Maximizing value and minimizing waste in clinical trials in swine: Selecting outcomes to build an evidence base

Jan M. Sargeant, DVM, PhD, FCAHS; Annette M. O'Connor BVSc, DVSc, FANZCVS; Terri L. O'Sullivan, DVM, PhD; Alejandro Ramirez, DVM, MPH, PhD, DACVPM

Summary

Researchers planning clinical trials should identify the primary trial outcome and adequately power the trial to detect clinically meaningful differences in this outcome. All primary and secondary outcomes and their measurement should be comprehensively described, and their results reported. There is evidence that trials on the same subject use different outcomes or measure the same outcome in different ways, making it difficult to compare intervention effectiveness across clinical trials. Consensus development of core outcome sets could improve consistency in outcome measures used across trials and aid in development of an evidence-based body of literature on intervention effectiveness in swine populations.

Resumen - Maximizando el valor y minimizando el desperdicio en pruebas clínicas en cerdos: Selección de resultados para construir una base de evidencia

Los investigadores que planifican pruebas clínicas deben identificar el resultado principal de la prueba y potenciar adecuadamente la prueba para detectar diferencias clínicamente significativas en este resultado. Todos los resultados primarios y secundarios y su medición deben describirse exhaustivamente y sus resultados deben ser informados. Existe evidencia de que las pruebas sobre el mismo tema usan diferentes resultados o miden el mismo resultado de diferentes maneras, lo que dificulta comparar la efectividad de la intervención entre las pruebas clínicas. El desarrollo de un consenso de los resultados centrales podría mejorar la consistencia en las medidas del resultado utilizadas en las pruebas, y ayudar al desarrollo de una compilación de literatura basado en evidencia sobre la efectividad de la intervención en poblaciones porcinas.

Résumé - Maximiser la valeur et minimiser le gaspillage dans les essais cliniques chez le porc: Sélectionner les résultats pour constituer une base de données probantes

Les chercheurs qui planifient des essais cliniques doivent identifier le résultat principal de l'essai et alimenter suffisamment l'essai pour détecter des différences cliniquement significatives dans ce résultat. Tous les résultats primaires et secondaires et leur mesure doivent être décrits de manière exhaustive et leurs résultats communiqués. Il existe des preuves que les essais sur le même sujet utilisent des résultats différents ou mesurent le même résultat de différentes manières, ce qui rend difficile la comparaison de l'efficacité des interventions entre les essais cliniques. L'élaboration d'un consensus sur les principaux ensembles de résultats pourrait améliorer la cohérence des mesures de résultats utilisées dans les essais et aider à l'élaboration d'un ensemble de documents fondés sur des données probantes sur l'efficacité des interventions dans les populations porcines.

The recent emphasis on evidence-based decision-making has led to a growth in literature on the design of clinical trials.[1] In this article, we use "clinical trials" as synonymous with "controlled trials" and define clinical trials as an experimental study intended to evaluate products or procedures in swine outside of a laboratory setting (ie, in a realistic-use setting).[2] When random allocation to an intervention group is applied in a clinical trial, the design is referred to as a randomized controlled trial. For clarity, we will use the term "clinical trial" throughout this article.

Clinical trials represent the primary research design with the highest evidentiary value when it is ethical and feasible to allocate animals to treatment groups.[3] Selecting appropriate outcomes is fundamental to clinical trial design because the difference in outcomes between intervention groups is inferred to be the result of the intervention.[4]

The word "outcome" encompasses different constructs. To clarify, we use the following vocabulary to describe the

JMS, TLO: Department of Population Medicine, Ontario Veterinary College, University of Guelph, Guelph, Ontario, Canada.

AMO: Department of Large Animal Clinical Sciences, College of Veterinary Medicine, Michigan State University, East Lansing, Michigan.

AR: University of Arizona College of Veterinary Medicine, Oro Valley, Arizona.

Corresponding author: Dr Jan M. Sargeant, 2536 Stewart Building, Ontario Veterinary College, University of Guelph, 50 Stone Road East, Guelph, Ontario, Canada N1G 2W1; Tel: 519-824-4120 Ext. 54045; Email: **sargeanj@uoguelph.ca**.

various components of an "outcome" with illustrative examples provided in Figure 1. The first level of outcome is the outcome domain; in swine research, relevant outcome domains include health, productivity, and animal welfare. Within an outcome domain there can be one or more conceptual outcomes. For instance, the domain of productivity would include conceptual outcomes such as growth performance or reproductive performance. Next, within a conceptual outcome, one or more operational outcomes may be relevant. Operational outcomes are something that can be measured. For the conceptual outcome of growth performance, operational outcomes may include feed:gain ratio or average daily gain. Finally, there is the outcome measure, which includes case definition, measurement tool, and time-at-risk. A conceptual outcome such as average daily gain could be measured in different ways. For example, the outcome measure for average daily gain may involve weighing a pen at the start and end of production or by calculating weight gains at the end of a production phase for surviving pigs. Decisions about trial outcomes require specification of each component of an outcome. These decisions then need to be conveyed to the reader.

