

The term 'Pilot Study' is misused in veterinary medicine: a critical assessment

Mark Rishniw,^{1,2} Maurice Edward White³

Abstract

Authors commonly use the term 'Pilot Study' in the veterinary literature. The term has a specific definition in medical literature, but is not defined in veterinary literature. Therefore, we sought to examine the frequency of the use of the term and the characteristics of studies using the term in the article title, and derive the intended meaning of the term. We identified all articles in veterinary literature using the term in the article title between 2008 and 2017. We then examined specific characteristics of articles published between 2008 and 2012. We found use of the term is increasing ($P < 0.0001$). Of articles using the term between 2008 and 2012, only 20 per cent led to a larger, more comprehensive verifying study. Most garnered few citations, but 75 per cent were cited in review articles. Pilot studies had a median sample size of 10 subjects. We found comparable studies for each pilot study that did not incorporate the term into their titles. None of the authors of any of the pilot studies defined the term or explained why their study was termed a 'pilot study'. Journals and authors used the term haphazardly. Our findings indicate that the term 'Pilot Study' is meaningless because it meets no specific, consistently adhered-to criteria. We believe that authors use the term as a means of 'Deficiency signaling' to editors, reviewers and readers. We recommend that authors and journals abandon the term in veterinary literature because it serves no purpose, is not used consistently and might harm veterinary medicine.

Introduction

In 1947 RW Parnell¹ used the term 'Pilot' to describe a survey. A year later, Dawson and Blagg² published the first article with the term 'Pilot Study' in the title. These appear to be the first uses of the term 'Pilot' to refer to a particular type of study, although neither author defined the term. 'Pilot Study' appeared for the first time in a veterinary article in 1967,³ followed a year later by Wallach and Frueh.⁴ These authors also failed to define the term.

The *Dictionary of Epidemiology* defines 'Pilot Study' as 'A small-scale test of the methods and procedures to be used on a larger scale'.⁵ The National Institutes of Health's National Center for Complementary

and Integrative Health states that 'The goal of pilot work is not to test hypotheses about the effects of an intervention, but rather, to assess the feasibility/acceptability of an approach to be used in a larger scale study'.⁶ Similarly, Leon *et al*⁷ state that 'A pilot study is not a hypothesis testing study. Safety, efficacy and effectiveness are not evaluated in a pilot [study]'. Therefore, a pilot study carries the expectations that a larger study will follow, and that the findings of the pilot study are mostly irrelevant, other than in determining feasibility, or fine-tuning methodological protocols, including intended statistical analyses. Occasionally, pilot studies can generate useable data, or alert investigators to unforeseen issues, variables or events that warrant consideration. Rarely, pilot studies can alert investigators to harm that warrants termination of the larger scale project.

We are not aware of any published definitions or guidelines for the term 'Pilot Study', or its use, in veterinary medicine. While authors appear to increasingly use the term 'Pilot Study' in veterinary articles, few, if any, cite references that define the term, define the term in their text or explain what characteristics of their study make it a 'Pilot Study'. The influential 20th-century philosopher Ludwig

Veterinary Record (2019) doi:10.1136/vetrec-2019-105377

¹Clinical Sciences, Cornell University, Ithaca, New York, USA

²Veterinary Information Network, Davis, California, USA

³Population Medicine and Diagnostic Sciences, Cornell University, Ithaca, New York, USA

E-mail for correspondence: Dr Mark Rishniw; mr89@cornell.edu

Provenance and peer review Not commissioned; externally peer reviewed.

