld residuals against transformed survival time for dag score deviates minorly from a straight line (Figure 4), but is not considered a violation that will impact the practical validity of these model results, in view of the general distribution of residuals and large study sample size. The same judgement is made regarding the plot from the summer rainfall environment (Figure 5).

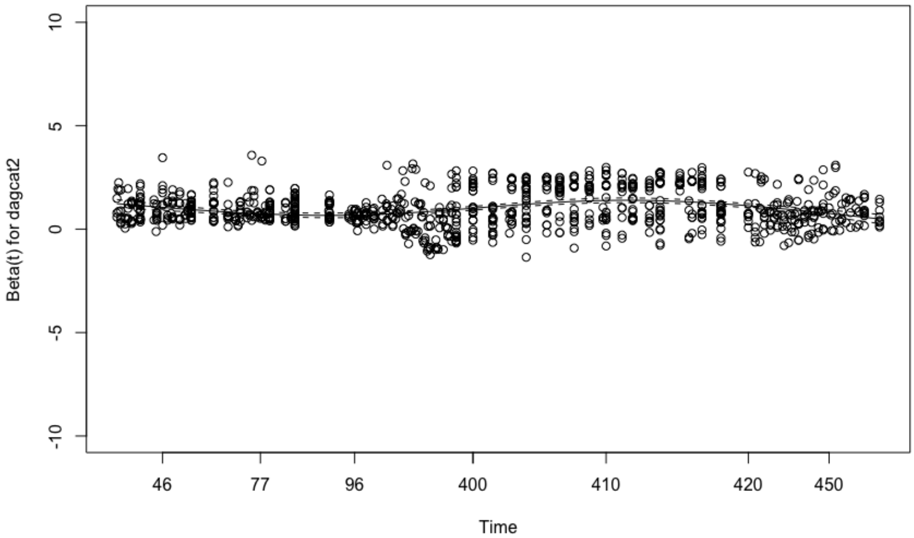


Figure 4 **Testing the proportional hazards assumption for the dag score model (winter rainfall environment): plot of scaled Schoenfeld residuals against transformed survival time**

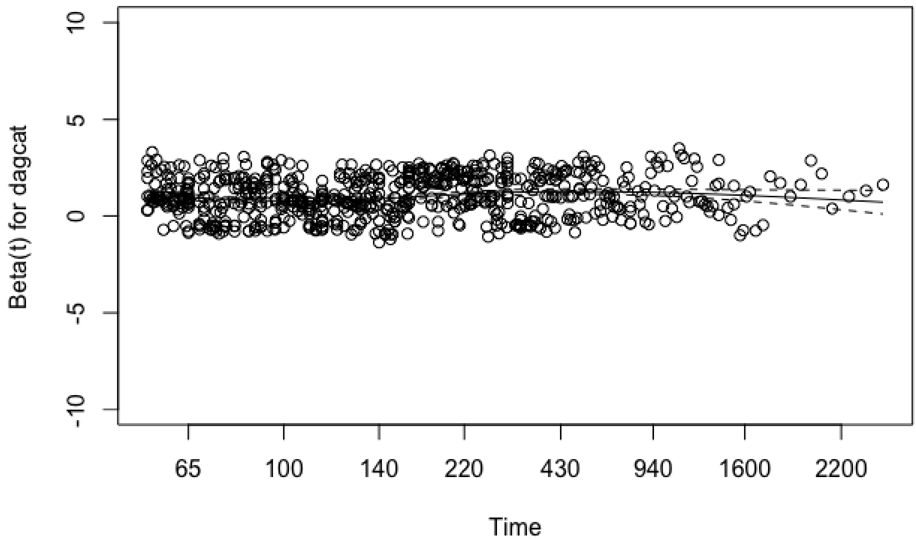


Figure 5 **Testing the proportional hazards assumption for the dag score model (summer rainfall environment): plot of scaled Schoenfeld residuals against transformed survival time**

Breech wrinkle score models

From the model for the winter rainfall environment, the smoothed curve of the plot of the scaled Schoenfeld residuals against transformed survival time supported the validity of the model (consistent with the assumption of proportional hazards: Figure 6; statistical hypothesis test p= 0.18).

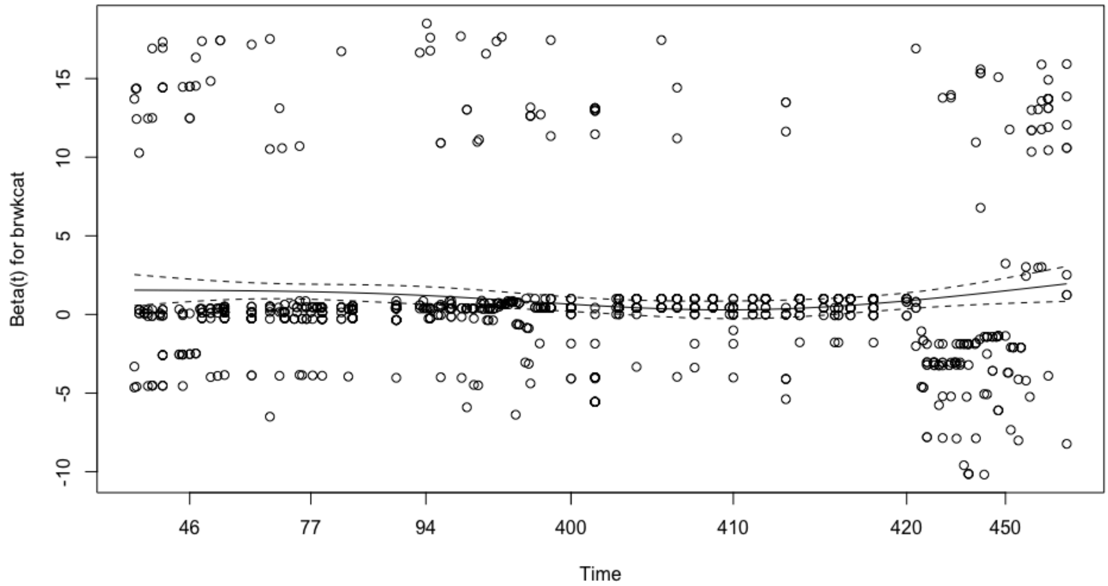


Figure 6 **Testing the proportional hazards assumption for the breech wrinkle score model (winter rainfall environment): plot of scaled Schoenfeld residuals against transformed survival time**

From the summer rainfall environment, while the smoothed curve of the plot of the scaled Schoenfeld residuals against transformed survival time for dag score deviates somewhat from a straight line (Figure 7). However, this is not considered a violation that will impact the practical validity of these model results, in view of the general distribution of residuals and study sample size.

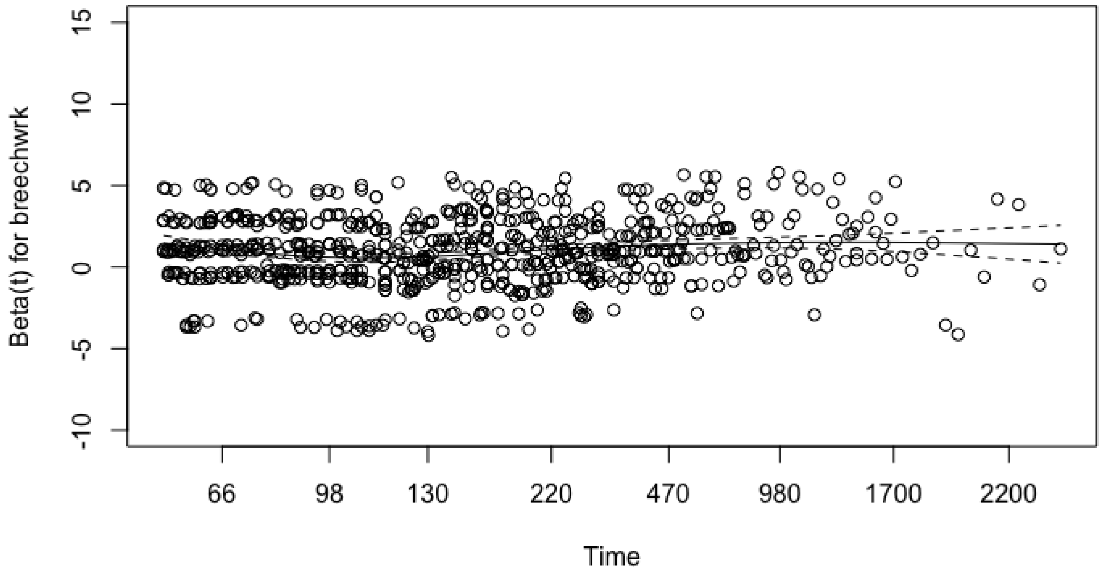


Figure 7 **Testing the proportional hazards assumption for the breech wrinkle score model (summer rainfall environment): plot of scaled Schoenfeld residuals against transformed survival time**

Breech cover score models

For both the winter rainfall environment and summer rainfall environment, the smoothed curve of the plot of the scaled Schoenfeld residuals against transformed survival time for breech cover score deviates somewhat from a straight line (Figure 8, Figure 9). However, the confidence limits and statistical hypothesis test (p= 0.16 and p=0.17, respectively) are consistent with the assumption of proportional hazards.