Considerable attention has been paid to potential sources of bias in clinical trials; however, the choice of outcome measures has received less attention.[1] This article will overview considerations when selecting outcomes and outcome measures for use in clinical trials, the importance of identifying the primary outcome, and the need for consistency in selecting outcomes across trials.

Considerations when selecting and reporting clinical trial outcomes

Once the researcher has determined the outcome domain and the conceptual outcome, operational outcomes need to be specified. The operational outcomes that are selected should be an expected benefit or harm of the intervention if that intervention is effective. The researcher must ask what they *expect* the intervention to do that is meaningful to those who might use the intervention.[1]

Relevant operational outcomes will differ as the intervention development research moves from proof-of-concept or safety trials to clinical trials evaluating efficacy in realistic-use settings.[5] Thus, as an example, in the early stages of vaccine development, the ability to produce antibodies to the target protein might be the most relevant operational outcome for a company considering whether to take the next step in product development by investing in a large-scale clinical trial. However, in clinical trials on the efficacy of that vaccine, the primary outcome measure should be of clinical relevance to the end-user of the vaccine. Therefore, in a clinical trial, outcome domains like health, production, and welfare should be operationalized with clinically relevant outcomes such as a mortality, morbidity, growth performance, or animal comfort.

Once an operational outcome is selected, the researcher must determine the associated outcome measure. There are a number of types of outcome measures

that can be used. Clinical outcomes are outcomes that reflect how an animal feels, functions, or survives.[5] Examples of clinical outcomes include measures of morbidity (disease occurrence) and mortality and outcome measures related to welfare. Outcomes also may be surrogates for a clinical outcome (eg, rectal temperature as a surrogate for morbidity) or may be biomarkers (biological measurements) used to predict a clinical outcome such as acute phase proteins as a biomarker for risk of morbidity. Composite endpoints represent a combination of correlated variables.[6,7] An example of a composite endpoint in swine could be the incidence of any clinical sign of disease (eg, at least one of diarrhea, lameness, weight loss, or coughing). Although composite endpoints may increase statistical power for rare outcomes, their use is not without issues. Interested readers are directed to other articles if composite outcomes are used.[6,8]

Determining the outcome measure pertains not only to what is measured and how it is measured, but also to the time at which it is measured. For instance, an operational outcome such as average daily gain could be measured over a specific period (eg, the 15 days following intervention administration), over a specific production period (eg, during the nursery phase), or over the entire period from weaning to market. In contrast, some outcome measures may logically only pertain to a single time or specific event; an example would be pigs born alive per litter, which is measured at a single time.

Figure 1: Flow chart of outcomes from domains to measures, with examples for swine research.

Examples for research conducted in swine

Outcome domain	Health	Production	Welfare
Conceptual outcome	Respiratory disease	Reproductive performance	Freedom from dicomfort
Operational outcome	Presence / absence of lung lesions	Fecundity	Pain
Outcome measure	At least 50% consolidation of lungs at post-mortem	Average pigs born alive per sow	Cortisol level 6 hours post castration

As an example of how the process of selecting an outcome might work, consider a researcher planning a trial of an intervention intended to reduce respiratory disease in finishing pigs. The trial could be designed to evaluate health as an outcome domain, occurrence of pneumonia as the conceptual outcome, lung lesions as an operational outcome representing a surrogate measure of pneumonia, and a specific scoring system of lung pathology at slaughter as the outcome measure. The trial could be designed to evaluate production as an outcome domain, growth performance as the conceptual outcome, average daily gain as an operational outcome, and average daily gain for 30 days post intervention as the outcome measure.

It is important to consider whether the selected outcome measure (including what, how, and when it is measured) is sensitive to the nature and degree of change expected from the intervention.[1] For instance, consider an intervention is applied to nursery swine where the outcome is measured as average daily gain at slaughter. It is possible that the intervention was successful at reducing production losses in the short term, but that the outcome at slaughter was not sensitive to the intervention due to compensatory growth. As another example, suppose that a researcher evaluated trembling as a welfare outcome. This outcome could be measured on a dichotomous scale (presence or absence during a defined period) or on a continuous scale (number of episodes of trembling during a defined period). An intervention might reduce the frequency of trembling, but not whether it occurred. Therefore, the researcher will need to decide whether the presence or absence of trembling is sensitive to the intervention or whether the outcome of number of episodes of trembling would be more appropriate.

