

Review

Identifying Sources of Potential Bias When Using Online Survey Data to Explore Horse Training, Management, and Behaviour: A Systematic Literature Review

Kate Fenner , Michelle Hyde, Angela Crean  and Paul McGreevy 

Sydney School of Veterinary Science, Faculty of Science, University of Sydney, Sydney, NSW 2006, Australia; michelle.hyde@sydney.edu.au (M.H.); angela.crean@sydney.edu.au (A.C.); paul.mcgreevy@sydney.edu.au (P.M.)

* Correspondence: kate@kadnoequine.com.au

Received: 31 July 2020; Accepted: 21 September 2020; Published: 22 September 2020



Abstract: Owner-reported behavioural observations form an essential part of the veterinarians' diagnosis and treatment plan. The way we train and manage horses affects their behaviour and, in turn, their health and welfare. Current horse training and management practices are largely driven by traditional techniques and longstanding methodologies. These approaches generally lack an evidence base for evaluation purposes. The absence of evidence and evaluation contributes to the persistent use of risky practices and this, in turn, increases risk of potential harms for both horse and rider, and fuels questioning of the equine industry's current social license to operate. Objective evidence is required to make training and management decisions based on demonstrable best practice. Large-scale experimental or intervention studies using horses are generally not practical because of the associated costs and logistics of gaining ethical approval. Small studies generally lack statistical power and are subject to the effects of many forms of bias that demand caution in the interpretation of any observed effects. An alternative to collecting large amounts of empirical data is the use of owner-reported observations via online survey. Horse owners are ideally placed to report on the domestic equine triad of training, management, and behaviour. The current article highlights three sources of potential bias in a systematic review of literature on large-scale online studies of horse owners' observational reports that met the following selection criteria: English-language, published, peer-reviewed articles reporting on studies with over 1000 respondents and open access to the survey instrument. The online surveys were evaluated for three common forms of bias: recall, confirmation, and sampling bias. This review reveals that online surveys are useful for gathering data on the triad of horse training, management, and behaviour. However, current use of online surveys to collect data on equitation science (including horse training, management, and behaviour) could be improved by using a standardised and validated tool. Such a tool would facilitate comparisons among equine and equitation science studies, thus advancing our understanding of the impacts of training and management on horse behaviour. The authors of the current review suggest the use of a standardised behavioural and management assessment tool for horses. Such a tool would help define what constitutes normal behaviour within geographically disparate populations of horses, leading to improvements in rider safety and horse welfare.

Keywords: owner-reported observations; horse behaviour; social license; survey bias; equine veterinary behaviour; domestic equine triad

1. Introduction

Domestic horses are generally very compliant animals [1], with an ethology that features exceptional behavioural flexibility and rapid habituation [2]. In the developed world, they are kept mainly as companion, leisure, or sporting animals rather than working animals and are handled by people with a diversity, and often paucity, of relevant skills [3]. Veterinary and husbandry interventions as simple as examinations or grooming can bring risks to human safety [4]. At the same time, traditional and contemporary training and management practices regularly compromise horse welfare across each of the Five Domains [5]. Shortfalls in horse welfare, and associated behavioural manifestations, can go unnoticed and lead to poor welfare outcomes for horses and increase hazards for riders, who are often unaware of the horse's emotional state [4,6].

There is a triadic relationship among training, management, and behaviour. When flaws in training and management practices are left unresolved, a cycle of poor training and management can develop. For example, chasing a horse in a round pen will elicit a flight response which the horse is likely to associate with the round-pen environment and thus offer each time the round pen is encountered, thus jeopardising both horse welfare and handler safety [7]. These training flaws can lead to further deterioration in the horse's behaviour and escalation in the use of force and coercion in training, followed by spiraling risks of flight responses, habituation to pressure cues and associated safety risks to riders and handlers. This cascade of 'knock-on' effects directly jeopardises horse welfare. Progress in understanding the influence of training and management practices on horse behaviour relies on the collection and analysis of large-scale data on each of these variables to establish a baseline and evaluate change.

