The Old School, Brewhouse Hill, Wheathampstead,
Hertfordshire AL4 8AN, UK
www.ufaw.org.uk

Animal Welfare 2013, 22: 267-275
ISSN 0962-7286
doi: 10.7120/09627286.22.2.267

The use of quantitative risk assessment to assess lifetime welfare outcomes for breech strike and mulesing management options in Merino sheep

AD Fisher*[†‡], A Giraudo[§], PAJ Martin[#] and MW Paton[#]

[†] Faculty of Veterinary Science and Animal Welfare Science Centre, 250 Princes Highway, Werribee, VIC 3030, Australia
[‡] CSIRO Livestock Industries, Locked Bag 1, Armidale, NSW 2350, Australia
[§] AgroParisTech, 16 rue Claude Bernard, 75006 Paris, France
[#] Department of Agriculture and Food, Western Australia, 3 Baron-Hay Court, Perth, WA 6151, Australia
* Contact for correspondence and requests for reprints: adfisher@unimelb.edu.au

Abstract

In Australia, flystrike can severely compromise sheep welfare. Traditionally, the surgical practice of mulesing was performed to alter wool distribution and breech conformation and thereby reduce flystrike risk. The aim of this study was to use published data to evaluate the effectiveness of an epidemiologically based risk assessment model in comparing welfare outcomes in sheep undergoing mulesing, mulesing with pain relief, plastic skin-fold clips, and no mulesing. We used four measures, based on cortisol, haptoglobin, bodyweight and behavioural change, across three farming regions in Australia. All data were normalised to a common scale, based on the range between the highest and lowest responses for each variable ('welfare impact'; I). Lifetime severity of welfare challenge (SWC) was estimated by summing annual SWCs (SWC = I × P, where P = probability of that impact occurring). The severity of welfare challenge during the first year of life was higher for mulesed animals compared to unmulesed. However, over five years of life, the highest severity of welfare challenge was for unmulesed animals, and the lowest was for the plastic skin-fold clips. The model produced estimates of SWC that are in broad agreement with expert consensus that, although mulesing historically represented a welfare benefit for sheep under Australian conditions, the replacement of mulesing with less invasive procedures, and ultimately genetic selection combined with anti-fly treatments, will provide a sustainable welfare benefit. However, the primary objective of this work was to evaluate the use of the risk assessment framework; not to compare welfare outcomes from mulesing and its alternatives.

Keywords: *animal welfare, flystrike, Merino, mulesing, risk assessment, sheep*

Introduction

Although science offers tools to measure aspects of an animal's welfare, the evaluation of overall animal welfare can be both subjective and qualitative. Animal welfare assessment often takes into account quantitative measurements of behaviour and physiological variables. However, no agreed scientific methodology exists to combine these different elements into an overall assessment of welfare, or for evaluating whether this welfare state is acceptable.

A science-based, objective determination of an animal's welfare will always be subject to differing interpretations based on individual ethical frameworks, however, there may be considerable opportunity to reduce the subjective component of animal welfare assessment.

The paper by Paton *et al* (2013; this issue) proposed a theoretical basis for the objective evaluation of animal welfare using a semi-quantitative approach based on risk assessment principles. In its Scientific Opinion (EFSA 2012) the European Food Safety Authority (EFSA) also details methods for evaluation of welfare using a risk assessment framework. It is recognised both by society and farming industries that it is important to develop systems which quantify, as much as possible, the effects of different environments and management practices on the welfare of animals. These systems may assist in making the science of animal welfare more quantitative, simpler to analyse and easier to communicate. More importantly, semi-quantitative risk assessment may allow comparison of animal welfare for different management strategies and environments.

In Australia, blowfly strike, and more particularly breech strike in sheep, is a serious disease problem and can severely compromise the welfare of animals. Blowfly eggs, laid on the skin of a sheep, hatch into larvae which feed on the sheep's tissue. Flystrike can produce inflammation, general systemic toxaemia, and even death.

The surgical husbandry practice of mulesing was developed to reduce an animal's risk of developing breech strike (Beveridge 1984). Mulesing is a procedure in which two strips of skin are cut from the hindquarters of Merino lambs in order to remove wool-bearing wrinkled skin, increase the perineal bare area, and thereby reduce the risk of breech strike throughout life thereafter. Although, in the years after its widespread adoption, mulesing was highlighted for its animal welfare benefits in reducing flystrike, the practice has more recently been the focus of concern because of the animal welfare costs of the procedure itself (Lee & Fisher 2007).

Pending the widespread availability of genetic lines of Merino sheep selected for resistance to breech strike, the Australian wool industry is developing alternatives to the conventional approach of surgical mulesing.

Assessing the overall animal welfare outcomes of different approaches to the problem requires the comparison of events that challenge the animal's welfare in different ways and at different times in the animal's life. There are some quantitative data on the probability and impact of these individual challenges, but there has been no way to compare the welfare outcomes of these different strategies objectively.