The number of outcomes to include in a trial also should be considered. It is common for trials to include multiple outcome domains (eg, health and welfare), and a single trial may include multiple conceptual and operational outcomes within each domain. Additionally, an operational outcome may be measured in multiple ways (eg, presence or absence of coughing, respiratory illness requiring treatment, or cough index). In studies on reporting of clinical trials for health and production and for on-farm food safety in livestock, 182 of 200 trials had multiple outcomes.[9,10] In the study

on food safety trials, [9] the mean number of outcomes per trial was 8.5 (range: 1-51) and in the study on livestock health and production trials,[10] the mean number of outcomes per trial was 9.5 (range: 1-41).

The outcomes selected should be those necessary for decision-making. Too many outcomes may lead to a lack of focus or difficulties in interpreting trial results, for instance when different outcome measures for the same conceptual outcome have different results or interpretation.[6] Additionally, as the number of outcome measures increases, so too does the probability of a type I error (a false positive finding).[6,11] If the authors are using null hypothesis significance testing with a type I error rate of 5% for each test, when there is no association, we would expect one type I error within each 20 independent tests. To illustrate the potential magnitude of this issue of multiplicity, the probability of *at least one* type I error in a population where the null is true, if testing for each outcome is at $P = .05$ and the outcomes are independent, can be calculated as $(1-[1-0.05]^k)$, where k = the number of outcome measures. Therefore, using the minimum (1), mean (9.5), and maximum (41) number of outcomes from the 100 trials evaluated in the study on reporting of livestock health trials,[10] and assuming an alpha of 0.05 for hypothesis testing, then the probability of at least one false positive result would be 5%, 38.6%, and 87.8%, respectively. Therefore, it is important to restrict the outcomes (and outcome measures) to those that are appropriate to the stage of the intervention development and evaluation and, in the case of clinical trials in the real-world, to those that are necessary for decision-making.

Further, when multiple outcomes are measured, causation should be used in interpreting the value of each additional outcome to the end-user, especially when the outcome measures are within the same operational outcome. For example, a randomly occurring type I error that impacts average daily gain will also randomly impact feed:gain ratio and feed conversion, as they are likely measuring much the same outcome. If two variables are highly correlated, not a lot of additional information is gained by including both. Therefore, evidence of an impact in multiple outcomes should not necessarily be interpreted as building a stronger evidence base. A stronger evidence base would exist if the impact of the intervention is observed

in multiple domains, ie, incidence of tail biting (welfare) and average daily gain (production). Therefore, when using multiple outcomes, these should be in different domains as much as feasible.[12]

The outcome measures must be comprehensively described or else the results of the trial cannot be interpreted. In an assessment of reporting in trials in livestock species, the measurement of all outcomes was described in 79% of trials, meaning that information with respect to all outcomes was not provided in approximately one-fifth of trials.[10] Guidance is available for the detail recommended when reporting outcomes, outcome measures, and results of a clinical trial.[13,14] There is a responsibility not only for authors to improve reporting of outcomes, but also for peer reviewers and journal editors to ensure that reporting is comprehensive.

It also is important that the results are reported for all outcome measures that were included in the trial, otherwise there is potential for selective outcome reporting.[15] There is evidence from human trials that outcomes associated with statistically significant results are more likely to be reported than those that are not significant.[16] Because it is uncommon to publish protocols for swine clinical trials, it is not possible to determine the extent to which this is an issue in swine research. However, if outcomes associated with statistically significant results are more likely to be presented in a manuscript (or, conversely, if outcomes associated with nonsignificant findings are excluded), it will lead to an exaggeration of intervention effectiveness and the probability of a type I error cannot be assessed. It may also mean that interventions that are not effective will continue to be researched.

Importance of defining the primary outcome

Regardless the number of outcomes, it is important that a primary outcome is identified. The primary outcome should be the outcome of most relevance to decision-making by the target audience, and is the outcome used to calculate the sample size required to ensure adequate power.[17] There may be situations where more than one outcome is considered of extremely high relevance. For instance, a researcher may be equally interested in a health outcome and a welfare outcome. In this instance, researchers should declare both outcomes as

primary and conduct sample size calculations for both, using the higher calculated sample size in the trial.[14]

Primary outcomes are not consistently identified in many veterinary trials; for trials published in veterinary journals in 2013, the primary outcome was identified in 19.3% of trials, compared to 98.3% of trials published in human medical journals.[18] In swine trials, this proportion has improved since the publication of the REFLECT reporting guidelines.[13,14] Prior to publication of REFLECT, the primary outcome was identified in 14% of vaccination trials in swine compared to 42% after 2010.[19] Although this improvement is encouraging, these results still suggest that the primary outcome is not identified in over half of the vaccine trials conducted in swine populations.