The concept of social license to operate (SLO) relates to community trust in a business, organisation, or entity operating within legal, cultural, social, and ethical norms, while striving for best practice. This is important because the traditional methods of animal handling, training, and management are increasingly being challenged [8,9]. Public attention has questioned whether industries, ranging from those using laboratory rodents to those using racing horses, maintain a SLO to use animals as they have done historically [10,11]. SLO is not a formal or written contract, rather society's permission for an entity to continue to operate. With the rise of social media, which fuels the distribution of troubling images, the equestrian community risks losing its SLO unless it can be seen to be taking proactive, affirmative action to improve participant safety and horse welfare [12]. In recent years, a series of ridden horse welfare issues have been prominent across the media, risking their SLO. These include the use of Rolkur [13], whip use in racing [14], the use of restrictive nosebands in various sporting disciplines [15], and wastage within the breeding and racing industries [16]. Some of these (such as restrictive nosebands) reflect the perceived need to reduce unwelcome behaviour while others (such as wastage) directly result from unwelcome behaviour.

Horse training and management practices currently lack a robust evidence base, despite decades of research into equine and equitation science [7]. Many traditional practices are untested empirically and even studies that have tested training techniques may suffer from small numbers of practitioners [17]. To date, data collection tools for the use of horses in sport and leisure contexts have been disparate and largely unvalidated [18].

Horse owners are ideally suited to contribute data on the domestic equine triad of training, and management practices and their horses' behaviour because they are generally very familiar with their animals [18]. Surveys have been used to collect data on horse behaviour, training, management, and health for over four decades [19] and online questionnaires for approximately twelve years [20]. Various disciplines, training and management practices have been investigated and owner-reported observations have already contributed considerably to our understanding of horse-human relationships [21–23]. However, these data have generally been collected with a narrow focus [24], small sample numbers [25], or a closed cohort of horses or riders, where inclusion was restricted to individuals meeting identified criteria [26], making it difficult to generalise their results across the equine population.

Nevertheless, online data collection has many benefits including cost effectiveness, the ability to recruit large numbers of respondents, and the ease and speed of global spread with the ability to target specific stakeholder groups. However, ease of use may have inadvertently led to online questionnaires lacking standardisation and validation, such that large numbers of rapidly produced surveys may have poor reliability, making results both difficult to evaluate individually and to compare to one another.

Historically, questionnaires for horse owners have collected both qualitative and quantitative data. The collection of qualitative data on horse owners' and riders' interpretation of behaviour has provided important insights into perceptions that potentially affect welfare. For example, research has shown that attitudes towards mares and geldings vary considerably, potentially putting welfare of mares at risk [27]. Furthermore, Visser and Van Wilk-Jensen [28] found significant differences among horse enthusiasts' (trainers, riders and coaches) beliefs and practices around welfare issues. While horse-owner attitudes and beliefs merit scientific scrutiny, the collection of qualitative data on attitudes and beliefs is unlikely to further our understanding of what constitutes normal behaviour in horses or be useful when making geo-cultural comparisons [29].

Defining what constitutes normal behaviour in horses necessitates the collection of quantitative data. It is important when investigating relationships among training, management and behaviour, that horse owners are asked to report their observations rather than their opinions or interpretations of behaviours. Free-text commentaries on observed behaviour require two interpretation procedures to take place: first the owner interprets the behaviour they observe; and second, the researcher interprets their text response. A survey instrument that examines the frequency of behavioural observations reduces the effects that arise from errors in interpretation.

The horse–human dyad is more complicated than that of most other human–other animal interactions due to horses being ridden. Clearly, there is no equivalent horse–horse interaction to mounting and riding astride the horse. This complexity begins with mounting the horse for riding [30]. Prior to this stage of the horse's education, much of the horse–human interaction—for example, stroking, grooming, and leading—can be plausibly aligned with conspecific interactions. All horses, especially those that are ridden or driven, require multifaceted management regimes and various levels of training. For example, an elite equine athlete might be housed in a stable full time, exercised using lunging, riding, and walking or even swimming. This same horse may have specialized tack and equipment, such as bits and headgear, which would not be used with horses in other disciplines. Alternatively, a horse may be housed in a large field and only interact with the owner once a month. However, in both cases, training and management, including interactions that handlers might not consider to be formal training, affect horse behaviour, which, in turn, reflects and contributes to the animal's health and welfare status.