Received January 23, 2019

Revised July 11, 2019

Accepted July 15, 2019

Wittgenstein stated: *'let the use of the words teach you their meaning'*.⁸ Given that an explicit meaning of 'Pilot Study' in veterinary literature seems not to exist, we analysed articles from the veterinary literature to see if we could deduce the meaning of 'Pilot Study' by examining the characteristics of articles using the term in their titles. We formulated several hypotheses. First, we hypothesised that the use of the term 'Pilot Study' is increasing in veterinary literature. Secondly, we hypothesised that pilot studies rarely lead to further investigation or validation by the authors. Thirdly, we hypothesised that pilot studies are infrequently cited by other investigators, but frequently garner citations in review articles. Fourthly, we hypothesised that authors use the term to denote studies that have small sample sizes or are not generalisable to a broader population. Fifthly, we hypothesised that pilot studies are mostly authored by inexperienced authors. Sixthly, we hypothesised that the term is inconsistently applied by authors. Finally, we hypothesised that authors do not define what they mean by 'Pilot Study' in their manuscripts.

To investigate these hypotheses, we critically assessed a set of veterinary articles that included the term 'Pilot Study' in the title.

Materials and methods

We conducted several identical searches using the PubMed database in January 2019. We used the following search string (with modifications of the publication date) for each search:

"Pilot Study"[TITLE] AND "YEAR"[Publication Date] AND ("veterinary"[Subheading] OR "veterinary"[All Fields] OR "veterinary medicine"[MeSH Terms] OR ("veterinary"[All Fields] AND "medicine"[All Fields]) OR "veterinary medicine"[All Fields]) where "YEAR" was the year of interest.

We conducted searches for the years 2008–2017 (10 identical searches).

We then excluded articles from further consideration if they were studies using animals as models for human disease or studies focused on human outcomes (with the exception of studies examining veterinary students or veterinarians). We excluded articles that had a publication year other than the year of interest (some articles had epub years that matched our search criteria, but final publication years that did not match—these were only counted once, in the year of the final publication date).

For all pilot studies published from 2008 to 2012, we recorded the following information: author names, journal details (journal, year, volume, issue, pages), whether the outcome was positive or negative, the type of study (development of a method, examination of a diagnostic test, examination of an intervention), target species and sample size.

To examine whether the frequency of the term 'Pilot Study' was increasing, we searched PubMed using the same search string as above, but without the first search term ('Pilot Study'[TITLE]), to get a count of all articles published each year in veterinary journals indexed in PubMed, from 2008 to 2017. We calculated the proportion of pilot studies for each year as the following:

$$\text{Proportion (\%)} = \frac{\# \text{ pilot studies}}{\# \text{ articles}} * 100$$

We then compared these proportions using a chi-squared test of independence.

To allow for sufficient time for a follow-up study to be published, we limited our analyses to pilot studies published in 2008–2012. We assumed that if a pilot study were to be followed up, the subsequent publication would appear within five years. We also assumed that the authors would cite their pilot study in a follow-up study. Therefore, to determine whether a follow-up study had been performed by the same author group, we examined articles that had cited the pilot study. We also examined the citations to determine whether any other investigators performed studies similar to the pilot study that they were citing.

To determine the number of citations for each pilot study, we searched for citations using Web of Science (V.5.31) in January 2019 using the article title as the topic search term. Web of Science reports the number of times an article has been cited and provides the details of the citing article. We recorded if the pilot study was cited in a review article, and whether the review article was written by the authors of the pilot study or by other authors.

To determine the publication experience of the first author, we searched for publications by the first author of each pilot study in PubMed and arranged them in order of publication. We considered authors inexperienced if they had fewer than four publications before the pilot study, or if they had only published within the two years before the pilot study. We compared the proportions of experienced and inexperienced authors who published follow-up studies using a chi-squared test of independence.

To examine the consistency of the use of the term 'Pilot Study', we searched each journal in which a pilot study had been published between 2008 and 2012 for comparable studies that did not include the term 'Pilot Study' in the title. If an article could not be identified in the same year as the pilot study, we searched adjacent years for comparable articles. We considered articles comparable if they had similar sample sizes, same species (where possible) and similar purpose (ie, assessment of diagnostic test, analysis of intervention and so on).

Finally, we examined each pilot study between 2008 and 2012 to determine if the authors defined the term in their manuscript.

