Uncontrolled pain: a call for better study design

Timothy H. Hyndman[1,2]*, Ross S. Bowden[3],
Andrew P. Woodward[4], Daniel S. J. Pang[5] and
Jordan O. Hampton[2,6]

[1]School of Veterinary Medicine, Murdoch University, Murdoch, WA, Australia, [2]Harry Butler Research
Institute, Murdoch University, Murdoch, WA, Australia, [3]School of Mathematics, Statistics, Chemistry
and Physics, Murdoch University, Murdoch, WA, Australia, [4]Faculty of Health, University of Canberra,
Canberra, ACT, Australia, [5]Faculty of Veterinary Medicine, University of Calgary, Calgary, AB, Canada,
[6]Faculty of Science, University of Melbourne, Parkville, VIC, Australia

Studies assessing animal pain in veterinary research are often performed primarily
for the benefit of animals. Frequently, the goal of these studies is to determine
whether the analgesic effect of a novel treatment is clinically meaningful,
and therefore has the capacity to improve the welfare of treated animals. To
determine the treatment effect of a potential analgesic, control groups are
necessary to allow comparison. There are negative control groups (where pain is
unattenuated) and positive control groups (where pain is attenuated). Arising out
of animal welfare concerns, there is growing reluctance to use negative control
groups in pain studies. But for studies where pain is experimentally induced, the
absence of a negative control group removes the opportunity to demonstrate
that the study methods could differentiate a positive control intervention from
doing nothing at all. For studies that are controlled by a single comparison
group, the capacity to distinguish treatment effects from experimental noise
is more difficult; especially considering that pain studies often involve small
sample sizes, small and variable treatment effects, systematic error and use pain
assessment measures that are unreliable. Due to these limitations, and with a
focus on farm animals, we argue that many pain studies would be enhanced
by the simultaneous inclusion of positive and negative control groups. This
would help provide study-specific definitions of pain and pain attenuation,
thereby permitting more reliable estimates of treatment effects. Adoption of our
suggested refinements could improve animal welfare outcomes for millions of
animals globally.

KEYWORDS

animal ethics, study design, randomized controlled trials, analgesia, farm animals

1 Introduction

Imagine a study where the goal is to assess the effectiveness of a new analgesic drug for
use in cattle. Assume the analgesic has relatively few adverse side effects and is cost-effective
in clinical applications. If shown to be sufficiently effective, this analgesic will likely be adopted
and used on millions of cattle for routine husbandry procedures that are painful, such as
surgical castration. You design a trial to assess the analgesic efficacy of this new drug in young
bulls. In your trial, you will have one control group of bull calves that will be castrated and
receive the gold standard analgesic (the most effective analgesic). You have a second group,
the experimental group, that will also be castrated but will receive the new analgesic. The

statistical analysis of the trial shows that the pain in both groups was similar. With this evidence that the new analgesic performed similarly to the gold standard, it is subsequently adopted and utilized for painful procedures in millions of young bulls.

Consider an alternate design with a third group of cattle where no analgesia is provided at the time of castration. You compare the pain in all three groups and find it is similar. Interpretation of the results is now more complicated than before because the outcome for the gold standard analgesic is now similar to no-analgesia. Perhaps the method of pain assessment was inadequate, meaning potentially relevant differences between the three groups was lost in the noise of the pain assessment results. That the pain response was similar between the gold standard analgesic and the novel analgesic is *consistent* with effectiveness of the novel analgesic, but not necessarily *evidence of* its effectiveness. Though the novel analgesic may have been similar to the gold standard in this study, without the no-analgesic control group, it is unwarranted to conclude that it is superior to doing nothing. If it in fact *is not* efficacious, and it is adopted for widespread use, then potentially millions of cattle will continue to suffer from the pain associated with castration. Worse, it may be presumed that these cattle are receiving effective pain relief from the new analgesic, which may dissuade others from further investigation into this area.

The central point here is that without a no-analgesia control group, erroneous conclusions with quantitatively enormous implications can easily be reached. Here, we dissect the need for pain in order to scientifically assess response to potential analgesics in animals that are used in veterinary research; research primarily for the benefit of animals. The purpose of this discussion is not to focus on the validity of statistical methods chosen to analyse data arising from pain experiments. Rather, the focus is on study design. To address this issue, we first review the background to modern pain research in animals.

2 Animal use committees

In most post-industrial nations, the use of animals for research, unlike most other forms of animal use, requires legal permission. To attain this permission, researchers must first submit an application with a detailed description of the project to an institutional animal use committee (1). These have various names globally, such as Institutional Animal Care and Use Committees (United States), Animal Care Committees (Canada), and Animal Ethics Committees (Australia and New Zealand) (2). Animal use committees assess applications proposing the use of animals for research through consideration of what might justify deliberately causing harm to sentient animals. This necessitates the application of ethical theories to these often-contentious questions (3). The ethical principles underpinning scientific animal use have been evolving over the last century, noting there are significant regional differences when it comes to the extent of the requirements made on researchers (4). The source of animals can also influence the decision-making process of an institutional animal use committee. Important distinctions are sometimes made between the use of pet, laboratory, farm or wild animals (5, 6). Our discussion will primarily consider farm animals but many of the principles discussed are relevant to all animal pain research.

3 Research ethics

When considering the use of animals for research purposes, the first requirement is that scientists should strive to minimize harm, typically expressed as the so-called "3Rs" (7), i.e., ways to Reduce the number of animals used to the minimum necessary for scientific validity; to Replace experiments that use live animals with alternative methods; and to Refine procedures for the remaining experiments so as to minimize suffering (8).

The second requirement is that animals should be used only when that use is likely to give rise to genuine benefits to humans (or animals) (9). Put another way, committees are expected to ensure that there is a proper balance between the harm imposed upon animals and the expected benefits of causing that harm. This is known as harm-benefit analysis, and is the core consequentialist ethical underpinning of animal-based research (10). The 3Rs have been implemented as an integral part of the way animal experiments are regulated and reviewed in many countries, and a requirement for consideration of harm-benefit analysis is also widely, but not universally, practiced (11).

The third requirement is to put an absolute cap on the suffering that animals may endure in the course of an experiment. According to this requirement, experiments should not be allowed if they involve *severe* suffering. Of course, this requirement could be seen as a special case of the requirement to Refine procedures. This is because the requirement to Refine is relative to what is possible without sacrificing the goal of the research, but the requirement to avoid severe suffering is absolute. This requirement has, with some limitations, been adopted by the European Union but has so far had little uptake outside Europe (12). All such restrictions placed on animal use must be applied in a way that still allows scientifically robust results to be produced from research, and that requires statistical validity.