This paper uses these existing data to assess the welfare impacts associated with flystrike and mulesing, using the epidemiologically based risk assessment approach (Paton *et al* 2013). The EFSA Opinion (2012) mentioned above detailed methods of handling expert opinion and to avoid heuristic bias in utilising this source of information in risk assessment. The issues outlined in the EFSA Opinion could be applied to the technique described in Paton *et al* (2013) and in the example described here. However, this is not a significant issue in this paper as the data used are from published work where data and not opinion is used. The comparisons in this paper are among mulesing, mulesing with pain relief, an alternative to mulesing, and no mulesing (with and without genetic selection for breech-strike resistance). In the past, assessments on the suitability of such options have depended on a subjective assessment of the relative degree of potential suffering of animals. A more quantitatively based assessment may enable more effective communication of the relative merits of different approaches.

This study does not propose to assign absolute welfare values to the five breech-strike management scenarios that would enable comparison to other situations without performing an additional welfare assessment. Rather, the study uses a risk analysis approach to compare existing data on sheep welfare across an animal's lifetime for these five scenarios, across different farming environments. There are considerable data on the responses of sheep to mulesing, alternative approaches to mulesing, and flystrike. This scenario was chosen as the subject of this example because it facilitated the methodology and welfare comparisons described. The EFSA Opinion (2012) details protocols for the appropriate assessment of data to include in risk analyses. As this is a simplified example using a defined set of data, the inclusion of such protocols are not appropriate in this paper. In undertaking a risk assessment for animal welfare, it is recognised that different observers have varying interpretations of the concept and best definition of animal welfare. It is not the intention of this study to try to resolve this issue. Rather, we have assumed a broad definition of animal welfare that also permits assessment at the group level (Fraser 2003).

This paper illustrates the use of a novel approach to welfare assessment with existing data, with the aim of examining the benefits of this approach in comparing welfare outcomes and identifying the assumptions and limitations involved.

Materials and methods

Management scenarios compared

The welfare assessment framework was developed for five management scenarios: 1) conventional surgical mulesing; 2) surgical mulesing with commercially available topical anaesthetic pain relief; 3) a mulesing alternative using clips to remove breech skin wrinkle through avascular necrosis (clips); 4) no mulesing with unselected animals; and 5) no mulesing with animals selected for a plain breech after three generations of flock selection.

The animal stress response data for use in the risk assessment framework were obtained from Paull *et al* (2007, 2008) for surgical mulesing with and without pain relief, and from Hemsworth *et al* (2009) for clips. All these three sources included conventional surgical mulesing and controls as common treatments. Animal stress response data for flystrike were sourced from Colditz *et al* (2005). Data on the relative risk of flystrike for mulesed and unmulesed sheep were obtained from Counsell (2001). Data on the relative risk of flystrike for sheep with and without three generations of flock selection for breech-strike resistance were obtained from genetic estimates arising from a selection project at The Commonwealth Scientific and Industrial Research Organisation (CSIRO) (J Smith, personal communication 2009), with an estimation of a reduction of one unit in breech wrinkle score resulting in a 50% reduction in breech strike (Breech Strike Genetics 2008).

It is also important to note that the scenarios were evaluated in the absence of any other possible strategies for managing breech-strike risk, such as increased surveillance of sheep for flystrike, increased crutching (shearing around the breech and hindquarters) and increased strategic use of preventative chemical insecticides.

Incorporation of uncertainty using stochastic simulation

For all scenarios, the data used were the means of individual animal responses to the procedure or scenario; more detailed descriptions of the data (standard errors or individual animal data) were not available. However, in order to illustrate the use and value of incorporating into the calculations uncertainty associated with estimates of welfare indicator values, we fabricated appropriate uncertainty measures in order to construct uncertainty distributions for each welfare indicator value and each probability estimate. Probabilities of sheep developing flystrike (see below and Table 1) were represented in stochastic simulation using Beta distributions parameterised using the expected value (point estimate) and the assumption that all such estimates were derived from observations of 1,000 sheep. Thus, for example, sheep in the pastoral zone are estimated to have a

https://doi.org/10.7120/09627286.22.2.267 Published online by Cambridge University Press

Table 1 Point estimates of the probability of flystrike for various classes of animal in a typical year (Counsell 2001).

Animal class		Farming zone		
		Pastoral	Wheat-sheep	High rainfall
Mulesed	Young	0.01	0.03	0.05
	Adult	0.01	0.03	0.05
Unmulesed	Young	0.15	0.40	0.70
	Adult	0.26	0.44	0.62
Unmulesed selected	Young	0.08	0.20	0.35
	Adult	0.13	0.22	0.31

Table 2 Values for welfare indicators used in the risk assessment framework for mulesing and clip procedures; actual means taken from research data, with fabricated standard errors, as used in stochastic simulation, in brackets.