While online surveys are inexpensive, expedient, and able to reach vast numbers of potential respondents, we must address the inherent risk of bias with this method of data collection. All online surveys are likely exposed to both self-selection bias, where only individuals with internet access who choose to visit a website and decide to participate can respond [31], and non-response bias, where participants are unwilling or unable to answer questions [32]. A further three types of bias are more readily addressed, including recall bias (participants fail to remember past events accurately), confirmation bias (participants interpret information in a way that is consistent with their existing beliefs), and sampling bias (the survey respondent sample does not represent the population). Finally, our ability to disseminate study results has a direct impact on their uptake and, thus, usefulness. The current systematic literature review investigates questionnaires that have been used to obtain data from horse owners. Its objective was to examine the current use of online surveys to understand how and where large datasets are being collected and review their reliability, by assessing recall, confirmation, and sampling bias, in achieving stated aims. The chief goal was to identify best practice for large-scale data collection and dissemination on the domestic equine triad of horse training, management, and behaviour.

2. Materials and Methods

A systematic literature review aims to minimise bias by pre-specifying the eligibility criteria to address a specific research question. To maximise the chance that all relevant studies were captured in the current review, the initial search was as broad and simple as possible. A title, abstract and keywords search of three large databases (Web of Knowledge, Scopus, and PubMed) using the search terms 'equine OR horse' AND 'survey OR question *' AND 'behav *' was conducted on 15 January 2020. After removal of duplicates, 813 unique articles were identified for screening. Titles and abstracts were reviewed by two independent reviewers (KF and AC), and 444 articles excluded from further review because they did not have an equine focus. The remaining 181 articles were retained for full-text review and were examined against our specified inclusion criteria (Table 1).

Table 1. Inclusion and exclusion criteria used for the literature review investigating the use of online surveys for data collection on horse training, management, and behaviour.

Criterion	Inclusion	Exclusion
Targeted respondents	Equine owner/caregiver/rider, any age or gender	Studies restricting participation by limiting availability to specific cohorts
Core demographic data	Breeds, disciplines, sex of horse, country, behaviour, training, management	
Questionnaire	Online, unrestricted url access, English-language, peer-reviewed original research articles	Non-English publications, review articles
Respondent numbers	Studies describing 1000 individual horses or more	Studies describing fewer than 1000 individual horses
Data collected	Quantitative/observational data reported on individual horses	Studies investigating horse owners' beliefs or opinions

Of the 181 articles identified for full-text review, 146 were removed due to being either a review article ($n = 26$) or because the surveys had fewer than 1000 respondents ($n = 113$). A further 11 articles were excluded because they were not written in English. Only English-language articles were included in this review because the research team lacked a reliable translation service and considered that, to appraise each instrument fully, such a service would be essential, given the nuances of describing behaviour and framing questions in different languages [33]. A further 13 studies were excluded as they were not open surveys. An open survey was defined as being available to any horse owner/caregiver and not restricted to a defined population, for example 'wind-sucking' horses. Of the 20 articles classified as open survey, seven were rejected because they were qualitative investigations examining horse owners' beliefs and opinions, rather than observational reports about an individual horse's behaviour, training, or management. An additional article [34] authored by the current research group and meeting our inclusion criteria was not identified by the online search but found during hand searching. The search was repeated on 1 July 2020 and two further articles [15,35], one of which was authored by the current research group, that were published after the initial literature search was conducted were also included, leaving a final set of 14 articles for full review and quality assessment.

Online surveys, as with other methods of data collection, are subject to various sources of bias [36]. When assessing reliability for our literature review, three potential sources of bias were considered as themes, and the included articles were given a subjective rating out of five on each theme (Table 2):

- (1) Recall bias may arise when participants are required to remember past events, with the length of the recall period increasing the likelihood of this bias occurring. In our review, we considered the lag between the behaviours or management practices occurring and their being reported. Studies that asked participants to report on their horses' behaviour or management in the 'previous week' [37–43] scored five out of five. In contrast, a study [44] that asked respondents to

report on behaviours witnessed during a specific event a year previously scored two out of five (see Table 2).