If there is no sample size calculation, or if there are secondary outcomes that are underpowered, then meaningful differences may not be detected as statistically significant at $P = .05$; the (arbitrary) cutpoint often used in clinical trials. This may result in meaningful differences being presented as "no difference between groups." To illustrate this concept, the minimum detectable risk ratio (RR) was calculated using data on mortality collected from 56 trials included in a systematic review and network meta-analysis on the comparative efficacy of swine bacterial respiratory vaccines.[20]

When calculating a sample size for a binary outcome, the researcher needs to define the proportion with the event in the baseline intervention group, the difference in the outcome that is clinically meaningful, and the desired confidence and power. In this example, we used data from completed trials to determine the smallest difference between treatment groups (expressed as a risk ratio) that could have been detected as statistically significant, given the baseline prevalence and the sample size used. The minimum detectable RR was calculated for each trial using the proportion of swine with the mortality outcome in the placebo group, the total sample size, power of 0.8, and confidence of 0.95 using epi.sscompb program in EpiR. The sample size corresponded to the individual animal level, and thus did not account for nonindependence of swine within pens. Figure 2 shows the distribution of minimum detectable RR, with the vertical dashed line representing the median of the minimum detectable RR of 2.0. The median proportion of swine mortality in the baseline intervention group was 0.06 (ie, 6%). A minimum detectable RR of 2.0 means that the proportion of pigs dying in the intervention group would need to be approximately double (or ½ for a preventive outcome) before detecting the RR as statistically significant. This may be a larger difference than what would be clinically

meaningful. Therefore, by designating a primary outcome and powering the study to detect clinically meaningful differences in this outcome, the researcher can ensure adequate power. The example illustrates that many current studies can only identify a relatively large difference as statistically significant.

Inconsistency of outcomes across trials

It is necessary to replicate the results across multiple trials to inform evidence-based decision-making because the results of a single trial are based on a sample from the source population and thus are subject to sampling error. Sampling error, also referred to as chance, occurs when the parameter of interest (eg, a mean or a proportion) is different between the source population and the study population. Combining the results of multiple trials, as is performed statistically with meta-analysis, increases precision of the estimate of effect size[15] and increases confidence that the results are not a reflection of sampling error.[21] However, trials evaluating the same intervention often do not use the same outcomes or outcome measures, precluding the ability to build a body of evidence across trials. Outcomes across trials may represent different outcome domains (eg, one trial measuring a welfare outcome and another measuring a production

Figure 2: Distribution of minimum detectable risk ratios for trials included in a systematic review and network meta-analysis on the comparative efficacy of swine bacterial respiratory vaccines.[20] Dotted vertical line represents the median value for the minimum detectable risk ratio.
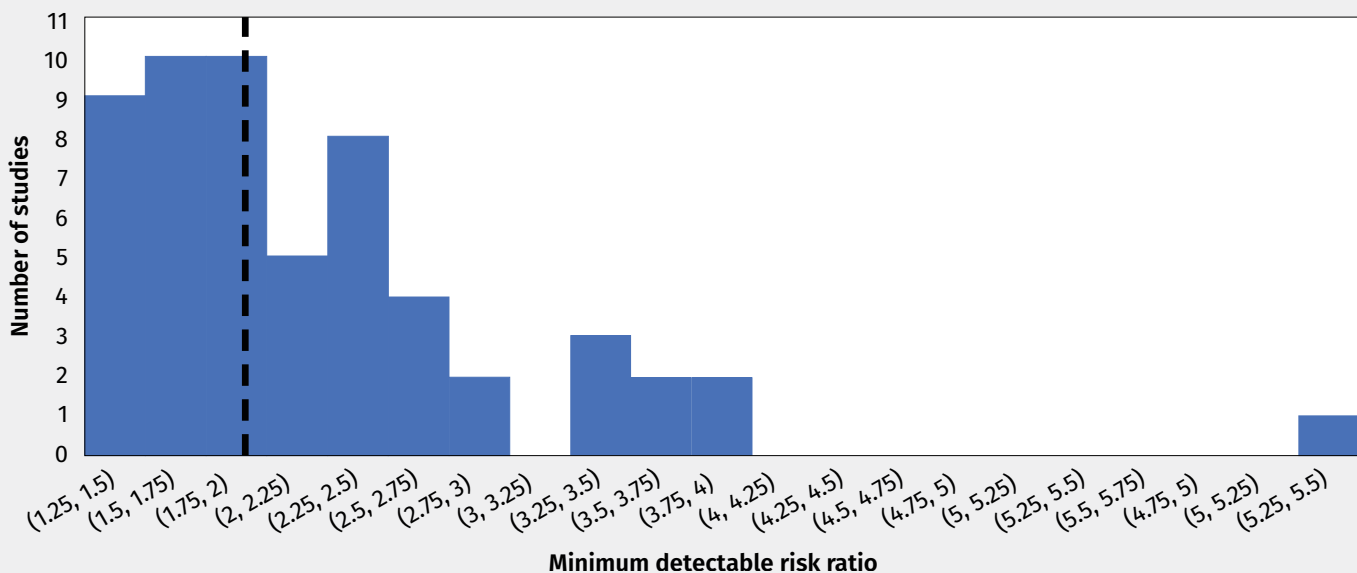
outcome), may represent the same conceptual outcome but with different operational outcomes (eg, one trial may measure pain using postural changes and another trial may measure pain using activity levels), or may represent the same operational outcome but with a different outcome measure (eg, average daily gain during the first 2 weeks post weaning in one trial and across the entire growing period in another).