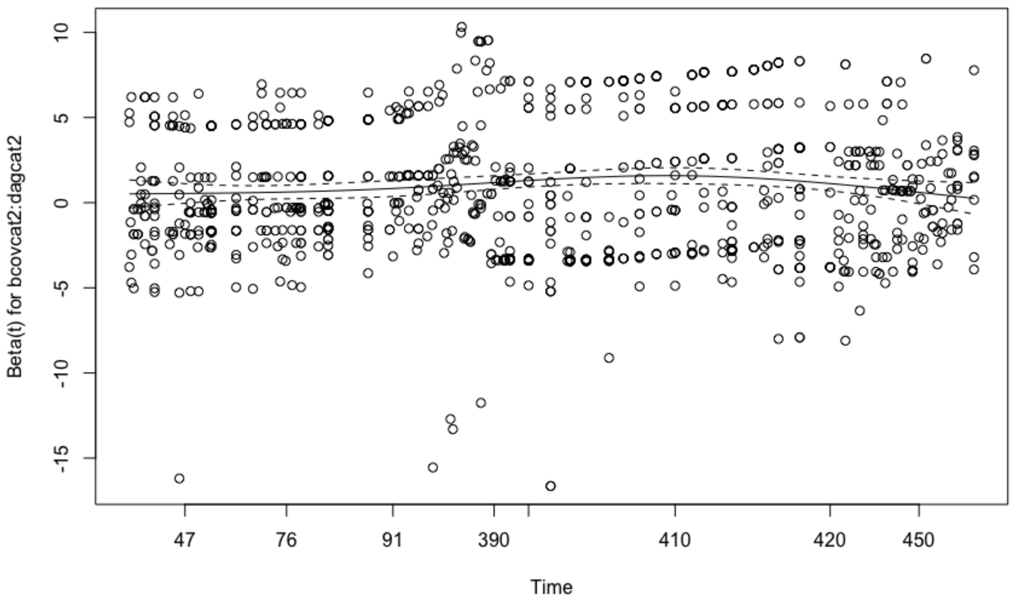


Figure 8 **Testing the proportional hazards assumption for the breech cover model (interacting with dag score) (winter rainfall environment): plot of scaled Schoenfeld residuals against transformed survival time**

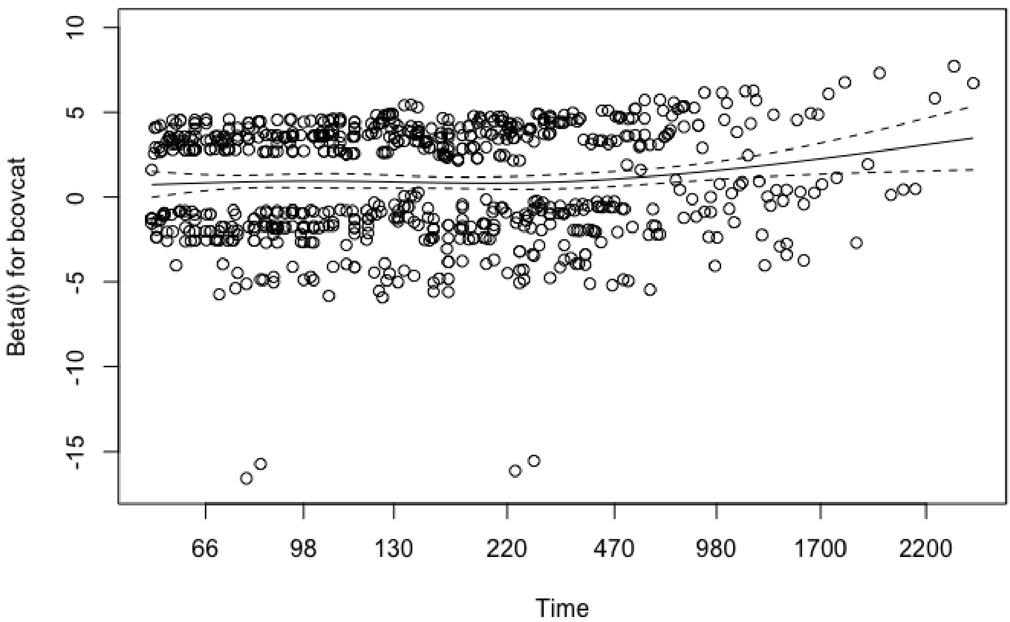


Figure 9 **Testing the proportional hazards assumption for the breech cover model (summer rainfall environment): plot of scaled Schoenfeld residuals against transformed survival time**

Appendix B Predictive and explanatory statistical models

Predictive and explanatory statistical modelling are useful and complementary tools when seeking to understand, influence and respond to disease occurrence.

Explanatory Models – does A cause B?

Explanatory modelling (Figure 10) is used to identify the causes of disease, and to estimate the magnitude of the contribution of various causes to disease occurrence. By doing so, strategies to prevent and/or treat disease can be identified, prioritised and optimised to reduce the occurrence of the disease in a population. For example, following the identification that smoking tobacco is an important cause of lung cancer, combined with knowledge that smoking was common amongst the community, strategies to reduce tobacco smoking were developed and implemented, and in doing so resulted in a considerable reduction in the rates of lung cancer[1]. Responsible choice of interventions is important from the perspectives of welfare and economic viability, and so understanding the contribution of a cause can help to decide between alternate management strategies.

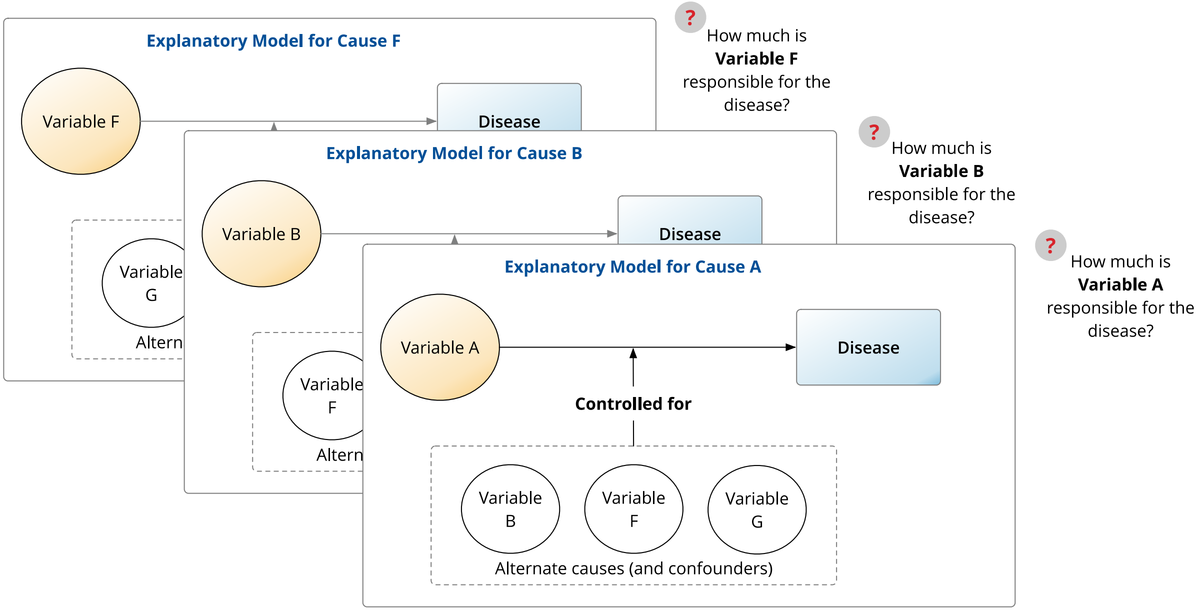


Figure 10 Explanatory models

In explanatory modelling, one model considers the putative association between a disease and a hypothesised cause of that disease. Evaluating additional hypothesised causes almost always requires construction of separate (tailored) models.

Predictive models – does a change in A impact B?

Predictive modelling (Figure 11) is used to identify a set of predictors that will predict, as accurately as possible, the occurrence of disease. Whether or not these predictors truly cause the disease, or are associated with the disease whilst not being causal, is irrelevant: as long as their inclusion improves the predictive accuracy of the model, they are an appropriate component[2]. For example, variables that can be used as predictors in a statistical model may actually be caused by the disease, or may proxy a causal factor that has not been captured in data available for the model (without being causal of itself). By identifying where disease is most likely to occur, predictive models can inform disease management strategies, such as ensuring adequate facilities and treatments are available where and when they are likely to be needed most. Predictive modelling can also have a role in breeding strategies, for example.