- (2) Confirmation bias may arise when a participant must rely on their preconceptions or beliefs to answer questions. We reviewed the survey items used and rated the studies according to their apparent objectivity. Studies that asked respondents only for simple observations [43,45], such as ‘does your horse wear a blanket?’ or ‘is your horse kept in a group?’, scored five out of five; whereas studies with questions that required considerable interpretation by the participant [37], such as ‘does the horse like to get its own way?’, scored three out of five (see Table 2).
- (3) Sampling bias may arise if a particular demographic of survey participants has a higher or lower probability of being included than others, thus affecting how representative the findings might be of the general horse population. For example, the work of Hartmann et al. (2017) reported on winter housing in a cold climate and generated a low score for sampling bias as the sampled population was located in a country with a climate not relevant to the global horse population [45]. Conversely, a high score would be obtained if the survey investigated behaviours that were unlikely to be geographically influenced [42] (see Table 2).

The scoring system (Table 2) was developed by the authors to rank bias themes between studies. Recall bias is a measure of how far back participants were required to remember when answering questions. Confirmation bias explored each question on the survey and rated it on a 5-point scale as objective (frequency with which a behaviour was observed or management practice used, score 5) to subjective (requiring considerable interpretation from participants, score 1). Finally, sampling bias assessed the relevance of all items within the survey to the entire equine population.

Table 2. Scoring system used for three themes that relate to potential bias in surveys.

Theme	Score				
	1	2	3	4	5
Recall bias	>1 year	up to 1 year	3 to 6 months	<1 and >3 months	Current or up to 1 month
Confirmation bias	Completely subjective interpretation	More subjective than observational	Both subjective and objective	More observational than subjective	Completely observational
Sampling bias	Relevant to very specific equine populations	Extremely limited to climatic or geographical relevance	Some climatic or geographical limitations	Partial climatic or geographic relevance	Relevant to all equine populations

3. Results

There were 14 studies meeting the inclusion criteria (Table 1). Data extracted from these studies included the behaviours and/or management practices investigated, the objectives of the studies and the conclusions drawn by the authors (Table 3). The included studies covered a wide range of behaviours and management techniques, horse breeds and ages and participant demographics. The scope of geographical coverage was somewhat limited, with the majority (8/14) of studies collecting data in the UK only. Many of the studies (6/14) came from a single UK-based research group. While all of the surveys were available online and open to the public, most (12/14) were advertised only in their own immediate geographical location and some (8/14) recruited participants solely from those geographical areas [United Kingdom (5), Scandinavia (2), Australia (1), and New Zealand (1)].

Bias scores, agreed by both assessors, for each of the three items were collated to give articles a total indicative score out of 15 (see Table 3).

Table 3. Included reviews (Ref.) characteristics and reliability assessment [(recall bias (RB), confirmation bias (CB), and sampling bias (SB) with total reliability score (Total)] of studies using online surveys of >1000 participants to quantify horse behaviour.

Ref.	Behaviours Investigated	Objectives (<i>n</i> = Number of Respondents)	Outcomes/Conclusions	RB	CB	SB	Total/15
[42]	Oral investigation, licking, searching, and biting/nipping	To investigate whether giving treats by hand, or the use of clicker training, was associated with oral behaviours including searching for food, licking and biting (<i>n</i> = 1067)	No significant associations between hand feeding and nipping/biting clothing were found. The use of clicker training was not associated with unwelcome oral behaviours.	5	4	5	14
[35]	Training, management, and behaviour	To explore behavioural differences between ridden mares and geldings (<i>n</i> = 1233)	No significant differences were found in ridden behaviour but mares were more likely to move away when being caught and geldings were more likely to chew on lead ropes when tied and rugs.	3	5	5	13
[43]	Human behaviour (housing practices)	How horse owners in these four countries keep their horses and their understanding of the importance of the horse's need for social contact. How management practices relate to horse breed and owner opinions (<i>n</i> = 3229)	Most horses were kept in groups for at least part of the 24 h period and those never grouped were primarily competition horses or stallions. Horses are usually kept in similar age and sex groups.	5	5	3	13
[39]	Stable-related and handling problems	To discover the prevalence of stable-related behavioural problems when handling horses (<i>n</i> = 1850)	A high prevalence of handling and stable-related problems was found, indicating that restricted turnout might have welfare costs.	5	4	4	13
[40]	15 ridden behavioural problems	To investigate the use of equipment and training practices and the prevalence of 15 behavioural problems in UK leisure horses (<i>n</i> = 1326)	The high prevalence of reported behavioural problems (91%) in ridden horses suggests a risk to rider safety and horse welfare.	5	4	4	13
[41]	Prefeeding behaviour (aggression, frustration, and stereotypies)	To generate data on diet and owners' perceptions of the use of supplements, feeding practices and the prevalence of behavioural problems encountered prior to feeding in UK leisure horses (<i>n</i> = 1324)	Current feeding and management regimes may compromise horse welfare.	5	4	4	13