Phase	Indicator	Mulesing	Alternatives		No mulesing
			Pain relief	Clips[3]	
Acute phase	*Cortisol*				
	Peak increase (nmol L^{-1})	115 (25)[1]	90 (20)[1]	48 (10)	0
	Mean increase (nmol L^{-1})	59 (10)	50 (12)	34 (8)	0
	Duration of increase (h)	6 (1)	6 (1.5)	1 (0.15)	0
	Abnormal behaviour day 1 (% of time)	19.7 (5.0)[1]	9.2 (2.5)[1]	1.0 (0.25)	0
Sub-acute phase	*Cortisol*				
	Mean increase (nmol L^{-1})	31.4 (7.6)[2]	31.4 (4.5)[1]	23.3 (6.0)	0
	Duration of increase (h)	66 (14)	66 (17)	48 (12)	0
	Haptoglobin				
	Day 3 increase (mg ml^{-1})	2.4 (0.5)[2]	2.4 (0.5)[1]	0.9 (0.1)	0
	Duration of increase (h)	72 (6.5)	72 (12)	72 (15)	0
	Abnormal behaviour				
	Day 2 to 3 presence (% of time)	12.9 (2)[3]	12.9 (2)[1,3]	3.1 (0.7)	0
	Duration of increase (h)	72+ (17)[1]	72+ (12)[1,3]	72 (15)	0
	Weight change				
	Day 1 to 7 (g per day)	–147 (25)[1]	–203 (35)[1]	+60 (10)	+87 (20)[1]

[1] Paull *et al* (2007); [2] Paull *et al* (2008); [3] Hemsworth *et al* (2009).
Notations in columns and at the top of cells apply to the entire columns/cell unless other notations are given.

1% chance of developing flystrike each year. This was represented using a Beta (11, 991) distribution (Vose 2008).

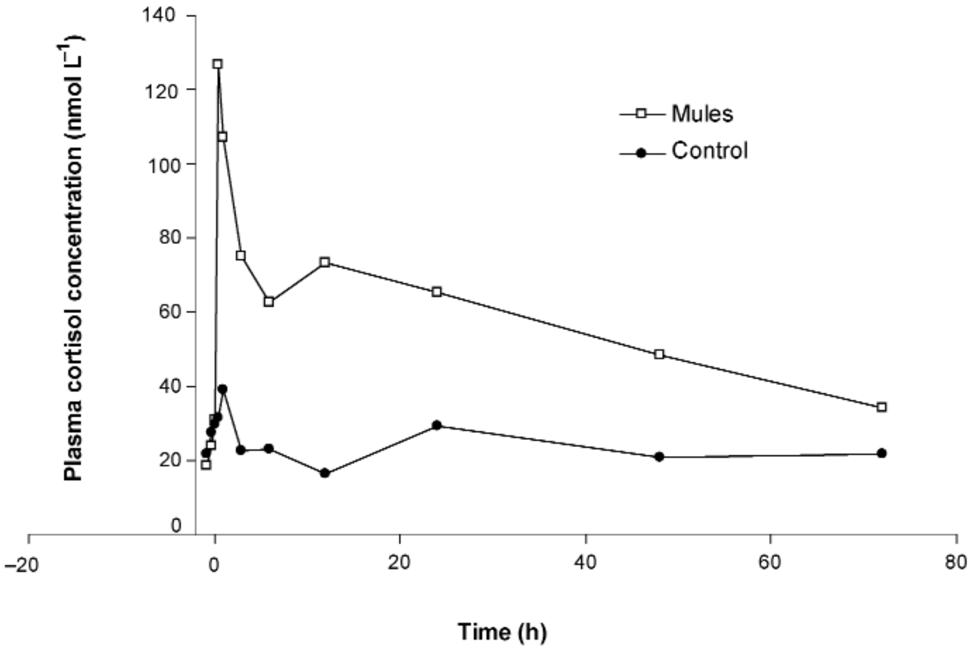
All other continuous variables were assumed to be normally distributed, and standard errors of the mean values derived from the relevant research papers were fabricated and used to represent uncertainty surrounding the mean using a normal distribution. For example, referring to Table 2, the peak increase in acute phase cortisol in sheep treated with clips was represented as Normal (48, 10) in stochastic simulation. Fabricated standard errors were kept deliberately low in order to minimise nonsensical (negative) values being generated by a normal distribution. While commonly assumed in presenting experimental findings, the normal assumption is not literally valid for such variables, and its use in simulation will generate invalid values; and in the absence of actual data we could not determine any more appropriate alternative.

Selection of welfare indicators

The adverse welfare outcomes of interest in this assessment are: (i) for mulesing and alternatives — pain associated with the intervention/procedure; and (ii) for flystrike — pain and illness arising from flystrike.