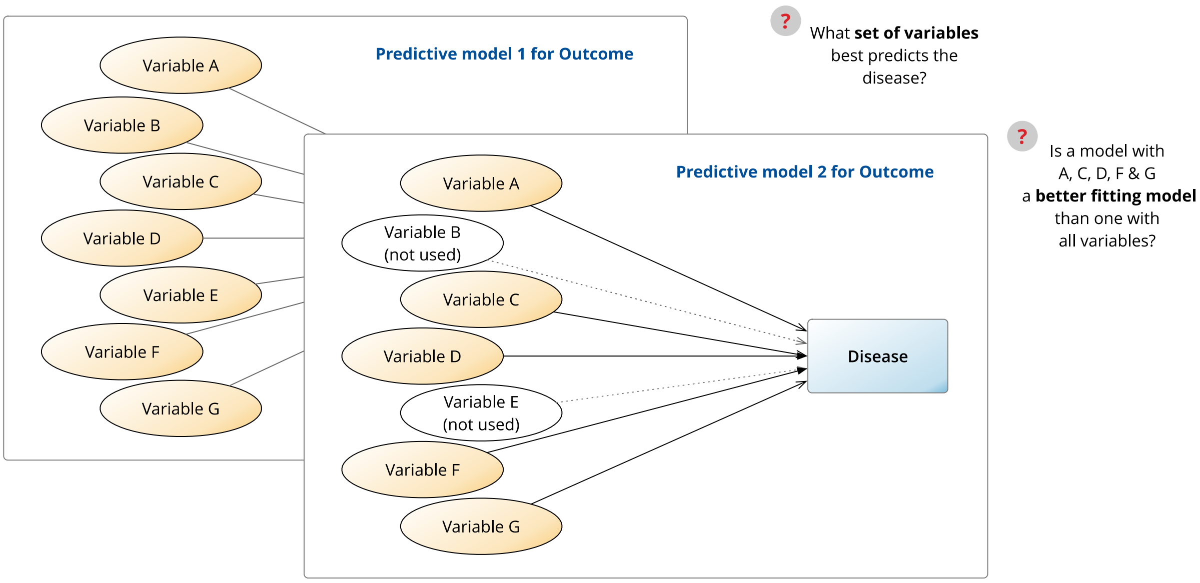


Figure 11 Predictive models

Are they not the same?

No. Regarding disease causation, predictive models have a role in generating hypotheses as to what may be causing a disease, but cannot be relied upon to accurately indicate whether, or to what extent, the predictors cause the disease[2]. As such, predictive models generally do not constitute adequate evidence on which to base costly interventions in an attempt to reduce the occurrence of disease. For example, premature hair greying in people was identified as a 'risk marker' for coronary artery disease[3]; despite it being a predictor of the disease, it is not causal—so interventions such as dying the hair of those with premature hair greying would not be effective at reducing the occurrence of the disease.

What are some key differences in model structure?

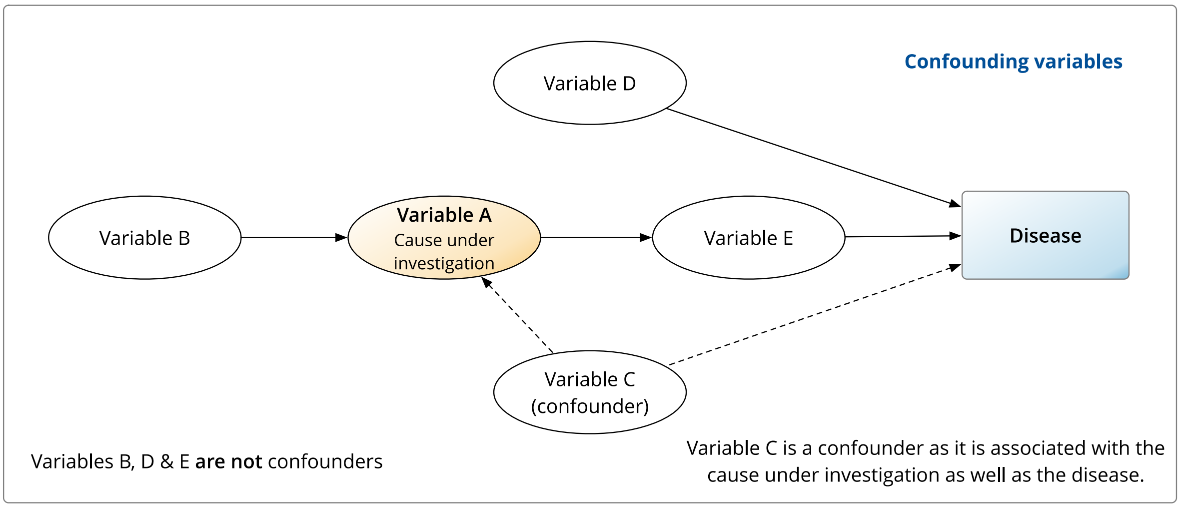
Some key differences between explanatory and predictive model structures include the approach to inclusion of covariates, and the impacts of collinearity.

How do we choose covariates for each type of model?

In explanatory modelling, the inclusion of covariates in a model considering the association between a hypothesised cause and a disease is undertaken to avoid bias by controlling for confounding.

1.5.1 What is confounding?

A confounding variable is a factor that is associated with the disease, and independently associated with the hypothesised cause (i.e. not on the causal pathway).



Statistically, if confounding is not appropriately addressed in a model structure, it will result in inaccuracies in the measurement of the association between a disease and a hypothesised cause.

Noting that inappropriately adjusting for covariates in statistical modelling can *induce* bias in the measure of magnitude of risk associated with a cause of disease (e.g. a coefficient or odds ratio)[4,5,6], in explanatory modelling it is important that covariate inclusion is undertaken in a systematic and considered manner, in view of known or biologically plausible causal relationships (as illustrated by directed acyclic graphs).

- For example, in the directed acyclic graph considering the association between dags and ovine breech flystrike (Figure 1), it is noted that crutching and shearing are on the causal pathway of the association between dags and breech flystrike, with no independent association with breech flystrike. Therefore, crutching and shearing is not plausibly confounding the association between dags and breech flystrike, and including it in a model considering this association may induce bias; so it is not included.
- In contrast, it can be observed that breech wrinkle is associated with dags, and independently associated with breech flystrike through other biological mechanisms. This makes it a plausible confounder of the relationship between dags and breech flystrike. Therefore, in measuring the association between dags and breech flystrike, breech wrinkle is an appropriate covariate for inclusion in a model to control for confounding bias.
- For complicated directed acyclic graphs, such as that in Figure 1, determining the appropriate sets of variables to include in a model to control for confounding can be complicated, and so specific programs such as DAGitty[7] are used for this purpose. There are usually multiple options of combinations of variables that will appropriately control for confounding bias in a complicated graph, as a result of the inter-relationships between causal chains and/or biasing pathways.

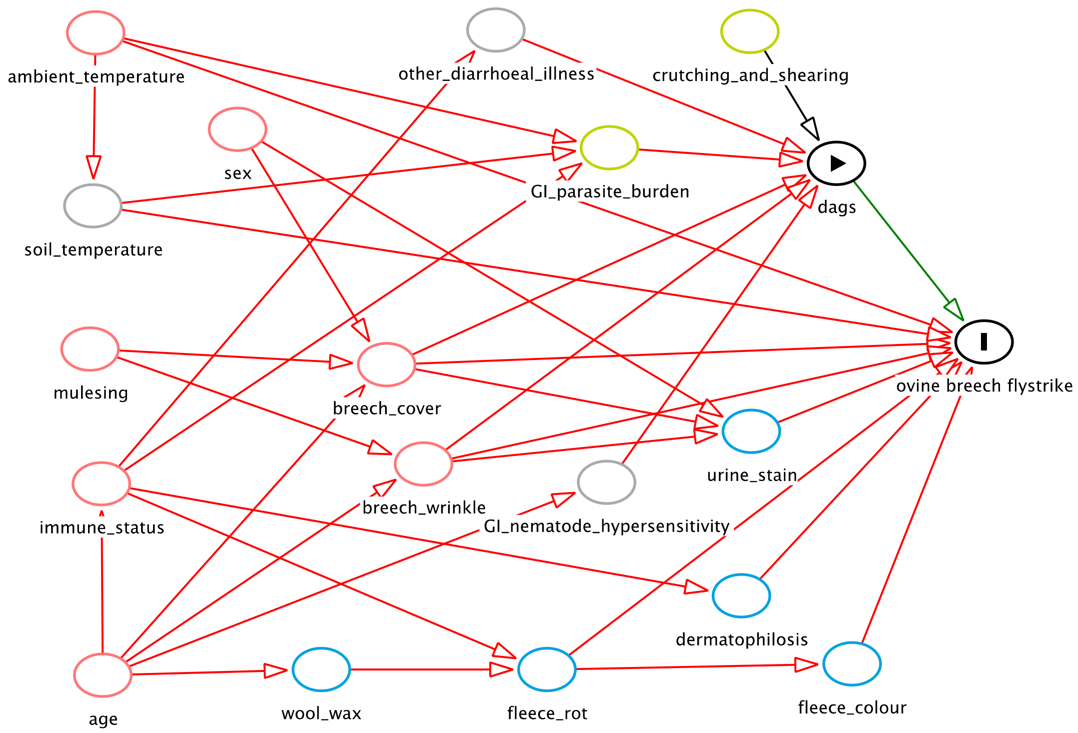


Figure 12 **Direct acyclic graph considering the association between dags and breech flystrike**
= hypothesised cause (dags); = disease (breech flystrike); = ancestor variables of the cause; = ancestor variables of the disease; = ancestor variables of both the cause and disease; = variables for which there are no data available; = causal path from dags to breech flystrike; = potentially biasing paths when considering the association between dags and breech flystrike; = non-biasing paths when considering the association between dags and breech flystrike.

Note that for explanatory modelling, one model considers the putative association between a disease and a hypothesised cause of that disease. Evaluation of further hypothesised causes almost always requires construction of separate (tailored) models (Figure 10)[8,9].