Table 3. Cont.

Ref.	Behaviours Investigated	Objectives (<i>n</i> = Number of Respondents)	Outcomes/Conclusions	RB	CB	SB	Total/15
[38]	Stereotypic behaviour and problem behaviour	To assess the risk factors associated with routine management practices on behavioural problems in UK leisure riding horses (<i>n</i> = 1226)	Different groups of behavioural problems were associated with different management risk factors. These leisure horse findings are similar to those found in performance horses.	5	4	4	13
[37]	Personality traits—excitability, anxiousness, dominance, and protection	To explore potential breed differences in horse personalities using a trait theory approach (<i>n</i> = 1223)	Horse breeds were found to differ in their personality types, and this finding was found to align with the traditional views on breed differences.	5	3	5	13
[45]	Human behaviour (blanketing/clipping)	The extent of blanketing and clipping horses in Sweden and Norway and owners' knowledge of the literature surrounding their use and heat loss (<i>n</i> = 6197)	While blanketing and clipping are very common practices in Sweden and Norway, owners understanding of the horse's thermoregulation and the literature behind these practices is limited.	5	5	2	12
[46]	Behavioural problems (stable related, handling and prefeeding)	The use of statistical techniques to identify associations and clusters of behaviours that could reflect an underlying welfare issue (<i>n</i> = 1681)	Behaviour clusters were found and may represent an underlying welfare issue, but this is not recognised by owners or traditional horsemanship methods. This leads to major welfare concerns. Once these patterns have been recognised by collecting data and analysing it, this information needs to be passed on to owners to enable them to take early action and improve horse welfare.	4	4	4	12
[34]	Human behaviour (equipment use)	To establish benchmark data on the prevalence of the use of an apparatus to apply aversive stimuli to ridden horses and ponies. The use of bits, nosebands, whips and spurs were investigated (<i>n</i> = 1101)	Crank nosebands in dressage and curb bits in Western disciplines could pose a risk to welfare. Whips and crops are used more in dressage and spurs by Western riders.	3	3	5	11

Table 3. Cont.

Ref.	Behaviours Investigated	Objectives (<i>n</i> = Number of Respondents)	Outcomes/Conclusions	RB	CB	SB	Total/15
[47]	Headshaking	To discover the prevalence of headshaking observed within the last year in horses in the UK and its relations to sex and breed (<i>n</i> = 1014)	The estimated prevalence of headshaking was 4.6% and 75.4% of respondents had attempted at least one form of treatment trial. Headshaking was not related to sex or breed.	3	3	5	11
[15]	Human behaviour (equipment choice, perceived effectiveness and complications)	The distribution of noseband use, owner-reported reasons for use, perceived effectiveness, design choices, tightness monitoring and detrimental consequences (<i>n</i> = 2332)	A plain cavesson was the most commonly used noseband (46.6%), and 18.6% of users reported at least one physical or behaviour complication from noseband use.	3	2	5	10
[44]	Behaviours—running, vocalisation, and breaking through fences	The effects of fireworks on horses, horse responses and management interventions employed (<i>n</i> = 1111)	A total of 39% of horses were rated as ‘anxious’ or ‘very anxious’, resulting in running (82%) and breaking through fences (35%). While 77% moved their horses as a management strategy, this was ineffective 30% of the time.	2	3	2	7

4. Discussion

The considerable number of studies found, in the first sweep of the current search, and the breadth of topics covered are testament to the endeavours of equine and equestrian scientists. The selected studies incorporated a wide range of topics including training interventions, management practices, and ridden and in-hand (on the ground) behavioural problems, with large sample sizes, demonstrating the important contribution such surveys make to our understanding of core elements within the horse–human dyad.