4 Sample size

There are a litany of critically-important aspects to sound study design such as randomization, blinding and data handling (13–16) but animal use committees will also consider the necessity of an appropriate sample size before animal-based studies can commence. Studies must be sufficiently large that their estimates are not overly influenced by random variation (17) and to reduce the occurrence of overestimates caused by chance (18). At the same time, researchers are expected to use the fewest animals capable of answering their question under the 3Rs principle of Reduction (19). This means that there is an unavoidable trade-off to achieve a sufficient but not excessive sample size (20). If Reduction is attempted without consideration for statistical rigour, a relatively small number of animals may be harmed, but it is likely that nothing useful will be learned. These ethical concerns become especially heightened when *painful* things are done to animals.

5 Pain and clinical research

Pain is defined in several ways that are generally variations on the theme of "an unpleasant sensory and emotional experience associated with, or resembling that associated with, actual or potential tissue damage" (21). There is a great deal of variation in how animals display

signs of pain (22) and given one aspect of it is emotion, assessing it objectively is challenging, especially in species that have evolved to hide or 'mask' signs of overt pain, such as ruminants (23). Before we explore the details of study design in pain research, a few definitions are necessary.
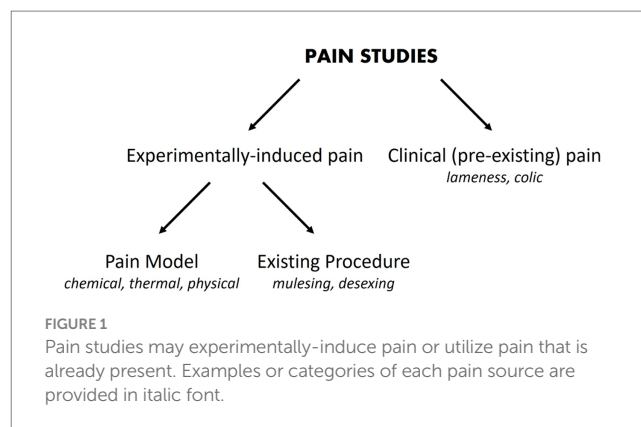
There are important differences between *hypo*algesia and *an*algesia with the former being an attenuation of pain and/or nociception whereas the latter is complete ablation of pain (24). For the purposes of this discussion, the terms analgesia and hypoalgesia are considered synonymous. For pain-based research, there is nearly always the requirement for an 'intervention', defined here as treatments applied to *attenuate* pain (e.g., administering an analgesic). We are not using the term 'intervention' to describe the procedure that *causes* pain in studies where the pain is experimentally-induced.

Then, there is the concept of a result being *clinically-relevant*. Statistical significance refers to whether a value of p is below a chosen level (typically $p < 0.05$) and is often used to determine differences between study groups, but it does not quantitate those differences. Knowing the magnitude of those differences is crucial when deciding whether that difference is clinically-relevant (25). While statistical methods can quantify what occurred, perceptions of clinical relevance are informed by values and norms, and require subjective value judgments. For example, small differences seen between an experimental group and a control group may be statistically significant but the difference may be clinically meaningless. Clinical relevance, and not statistical significance, generally informs clinical practice standards and policy decisions (26). Moreover, increasing calls are being made to stop using statistical significance and hypothesis testing as the primary tool from which a study draws its conclusions. This is due to the arbitrary nature of the convention to have $p < 0.05$ (or any other value) define the level of significance, and because many studies (especially in the biomedical sciences) have small and variable effects, systematic error and noisy measurements (27, 28). The reporting of estimated effect sizes is included in the guidelines of the Consolidated Standards of Reporting Trials (CONSORT) (29).

Policy decisions, such as standards imposed on agricultural producers for routine husbandry practices in farm animals, have the capacity to impact upon orders of magnitude more animals than were involved in any study assessing these practices. For example, sample sizes of ~10 are not uncommon for studies in farmed chickens that assess pain associated with slaughter procedures (30). The results of these studies may then be applied to the billions of chickens that are slaughtered for human consumption every year (31). Therefore, scientific rigour in animal pain studies (especially in livestock) is extremely important for shaping practices that affect large numbers of animals.

6 Pain study design

Investigators may be motivated to perform a pain study for a variety of reasons. There may be an imperative to alleviate pain following a common surgical procedure [e.g., desexing (32)] or a routine husbandry practice [e.g., mulesing (33)], or the primary interest may be in an analgesic that could be used for many causes of pain (34). Regardless of the impetus for the study, a number of choices

PAIN STUDIES

Experimentally-induced pain   Clinical (pre-existing) pain
lameness, colic

Pain Model   Existing Procedure
chemical, thermal, physical   *mulesing, desexing*

FIGURE 1
Pain studies may experimentally-induce pain or utilize pain that is already present. Examples or categories of each pain source are provided in italic font.

need to be made about how the study will be designed. The source of pain (Figure 1) and a suitable way to measure it need to be chosen.

The source of pain may be pre-existing and will be seen commonly in clinical practice (35). Examples include pain arising from lameness (36) and colic (37). Pain may also be experimentally-induced, which can be subdivided into pain caused using standardized approaches (pain models), where the noxious stimulus may be chemical (e.g., formalin, capsaicin), thermal (e.g., heat probes) or physical (e.g., skin incisions) (34, 38), or pain that is induced by existing procedures such as mulesing or surgical desexing.

The method of pain assessment plays an important part in interpreting the results of a pain study. Surrogate measures are often used, especially in animals, as pain is difficult to quantify. Examples include force plate analysis for lameness studies (39), nociceptive threshold testing using a noxious thermal stimulus (40) and the evaluation of acute pain by electroencephalograms following noxious chemical stimulus (41). The quantification of pain will rely on methods that were validated previously, or are validated as part of the study (42). Ideally, pain assessment should be at least species-specific, context-specific, validated and potentially composite (using a suite of measures) (22, 43–45). For simplicity, this discussion will assume that a single outcome measure of pain has been used in a study and that different levels of pain can be distinguished.