In predictive modelling, the inclusion of covariates (predictors) in a model is driven by improving the predictive performance of the model—as measured, for example, by an r-squared value or Akaike information criterion. Therefore, if a covariate improves model fit, it is generally retained: a model of improved fit will provide more accurate statistical information to predict disease occurrence[2].

Managing collinearity

Collinearity (the situation where two or more variables are themselves highly correlated) is another model structure consideration that varies between explanatory and predictive models. In explanatory models, it is crucial to avoid collinearity within a model structure to obtain accurate estimates of the magnitude of a causal association[10,11]. In contrast, the influences of collinearity are relatively unimportant in predictive models[2,12].

References

1. Moolgavkar SH, Holford TR, Levy DT, Kong CY, Foy M, Clarke L, Jeon J, Hazelton WD, Meza R, Schultz F, McCarthy W, 2012. Impact of reduced tobacco smoking on lung cancer mortality in the United States during 1975–2000. *Journal of the National Cancer Institute*, 104, 541–548.
2. Shmueli G, 2010. To explain or to predict? *Statistical science*, 25, 289–310.
3. Kocaman SA, Çetin M, Turan E, Çanga A, Çiçek Y, Şahin İ, Bostan M, 2012. The degree of premature hair graying as an independent risk marker for coronary artery disease: a predictor of biological age rather than chronological age. *Anatolian Journal of Cardiology/Anadolu Kardiyoloji Dergisi*, 12.
4. Greenland S, Pearl J, Robins JM, 1999. Causal diagrams for epidemiologic research. *Epidemiology*, 10, 37–48.
5. Shrier I, Platt RW, 2008. Reducing bias through directed acyclic graphs. *BMC medical research methodology*, 8, 70.

6. Bello NM, Ferreira VC, Gianola D, Rosa GJ, 2018. Conceptual framework for investigating causal effects from observational data in livestock. *Journal of Animal Science,* 96, 4045–4062.
7. Textor J, Hardt J, Knüppel S, 2011. DAGitty: a graphical tool for analyzing causal diagrams. *Epidemiology*, 22, 745.
8. Westreich D, Greenland S, 2013. The table 2 fallacy: presenting and interpreting confounder and modifier coefficients. *American journal of epidemiology*, 177, 292–298.
9. Bandoli G, Palmsten K, Chambers CD, Jelliffe-Pawlowski LL, Baer RJ, Thompson CA, 2018. Revisiting the Table 2 fallacy: A motivating example examining preeclampsia and preterm birth. *Paediatric and perinatal epidemiology*, 32, 390–397.
10. Midi H, Sarkar SK, Rana S, 2010. Collinearity diagnostics of binary logistic regression model. *Journal of interdisciplinary mathematics*, 13, 253–267.
11. Schisterman EF, Perkins NJ, Mumford SL, Ahrens KA, Mitchell EM, 2017. Collinearity and causal diagrams–a lesson on the importance of model specification. *Epidemiology*, 28, 47.
12. Vaughan TS, Berry KE, 2005. Using Monte Carlo Techniques to Demonstrate the Meaning and Implications of Multicollinearity. *Journal of Statistics Education*, 13.

Appendix C Comparison of the FlyBoss model estimated breech strike risk with the results in “Exploratory data analysis of two research flocks”.

Brian Horton, Tasmanian Institute of Agriculture, University of Tasmania

SUMMARY

The FlyBoss program uses a model for estimating the risk of breech strike from breech scores. This model is compared with a recent report by Hillman, Sadler and Madin (referred to as HSM here). This used the same flocks as those for the FlyBoss model, but included more data and used a different method of analysis. Both use breech wrinkle, breech cover and dag score to indicate the hazard ratio of sheep with a specified breech score, compared with sheep at a very low breech score.

At low to medium scores the methods give very similar results, but there is some deviation at high scores for dags and for breech wrinkle. The results for breech cover are comparable, although the HSM model reported an interaction of breech cover with dag score. At high dag scores this interaction results in a reduction in breech strike risk with increasing breech cover, which is not consistent with other data on breech cover and may not be significant.

The new analysis separated results into winter rainfall and summer rainfall, but most results were within overlapping ranges for these groups. The FlyBoss model is intended for all regions of Australia, so these flocks were combined in that analysis, and for most purposes given similar results to the HSM model.

The FlyBoss model may be over- or under-estimating the risk of strike at very high dag or wrinkle scores, but the differences between models may be due to the method of analysis and the animals included or excluded from analysis.

The conclusions derived by the FlyBoss model agree closely with the conclusions provided by the HSM analysis. There is no need to revise the FlyBoss model at this time, but it could be improved by the addition of more data that was not available for the original analysis.

BACKGROUND

The report by Hillman, Sadler and Madin (2020) used data from flocks established to examine the risk of breech strike in mulesed and unmulesed sheep, at Armidale, NSW, and Mt Barker, WA. This study used explanatory modelling to take into account the causal framework of the association between a range of factors and breech strike.

An example from that study is shown in Figure 1, where ovine breech flystrike is the outcome to be explained, dags are the exposure if interest and the other factors shown are ancestor variables of the exposure or of the outcome. Similar directed acyclic graphs were used for breech wrinkle and for breech cover. This method can avoid the inclusion in a model of inappropriate factors, that are common in standard statistical methods. However, the procedure may depend on correct inclusion of relevant ancestors and assignment of all associations. In this case the model indicates that all effects of mulesing occur through its reduction of breech cover and breech wrinkle. However, mulesing does have some additional effects, and this may affect the model, as noted later.

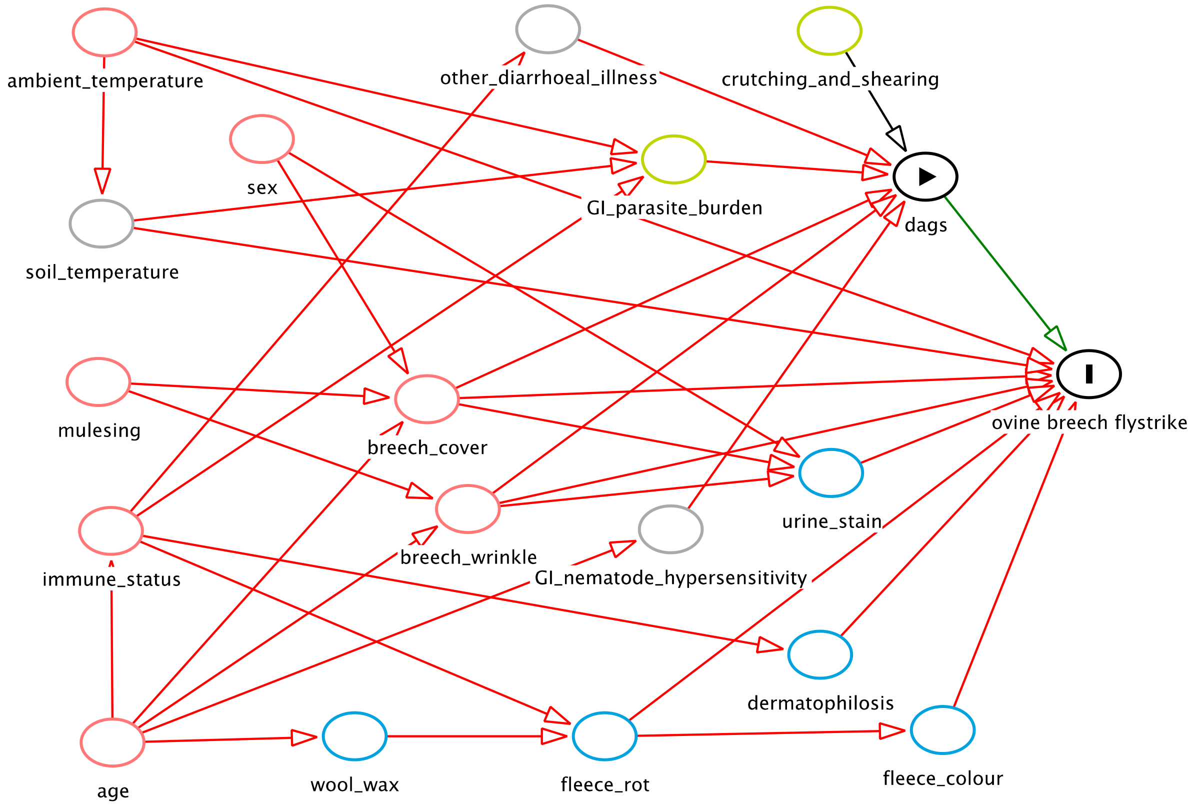


Figure 13. Directed acyclic graph considering the putative association between dags and breech flystrike in sheep {Hillman, 2020 #386}.

The same two flocks used by Hillman, Sadler and Madin (2020) provided data used for an analysis (Horton et al. 2020), which has been used to estimate the risk of breech strike for the FlyBoss Flystrike Risk Management Planner. This was a purely statistical analysis that allowed for effects of mulesing independent of its effects on wrinkle, breech cover and dag, while allowing for any interactions between these factors. These two different studies will be referred to here as the HSM model and the FlyBoss model. This report compares the results found by each model.

Differences in the data analysis

The FlyBoss model used data from sheep born in 2005 to 2009, whereas the HSM model used data from 2006 to 2014. There were some changes over time in the risk of strike, due to seasonal differences, and the selection of some sheep for resistance to breech strike. Some differences would also be expected due to the larger amount of data in the HSM model and to the changes in strike risk over this period.