The search and selection criteria for the current review revealed 14 relevant, original research articles. Critical analysis of these articles confirmed the potential for three main sources of bias that reduced, at least in some part, the reliability of their results. Some of the identified limitations are more obstructive to the drawing of conclusions than others and are more readily overcome in future attempts to embrace best practice. For example, sample bias is somewhat easier to address in survey design and participant selection. Confirmation bias can be avoided by not requiring participants to interpret their horses' behaviour or motivations for performing a behaviour. For example, rather than asking respondents whether their horse was 'anxious', simply asking them to report how frequently the horse performed a given behaviour, such as running along the fence, requires no interpretation and is thus less likely to reflect confirmation bias.

The authors of the current review recognise and work within the described restrictions inherent in large-scale data collection themselves and, as such, acknowledge and applaud the efforts made by the designers of the studies discussed here. There is little doubt that these studies have significantly advanced knowledge in the field. However, our review also reveals that the quality of data collected via online surveys varies. This reflects the dearth of survey instruments that are standardised, validated, and longitudinal. Standardisation across instruments facilitates valid comparisons among populations of horses. The use of a survey instrument that has been subjected to inter- and intra-reporter reliability increases confidence in both the survey and its findings.

The formulation of objective questions that simply ask participants to report on the frequency with which they observe behaviours should be a top priority for questionnaire designers and a deciding factor on the reliability of the resulting behavioural data. For example, in their survey of the management of horses during fireworks, Gronqvist et al. [44] acknowledged this limitation, but still chose to ask participants to rate '*how anxious their horses were*' when exposed to fireworks. Perhaps if they had asked participants about the behaviours they observed, such as running the fence, vocalising and so on, their survey would have produced a more robust, and thus generalisable, result. In the same vein, attempting to ascribe anthropomorphic terminology such as personality traits to horses may not be useful. For example, Ross et al.'s [47] participants were required to include interpretation in their responses, e.g., to report whether their horse displayed '*an anxious expression*' when headshaking.

Attempting to guess a horse's motivations for behaving in a certain way can lead to erroneous assumptions and jeopardise welfare [48]. For example, Lloyd et al. [49] asked participants to assess their horses' personality using a 7-point Likert scale with pre-defined adjectives such as '*Motherly—Provides warm receptive secure base for others, is tender and caring*', and concluded that, despite the answer requiring considerable participant interpretation, the use of the anthropomorphic terminology was more useful than an objective description of the animal's behaviour. However, as respondents likely differ in their definition of 'tender and caring', perhaps the insertion of requests for objective behavioural observations might be a useful addition to such inquiries. In such cases, simply asking horse owners or caregivers to report on the behaviours observed (that may or may not have led them to conclude that the horse was 'tender and caring') would be useful information to collect.

As there is currently no validated, standardised tool available for exploring the domestic equine triad, researchers have had to start afresh each time a behavioural survey is undertaken, as demonstrated in our results. This results in a collection of dissimilar survey tools, making comparisons between studies impossible. A single questionnaire, tested for construct validity and reliability, would solve this problem while building a behavioural database that is greatly needed. With these baseline data in place,

researchers could ask for additional information on topics of interest or interventions, providing validity and enabling comparison between populations.

Online survey tools can provide useful cross-sectional data but, when building an evidence base from which to establish a description of normal behaviour and investigate the effect of training and management regimes, it is important to source data at different time points. Longitudinal studies are required to investigate the consequences of training and management over time. By retesting the same population at regular intervals, using the same standardised and validated instrument, one can draw evidence-based conclusions on topics that affect rider safety and horse welfare. Unfortunately, none of the studies in the current review were longitudinal. While this may reflect the current nature of scientific endeavour in the equine and equitation science fields and the reality that these disciplines are essentially still in their infancy [50], the current authors suggest that the collection of longitudinal data should be seen as an opportunity to enhance the disciplines.