7 Control groups

To determine the effect that an intervention (e.g., a novel drug) has on pain, comparisons should be made between a group of study participants that received this intervention (the experimental group) and one or more other groups that provide measures of analgesia and/or no analgesia [the control group(s)] (46). Where the intervention is a drug, there should be evidence that it has an acceptable safety profile in the target species. For a study that contains its own control group(s), the study is internally controlled and inferences can be made about its internal validity (47), such as were the methods used in that study capable of detecting pain at all? Where a study is uncontrolled, control groups from other studies are often used, which are referred to as external, or historical, controls; noting that the use of historical controls is often prone to bias (48, 49). For studies that have a single control group, the control group may be a *positive control* or a *negative control* (50). For a pain study, a positive control involves the provision

of a drug or method that provides measurable and meaningful pain attenuation. A negative control represents the absence of any pain attenuation methods. In drug trials, placebos are often used for the negative control (51), noting that a placebo control group should not be considered interchangeable with a no treatment group (52). The same can be said about sham-procedures being different to no intervention (53, 54).

Before beginning a study that compares pain in an experimental group to pain in a single control (positive or negative) group, it will not be known whether there will be a clinically-relevant difference between the two groups (if it *was* known, there would not be scientific motivation to do the study). This means the investigators need to consider the possibility that there will not be a clinically-relevant difference.

For single control group studies, there are three possible outcomes – the outcome measure of pain in the experimental group can be below, above, or not relevantly different to the outcome measure from the control group. For the scenarios where the outcome measure is above or below the outcome measure of the control group, it will often be assumed that because a (statistically) clear result was obtained, the methods of pain assessment were appropriate to distinguish the groups in a way that then permits a judgment as to whether that difference is clinically-relevant. When there is a null result (i.e., no meaningfully-significant difference between the two groups), the interpretation of the results needs to be more nuanced (55). A true difference between the groups may have been missed due to an inadequate sample size that meant imprecise estimates of the outcome measure in one or both groups prevented the true difference from being distinguished from experimental noise (56).

In a negative (placebo)-controlled study (Figure 2A), a null result is more likely to occur if the assessment of pain is unreliable (and is without bias toward either group). Assuming the experimental intervention truly provides some level of pain relief, and if unreliable methods are the cause of the null result, then the conclusion that the intervention used in the experimental group does *not* provide analgesia, will be false.

In studies controlled only by a positive control group, it is tempting to conclude from a pain study that compares a novel intervention in the experimental group to only a positive control group that a null result (Scenarios 3 and 4 in Figure 2B) reflects equivalence to a positive control; ergo, the novel intervention in the experimental group decreased pain. However, this conclusion is based on an assumption that the positive control is itself superior to a negative control, the so-called 'historical control assumption' (57). The method of pain assessment also needs to be considered. If it is unreliable (and unbiased), then as for the negative-controlled trial, a null result of no relevant difference becomes more likely. This immediately clouds the capacity to distinguish between (a) the true absence of any relevant difference between the experimental group and the positive control group, and (b) whether the method of pain assessment was unreliable. If it is assumed that the intervention does not provide pain relief (and therefore there is truly a difference between the experimental group and the positive control group), then if unreliable methods of pain assessment were the cause of a null result (i.e., the methods have prevented a relevant difference being observable), then the conclusion that the novel intervention *did* work will be false.
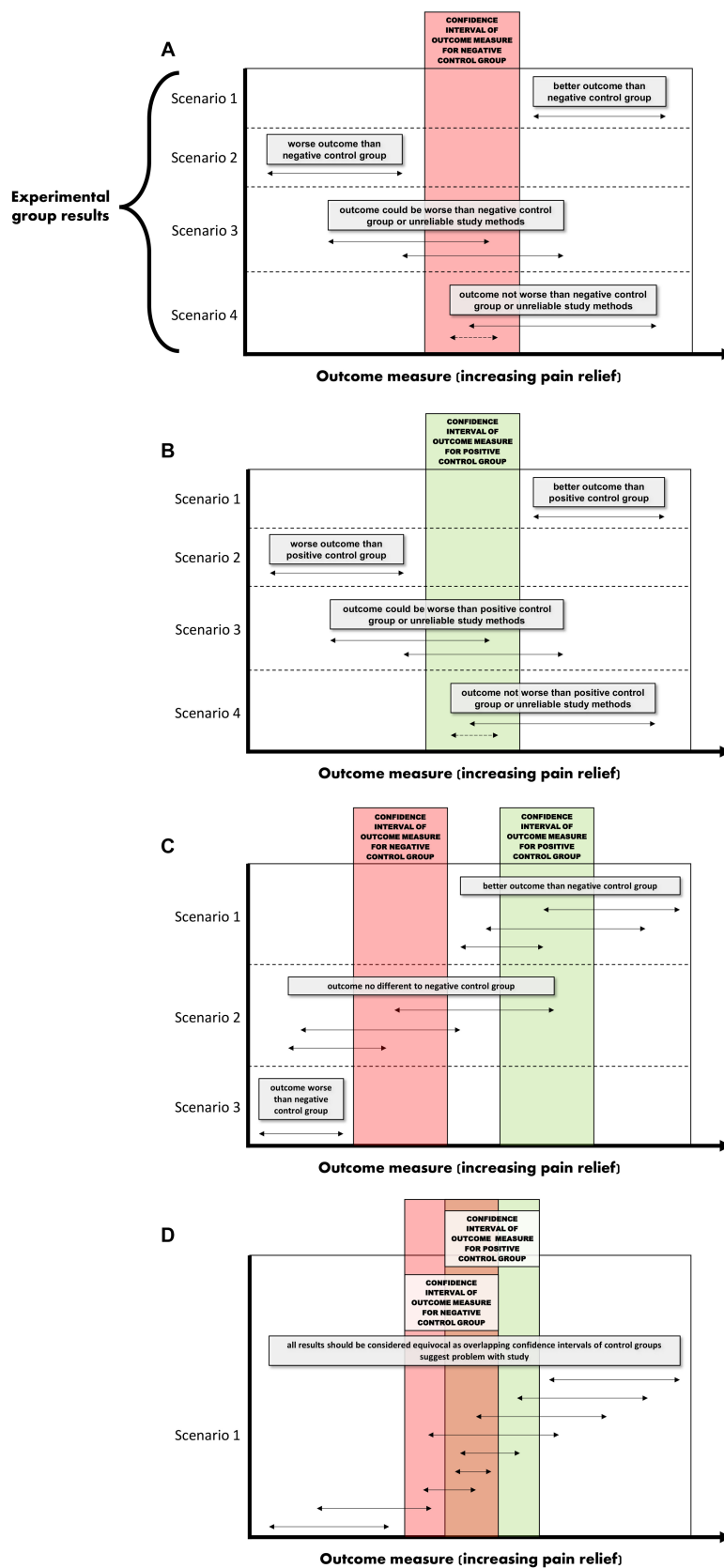
8 Discussion

8.1 Suggestions for improvement

So what does this all mean? If a study only includes a single control group, and if a statistically significant difference is observed between the experimental group and the control group, it provides some evidence that the pain assessment methods used were sufficient to reach the conclusion that there is a difference between the two groups (Scenarios 1 and 2 for Figures 2A,B). In strong contrast to this, when there is no difference between the experimental group and the single control group, then there are two conclusions that cannot be easily distinguished; either the intervention in the experimental group performs similarly to the control, or that the methods of pain assessment were inadequate (Scenarios 3 and 4 for Figures 2A,B).