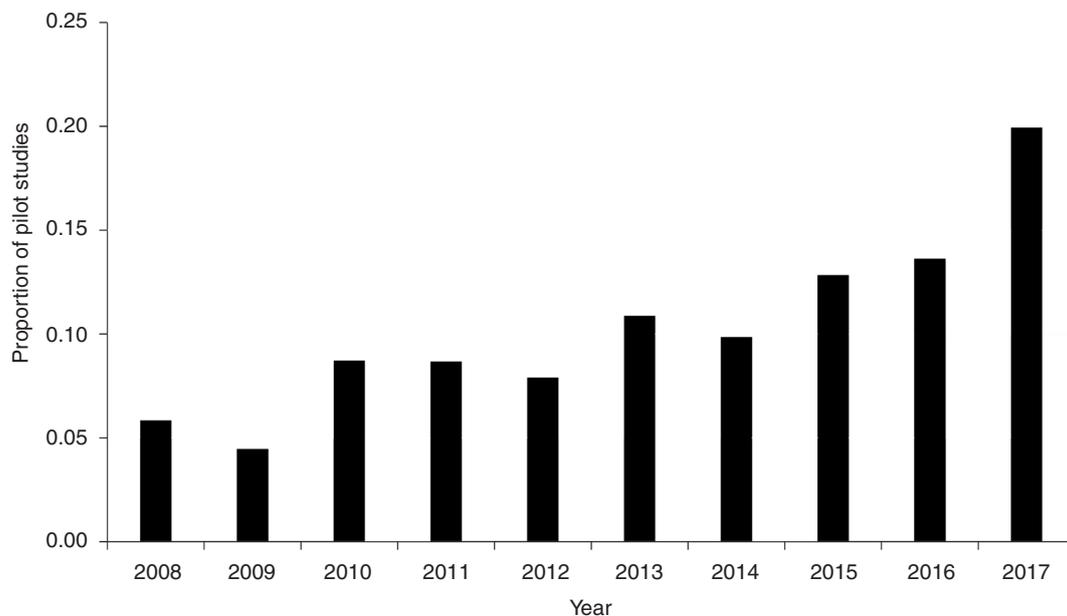


Figure 1 Use of the term 'Pilot Study' is increasing in veterinary medicine. The column graph illustrates the proportion of articles per year using the term ($P=0.0001$).

All the data are available as online supplementary file 1.

Results

We identified 73 articles with the term 'Pilot Study' matching our criteria that were published between January 2008 and December 2012 (online supplementary file 1). We identified an increasing number of pilot studies from 2008 to 2017 ($P=0.0001$; figure 1).

We identified 17 of 73 (23 per cent) pilot studies with follow-up studies by the same investigator group that verified or further assessed the findings of the pilot study. Another five pilot studies were replicated by investigators not associated with the original investigators. We subjectively determined that the sample size of the follow-up studies substantially exceeded that of the pilot study in 12 of 22 studies. Five follow-up studies had sample sizes that were either smaller or similar to those of the pilot study. None of the follow-up studies justified their sample size with a sample size calculation.

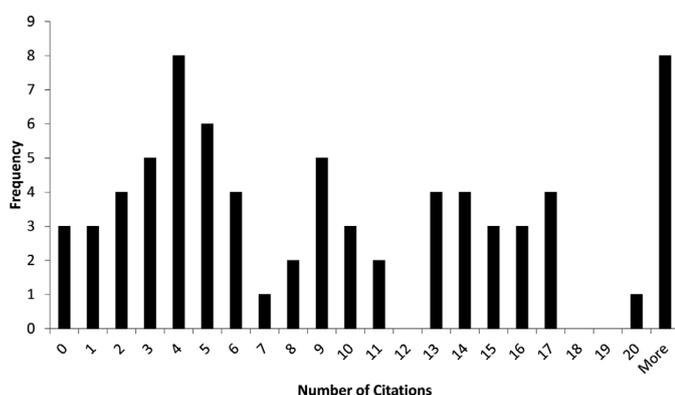


Figure 2 Most pilot studies garner few citations. The histogram illustrates the number of pilot studies ($n=73$) with k citations. The median number of citations was 9.