The current review reveals that data on the domestic equine triad of training, management, and behaviour for focal horses need to be collected concurrently, and repeated measurements need to be taken longitudinally to reveal how the elements of the triad interact. A longitudinal survey tool will allow collection of data that can be analysed to reveal how changes in training and management manifest as behaviour, and how these behavioural changes influence horse welfare. While not all survey objectives demand the collection of longitudinal data, our search results indicate that no large-scale longitudinal data on the domestic equine triad have been collated.

The collection of longitudinal data may also assist when considering patterns in responded data over time. Recall bias is largely a function of the specific timeframe participants are invited to report on [51]. Inaccuracies in recall are an important limitation when surveying participants in many areas of science, and are particularly notable in the field of nutrition [52]. Hockenull and Creighton [38–41] managed this by instructing their participants to report on the previous week's observations, whereas Gronqvist et al.'s [44] participants were asked to report on events that occurred a full twelve months previously. In contrast, Hartmann et al.'s [43,45] management practice surveys specified that respondents were to report on horses currently in their care, thus maximising the possibility of returning accurate results. By allowing respondents to report on horses that were no longer in the possession of the participants, either having been sold or deceased, Ross et al. [47] were, arguably, relying too heavily on owners' recollections.

The quality of data collected is limited by the validity and reliability of the survey instrument applied. While the risks of bias, as discussed in this review, can be minimised by a carefully crafted questionnaire, this does not validate the instrument or ensure it collects reliable data among various respondents or over time. The current authors suggest the need for a three-part validation and reliability assessment process [53]. Construct validity can be assessed by asking respondents to give a general assessment of their horse and compare that to their results. This has been used to validate the Feline Behavior Assessment and Research Questionnaire (Fe-BARQ) [54], wherein respondents were asked "Are you currently experiencing any problems with this cat's behavior or temperament?", and answers were found to align with the results of the Fe-BARQ, providing some construct validity. Reliability testing requires two assessments, intra-rater and inter-rater reliability assessments. Intra-rater reliability can be assessed by the same respondent reporting on a horse twice over a given period. To show intra-rater reliability, results from both assessments must align. Finally, inter-rater reliability testing involves two respondents reporting on the same horse. Questionnaires containing objective questions, reporting on frequency of behaviours observed, should reveal significant alignment between observers. Inter-rater reliability testing will likely reveal, notably where there is a lack of correlation, those questionnaire items that have required respondents to interpret behaviour. It is recommended that future large-scale data collection endeavours incorporate each of these validation and reliability testing protocols to assure the quality of findings and enhance the scientific knowledgebase.

Dissemination of results may be greatly assisted by two elements that should be incorporated into any future questionnaires. The first is making questionnaires available to a global population.

This aim can be greatly advanced by translating the survey into several different languages and thus avoid an Anglo-centrally focused database. Translations, to overcome language barriers, need to be conducted by content experts to ensure precision and accuracy [55]. Social media and extensive networking tools that are currently available make distributing online questionnaires easier than ever before. While one survey tool will never answer every question researchers hope to ask, a standardised and eventually validated method can provide scientists with baseline behavioural data from which to contextualise their investigation.

Administering a standardised behavioural instrument across different experimental settings will allow results to be generalisable, contribute to our understanding of what constitutes normal behaviour in horses and reveal relationships among variables not previously investigated. Baseline behavioural indices allow researchers to understand the influence of the age and breed of horses on behaviour. Further, once that a baseline has been established, one can begin to discern the effect of management and training practices. Perhaps Hartmann's et al. [43,45] studies on management practices, while geographically limited to Nordic countries with typically cool climates, may have benefitted from such baseline behavioural indices.