Confidence in the study conclusions will be enhanced if the study utilises pain assessment methods that have been validated in previous studies (e.g., those reviewed by Muley, Krustev and McDougall (58) for experimental models of inflammatory pain). Adoption of these validated methods in a study with only one control group will offset some (but not all) of the concerns of the diminished internal validity of the study (i.e., by not having positive *and* negative control groups). However, caution needs to be exercised as some of these methods have been shown to have limited reproducibility, or are not broadly translatable (59).
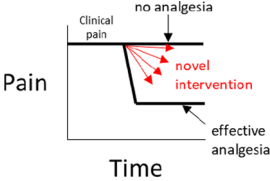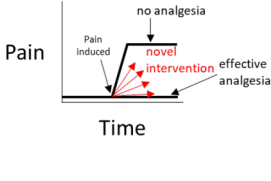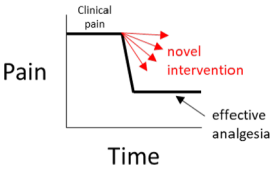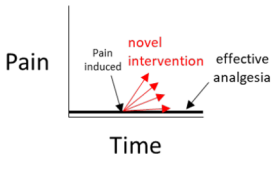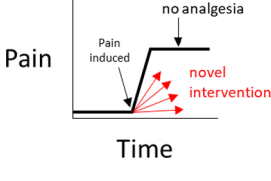
What is more resilient to scientific criticism than using a single control group with a validated method, is the inclusion of a positive control group *and* a negative control group (Figure 2C) (46, 60–65). The inclusion of both a positive and a negative control group provides an opportunity to assess the internal validity of the method of pain assessment (Figures 2C, D). That is, is the study of sufficient internal validity that it can provide study-specific definitions of pain attenuation and no pain attenuation? If so, conclusions can be more confidently drawn about the pain attenuating effect of the novel intervention by comparing the experimental group to these two control groups.

There are certain situations where single-controlled studies may still provide study-specific definitions of pain attenuation and no pain attenuation, while also providing an assessment of the internal validity of a study (Table 1). As described earlier, the source of pain in a study may be experimentally-induced or pre-existing (Figure 1). If pain is experimentally-induced in pain-free animals, then including a negative (placebo) control group would provide a study-specific definition of pain following pain induction (66, 67). Those animals that received the novel intervention could be compared to this negative control group to see if there was pain attenuation, but comparisons between the experimental group and a 'gold standard' level of pain attenuation are absent in this study as there is no positive control group. In contrast, if the pain is pre-existing, then including a positive control group would provide a study-specific definition of pain attenuation (68). Those animals that received the novel intervention could be compared to both the positive control group and their own pre-intervention pain outcome measures to see if there was pain attenuation. However, the absence of a negative control group means that clinical pain that naturally subsides (after the novel intervention is administered) is harder to identify, and the natural phenomenon of 'regression to the mean' remains a problem (69). As a generalization for single-controlled studies, negative-controlled

FIGURE 2

(A–D) Interpretations for hypothetical scenarios where the outcome measure of pain in the experimental group (represented as arrows to reflect confidence intervals) exceeds, is no different, or is below the confidence interval of the outcome measure of pain in the control group(s). Interpretations are included in grey text boxes that overlay each plot and are made in the context of whether the study has a negative control group only (plot **A**), a positive control group only **(B)**, negative and positive control groups that have non-overlapping results **(C)**, or negative and positive control groups with results that overlap **(D)**. The confidence intervals are assumed to be constructed such if they do not overlap then this indicates a statistically significant difference.

TABLE 1 Control groups used in pain studies according to the source of pain.

Study type	Positive-control group[a]	Negative (placebo)-control group[b]	Source of pain	
			Clinical pain (pre-existing pain)	Experimentally-induced pain
Double-controlled study	Yes	Yes	The novel intervention is compared to study-specific definitions of no analgesia and effective analgesia. As the pain was pre-existing, there is no definition of pain-free.	The novel intervention is compared to study-specific definitions of no analgesia and effective analgesia. As the pain was induced, baseline measurements can define pain-free.
Positive-controlled study	Yes	No	The novel intervention is compared only to effective analgesia. There is no negative control group to control for clinical pain that naturally subsides.	The novel intervention is compared only to effective analgesia. There is no negative control group to permit measurement of no pain attenuation or to assess whether pain was induced at all (this assumes the positive control was highly effective).
Negative-controlled study	No	Yes	The novel intervention is compared only to no analgesia. There is no positive control group to permit measurement of pain attenuation.	The novel intervention is compared only to no analgesia. There is no positive control group to permit measurement of pain attenuation.

Novel interventions are compared to negative and/or positive controls. Novel interventions may include surgical procedures and drugs where the pain-attenuating effect of the intervention is unknown. [a]Positive controls may include established surgical procedures and analgesic drugs that are known to provide pain relief. [b]Negative controls are known to not provide any pain relief and may include the deliberate omission of any intervention or the administration of a placebo.

studies are better suited for experimentally-induced pain, and positive-controlled studies are better suited for pre-existing pain.

To this point, it has been assumed that a positive control is available for a pain study. But what if it is not? Remember that the purpose of the positive control is to provide a study-specific definition of pain attenuation. The corollary to this is that the positive control provides a level of analgesia that is meaningfully different to doing nothing. Therefore, positive controls can be found by comparing potential analgesics to negative controls. A value judgment can then be used as to whether the treatment effect of the positive control will make it appropriate to use as a positive control in studies assessing other potential analgesics. In other words, the newly discovered positive control may outperform doing nothing, but is its analgesic efficacy clinically relevant? If it is, then its ongoing use as a positive control is more easily justified for studies that intend to assess pain in the same context (e.g., where the target

species, source of pain and pain measurement method are the same). From a set of positive controls, the so called 'gold standard' (or criterion standard) can be defined as the one with an acceptable safety profile and the greatest analgesic efficacy. It should be noted that a 'gold standard' would not necessarily be used routinely in field settings as it may be cost prohibitive or not registered in a particular species.