The 73 pilot studies garnered a median of 9 citations (range: 0–42) (figure 2). However, 54 of 73 (74 per cent) pilot studies had been cited at least once in a review article; 21 per cent of authors of a pilot study cited their pilot study in a review article.

The 73 pilot studies had a median sample size of 10 (range: 2–614) (figure 3). The authors considered the outcome to be positive in 62 of 73 (85 per cent) pilot studies. The study with the largest sample size ($n=614$) involved an investigation in one dairy herd, which the authors correctly identified as a single experimental unit with 614 replicates.

Of the 73 pilot studies, 30 (41 per cent) had a first author who met our definition of 'inexperienced'. Inexperienced authors were no less likely to publish studies lacking follow-up than experienced authors (6 of 30 pilot studies by 'inexperienced' authors were followed up by the investigator groups, versus 9 of 43 pilot studies by 'experienced' authors; $P=0.84$).

For 72 of 73 pilot studies, we identified a comparable study that did not use the term 'Pilot Study' in the title. Furthermore, we identified an author who used the term

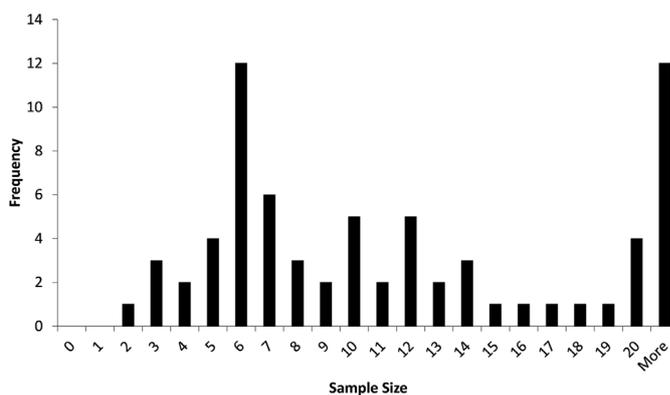


Figure 3 Most pilot studies have small sample sizes. The histogram illustrates the number of pilot studies ($n=73$) with k experimental units. The median sample size was 10.

'Pilot Study' inconsistently for several methodologically identical studies.

We could not find any authors who defined the term 'Pilot Study' in their manuscripts, explained why they used the term to describe their work or cited the relevant definitions provided in the medical literature.

Discussion

Our study shows that the term 'Pilot Study' is used with increasing frequency in the veterinary literature. No authors who used the term defined it or explained why they used the term. Furthermore, none of the pilot studies we identified adhered to the medical definitions of 'Pilot Study'.⁵⁻⁷ Instead, almost all authors provided quantitative results, interpreted these results and generalised the conclusions of their findings. Most pilot studies had small sample sizes. Most pilot studies garnered relatively few citations, but most were cited in review articles. Only one in five pilot studies was followed up by the original investigators. Authors and journals applied the term 'Pilot Study' haphazardly—we found comparable studies that failed to use the term for almost every pilot study in our analysis. We even identified an author who used the term inconsistently for his own virtually identical studies.

We found authors increasingly use the term 'Pilot Study' (from 10 articles in 2008 to 68 in 2017). We limited our search to use of the term within the article title. However, we have observed the term 'Pilot Study' within the abstract or the body of the manuscript, but omitted from the title, in multiple articles. Consequently, we likely underestimated the frequency of the use of the term 'Pilot Study'.