For each type of welfare challenge to the sheep (mulesing, clips, and flystrike), welfare indicators were chosen, each of which represented a single systemic response to the

Figure 1



Typical cortisol response to the mulesing procedure in sheep (data from Paull *et al* 2008).

challenge. Thus, for mulesing and the clip procedure, the welfare indicators were: the physiological stress-responsive hormone, cortisol; the inflammatory marker, haptoglobin; abnormal behaviour following the procedure; and changes in animal weight. These measures broadly reflect the recognised key components of animal welfare, namely behaviour, physiology and health, and productivity (Fraser 2003). It is recognised that none of these measures provides a direct assessment of pain, such as might be provided by neurophysiological testing. However, as for mulesing and flystrike, it is unlikely that most scenarios that could be examined using this method for welfare assessment would have a complete set of direct pain measures available.

To calculate the welfare impact of the mulesing operation, it was necessary to examine the welfare indicators over both acute and sub-acute time-frames. This is because the stress response to mulesing typically exhibits a biphasic pattern (Paull *et al* 2008), in which there is an initial response, apparently to the pain of the incision, followed by a response likely to reflect the inflammation and tissue trauma resulting from the procedure (Figure 1). Furthermore, the benefits of topical anaesthetic pain relief are only present during the acute phase (Paull *et al* 2007). The welfare indicators for the acute period were cortisol (0 to 6 h) and abnormal behaviour (for a 12-h period on day 1), whereas the welfare indicators for the sub-acute period were cortisol (6 to 72 h), abnormal behaviour (during days 2 and 3), haptoglobin concentration (mean on day 3 and duration of haptoglobin increase) and bodyweight change (between days 1 and 7). Differences in data collection patterns in the reference material meant that it was not possible to align the time-periods for each variable perfectly. The welfare indicators for the clip procedure were the same as for mulesing.

For flystrike, the welfare indicators were also cortisol as an indicator of physiological stress, haptoglobin data as a marker of tissue trauma and inflammation, changes in behaviour, and weight change. Data on cortisol, haptoglobin and bodyweight were obtained from Colditz *et al* (2005). Although Colditz *et al* (2005) did not measure animal behaviour, their study recorded changes in body temperature, and we made the assumption, based on clinical experience (IG Colditz, personal communication 2007), that the duration of abnormal behaviour of an animal would correspond to the duration of increased body temperature induced by the flystrike.

Because the study of Colditz *et al* (2005) used an experimental infection and treated sheep with insecticides after ten days, some further assumptions were made on the typical duration of natural flystrike infections of sheep, based on the data of Counsell (2001) for different sheep farming zones within Australia. Counsell (2001) categorised flystrike risk for different areas depending on the climate, the size of the farms and the farming system. The areas are termed the high rainfall zone, the wheat-sheep belt zone, and the pastoral zone. The high rainfall zone, as its name implies, is wetter, and contains more closely settled farming areas. The wheat-sheep zone is drier, with wool production mixed with cropping on farms. The pastoral zone is drier again, with larger farms where sheep and/or

https://doi.org/10.7120/09627286.22.2.267 Published online by Cambridge University Press

Table 3 Mean values for the welfare indicators used in the risk assessment framework for flystrike, with fabricated standard errors, as used in stochastic simulation, in brackets (Colditz *et al* 2005).

Indicator	Pastoral zone	Wheat-sheep zone	High rainfall zone
Cortisol			
Peak increase (nmol L^{-1})[1]	189 (40)	189 (40)	189 (40)
Mean increase (nmol L^{-1})[2]	157.6 (20)	157.6 (30)	129.3 (28)
Duration of increase (days)[3]	6 (1.05)	6 (1)	3 (0.7)
Haptoglobin			
Peak increase (mg ml^{-1})[4]	4.5 (1)	4.5 (1.2)	3.2 (0.6)
Mean increase (mg ml^{-1})[5]	2.9 (0.8)	2.9 (0.55)	1.8 (0.4)
Duration of increase (days)	13 (2)	13 (3)	8 (1.7)
Abnormal behaviour[6]			
Duration of increase (days)	9 (2)	9 (1.6)	6 (1)
Weight change			
Day 1 to 7 (g per day)	–374 (75)	–374 (75)	–374 (75)

[1] Day 4; [2] Days 4 to 9; [3] Based on detection scenarios explained in text; [4] Day 9; [5] Days 3 to 15; [6] Assumed from duration of pyrexia.

cattle are managed under very extensive conditions. We assumed that in the pastoral zone and wheat sheep belt, the farmers check sheep once a week (J Smith, personal communication 2008), so flystrike would be detected after a maximum of seven days of visible infection. We assumed that in these two zones, on average, flystruck sheep underwent the same consequences as described by Colditz *et al* (2005) (ie after ten days of actual infection). In the high rainfall zone, we made the assumption that the sheep would be checked twice a week (J Smith, personal communication 2008), so flystrike would be detected after 5 days of infection at worst. The durations of abnormal values for welfare indicators for flystrike were therefore based on these durations, together with the findings of Colditz *et al* (2005) that when sheep were treated, cortisol returned to baseline within one day, haptoglobin within 3 to 5 days, and pyrexia (elevated body temperature) within two days.