To illustrate, we used data from 61 lung lesion outcome measures reported in 58 trials evaluating nonspecific lung lesions at slaughter from a systematic review of the efficacy of bacterial respiratory vaccines (Table 1).[20] Not only were different outcome measures used across the trials, but key features of the measurement of the specific outcomes often were not provided. For example, the outcome "general appearance" often did not include a comprehensive description of the criteria for determining whether the general appearance corresponded to a positive or negative result. This limits the ability to combine results across trials and thereby build a body of evidence. The example provided represents only one type of intervention (vaccination against bacterial pathogens) and one operational outcome (lung lesions). However, the example serves to highlight the inadequate reporting and inconsistency in outcome measures across trials, and the resulting challenges in synthesizing research results.

To reduce inconsistencies in outcomes across trials, individual researchers should be familiar with the literature in their area and select operational outcomes and outcome measures that have been used in previous trials. Ideally, outcome measures should be validated or agreed upon by consensus of experts in the area; otherwise, outcomes with poor reliability or validity might be selected based on use in a previous trial. At the industry level, a possible solution to inconsistency and selective reporting of outcome measures is the creation of core outcome sets for specific topic areas within swine research. Core outcome sets represent an agreed minimum set of outcomes and outcome measures that should be reported in all trials that are conducted on a specific disease or condition.[4,29] Although the core outcomes should be included in all trials, researchers may include other primary or secondary outcomes that are of interest in their specific trial.[4,30] Core outcome sets also may need to be updated as technologies and diagnostic tests are developed and validated. Guidelines are available for developing core outcome sets in the COMET initiative handbook.[4] The COMET initiative was launched in 2010 with a key objective of encouraging the development and updating of core outcome sets.[4] The COMET initiative was developed for human health outcomes, and the relevant outcome domains may differ for swine. Nonetheless, the process for developing core outcome sets would be relevant for swine applications. The process of developing a core outcome set involves a decision as to the topic (eg, a disease, a domain such as welfare, a conceptual outcome such as pain, or a type of intervention and a disease), evaluation of the existing literature on trials to determine what outcome domains, conceptual outcomes, operational outcomes, and outcome measures have previously been used, and a consensus process to determine which of these to include in the core outcome set.[4,31] The creation of core outcome sets should include the

Table 1: Lung lesion outcome measures reported in 58 trials evaluating nonspecific lung lesions at slaughter from a systematic review of the efficacy of bacterial respiratory vaccines[20]

Lung lesion scoring system	Range of scores for scoring system	# trials (dichotomous outcome)	# trials (continuous outcome)
Christensen et al,[22] 1999			1
Madec and Kobisch,[23] 1982	0 - 28	2	
Madec and Kobisch,[23] 1982	0 – 24	1	
Madec and Kobisch,[23] 1982	Not reported	1	1
Goodwin et al,[24] 1969	Not reported	1	
Goodwin and Whittlestone,[25] 1973	0-55	1	10
Goodwin and Whittlestone,[25] 1973	Not reported	1	1
Piffer and Brito,[26] 1991	Not reported	1	
Hannan et al,[27] 1982	0 -55	1	
Morrison et al,[28] 1985	Percentage of pneumonia in different lung lobes	1	1
None reported	0 - 14	2	
None reported	0 – 28	1	1
None reported	0 - 35		1
None reported	Not reported	1	3
General appearance		28	