Because no definition exists for 'Pilot Study' in the veterinary literature, we followed Wittgenstein's dictum to '*Let the use teach you the meaning*'.⁸ Our review of the use of the term 'Pilot Study' in veterinary medicine indicates it is meaningless. Authors do not define what they mean by 'Pilot Study', they do not cite a source for the definition of the term, nor do they specifically report why their work meets the definition of a pilot study. Authors almost never use the term to describe exploratory 'putting a toe in the water' studies to determine if the 'water temperature warrants diving in'. They occasionally use the term for large, well-designed studies that are not necessarily generalisable to a larger population.⁹ Authors commonly use the term for studies that are never followed up more comprehensively; only occasionally is the term used for studies that are followed up by the authors or by others. Authors commonly use it to indicate small studies, but use the term haphazardly—similar small studies use or do not use 'Pilot Study' in the title (online supplementary file 1). Even multiple, virtually identical studies by the same author published in the same journals are variably described as 'Pilot Studies' without explanation.¹⁰⁻¹⁴ Such inconsistent use implies that authors, editors

and reviewers feel that adding or removing the term is inconsequential, an attitude consistent with our view that the term is meaningless. Unlike terms such as 'randomized', 'case report', 'case series', 'survey' and so on that convey information to the reader, the meaning of the term 'Pilot Study' must be reverse-engineered—after scrutinising the paper the reader is left to guess why the authors described it as a 'Pilot Study'.

Why, then, do authors include the term ever more frequently? Clearly, for authors, the term signifies *something*, perhaps 'We know our study is small, but we promise to do better by producing a more definitive study'; however, such promises are rarely kept. The more covert signal, in our opinion, is 'We know our study is small, likely underpowered, and the results of very limited value, so please be lenient in your evaluation'. This 'Deficiency Signaling' is a tacit acknowledgement of the low quality of the paper.

One can approach the term from a help versus harm analysis. If the term is 'helpful', then it should be retained and applied consistently to particular types of studies, based on some definition. If the term is neutral (neither helpful nor harmful), then it is redundant and warrants removal—it provides no useful information to the reader. If the term is 'harmful', it should also be removed.

How could the term be 'helpful'? If, as we suggest, the term is a form of 'Deficiency Signaling', then it can act as a red flag, or warning, to readers, reviewers and editors that the study should be interpreted extremely cautiously, or even avoided. Reviewers can be additionally careful in scrutinising the study, especially the authors' interpretations of their findings. They can request that the authors justify the use of the term. Indeed, if it is meant to be helpful in this manner, the term 'Pilot Study' could just as easily be replaced by a red flag icon.

How could the term be 'harmful'? Studies with sample populations in the single digits have a high probability of observing a false positive finding or an exaggerated effect size.^{15 16} None of the pilot studies in our analysis considered the possibility of a false positive observation, or an exaggerated effect size as the cause of their findings. Instead, in almost all cases, authors reported their findings as conclusive, and often suggested that their data be considered so by other investigators. If left uncited, a pilot study would cause little harm. However, our data show that 75 per cent of pilot studies are cited in review articles, where they are not necessarily critically evaluated, thereby gaining legitimacy, effectively being 'laundered' by their inclusion. Given that pilot studies have limited value, this legitimising by authors of the review article allows weak evidence to enter the mainstream veterinary literature with the risk that it will be adopted into clinical practice. Once such studies are cited in review articles, the reader commonly assumes the finding is

real, rarely going back to the original literature and evaluating it critically.

Because the term is meaningless, used haphazardly and, when used, causes more harm than good, we suggest that the term 'Pilot Study' be abandoned in veterinary medicine. We are not the first to suggest that the term 'Pilot Study' be removed from manuscripts. A decade ago, the editors of *Journal of Clinical Nursing* decided to remove the term from manuscripts with few exceptions.¹⁷ These editors recognised the misuse of the term, and suggested that in some instances the 'author has decided to apply the label 'pilot', seemingly to make it more acceptable [to the readers and reviewers]'. Other authors have also recognised that the term has been 'misrepresented as an excuse for not having enough of a sample'.^{18 19}

What alternatives exist for authors of what are currently referred to as 'Pilot Studies'? Because the term is meaningless, no title will be the worse for its absence. Authors could adopt substantive, declarative titles that inform the reader of what they might expect in the paper they are about to read, for example, by indicating the species, study design, main outcome and sample size.