It should be noted that there are other methods that can be used to improve the credence of the study conclusions and/or reduce animal suffering. Bayesian analyses can include so-called prior knowledge (produced by other studies) about the expected outcomes of the positive or negative controls (70). Adaptive study designs are where the design of the study is altered (adapted) after experimentation has started (71). For example, response-adaptive randomization is where the treatment allocation ratio(s) is changed in favor of an intervention with demonstrated

beneficial effects (72); or the study is stopped early following an interim analysis that concludes there is sufficient evidence to answer the research question, e.g., a sufficiently precise estimate of the treatment effect has been obtained (73).

8.2 Animal welfare harms and benefits

For pain studies, the inclusion of a positive control (analgesia) group is easily justified if an analgesic is available with clinically-relevant effects that can be reliably quantified by the methods being used in the study. The inclusion of a negative (placebo) control (no-analgesia) group is more problematic because this raises the obvious conflict of balancing the ethical use of research animals against the scientific rigour of the studies they are used for. In humans, ethical objections have been raised to this practice, especially in neonates (74, 75). However, the World Medical Association Declaration of Helsinki justifies the use of a placebo control group if there are "compelling and scientifically sound methodological reasons the use of any intervention less effective than the best proven one, the use of placebo, or no intervention is necessary to determine the efficacy or safety of an intervention" (76). For the majority of pain studies in animals, we argue that there *are* "compelling and scientifically sound methodological reasons" to include positive and negative (no analgesia) control groups. This is especially pertinent where study results will be applied to very large numbers of animals, such as farm animals.

Numerous husbandry practices in livestock management are considered to impose unacceptable pain levels by the standards set by animal use committees or scientific journals. Various arguments have been raised to support the position that analgesia is not required or justified for these procedures (77–79). We acknowledge that the validity of these arguments will rest upon chosen ethical theories, and accompanying ethical commitments and priorities. Nonetheless, some of these reasons may be regarded by some as fallacious (or at least ethically contestable), but some are more widely understandable: the financial value of the farm animal relative to the cost of treatment; a belief that the procedure is not painful; the assertion that young animals feel less pain than adults; logistics of drug administration; a limited understanding of farm animal pain behaviors; a fear of adverse side effects following the provision of analgesia; a limited number of analgesics approved for use in farm animals; and worry about violating food withholding periods resulting in drug residues entering the food chain. As a consequence, massive numbers of farm animals continue to be exposed to painful husbandry procedures—such as debudding, dehorning, hoof trimming, castration, tail docking, ear tagging, ear notching and mulesing—without the provision of analgesia (80–84).

Rather than preventing these (likely) painful husbandry practices from being included in pain assessment studies as comparator groups, we argue that there should be encouragement to do so, to facilitate direct comparisons with alternative methods being investigated. These existing husbandry practices are already being done on large numbers of animals and so utilising some of these animals in experimental or observational studies is a pragmatic approach for performing translational research that is likely to be adopted by industry. This amounts to applying a knowledge-based ethic to learn as much as possible from existing practices (85). In more simple terms, it could be called making the best of a bad situation.

Enumerating the animal welfare implications of our proposed approach is worth exploring. Returning to the hypothetical case study with which we began this paper – we are studying a new analgesic for use in farmed cattle. But this time, we assume that the newly developed analgesic (referred to herein as the experimental analgesic) is truly capable of reducing pain by 50% in animals but this effect has not yet been established by controlled studies. A study is proposed that includes a positive control group where animals receive an analgesic known to be effective (herein referred to as the control analgesic) but is prohibitively expensive for large-scale management practices. The use of a negative (no analgesia) control group is not allowed due to the decision of an animal use committee or the editorial policy of a journal. Furthermore, assume that the pain attenuation observed in the experimental group (given the experimental analgesic) was below that for the positive control group (given the control analgesic). Without a negative control group, it is unclear whether the experimental analgesic did indeed attenuate pain, but less so than the control analgesic, or whether there was no observed analgesic effect at all. Without an appreciation of the important distinction that exists between these two interpretations, it is conceivable that the investigators and/or the industry leaders may (erroneously) conclude that the experimental analgesic should not be used and the industry maintains its position to perform the procedure without any analgesia at all.

The animal welfare cost of the experimental analgesic remaining unused increases the pain experienced by millions of cattle; it is now twice as much compared to if the experimental analgesic was used: this is the forgone benefit from the harm-benefit analysis. Thought of another way, this is the animal welfare cost of *not* using the negative control group. The animal welfare benefit of the decision to disallow an observational negative control group is zero – the animals were subjected to the painful procedure anyway as part of routine management. Clearly, the animal welfare outcomes of the decision to disallow a negative control group are profoundly negative.

9 Conclusion

For situations where there is a reluctance to include a negative control group, we argue that better net animal welfare outcomes will usually result if well-designed studies harm a (relatively) small number of animals through the inclusion of this group. We acknowledge that researchers need to protect their reputations and preserve 'social licence' in an era where animal welfare scrutiny from society can be intense, but that should not come at the cost of scientific rigor. Any attempt to compromise the statistical robustness of pain studies in the name of animal welfare may instead result in worsened animal welfare outcomes for millions of animals.

Author contributions

TH: Conceptualization, Writing – original draft, Writing – review & editing. RB: Writing – review & editing. AW: Writing

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Retnam L, Chatikavanij P, Kunjara P, Paramastri YA, Goh YM, Hussein FN, et al. Laws, regulations, guidelines and standards for animal care and use for scientific purposes in the countries of Singapore, Thailand, Indonesia, Malaysia, and India. *ILAR J*. (2016) 57:312–23. doi: 10.1093/ilar/ilw038

2. Hampton JO, Hyndman TH. Underaddressed animal-welfare issues in conservation. *Conserv Biol*. (2019) 33:803–11. doi: 10.1111/cobi.13267

3. Palmer C, Fischer B, Gamborg C, Hampton JO, Sandøøe P. *Wildlife research: Toe-clipping. Wildlife ethics: Ethics in wildlife management and conservation*. Oxford, United Kingdom: Wiley-Blackwell. (2023). 216–229.