Cortisol data differed between references for the common procedure of mulesing (Paull *et al* 2007, 2008; Hemsworth *et al* 2009). To account for these differences, cortisol data from Paull *et al* (2008) and Hemsworth *et al* (2009) were linearly adjusted to align with the mean data in Paull *et al* (2007). For example, the cortisol peak for mulesing in Hemsworth *et al* (2009) was 159.7 nmol L^{-1}, compared with 133.1 nmol L^{-1} in Paull *et al* (2007). Accordingly, in order to maintain treatment relativity, all the cortisol values extracted from Hemsworth *et al* (2009) were multiplied by 133.1/159.7. Finally, all cortisol values were expressed as cortisol increase above the recognised baseline (20 nmol L^{-1}; Paull *et al* 2007, 2008).

Table 2 presents the values for the welfare indicators used in the welfare assessment for the mulesing and clip procedures. Table 3 presents the values for the welfare indicators used in the welfare assessment for flystrike.

Scaling of welfare indicators

For each welfare indicator, a scaled welfare measure (SWM) was calculated on a linear and continuous ten-point scale. For variables ranging from a potential 0 value to a measured maximum in the source references (haptoglobin, cortisol and abnormal behaviour), the SWM was calculated by scaling the value for a particular procedure against the rounded maximum value measured across all the reference papers. The ranges used to calculate the SWM were thus 0 to 5.0 mg ml^{-1} for haptoglobin (Colditz *et al* 2005), and 0 to 20% of total time spent in abnormal behavioural postures (eg hunched standing) (Paull *et al* 2007). For change in bodyweight, the range was +100 to –400 g per day (Colditz *et al* 2005; Paull *et al* 2007). For cortisol, a range of 0 to 300 nmol L^{-1} was used, based on the maximum value recorded in the reference studies (Colditz *et al* 2005).

Because the impact of a procedure is also reflected by the duration of the animal's stress response, SWM were also calculated for the duration of increases in cortisol and haptoglobin and the duration of changes in behaviour. The ranges used were 0 to 312 h for haptoglobin, 0 to 144 h for cortisol and 0 to 216 h for behaviour, and were based on the data from Colditz *et al* (2005) and the estimates for the duration of flystrike infection in the three farming zones, as described above.

Several variables thus had multiple contributing elements (eg haptoglobin), and the SWMs for each of these elements were averaged to derive the overall SWM for the variable. Table 4 presents the actual SWMs used for the different mulesing and clip scenarios, and Table 5 presents the SWMs for flystrike. The 'Impact' of a particular challenge (eg mulesing) was then calculated as the mean of the SWMs (Tables 4 and 5).

Table 4 Scaled welfare measures (SWM) for mulesing and clip procedures.

Phase	Indicator	Mulesing	Alternatives		No mulesing
			Pain relief	Clips	
Acute phase	*Cortisol*				
	Peak increase	3.8	3.0	1.6	0.0
	Mean increase	2.0	1.7	1.1	0.0
	Duration of increase	0.4	0.4	0.1	0.0
	Mean for cortisol	2.1	1.7	0.9	0.0
	Abnormal behaviour day 1	9.9	4.6	0.5	0.0
	Overall mean for acute phase	6.0	3.2	0.7	0.0
Sub-acute phase	*Cortisol*				
	Mean increase	1.1	1.1	0.8	0.0
	Duration of increase	4.6	4.6	3.3	0.0
	Mean for cortisol	2.8	2.8	2.1	0.0
	Haptoglobin				
	Day 3 increase	4.8	4.8	1.8	0.0
	Duration of increase	2.3	2.3	2.3	0.0
	Mean for haptoglobin	3.6	3.6	2.1	0.0
	Abnormal behaviour				
	Day 2 to 3 presence	6.5	6.5	1.6	0.0
	Duration of increase	3.3	3.3	3.3	0.0
	Mean for behaviour	4.9	4.9	2.4	0.0
	Weight change				
	Day 1 to 7	4.9	6.1	0.8	0.3
	Overall mean for acute phase	4.1	4.3	1.8	0.1
Overall mean		5.0	3.7	1.3	0.03

Table 5 Scaled welfare measures (SWM) for flystrike.

Indicator	Pastoral zone	Wheat-sheep zone	High rainfall zone
Cortisol			
Peak increase	6.3	6.3	6.3
Mean increase	5.3	5.3	4.3
Duration of increase	10.0	10.0	5.0
Mean for cortisol	7.2	7.2	5.2
Haptoglobin			
Peak increase	9.0	9.0	6.4
Mean increase	5.8	5.8	3.6
Duration of increase	10.0	10.0	6.2
Mean for haptoglobin	8.3	8.3	5.4
Abnormal behaviour			
Duration of increase	10.0	10.0	6.7
Weight change			
Day 1 to 7	9.5	9.5	9.5
Overall mean	8.7	8.7	6.7

Calculation of severity of welfare challenge

The severity of welfare challenge (SWC) for a particular situation was defined as SWC_x = Impact_x × Pr(x), where Pr(x) was defined as the probability of challenge x occurring. Therefore, the probability of mulesing for mulesed animals was 1, but the probability of flystrike varied according to farming zone and mulesing status. The estimates of probability of flystrike in a typical year were obtained from Counsell (2001) and are presented in Table 1. The benefits of flock selection for breech-strike resistance were incorporated as a halving in the incidence of flystrike in response to three generations of selection. The efficacy of clip treatment in reducing flystrike was taken to be the same as that of mulesing.