In conclusion, if authors choose to include the term 'Pilot Study', we suggest that reviewers and editors require the authors to explain in the manuscript exactly why the study warrants this largely meaningless term, and readers should recognise the warning implicit in its use.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Data availability statement All data relevant to the study are included in the article.

© British Veterinary Association 2019. No commercial re-use. See rights and permissions. Published by BMJ.

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/vr.105377>).

References

- 1 Parnell RW. Health examinations of students; pilot survey in Oxford. *Lancet* 1947;2:939–41.
- 2 Dawson CE, Blagg W. The effect of human saliva on the cholera Vibrio in vitro; a pilot study. *J Dent Res* 1948;27:547–52.
- 3 Eisengarten HJ. [On the spreading of avian encephalomyelitis in Thuringen. 1. Pilot study from the Gera district]. *Monatsh Veterinarmed* 1967;22:19–22.
- 4 Wallach JD, Frueh R. Pilot study of an organophosphate anthelmintic in camels and primates. *J Am Vet Med Assoc* 1968;153:798–9.
- 5 Porta M, ed. A dictionary of epidemiology. 5th edn. New York, NY: Oxford University Press, 2008.
- 6 Anon. National Center for Complementary and Integrative Health. *Natl Institutes Heal*, 2017. Available: https://nccih.nih.gov/grants/whattncihfunds/pilot_studies [Accessed 15 Jan 2019].
- 7 Leon AC, Davis LL, Kraemer HC. The role and interpretation of pilot studies in clinical research. *J Psychiatr Res* 2011;45:626–9.
- 8 Wittgenstein L. Philosophical investigations. 3rd edn. Malden, MA: Blackwell Publishing, 2001.
- 9 Caixeta LS, Bicalho RC. A pilot study to determine the production and health benefits of milking visibly lame cows twice daily compared with three times daily. *Can J Vet Res* 2011;75:233–6.
- 10 Giorgi M, Del Carlo S, Saccomanni G, et al. Pharmacokinetics of tramadol and its major metabolites following rectal and intravenous administration in dogs. *N Z Vet J* 2009;57:146–52.
- 11 Kim TW, Della Rocca G, Di Salvo A, et al. Evaluation of pharmacokinetic and pharmacodynamic properties of cimicoxib in fasted and fed horses. *N Z Vet J* 2015;63:92–7.
- 12 Giorgi M, Saccomanni G, Łebkowska-Wieruszewska B, et al. Pharmacokinetic evaluation of tramadol and its major metabolites after single oral sustained tablet administration in the dog: a pilot study. *Vet J* 2009;180:253–5.
- 13 Giorgi M, Yun H. Pharmacokinetics of mirtazapine and its main metabolites in beagle dogs: a pilot study. *Vet J* 2012;192:239–41.
- 14 Giorgi M, Cuniberti B, Ye G, et al. Oral administration of tepoxalin in the horse: a PK/PD study. *Vet J* 2011;190:143–9.
- 15 Button KS, Ioannidis JPA, Mokrysz C, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci* 2013;14:365–76.
- 16 Ioannidis JPA. Why most discovered true associations are inflated. *Epidemiology* 2008;19:640–8.
- 17 Watson R, Atkinson I, Rose K. Editorial: pilot studies: to publish or not? *J Clin Nurs* 2007;16:619–20.
- 18 Kim Y. The pilot study in qualitative inquiry. *Qual Soc Work Res Pract* 2011;10:190–206.
- 19 Perry SE. Appropriate use of pilot studies. *J Nurs Scholarsh An Off Publ Sigma Theta Tau Int Honor Soc Nurs* 2001;33.