4. Alberthsen C, Waudby HP, Wilkinson LL, Lunney D, Bathurst M, Smith BP. Animal ethics committees In: BP Smith, HP Waudby, CA Alberthsen and JO Hampton, editors. *Wildlife research in Australia: Practical and applied methods*. Melbourne, Australia: CSIRO (2022). 19–28.

5. Taylor P, Meyer RE. Veterinary clinical research or experiments on pets. *Vet Anaesth Analg*. (2023) 50:383–5. doi: 10.1016/j.vaa.2023.07.006

6. Ashall V, Morton D, Clutton E. A declaration of Helsinki for animals. *Vet Anaesth Analg*. (2023) 50:309–14. doi: 10.1016/j.vaa.2023.03.005

7. Russell WMS, Burch RL. *The principles of humane experimental technique*. London: Methuen (1959).

8. Guhad F. Introduction to the 3Rs (refinement, reduction and replacement). *J Am Assoc Lab Anim Sci*. (2005) 44:58–9.

9. Pound P, Nicol CJ. Retrospective harm benefit analysis of pre-clinical animal research for six treatment interventions. *PLoS One*. (2018) 13:e0193758. doi: 10.1371/journal.pone.0193758

10. Grimm H, Olsson IAS, Sandøe P. Harm–benefit analysis–what is the added value? A review of alternative strategies for weighing harms and benefits as part of the assessment of animal research. *Lab Anim*. (2019) 53:17–27. doi: 10.1177/0023677218783004

11. Brønstad A, Newcomer CE, Decelle T, Everitt JI, Guillen J, Laber K. Current concepts of harm–benefit analysis of animal experiments–report from the AALAS–FELASA working group on harm–benefit analysis–part 1. *Lab Anim*. (2016) 50:1–20. doi: 10.1177/0023677216642398

12. Olsson IAS, Nicol CJ, Niemi SM, Sandøe P. From unpleasant to unbearable—why and how to implement an upper limit to pain and other forms of suffering in research with animals. *ILAR J*. (2020) 60:404–14. doi: 10.1093/ilar/ilz018

13. Landis SC, Amara SG, Asadullah K, Austin CP, Blumenstein R, Bradley EW, et al. A call for transparent reporting to optimize the predictive value of preclinical research. *Nature*. (2012) 490:187–91. doi: 10.1038/nature11556

14. Munro BA, Bergen P, Pang DS. Randomization, blinding, data handling and sample size estimation in papers published in veterinary Anaesthesia and analgesia in 2009 and 2019. *Vet Anaesth Analg*. (2022) 49:18–25. doi: 10.1016/j.vaa.2021.09.004

15. Macleod MR, Lawson McLean A, Kyriakopoulou A, Serghiou S, de Wilde A, Sherratt N, et al. Risk of Bias in reports of in vivo research: a focus for improvement. *PLoS Biol*. (2015) 13:e1002273. doi: 10.1371/journal.pbio.1002273

16. Leung V, Rousseau-Blass F, Beauchamp G, Pang DSJ. ARRIVE has not ARRIVEd: support for the ARRIVE (animal research: reporting of in vivo experiments) guidelines does not improve the reporting quality of papers in animal welfare, analgesia or anesthesia. *PLoS One*. (2018) 13:e0197882. doi: 10.1371/journal.pone.0197882

17. Hampton JO, MacKenzie DI, Forsyth DM. How many to sample? Statistical guidelines for monitoring animal welfare outcomes. *PLoS One*. (2019) 14:e0211417. doi: 10.1371/journal.pone.0211417

18. Vasishth S, Mertzen D, Jäger LA, Gelman A. The statistical significance filter leads to overoptimistic expectations of replicability. *J Mem Lang*. (2018) 103:151–75. doi: 10.1016/j.jml.2018.07.004

19. Kramer M, Font E. Reducing sample size in experiments with animals: historical controls and related strategies. *Biol Rev*. (2017) 92:431–45. doi: 10.1111/brv.12237

20. Dell RB, Holleran S, Ramakrishnan R. Sample size determination. *ILAR J*. (2002) 43:207–13. doi: 10.1093/ilar.43.4.207

21. Raja SN, Carr DB, Cohen M, Finnerup NB, Flor H, Gibson S, et al. The revised IASP definition of pain: concepts, challenges, and compromises. *Pain*. (2020) 161:1976–82. doi: 10.1097/j.pain.0000000000001939

22. Sneddon LU, Elwood RW, Adamo SA, Leach MC. Defining and assessing animal pain. *Anim Behav*. (2014) 97:201–12. doi: 10.1016/j.anbehav.2014.09.007

23. Hudson C, Whay H, Huxley J. Recognition and management of pain in cattle. *In Pract*. (2008) 30:126–34. doi: 10.1136/inpract.30.3.126

24. Loeser JD, Treede R-D. The Kyoto protocol of IASP basic pain terminology. *Pain*. (2008) 137:473–7. doi: 10.1016/j.pain.2008.04.025

25. Ranganathan P, Pramesh CS, Buyse M. Common pitfalls in statistical analysis: clinical versus statistical significance. *Perspect Clin Res*. (2015) 6:169–70. doi: 10.4103/2229-3485.159943

26. Sedgwick P. Clinical significance versus statistical significance. *BMJ*. (2014) 348:g2130. doi: 10.1136/bmj.g2130

27. Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol*. (2016) 31:337–50. doi: 10.1007/s10654-016-0149-3

28. McShane BB, Gal D, Gelman A, Robert C, Tackett JL. Abandon statistical significance. *Am Stat*. (2019) 73:235–45. doi: 10.1080/00031305.2018.1527253

29. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ*. (2010) 340:c332. doi: 10.1136/bmj.c332

30. Raj AM, O'Callaghan M. Evaluation of a pneumatically operated captive bolt for stunning/killing broiler chickens. *Br Poult Sci*. (2001) 42:295–9. doi: 10.1080/00071660120055232

31. Marchant-Forde JN, Boyle LA. COVID-19 effects on livestock production: a one welfare issue. *Front Vet Sci*. (2020) 7:585787. doi: 10.3389/fvets.2020.585787

32. Davis J, Musk GC. Anesthesia and analgesia in dogs and cats In: MC Dyson, P Jirkof, J Lofgren, EA Nunamaker and D Pang, editors. *Anesthesia and analgesia in laboratory animals. 3rd* ed. London: Elsevier (2023). 481–513.