input of relevant stakeholders.[4,29] For instance, individuals from academia, industry, and other relevant stakeholders might decide to identify a core outcome set for trials evaluating interventions to prevent respiratory disease in swine or a core outcome set for trials related to improving swine welfare. Creating the core outcome sets would involve identifying relevant domains, then relevant conceptual outcome within domains, followed by specific operational outcomes within each conceptual outcome and finally the outcome measure for each conceptual outcome, including case definition, measurement tool, and period at risk. Defining the core outcomes also may involve defining normal or abnormal cut points for outcomes measured on a continuous scale for which a qualitative label is desired. This may be more challenging for welfare or other domains that are more recently included in trials where there has not been a long history of using, validating, and interpreting relevant outcome measures. The selection of core outcomes would need to take into consideration the validity of the outcomes in measuring the construct that they are intended to measure. The cost associated with collecting the outcome data also may be a consideration. The COMET initiative handbook for development of core outcome sets[4] does not provide specific input on the number of outcomes that should be included in a core outcome set; however, the number will need to be a balance between feasibility, probability of type I error, and information required for clinical decision-making.

Core outcome sets increasingly are being developed for use in human trials; as of 2018, there were 410 core outcome sets for a wide range of human trial topic areas including cancer, urology, and child health.[32] Veterinary medicine has been slower to adopt core outcome sets; to date, there is a core outcome set published for trials in feline chronic kidney disease[33] and one for therapeutic trials for canine atopic dermatitis.[34] The development of core outcome sets is an area in which the swine industry could provide leadership. In swine research, core outcome sets could include outcomes from domains such as health, production, and welfare. Stakeholders could include swine producers and veterinarians, industry groups, researchers, and research funders. Although consensus can be challenging, there is precedent in swine research; naming of the disease

periweaning failure to thrive syndrome was reached by consensus,[35] as were standardized systems for classifying herd level status for porcine reproductive and respiratory syndrome[36] and for *Mycoplasma hyopneumoniae* in breeding herds.[37] Recently, a consortium of researchers, industry, veterinarians, and regulatory agencies developed a methodology to measure pain associated with surgical castration in piglets.[38]

These prior initiatives suggest that the swine industry could be successful in coming to consensus on core outcome sets. Creating core outcome sets will aid individual researchers in identifying outcomes and outcome measures to use in their trial and will facilitate synthesis of results from multiple trials. This will allow a body of evidence to be developed to determine the effectiveness of specific interventions for a disease or condition, to identify when further trials will not increase our knowledge of the effectiveness of an intervention, and to determine the relative efficacy of multiple intervention options for the same disease or conditions. This will maximize the utility of research trials conducted in swine populations.

Implications

- Primary and secondary outcomes should be defined and clearly reported.
- Primary outcomes determine sample size; many swine trials are underpowered.
- Core outcome sets can improve consistency in outcome measures used across trials.

Acknowledgments

Conflict of interest
None reported.

Disclaimer
Drs O'Sullivan and Ramirez, the journal's executive editor and editorial board member, respectively, were not involved in the editorial review of or decision to publish this article.

Scientific manuscripts published in the *Journal of Swine Health and Production* are peer reviewed. However, information on medications, feed, and management techniques may be specific to the research or commercial situation presented in the manuscript. It is the responsibility of the reader to use information responsibly and in accordance with the rules and regulations governing research or the practice of veterinary medicine in their country or region.

References

1. Coster WJ. Making the best match: Selecting outcome measures for clinical trials and outcome studies. *Am J Occup Ther*. 2013;67(2):162-170. **https://doi.org/10.5014/ajot.2013.006015**

2. Dohoo I, Martin W, Stryhn H. *Veterinary Epidemiologic Research*. 2nd ed. VER Inc; 2009.

3. Sargeant JM, Kelton DF, O'Connor AM. Study designs and systematic review of interventions: Building evidence across study designs. *Zoonoses Public Health*. 2014;61(Suppl1):10-17. **https://doi.org/10.1111/zph.12127**

4. Williamson PR, Altman DG, Bagley H, Barnes KL, Blazeby JM, Brookes ST, Clarke M, Gargon E, Gorst S, Harman N, Kirkham JJ, McNair A, Prinsen CAC, Schmitt J, Terwee CB, Young B. The COMET Handbook: Version 1.0. *Trials*. 2017;18(Suppl3):1-50. **https://doi.org/10.1186/s13063-017-1978-4**

*5. US Food and Drug Administration. Biomarkers and surrogate endpoints in clinical studies to support effectiveness of new animal drugs. Guidance for Industry #267. Published October 2021. Accessed November 8, 2021. **https://www.fda.gov/media/138160/download**

6. Vetter TR, Mascha EJ. Defining the primary outcomes and justifying secondary outcomes of a study: Usually, the fewer, the better. *Anesth Analg*. 2017;125(2):678-681. **https://doi.org/10.1213/ANE.0000000000002224**