In the first three years after the flocks were established, half the sheep were mulesed and half left unmulesed, but for the remaining period all new lambs were unmulesed. The FlyBoss model compared mulesed and unmulesed sheep and estimated the effect of different breech scores in both mulesed sheep and unmulesed sheep for both scores. However, being later calendar years, the HSM model had a higher proportion of unmulesed sheep. For wrinkle and breech cover analyses the mulesed sheep were omitted, so the results are only for unmulesed sheep. For dag effects the mulesed sheep were included, but the graphs showing putative association indicate that the effect of mulesing was only expected to result in a change in breech wrinkle and breech cover, and these would lead to changes in breech strike. This model did not allow for other effects of mulesing on breech strike, apart from these changes in breech scores. But the FlyBoss model found that mulesing influences breech strike, separate from the effect on breech scores. It was shown that if two groups of sheep have identical scores for breech wrinkle, breech cover and dag score, but one group has been mulesed and the other is unmulesed, then the unmulesed group, under the same conditions, may have up to three times the risk of breech strike compared with the mulesed group. This calculation was confirmed separately by Johan Greeff using different analytical methods. It was suggested that mulesing may reduce the moisture level in the breech area, in addition to its direct effect on breech wrinkle and breech cover.

The effect of including some mulesed sheep in the analysis, but assuming that the only effect of mulesing would be through changes in breech scores, would be to have a high proportion of mulesed sheep in the low scoring group, with a lower risk in that group due to the separate protective effect of mulesing. The high scoring groups would have fewer mulesed sheep, so the hazard ratio would be overestimated for high score groups compared with low score groups.

The HSM model used explanatory modelling, using directed acyclic graphs, to model the biological basis of all factors included in the model. This method found an interaction of dags with breech cover and analysed breech cover separately for four different levels of dag score. This interaction was not found in the FlyBoss model, although tests were made for interactions between traits. There may not have been enough data in the FlyBoss model to detect this interaction. However, using separate dag score groups for analysis would result in smaller groups for the study of breech cover and the results for breech cover are not consistent with the expected model.

The HSM model provided a separate analysis for high summer rainfall and for low summer rainfall. However, this differentiation was based on only a single property in each rainfall zone. Although some replication may apply due to using many years of data, there could be other differences between the Armidale and Mt Barker flocks that are not due only to the timing of rainfall. For example, the foundation ewes selected for the study may not have been fully comparable, the management would be consistently different, and the local environment of neighbouring flocks could affect flystrike in that specific area. Therefore, it may not be safe to draw specific conclusions related to the timing of rainfall.

The HSM model used only the first occurrence of breech strike, since treatment of that strike might affect later strike. The FlyBoss model counted all strikes, so a sheep struck three times was considered to have three times the risk of a sheep struck once. This method was used because all treatments provided only short-term protection, so treatment for strike may have had limited effect on later strikes. This would increase the calculated hazard ratio for very high-risk sheep.

The HSM model analysed each block of scores separately, and this avoided forcing the model artificially to increase risk at higher scores, and could result in some non-linear effects. However, this makes it more difficult to estimate results for intermediate scores. The method used for the FlyBoss model was intended to provide a smooth change in breech strike risk as the scores changed, to allow interpolation to any intermediate score. This may have caused some smoothing at the extremes with very low or very high scores. Where there were few high scores, the results may have been influenced by the next lower scores.

The FlyBoss decision support program required a model that would be generally applicable to sheep farms in any area of Australia, so the analysis for the FlyBoss model did not try to include specific property effects, (except in terms of the relative risk of strike from one flock compared with another), while trying to find a model of the effects of breech scores that applied across both flocks.
The combined result of these differences in analysis mean that it should not be expected that the findings will be identical, and the HSM model may be more precise where more data was available for the specific conditions examined.

METHODS

The FlyBoss decision support program assumes that the user has a flock of sheep with a given mean score for breech wrinkle, breech cover and dag score and an estimated risk of breech strike. This flock average is made up of individual sheep with a range of scores, but the higher scoring sheep have the greatest risk of strike, so the actual risk of strike for a flock of sheep of wrinkle score 3.0 is higher than a subgroup of sheep all of wrinkle score 3.0. The flock with a wider spread of scores would have some with score 4 or higher, that would contribute most of the strike. For this comparison, the FlyBoss model was used for individual sheep of the specified score, not for a flock of that score. The risk of strike in the subgroup of sheep was compared with the results in the HSM model, where the results were based on a narrow range of scores.

Since the FlyBoss model does not include an interaction between wrinkle, breech score and dag, the scores not being tested were set to 1.0 and the other scores set to the values tested for the HSM model

RESULTS

Dag Score

Figure 2A shows the estimated hazard ratio calculated by the FlyBoss model with the upper and lower range given by the HSM model for winter rainfall, while Figure 2B shows the same FlyBoss model risk, but compared with the HSM model estimate for summer rainfall. The FlyBoss model estimates are similar to those in the HSM model at lower dag scores, but the HSM model indicates a much higher risk at the highest dag score.

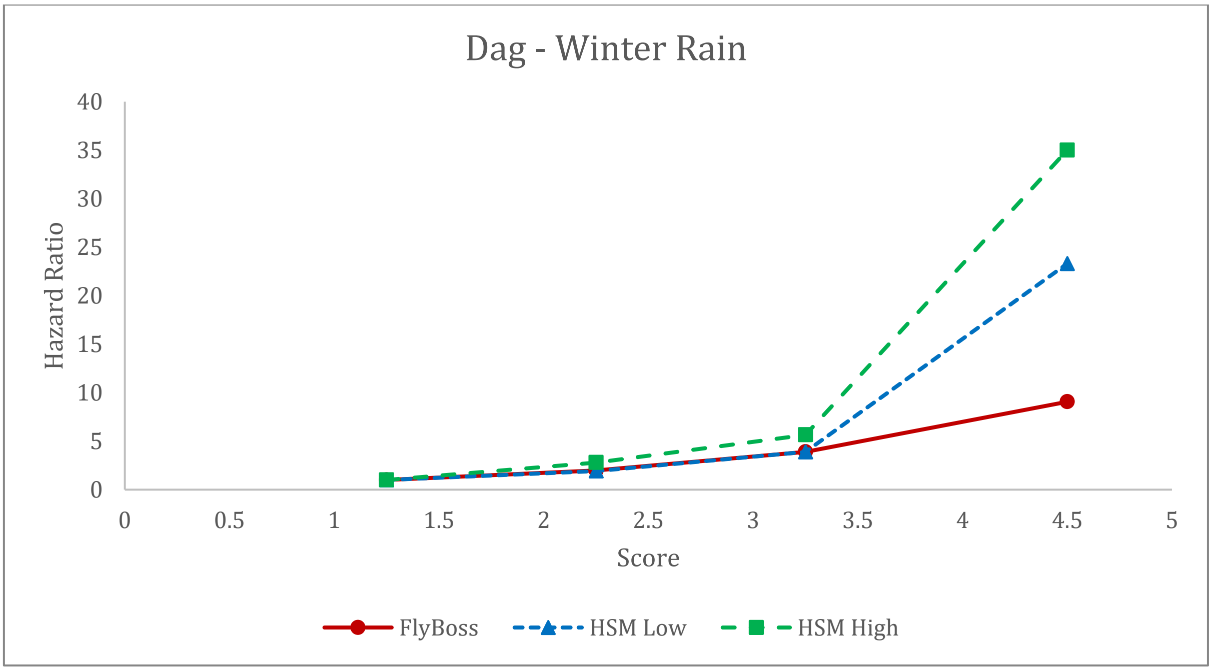


Figure 2A. Breech strike hazard ratio of different dag scores, compared with dag score 1.25 for FlyBoss and the low and high ranges from the HSM model for winter rainfall.