The SWC was calculated for mulesed, clip-treated, mulesed with pain relief, unmulesed and unmulesed selected animals. Because mulesing (or equivalent procedures) only occur during the first year of life, but flystrike can occur every year, the SWC was calculated both for mulesing and equivalents, and for the lifetime of the animal (assumed five years), whereby SWC_lifetime = (SWC_mulesing × 1) + (SWC_flystrike × 5).

https://doi.org/10.7120/09627286.22.2.267 Published online by Cambridge University Press

Table 6(a) Severity of welfare challenge (SWC) results for five mulesing and flystrike management scenarios and three farming zones.

	Mulesing	Alternatives		No mulesing	
		Pain relief	Clips	Unselected	Selected
Year 1					
Pastoral	5.1	3.8	1.4	1.3	0.7
Wheat-sheep	5.3	4.0	1.5	3.5	1.8
High rainfall	5.3	4.1	1.6	4.7	2.4
Lifetime					
Pastoral	5.4	4.2	1.7	10.4	5.2
Wheat-sheep	6.3	5.1	2.6	18.9	9.5
High rainfall	6.7	5.4	3.0	21.3	10.7

Table 6(b) Severities of welfare challenge (SWC) for five mulesing and flystrike management scenarios and three farming zones; median (in bold) (95% certainty interval) of stochastic simulation output distributions.

	Mulesing	Alternatives		No mulesing	
		Pain relief	Clips	Unselected	Selected
Year 1					
Pastoral	**5.0** (3.9–5.4)	**3.8** (3.2–4.5)	**1.4** (1.2–1.6)	**1.3** (1.0–1.5)	**0.7** (0.5–0.8)
Wheat-sheep	**5.1** (4.0–5.6)	**4.0** (3.3–4.7)	**1.5** (1.3–1.7)	**3.4** (2.9–3.8)	**1.7** (1.4–2.0)
High rainfall	**5.2** (4.1–5.6)	**4.1** (3.4–4.8)	**1.6** (1.4–1.8)	**4.7** (4.0–5.3)	**2.4** (2.0–2.7)
Lifetime					
Pastoral	**5.3** (4.2–5.8)	**4.2** (3.5–4.9)	**1.7** (1.5–2.0)	**9.9** (8.4–11.3)	**5.0** (4.1–5.9)
Wheat-sheep	**6.1** (5.0–6.8)	**5.0** (4.2–5.8)	**2.6** (2.1–3.0)	**18.0** (15.6–19.9)	**9.0** (7.7–10.3)
High rainfall	**6.5** (5.4–7.2)	**5.4** (4.6–6.2)	**3.0** (2.5–3.5)	**21.1** (17.9–23.8)	**10.5** (8.9–12.1)

Results

The SWC for the five management scenarios for the first year of life and for the whole of life are presented in Table 6(a). Because mulesing is performed on lambs, the SWC for the first year of life was greatest for the mulesing scenario (5.1 to 5.3), regardless of farming zone. The highest lifetime SWC (21.3) was for unselected, unmulesed animals in the high rainfall farming zone. The mulesing alternatives, particularly clips, produced lifetime SWC values that were lower than those for mulesed animals. Table 6(b) shows summaries of the output distributions for SWCs (medians, 2.5 and 97.5 percentiles) following simulation incorporating uncertainty distributions for welfare indicator estimates and probability estimates. These medians differ slightly from the deterministic calculations in Table 6(a), but the relationships among the SWCs are the same.

Discussion

In this study, the welfare assessment model produced results that were broadly in agreement with expert consensus on mulesing and flystrike welfare risks in the Australian Merino (Beveridge 1984; James 2006; Lee & Fisher 2007). This does not necessarily indicate that a model such as this is 'right', but that it may serve a useful purpose in providing an ancillary viewpoint in the evaluation of a welfare issue.

Incorporating uncertainty surrounding estimates for welfare indicators and probabilities of flystrike generated output distributions for SWCs whose medians were, predictably, closely comparable to the deterministic estimates. The uncertainty distributions used in this simulation were fabricated due to lack of adequate information to determine the 'true' distributions suggested by the data. It is clear from Table 6(b), however, that uncertainty associated with welfare indicator and probability estimates results in uncertainty in interpretation of the lifetime SWCs. We deliberately used low standard errors to represent uncertainty in these continuous variables. Sensitivity analysis can be used to evaluate the significance of the uncertainty surrounding specific welfare indicator estimates, thus determining where additional research data are critical for the evaluation of lifetime SWCs for the different scenarios; we have not included such analysis here since the stochastic simulation was included for illustrative purposes only, using fabricated distributions.