7. Oyama MA, Ellenberg SS, Shaw PA. Clinical trials in veterinary medicine: A new era brings new challenges. *J Vet Intern Med*. 2017;31(4):970-978. **https://doi.org/10.1111/jvim.14744**

8. Freise KJ, Lin T-L, Fan TM, Recta V, Clark TP. Evidence-based medicine: The design and interpretation of noninferiority clinical trials in veterinary medicine. *J Vet Intern Med*. 2013;27(6):1305-1317. **https://doi.org/10.1111/jvim.12211**

9. Sargeant JM, Saint-Onge J, Valcour J, Thompson A, Elgie R, Snedeker K, Marcynuk P. Quality of reporting in clinical trials of preharvest food safety interventions and associations with treatment effect. *Foodborne Pathog Dis*. 2009;6:989-999. **https://doi.org/10.1089/fpd.2009.0321**

10. Sargeant JM, Elgie R, Valcour J, Saint-Onge J, Thompson A, Marcynuk P, Snedeker K. Methodological quality and completeness of reporting in clinical trials conducted in livestock species. *Prev Vet Med*. 2009;91:107-115. **https://doi.org/10.1016/j.prevetmed.2009.06.002**

11. Tukey JW. Some thoughts on clinical trials, especially problems of multiplicity. *Science*. 1977;198(4318):679-684. https://doi.org/10.1126/science.333584

12. Li T, Mayo-Wilson E, Fusco N, Hong H, Dickersin K. Caveat emptor: The combined effects of multiplicity and selective reporting. *Trials*. 2018;19:497. https://doi.org/10.1186/s13063-018-2888-9

13. O'Connor AM, Sargeant JM, Gardner I, Dickson J, Torrence M, Dewey CE, Dohoo I, Evans R, Gray J, Greiner M, Keefe G, Lefebvre S, Morley P, Ramirez A, Sischo W, Smith D, Snedeker K, Sofos J, Ward M, Wills R. The REFLECT statement: Methods and processes of creating reporting guidelines for randomized controlled trials for livestock and food safety. *J Swine Health Prod*. 2010;18:18-26.

14. Sargeant JM, O'Connor AM, Gardner IA, Dickson JS, Torrence ME, Dohoo IR, Lefebvre SL, Morley PS, Ramirez A, Snedeker K. The REFLECT statement: Reporting guidelines for randomized controlled trials in livestock and food safety: explanation and elaboration. *J Food Prot*. 2010;73(3):579-603. https://doi.org/10.4315/0362-028x-73.3.579

15. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, eds. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 6.2. Updated February 2021. Cochrane; 2021.

16. Dwan K, Gamble C, Williamson PR, Kirkham JJ, Reporting Bias Group. Systematic review of the empirical evidence of study publication bias and outcome reporting bias - an updated review. *PLoS One*. 2013;8(7):e66844. https://doi.org/10.1371/journal.pone.0066844

17. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, Elbourne D, Egger M, Altman DG, CONSORT. CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *Int J Surg*. 2012;10(1):28-55. https://doi.org/10.1016/j.ijsu.2011.10.001

18. Di Girolamo N, Meursinge Reynders R. Deficiencies of effectiveness of intervention studies in veterinary medicine: A cross-sectional survey of ten leading veterinary and medical journals. *PeerJ*. 2016;4:e1649. https://doi.org/10.7717/peerj.1649

19. Moura CAA, Totton SC, Sargeant JM, O'Sullivan TL, Linhares DCL, O'Connor AM. Evidence of improved reporting of swine vaccination trials in the post-REFLECT statement publication period. *J Swine Health Prod*. 2019;27(5):265-277.

20. Sargeant JM, Deb B, Bergevin MD, Churchill K, Dawkins K, Dunn J, Hu D, Moody C, O'Connor AM, O'Sullivan TL, Reist M, Wang C, Wilhelm B, Winder CB. Efficacy of bacterial vaccines to prevent respiratory disease in swine: A systematic review and network meta-analysis. *Anim Health Res Rev*. 2019;20(2):274-290. https://doi.org/10.1017/S1466252319000173

21. Valentine JC, Biglan A, Boruch RF, Castro FG, Collins LM, Flay BR, Kellam S, Mościcki EK, Schinke SP. Replication in prevention science. *Prev Sci*. 2011;12:103-117. https://doi.org/10.1007/s11121-011-0217-6

22. Christensen G, Sorensen V, Mousing J. Diseases of the Respiratory System. In: Straw BE, D`Allaire S, Mengelind WL, Taylor DJ, eds. *Diseases of Swine*. 8th ed. Iowa State University Press; 1999:913-940.