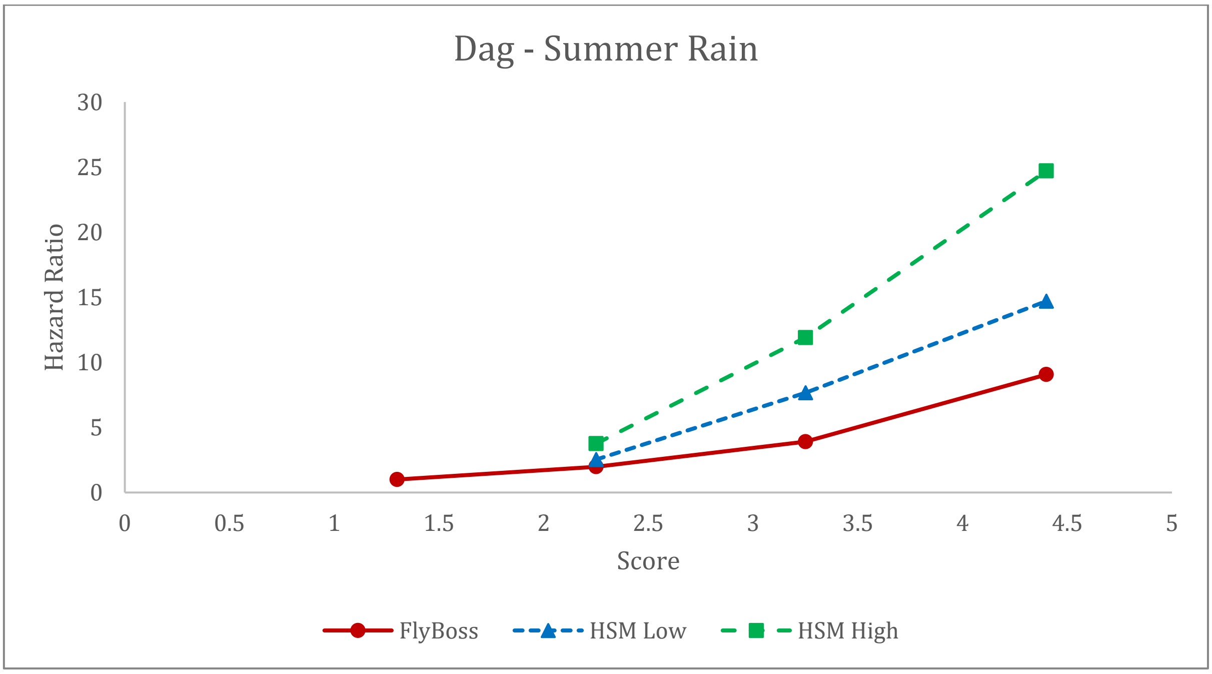


Figure 2B. Breech strike hazard ratio of different dag scores, compared with dag score 1.25 for FlyBoss and the low and high ranges from the HSM model for summer rainfall.

Breech Wrinkle
Figure 3A compares the FlyBoss model risk with the HSM model risk for a range of breech wrinkle scores for winter rainfall, while Figure 3B shows the comparison for summer rainfall. In this case the FlyBoss model risk was higher than estimated by the HSM model at the highest scores.

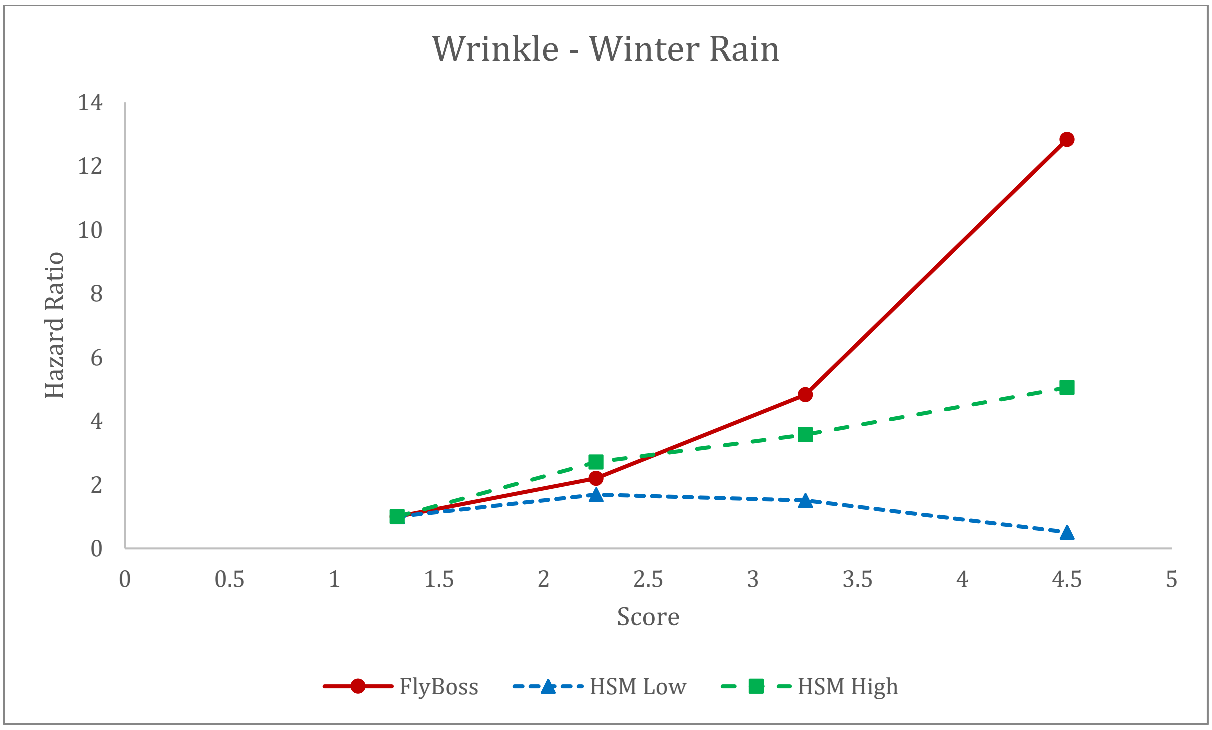


Figure 3A. Breech strike hazard ratio of different breech wrinkle scores, compared with breech wrinkle 1.25 for FlyBoss and the low and high ranges from the HSM model for winter rainfall.

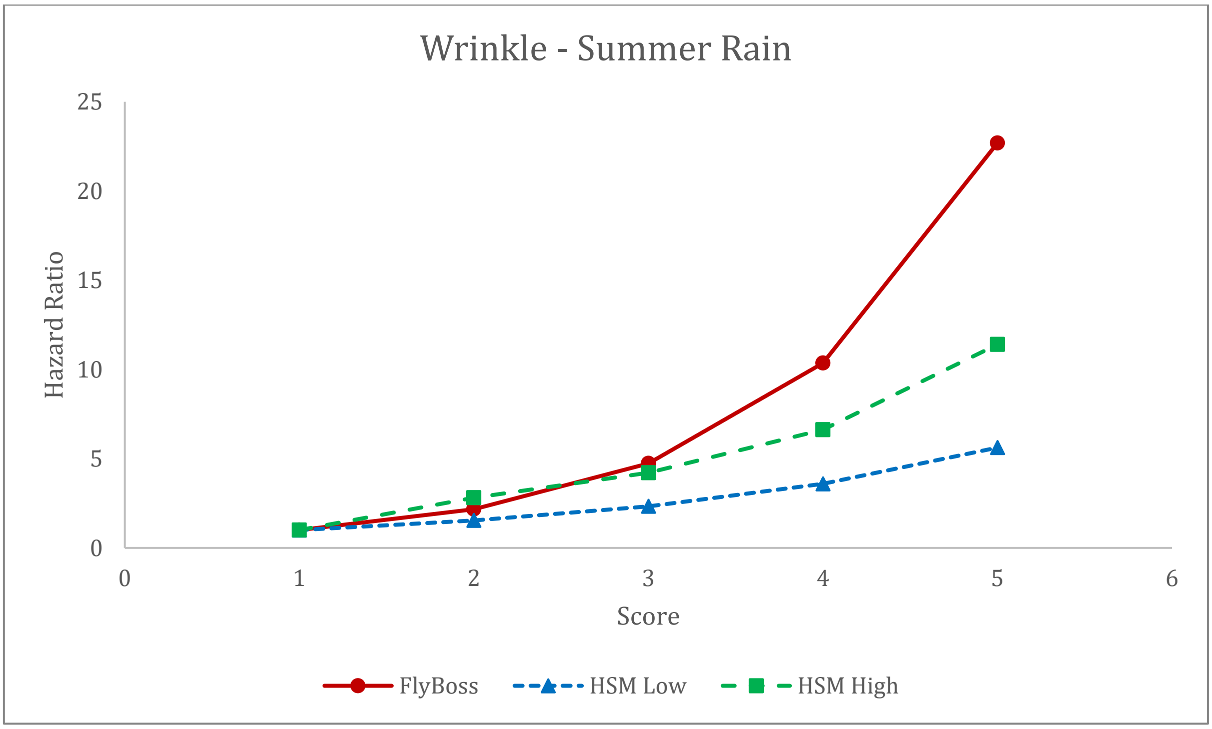


Figure 3B. Breech strike hazard ratio of different breech wrinkle scores, compared with breech wrinkle 1.25 for FlyBoss and the low and high ranges from the HSM model for summer rainfall.

Breech Cover
Figures 4A, 4B, 4C and 4D show the estimated risk of breech strike at different breech cover scores in the winter rainfall area. In this case the HSM model found an interaction with dag score, so separate results were given for dag score 1.25, 2.25, 3.25 and 4.5. The FlyBoss model does not include an interaction between dag

score and breech cover effects on breech strike, so the FlyBoss result is the same in all cases. The fit between the FlyBoss model and the HSM model is good at low dag score, but at higher dag scores, the HSM model has a reduction in breech strike risk as breech cover increases, so the FlyBoss model is outside the HSM model range for the highest breech cover.

Figure 4E shows the results for the summer rainfall area, where the FlyBoss model fits in the range given by the HSM model.