Although this study was conducted to evaluate the welfare assessment approach of Paton *et al* (2013) for a practical situation, the results were in alignment with the viewpoint that, historically, in the absence of alternative strategies, mulesing represented a lifetime welfare improvement for Merino sheep under high flystrike challenge. This is not to imply that mulesing was the best strategy that could have been developed, or that it is currently a sustainable practice from an animal welfare standpoint. The results of the model are also in alignment with the position that pending the widespread identification, selection and dissemination of Australian Merino sheep with higher levels of breech-strike resistance, low impact methods of altering the breech conformation of Merinos to induce breech-strike resistance represent an improvement both on conventional mulesing and on having sheep of poor breech-strike genetics in a high fly risk environment.

The data used in this study were collected for a variety of purposes and their use to compare animal welfare outcomes depends on developing a framework where semi-quantitative scales can be produced to estimate welfare outcomes. The data may not be ideal for making welfare comparisons but are likely to be the best available for this purpose. The data were collected from Australian commercial sheep flocks, so management practices, such as how often sheep can be checked, are limited by the practicalities of these extensive systems.

The EFSA panel on Animal Health and Welfare (2012) in its *Guidance on Risk Assessment for Animal Welfare* uses an example to compare two scenarios with a series of possible different welfare outcomes. The authors use an example where a series of sequential disease events, (arranged in a scenario tree) stemming from a mutually exclusive difference in management (using sexed or unsexed semen in cattle), to estimate the effects of a series of outcomes. This scenario requires a different risk calculation methodology than used as an example in this paper. Estimated welfare scores (based on expert opinion) for each outcome are added and probabilities are multiplied, then the cumulative probability and welfare score outcomes are multiplied together to give an expected welfare score. These scores are then added for all outcomes of each scenario and compared. In this paper, probability and impact are multiplied to give likely impact for that welfare outcome. These are averaged within welfare effect categories (for this example, acute and sub-acute physiological, behavioural and production) and then added to give the SWC to be compared.

It is worth noting, however, that even for an example such as flystrike and mulesing where there exist considerable data, a number of assumptions were required to place the data into a common framework for calculation and comparison. Although common welfare-related variables were measured across studies examining conventional mulesing and other options for altering existing breech characteristics such as mulesing with pain relief and clips (Paull *et al* 2007, 2008; Hemsworth *et al* 2009), there was not a complete match with variables measured for flystrike (Colditz *et al* 2005). This necessitated the assumption that abnormal behaviour in the flystruck sheep would be associated with the presence of pyrexia (IG Colditz, personal communication 2007). Although this assumption was not unreasonable, it would be preferable to have common variables across the scenarios being compared.

Although this may seem to be an arbitrary approach to using data which are not directly comparable, it may nevertheless be reasonable in the context of this framework. A series of rules will need to be developed, as exists for other areas where risk analysis is applied (Paton *et al* 2013), to ensure this methodology is applied consistently and therefore provides the most value to animal welfare science. This may be particularly important in the case of behavioural data which are likely to show more variation in the methods of measurement. Another important area for the development of rules for welfare assessment is the independence of different data. As data collected on physiological, behavioural and production parameters are always going to be in some way interdependent, the development of rules to overcome undue weighting of particular welfare measures will be particularly important. These issues are discussed in more detail by Paton *et al* (2013).

This leads to a necessary consideration of a principle relevant to all data models, including the welfare risk assessment approach described by Paton *et al* (2013) and utilised here. The principle of 'garbage in, garbage out' warns us that the usefulness of a model is dependent on the relevance and accuracy of the data used in its construction. In the current study, the lifetime welfare challenge of flystrike was substantially greater than that of the breech conformation operations, leading to the fact that relatively minor changes in flystrike risk values would result in significant changes to the final SWC results. It is well known that flystrike risk for Merino sheep in Australia varies substantially, not just from farming zone to zone, but in relation to both weather and season (Watts *et al* 1979). Accordingly, a particularly bad (or good) flystrike season could cause the results of the model to be of little relevance for a particular group of sheep and a particular time. Similarly, the use of the risk assessment approach to compare animal welfare across different farming systems for a particular species, such as housing for pigs, is dependent on data derived from a consistent quality of management and stockmanship. Because of the importance of stockmanship for animal welfare, individual instances of poor management would result in poorer welfare outcomes for particular housing systems, even if these housing types were associated with lower SWC values from a welfare assessment calculation. Variability in flystrike incidence and in other model parameters might be incorporated into the model using stochastic modeling techniques, but in this study we were unable to do so since the papers from which we took the data for the analysis did not contain sufficient information.