33. Fisher AD. Addressing pain caused by mulesing in sheep. *Appl Anim Behav Sci*. (2011) 135:232–40. doi: 10.1016/j.applanim.2011.10.019

34. Abboud C, Duveau J, Bouali-Benazzouz R, Massé K, Mattar J, Brochoire L, et al. Animal models of pain: diversity and benefits. *J Neurosci Meth*. (2021) 348:108997. doi: 10.1016/j.jneumeth.2020.108997

35. Simon BT, Scallan EM, Carroll G, Steagall PV. The lack of analgesic use (oligoanalgesia) in small animal practice. *J Small Anim Pract*. (2017) 58:543–54. doi: 10.1111/jsap.12717

36. Shearer JK, Stock ML, Van Amstel SR, Coetzee JF. Assessment and management of pain associated with lameness in cattle. *Vet Clin North Am Food Anim Pract*. (2013) 29:135–56. doi: 10.1016/j.cvfa.2012.11.012

37. Boesch JM. Anesthesia for the horse with colic. *Vet Clin North Am Equine Pract*. (2013) 29:193–214. doi: 10.1016/j.cveq.2012.11.005

38. Le Bars D, Gozariu M, Cadden SW. Animal models of nociception. *Pharmacol Rev*. (2001) 53:597–652.

39. Voss K, Imhof J, Kaestner S, Montavon PM. Force plate gait analysis at the walk and trot in dogs with low-grade hindlimb lameness. *Vet Comp Orthopaedics Traumatol*. (2007) 20:299–304. doi: 10.1160/VCOT-07-01-0008

40. Love EJ, Murrell J, Whay HR. Thermal and mechanical nociceptive threshold testing in horses: a review. *Vet Anaesth Analg*. (2011) 38:3–14. doi: 10.1111/j.1467-2995.2010.00580.x

41. Kells NJ, Beausoleil NJ, Sutherland MA, Johnson CB. Electroencephalographic responses of anaesthetised pigs to intraperitoneal injection of sodium pentobarbital. *Anim Welf*. (2018) 27:205–14. doi: 10.7120/09627286.27.3.205

42. Mogil JS, Crager SE. What should we be measuring in behavioral studies of chronic pain in animals? *Pain*. (2004) 112:12–5. doi: 10.1016/j.pain.2004.09.028

43. Benato L, Murrell J, Knowles TG, Rooney NJ. Development of the Bristol rabbit pain scale (BRPS): a multidimensional composite pain scale specific to rabbits (*Oryctolagus cuniculus*). *PLoS One*. (2021) 16:e0252417. doi: 10.1371/journal.pone.0252417

44. Stasiak KL, Maul D, French E, Hellyer PW, VandeWoude S. Species-specific assessment of pain in laboratory animals. *Contemp Top Lab Anim Sci*. (2003) 42:13–20.

45. Martin MS, Kleinhenz MD, White BJ, Johnson BT, Montgomery SR, Curtis AK, et al. Assessment of pain associated with bovine respiratory disease and its mitigation with flunixin meglumine in cattle with induced bacterial pneumonia. *J Anim Sci*. (2022) 100:1–13. doi: 10.1093/jas/skab373

46. Slingsby L. Considerations for prospective studies in animal analgesia. *Vet Anaesth Analg*. (2010) 37:303–5. doi: 10.1111/j.1467-2995.2010.00544.x

47. Dusetzina SB, Brookhart MA, Maciejewski ML. Control outcomes and exposures for improving internal validity of nonrandomized studies. *Health Serv Res*. (2015) 50:1432–51. doi: 10.1111/1475-6773.12279

48. Sacks H, Chalmers TC, Smith H Jr. Randomized versus historical controls for clinical trials. *Am J Med*. (1982) 72:233–40. doi: 10.1016/0002-9343(82)90815-4

49. Viele K, Berry S, Neuenschwander B, Amzal B, Chen F, Enas N, et al. Use of historical control data for assessing treatment effects in clinical trials. *Pharm Stat*. (2014) 13:41–54. doi: 10.1002/pst.1589

50. Gross AJ, Mantel N. The effective use of both positive and negative controls in screening experiments. *Biometrics*. (1967) 23:285. doi: 10.2307/2528162

51. Turner JA, Deyo RA, Loeser JD, Von Korff M, Fordyce WE. The importance of placebo effects in pain treatment and research. *J Am Med Assoc*. (1994) 271:1609–14. doi: 10.1001/jama.1994.03510440060036

52. Charlesworth JEG, Petkovic G, Kelley JM, Hunter M, Onakpoya I, Roberts N, et al. Effects of placebos without deception compared with no treatment: a systematic review and meta-analysis. *J Evid Based Med*. (2017) 10:97–107. doi: 10.1111/jebm.12251

53. Miller FG, Kaptchuk TJ. Sham procedures and the ethics of clinical trials. *J R Soc Med*. (2004) 97:576–8. doi: 10.1177/014107680409701205

54. Kruger MC, Morel PC. Experimental control for the Ovariectomized rat model: use of sham versus nonmanipulated animal. *J Appl Anim Welf Sci*. (2016) 19:73–80. doi: 10.1080/10888705.2015.1107727

55. Ranganathan P, Pramesh CS, Buyse M. Common pitfalls in statistical analysis: "no evidence of effect" versus "evidence of no effect". *Perspect Clin Res*. (2015) 6:62–3. doi: 10.4103/2229-3485.148821

56. Festing MFW. On determining sample size in experiments involving laboratory animals. *Lab Anim*. (2018) 52:341–50. doi: 10.1177/0023677217738268

57. Makuch R, Johnson M. Issues in planning and interpreting active control equivalence studies. *J Clin Epidemiol*. (1989) 42:503–11. doi: 10.1016/0895-4356(89)90146-7

58. Muley MM, Krustev E, McDougall JJ. Preclinical assessment of inflammatory pain. *CNS Neurosci Ther*. (2016) 22:88–101. doi: 10.1111/cns.12486

59. Rosier EM, Iadarola MJ, Coghill RC. Reproducibility of pain measurement and pain perception. *Pain*. (2002) 98:205–16. doi: 10.1016/S0304-3959(02)00048-9

60. Musk GC, Jacobsen S, Hyndman TH, Lehmann HS, Tuke SJ, Collins T, et al. Objective measures for the assessment of post-operative pain in *Bos indicus* bull calves following castration. *Animals*. (2017) 7:1–15. doi: 10.3390/ani7100076