23. Madec F, Kobisch M. Bilan lesionnel des poumons de porcs charcutiers a l'abattoir [Lesion assessment of the lungs of slaughterhouse pigs]. *Journées de la Recherche Porcine*. 1982;14:405-12.

24. Goodwin RF, Hodgson RG, Whittlestone P, Woodhams RL. Some experiments relating to artificial immunity in enzootic pneumonia of pigs. *J Hyg(Lond)*. 1969;67(3):465-476. https://doi.org/10.1017/s0022172400041887

25. Goodwin RF, Whittlestone P. Enzootic pneumonia of pigs: Immunization attempts inoculating *Mycoplasma suipneumoniae* antigen by various routes and with different adjuvants. *Br Vet J*. 1973;129:456-464. https://doi.org/10.1016/s0007-1935(17)36386-8

*26. Piffer IA, Brito JRF. Descricao de um modelo para avaliacao e quantificacao de lesoes pulmonares de suinos e formulacao de um indice para classificacao de rebanhos [Description of a model for evaluation and quantification of lung lesions in swine and formulation of an index for classification of herds]. Concordia: EMBRAPA-CNPSA.

27. Hannan PC, Bhogal BS, Fish JP. Tylosin tartrate and tiamutilin effects on experimental piglet pneumonia induced with pneumonic pig lung homogenate containing mycoplasmas, bacteria and viruses. *Res Vet Sci*. 1982;33:76-88. https://doi.org/10.1016/S0034-5288(18)32364-6

28. Morrison RB, Hilley HD, Leman AD. Comparison of methods for assessing the prevalence and extent of pneumonia in market weight swine. *Can Vet J*. 1985;26:381-384.

29. Williamson PR, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, Tugwell P. Developing core outcome sets for clinical trials: issues to consider. *Trials*. 2012;13:132. https://doi.org/10.1186/1745-6215-13-132

30. Webbe J, Sinha I, Gale C. Core outcome sets. *Arch Dis Child Educ Pract Ed*. 2018;103(3):163-166. https://doi.org/10.1136/archdischild-2016-312117

31. Prinsen CA, Vohra S, Rose MR, Boers M, Tugwell P, Clarke M, Williamson PR, Terwee CB. How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" - a practical guideline. *Trials*. 2016;17(1):449. https://doi.org/10.1186/s13063-016-1555-2

32. Gargon E, Gorst SL, Williamson PR. Choosing important health outcomes for comparative effectiveness research: 5th annual update to a systematic review of core outcome sets for research. *PLoS One*. 2019;14(12):e0225980. https://doi.org/10.1371/journal.pone.0225980

33. Doit H, Dean RS, Duz M, Finch NC, Brennan ML. What outcomes should be measured in feline chronic kidney disease treatment trials? Establishing a core outcome set for research. *Prev Vet Med*. 2021;192:105348. https://doi.org/10.1016/j.prevetmed.2021.105348

34. Olivry T, Bensignor E, Favrot C, Griffin CE, Hill PB, Mueller RS, Plant JD, Williams HC, International Committee of Allergic Diseases of Animals (ICADA). Development of a core outcome set for therapeutic clinical trials enrolling dogs with atopic dermatitis (COSCAD'18). *BMC Vet Res*. 2018;14(1):238. https://doi.org/10.1186/s12917-018-1569-y

35. Huang Y, Henry S, Friendship R, Schwartz K, Harding J. Clinical presentation, case definition, and diagnostic guidelines for porcine periweaning failure to thrive syndrome. *J Swine Health Prod*. 2011;19:340-344.

36. Holtkamp D, Torremorell M, Corzo CA, Linhares DCL, Almeida MN, Yeske P, Polson DD, Becton L, Snelson H, Donovan T, Pittman J, Johnson C, Vilalta C, Silva GS, Sanhueza J. Proposed modifications to porcine reproductive and respiratory syndrome virus herd classification. *J Swine Health Prod*. 2021;29:261-270.

37. Clavijo MJ, Galina Pantoja L, Holtkamp DJ, Yeske P, Johnson C, Sprague M, Fano E, Main R, McDowell E, Painter T, Becton L, Baumert D, Glowsenski L, Snelson H, Maschhoff A. Establishing *Mycoplasma hyopneumoniae* herd status classification criteria for breeding herds. *J Swine Health Prod*. 2021;29:319-326.

38. Baysinger A, Webb S, Brown J, Coetzee J, Crawford S, DeDecker A, Karriker LA, Pairis-Garcia M, Sutherland MA, Viscardi A. Proposed multidimensional pain outcome methodology to demonstrate analgesic drug efficacy and facilitate future drug approval for piglet castration. *Anim Health Res Rev*. 2021;22:163-176. https://doi.org/10.1017/S1466252321000141

* Non-refereed reference.