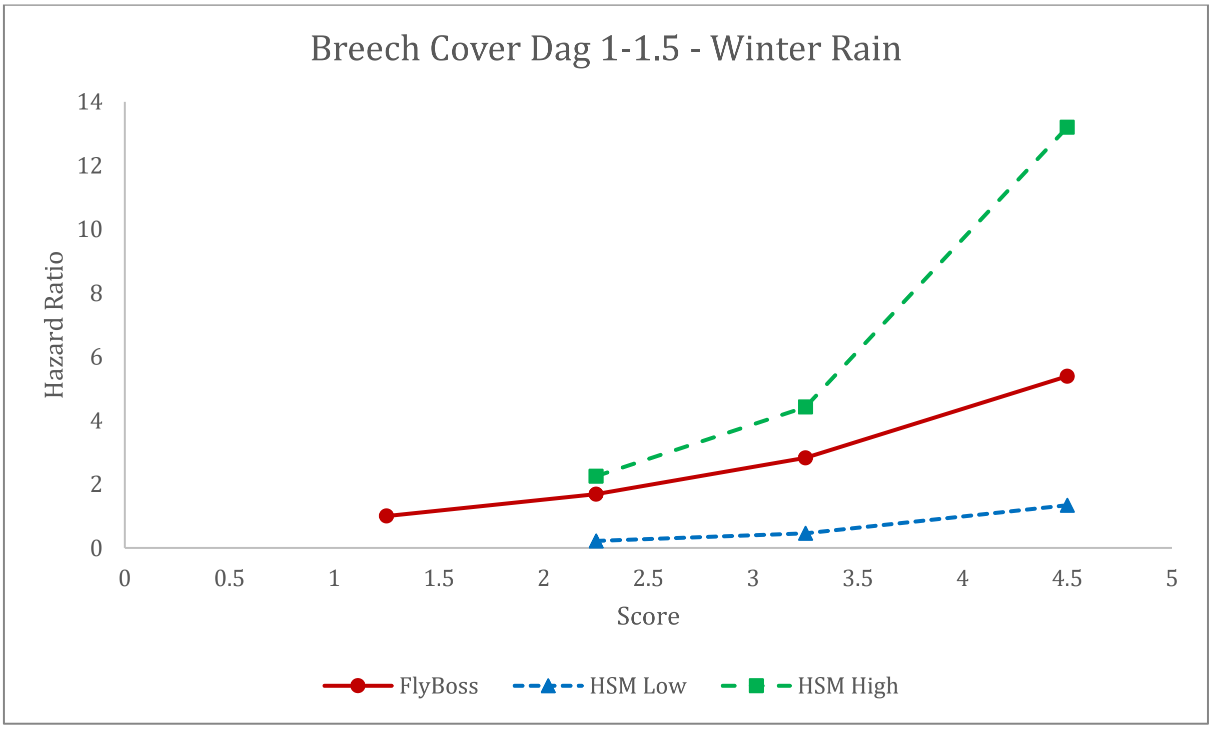


Figure 4A. Breech strike hazard ratio of different breech cover scores at dag score 1.25, compared with breech cover 1.25 for FlyBoss and the low and high ranges from the HSM model for winter rainfall.

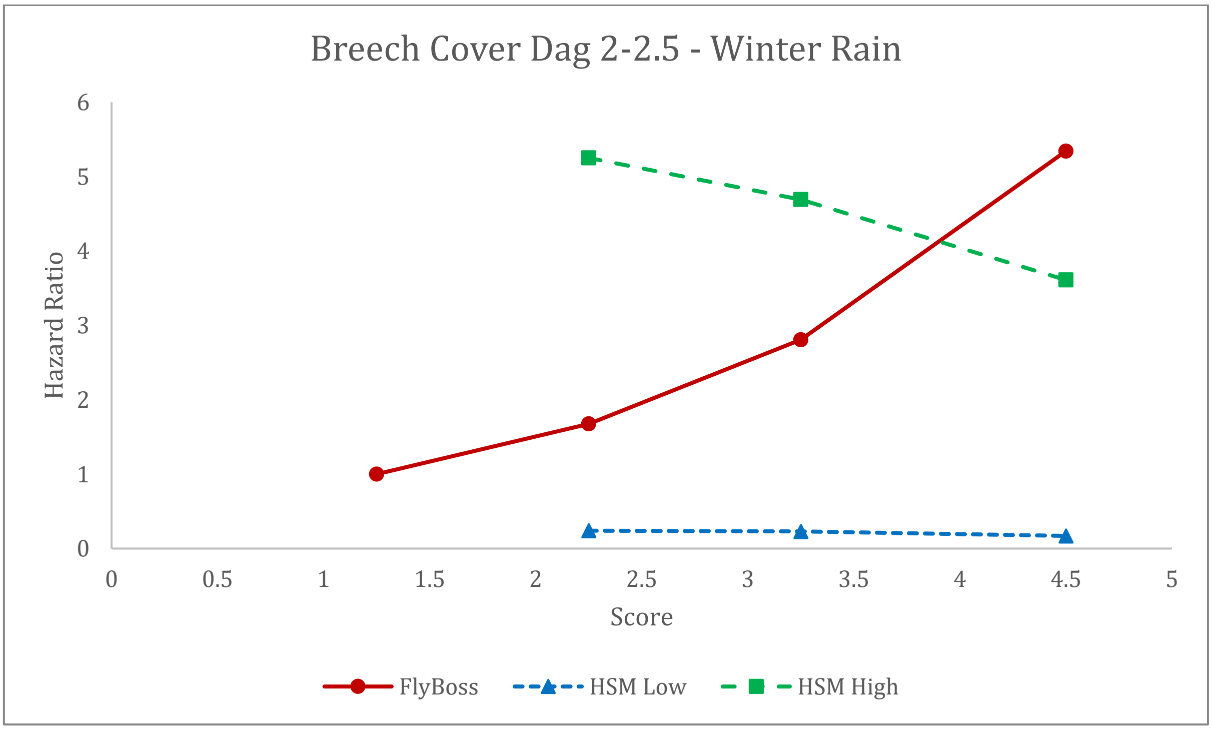


Figure 4B. Breech strike hazard ratio of different breech cover scores at dag score 2.25, compared with breech cover for FlyBoss and the low and high ranges from the HSM model for winter rainfall.

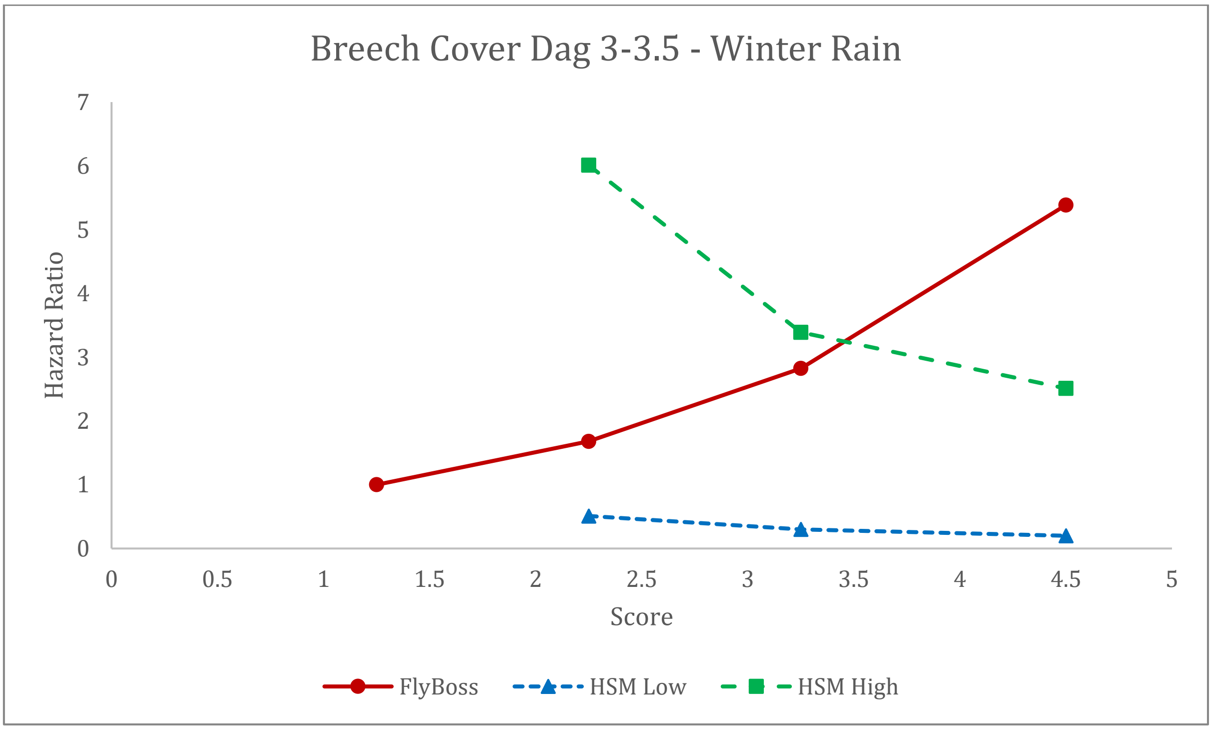


Figure 4C. Breech strike hazard ratio of different breech cover scores at dag score 3.25, compared with breech cover for FlyBoss and the low and high ranges from the HSM model for winter rainfall.