61. Gottardo F, Scollo A, Contiero B, Ravagnani A, Tavella G, Bernardini D, et al. Pain alleviation during castration of piglets: a comparative study of different farm options. *J Anim Sci*. (2016) 94:5077–88. doi: 10.2527/jas.2016-0843

62. Grant EP, Wickham SL, Anderson F, Barnes AL, Fleming PA, Miller DW. Preliminary findings on a novel Behavioural approach for the assessment of pain and analgesia in lambs subject to routine husbandry procedures. *Animals*. (2020) 10:1–13. doi: 10.3390/ani10071148

63. Cutler JH, Cramer G, Walter JJ, Millman ST, Kelton DF. Randomized clinical trial of tetracycline hydrochloride bandage and paste treatments for resolution of lesions and pain associated with digital dermatitis in dairy cattle. *J Dairy Sci*. (2013) 96:7550–7. doi: 10.3168/jds.2012-6384

64. Kleinhenz M, Van Engen N, Smith J, Gorden P, Ji J, Wang C, et al. The impact of transdermal flunixin meglumine on biomarkers of pain in calves when administered at the time of surgical castration without local anesthesia. *Livest Sci*. (2018) 212:1–6. doi: 10.1016/j.livsci.2018.03.016

65. Young JM, Schoonover MJ, Kembel SL, Taylor JD, Bauck AG, Gilliam LL. Efficacy of orally administered gabapentin in horses with chronic thoracic limb lameness. *Vet Anaesth Analg*. (2020) 47:259–66. doi: 10.1016/j.vaa.2019.11.003

66. Smith LJ, Bentley E, Shih A, Miller PE. Systemic lidocaine infusion as an analgesic for intraocular surgery in dogs: a pilot study. *Vet Anaesth Analg*. (2004) 31:53–63. doi: 10.1111/j.1467-2995.2004.00142.x

67. Gao Z, Cui F, Cao X, Wang D, Li X, Li T. Local infiltration of the surgical wounds with levobupivacaine, dexibuprofen, and norepinephrine to reduce postoperative pain: a randomized, vehicle-controlled, and preclinical study. *Biomed Pharmacother*. (2017) 92:459–67. doi: 10.1016/j.biopha.2017.05.038

68. Wright A, Amodie DM, Cernicchiaro N, Lascelles BDX, Pavlock AM, Roberts C, et al. Identification of canine osteoarthritis using an owner-reported questionnaire and treatment monitoring using functional mobility tests. *J Small Anim Pract*. (2022) 63:609–18. doi: 10.1111/jsap.13500

69. Barnett AG, van der Pols JC, Dobson AJ. Regression to the mean: what it is and how to deal with it. *Int J Epidemiol*. (2005) 34:215–20. doi: 10.1093/ije/dyh299

70. Berry DA. Bayesian clinical trials. *Nat Rev Drug Discov*. (2006) 5:27–36. doi: 10.1038/nrd1927

71. Bothwell LE, Avorn J, Khan NF, Kesselheim AS. Adaptive design clinical trials: a review of the literature and ClinicalTrials.gov. *BMJ Open*. (2018) 8:e018320. doi: 10.1136/bmjopen-2017-018320

72. Grieve AP. Response-adaptive clinical trials: case studies in the medical literature. *Pharm Stat*. (2017) 16:64–86. doi: 10.1002/pst.1778

73. Stevely A, Dimairo M, Todd S, Julious SA, Nicholl J, Hind D, et al. An investigation of the shortcomings of the CONSORT 2010 statement for the reporting of group sequential randomised controlled trials: a methodological systematic review. *PLoS One*. (2015) 10:e0141104. doi: 10.1371/journal.pone.0141104

74. Bellieni CV, Johnston CC. Analgesia, nil or placebo to babies, in trials that test new analgesic treatments for procedural pain. *Acta Paediatr*. (2016) 105:129–36. doi: 10.1111/apa.13210

75. Waisel DB. Ethics of research for patients in pain. *Curr Opin Anaesthesiol*. (2017) 30:205–10. doi: 10.1097/ACO.0000000000000438

76. World Medical Association. World medical association declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA*. (2013) 310:2191–4. doi: 10.1001/jama.2013.281053

77. Steagall PV, Bustamante H, Johnson CB, Turner PV. Pain Management in Farm Animals: focus on cattle, sheep and pigs. *Animals*. (2021) 11:1–17. doi: 10.3390/ani11061483

78. Robles I, Arruda AG, Nixon E, Johnstone E, Wagner B, Edwards-Callaway L, et al. Producer and veterinarian perspectives towards pain management practices in the US cattle industry. *Animals*. (2021) 11:1–13. doi: 10.3390/ani11010209

79. Kleinhenz MD, Viscardi AV, Coetzee JF. Invited review: on-farm pain management of food production animals. *App Anim Sci*. (2021) 37:77–87. doi: 10.15232/aas.2020-02106

80. Hempstead MN, Lindquist TM, Shearer JK, Shearer LC, Plummer PJ. Health and welfare survey of 30 dairy goat farms in the Midwestern United States. *Animals*. (2021) 11:1–16. doi: 10.3390/ani11072007

81. Larrondo C, Bustamante H, Gallo C. Sheep Farmers' perception of welfare and pain associated with routine husbandry practices in Chile. *Animals*. (2018) 8:1–14. doi: 10.3390/ani8120225

82. Gottardo F, Nalon E, Contiero B, Normando S, Dalvit P, Cozzi G. The dehorning of dairy calves: practices and opinions of 639 farmers. *J Dairy Sci*. (2011) 94:5724–34. doi: 10.3168/jds.2011-4443

83. Cozzi G, Gottardo F, Brscic M, Contiero B, Irrgang N, Knierim U, et al. Dehorning of cattle in the EU member states: a quantitative survey of the current practices. *Livest Sci*. (2015) 179:4–11. doi: 10.1016/j.livsci.2015.05.011

84. Scollo A, Contiero B, De Benedictis GM, Galli MC, Benatti D, Gottardo F. Analgesia and/or anaesthesia during piglet castration–part I: efficacy of farm protocols in pain management. *Ital J Anim Sci*. (2021) 20:143–52. doi: 10.1080/1828051X.2021.1873707

85. Warburton B, Norton BG. Towards a knowledge-based ethic for lethal control of nuisance wildlife. *J Wildl Manag*. (2009) 73:158–64. doi: 10.2193/2007-313