One area where the results of the model and the consensus of scientific and industry experts appear to differ is in the benefits of genetic selection for breech-strike resistance in improving overall welfare. The SWC value for no mulesing with genetic selection for sheep in the high rainfall zone

https://doi.org/10.7120/09627286.22.2.267 Published online by Cambridge University Press

was 10.76, compared with the corresponding value for mulesed sheep of 6.68. In contrast, it is the consensus of animal welfare interest groups, scientists and industry representatives that genetic selection for enhanced breech-strike resistance through reducing breech wrinkle and increasing wool-free skin will produce the best animal welfare benefit. However, this apparent lack of consistency should be interpreted with caution. In our model, we assumed a 50% reduction in flystrike risk after a fixed period of selection. However, continued selection for a plain breech would lead to further progress in the trait. It was estimated by James (2006), based on the review of a number of studies, that the heritability for breech skin wrinkle was moderate to high, offering opportunity for substantial progress. Furthermore, it is worth noting that mulesing was developed in an era where there were few anti-fly preventative strategies available. Although we did not include this option in the model, it is reasonable to suggest that the strategic use of modern insecticidal flystrike prevention treatments could be used to greatly reduce the flystrike risk of sheep that represent a midpoint in genetic progress toward breech-strike resistance.

The output from this risk assessment process can provide an estimate of comparative animal welfare outcomes. However, there are issues, such as assumptions of equivalence, normalisation of data and aspects of the calculation of the outputs which should be the subject of the development of rules to ensure this methodology can be consistently applied.

In conclusion, this study suggests that the careful use of a semi-quantitative risk assessment model for situations where there are robust data with extensive measurement of common variables may permit the evaluation of animal welfare across different management scenarios for the same species. The study proposes one approach to using this framework which provides an example of how data from different sources could be used. Like most new methodologies, the refinement of its use will further enhance its value as a tool for comparing animal welfare outcomes. Like almost all types of welfare assessment, the risk assessment approach is probably best used in combination with other information, rather than being relied upon as the sole determinant of what is right and what is inappropriate for animal welfare.

References

Beveridge WIB 1984 The origin and early history of the mules operation. *Australian Veterinary Journal 61*: 161-163. http://dx.doi.org/10.1111/j.1751-0813.1984.tb07222.x

Breech Strike Genetics 2008 Flystrike results 2007-08. *Breech Strike Genetics Newsletter Issue 2*: 2. http://www.wool.com/Grow_Animal-Health_Flystrike-prevention_Genetic-and-breeding.htm

Colditz IG, Walkden-Brown SW, Daly BL and Crook BJ 2005 Some physiological responses associated with reduced wool growth during blowflystrike in Merino sheep. *Australian Veterinary Journal 11*: 695-699. http://dx.doi.org/10.1111/j.1751-0813.2005.tb13053.x

Counsell D 2001 A benefit-cost analysis of the mules operation to the Australian sheep flock. In: Larsen J and Marshall J (eds) *Proceedings of the Australian Sheep Veterinary Society Volume 11*: 10-14

EFSA 2012 Guidance on risk assessment for animal welfare. *EFSA Journal 10(1)*: 2513

Fraser D 2003 Assessing animal welfare at the farm and group level: the interplay of science and values. *Animal Welfare 12*: 433-443

Hemsworth PH, Barnett JL, Karlen GMA, Fisher AD, Butler KL and Arnold NA 2009 Effects of mulesing and alternative procedures to mulesing on the behaviour and physiology of lambs. *Applied Animal Behaviour Science 117*: 20-27. http://dx.doi.org/10.1016/j.applanim.2008.12.007

James PJ 2006 Genetic alternatives to mulesing and tail docking in sheep: a review. *Australian Journal of Experimental Agriculture 46*: 1-18. http://dx.doi.org/10.1071/EA05100

Lee C and Fisher AD 2007 Welfare consequences of mulesing of sheep. *Australian Veterinary Journal 85*: 89-93. http://dx.doi.org/10.1111/j.1751-0813.2007.00114.x

Paton MW, Martin PAJ and Fisher AD 2013 Risk assessment: a useful tool to evaluate animal welfare. *Animal Welfare 22*: 277-285. http://dx.doi.org/10.7120/09627286.22.2.277

Paull DR, Lee C, Colditz IG, Atkinson SJ and Fisher AD 2007 The effect of a topical anaesthetic mixture, and systemic flunixin or carprofen, on modifying the pain and stress responses of Merino lambs to mulesing. *Australian Veterinary Journal 85*: 98-106. http://dx.doi.org/10.1111/j.1751-0813.2007.00115.x

Paull DR, Lee C, Atkinson SJ and Fisher AD 2008 The effects of meloxicam or tolfenamic acid administration on the pain and stress responses of Merino lambs to mulesing. *Australian Veterinary Journal 86*: 303-311. http://dx.doi.org/10.1111/j.1751-0813.2008.00325.x

Vose D 2008 *Risk Analysis: A Quantitative Guide, Third Edition.* John Wiley & Sons: Chichester, UK

Watts JE, Murray MD and Graham NPH 1979 The blowfly strike problem of sheep in New South Wales. *Australian Veterinary Journal 55*: 325-334. http://dx.doi.org/10.1111/j.1751-0813.1979.tb00419.x