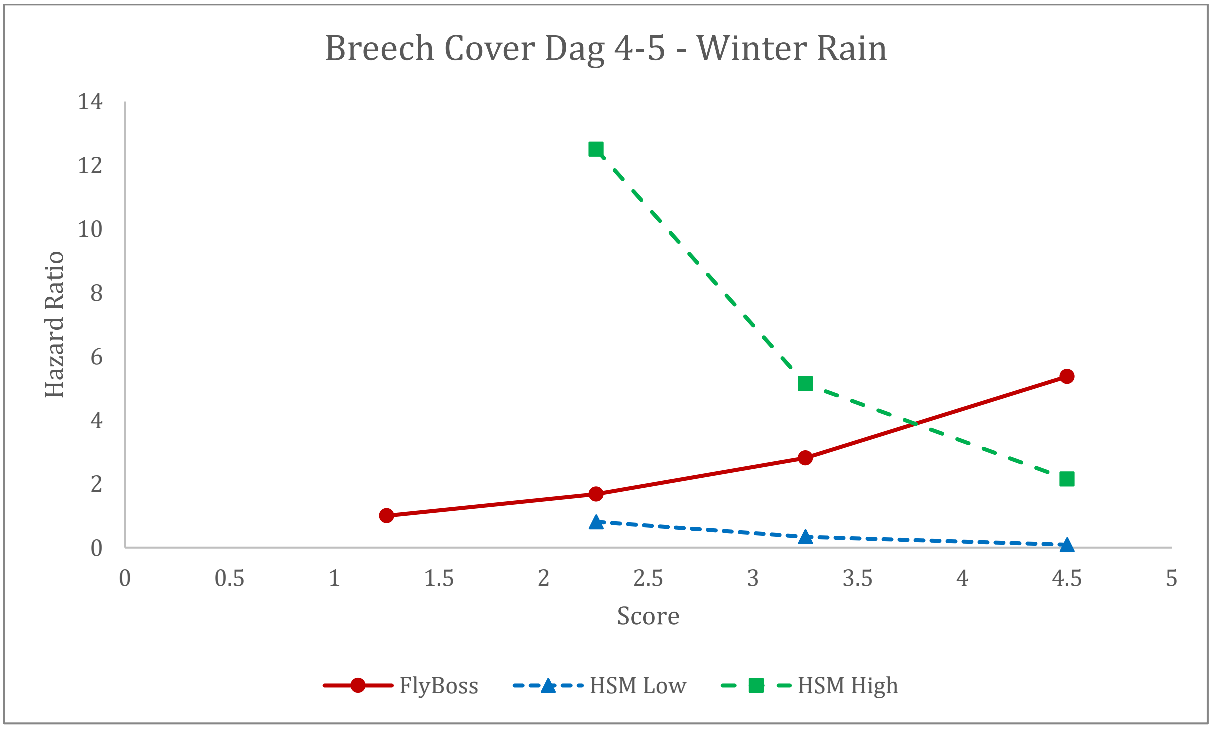


Figure 4D. Breech strike hazard ratio of different breech cover scores at dag score 4.5, compared with breech cover for FlyBoss and the low and high ranges from the HSM model for winter rainfall.

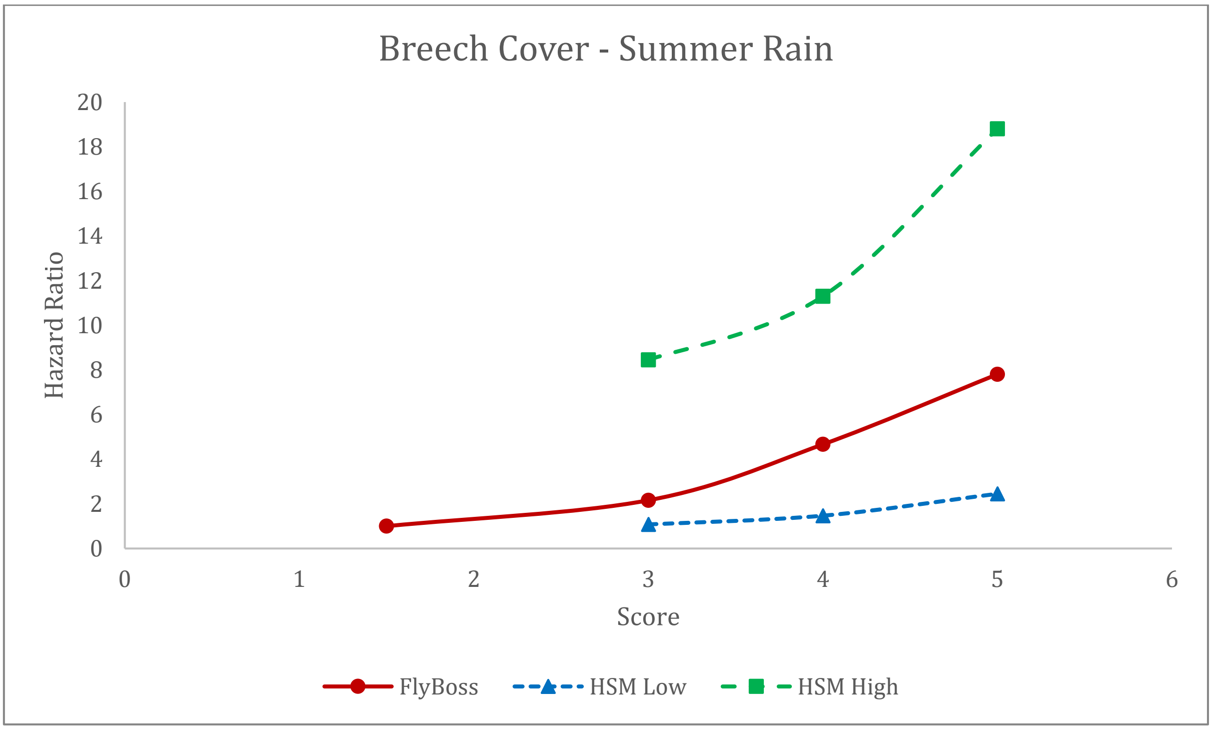


Figure 4E. Breech strike hazard ratio of different breech cover scores, compared with breech cover for FlyBoss and the low and high ranges from the HSM model for summer rainfall.

DISCUSSION

In all cases the FlyBoss and HSM models agree closely at low to medium scores for dag, wrinkle and breech cover. However, at high dag scores the FlyBoss model underestimates the risk compared with the HSM model, while at high breech wrinkle score the FlyBoss model overestimates the risk. For breech cover the FlyBoss model is in close agreement for summer rainfall areas, and for winter rainfall at low dag scores.

Dag score

Mulesed sheep were not excluded from the HSM model dag score analysis. Most of the mulesed sheep would have been at low dag scores and very few at high dag scores. As noted above, the breech strike risk for mulesed sheep at any given dag score is less than for unmulesed sheep at the same dag score. The effect of including mulesed sheep would be to lower the apparent risk for low dag scores and increase the relative risk at higher scores. The FlyBoss model includes mulesed vs not mulesed in each analysis. This may explain why the FlyBoss model has a lower hazard ratio than the Hillman model at the highest dag scores.

Breech wrinkle

Only 1% of the winter rainfall results (61 cases) had a breech wrinkle in the range 4 to 5, so the deviation at this score may not be as reliable as the other results. The summer rainfall group had 7% of results in this range. The results for the FlyBoss model may have been influenced by the larger number of lower scores, since it was designed to provide a smooth transition between scores. The FlyBoss model allotted a higher risk to sheep that suffered multiple strikes, whereas the HSM model only tested the first strike for each sheep. This would have allocated a higher risk in the FlyBoss model to sheep that had the highest scores resulting in repeated strikes to the same sheep.

Breech cover

For breech cover, the HSM and FlyBoss models agree closely for summer rainfall and also for winter rainfall at low dag scores. At high dag scores, the HSM breech cover model appears to decrease risk at high scores compared with the risk at low scores. The FlyBoss model did not obtain this result and it is not clear why it should happen. Hillman et al (2020) note that there is no biological explanation for a decreased risk at high breech cover and suggest the result should be interpreted cautiously. The possible range of values, from low to high score, is wide, and does not preclude a model with increasing risk as breech cover increases. Since the

FlyBoss and HSM models agree for summer rainfall and for winter rainfall at low dag score, there does not appear to be any reason to alter the FlyBoss model as a result of these findings.

Winter vs summer rainfall

The two flocks used for analysis were in different rainfall zones and were analysed separately in the HSM model, but combined for the FlyBoss model. The differences between the two rainfall zones found for the HSM model in most cases have overlapping ranges or are only marginally significantly different. This confirms that combining the data for the FlyBoss model, which is intended to apply to all regions, has not introduced substantial errors.

Data available for the models

The HSM model was based on a larger sample of data than the FlyBoss model, which only used the first four years of the data from the two flocks studied. However, the results of the two analyses agree at all low to moderate scores.

The FlyBoss model could be improved by inclusion of the additional data used in the HSM model, but there is no need to alter the current FlyBoss model at this time. Users of the FlyBoss model are expected to enter flock averages, not results for individual sheep. Therefore, there should be few flocks in Australia with a flock average at the extremely high breech scores, where the deviation between models occurs. If flocks do have extreme scores then both models indicate that substantial reductions in risk will occur with any reduction in those breech scores, to reach intermediate levels, where the models give similar results.

REFERENCES

Hillman, A., Sadler, R. and Madin, B. 2020. Explanatory modelling (ON-00715 extending ON-00555: Exploratory data analysis of two research flocks), Ausvet, ACT, Australia.

Horton, B.J., Corkrey, R., Smith, J., Greeff, J. and Karlsson, L.J.E. 2020. Modelling of breech strike risk and protective efficacy of mulesing in adult Merino sheep. Animal Production Science, 60: 1051-1060.