The dissemination of results from citizen science activities, such as pooling of horse-owner observations, is critical when feeding back to participants [56]. Dissemination is greatly assisted by open access publication of study results. It is interesting to note that of the 14 papers reviewed here, only three, [15,35,44], were published in an open access journal. This is concerning because horse welfare relies on the dissemination of clear and accurate research results, particularly to horse owners and caregivers who are responsible for the largest group of domestic horses, the pleasure riding community [57]. It is interesting to consider that a possible source of publication bias arises due to the higher costs of open access publication restricting this type of publication to large institutions or well-funded research groups. We recognise that many authors do not have the resources to fund open access.

5. Conclusions

Collecting data on horse training, management and behaviour using owner-reported observations is a simple, efficient, and inexpensive procedure. The use of online surveys vastly increases the potential reach and participation rate of studies, greatly increasing the ability to obtain robust datasets. Current large-scale online surveys have greatly strengthened our understanding of the use of equipment, how management practices are adapted in various geographical regions, and how training and management interact with behaviour. The data obtained from owners could be optimised by ensuring that behavioural questions do not need to be interpreted by owners, thus reducing confirmation bias. The prospects for integrating the collection of behavioural data into a global citizen science project are strong. A standardised and repeatedly accessible questionnaire could potentially change the way numerous stakeholder groups benchmark and understand horse behaviour and provide a baseline description of what constitutes normal behaviour across all geographic locations.

Author Contributions: Conceptualization: K.F., M.H., A.C., P.M.; Data Curation: K.F., A.C.; Investigation: K.F., M.H., A.C.; Methodology: A.C., K.F., P.M.; Project administration: P.M.; Resources: A.C.; Software: A.C.; Supervision: P.M.; Validation: K.F., A.C., M.H.; Visualization: K.F.; Writing—original draft preparation: K.F., A.C., M.H., P.M.; Writing—review and editing: K.F., M.H., A.C., P.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors would like to thank the anonymous reviewers for their assistance in the preparation of this manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Munsters, C.C.B.M.; Visser, K.E.K.; van den Broek, J.; Sloet van Oldruitenborgh-Oosterbaan, M.M. The influence of challenging objects and horse-rider matching on heart rate, heart rate variability and behavioural score in riding horses. *Vet. J.* **2012**, *192*, 75–80. [[CrossRef](#)] [[PubMed](#)]
2. McGreevy, P. *Equine Behavior: A Guide for Veterinarians and Equine Scientists*, 2nd ed.; W. B. Saunders: London, UK, 2012.
3. Parkin, T.; Brown, J.; Mcacdonald, E. Occupational risks of working with horses: A questionnaire survey of equine veterinary surgeons. *Equine Vet. Educ.* **2018**, *30*, 200–205. [[CrossRef](#)]
4. Lansade, L.; Bonneau, C.; Parias, C.; Biau, S. Horse's emotional state and rider safety during grooming practices, a field study. *Appl. Anim. Behav. Sci.* **2019**, *217*, 43–47. [[CrossRef](#)]
5. McGreevy, P.; Berger, J.; de Brauwere, N.; Doherty, O.; Harrison, A.; Fiedler, J.; Jones, C.; McDonnell, S.; McLean, A.; Nakonechny, L.; et al. Using the Five Domains Model to Assess the Adverse Impacts of Husbandry, Veterinary, and Equitation Interventions on Horse Welfare. *Animals* **2018**, *8*, 41. [[CrossRef](#)]
6. McLean, A.N.; McGreevy, P.D. Horse-training techniques that may defy the principles of learning theory and compromise welfare. *J. Vet. Behav. Clin. Appl. Res.* **2010**, *5*, 187–195. [[CrossRef](#)]
7. Fenner, K.; McLean, A.; McGreevy, P. Cutting to the chase: How round-pen, lunging and high-speed liberty work may compromise horse welfare. *J. Vet. Behav. Clin. Appl. Res.* **2019**, *29*, 88–94. [[CrossRef](#)]
8. Todd, Z. Barriers to the adoption of humane dog training methods. *J. Vet. Behav.* **2018**, *25*, 28–34. [[CrossRef](#)]
9. Perez, B.; Mehrkam, L.; Foltz, A.; Dorey, N. Effects of Enrichment Presentation and Other Factors on Behavioral Welfare of Pantropical Spotted Dolphin (*Stenella attenuata*). *J. Appl. Anim. Welf. Sci.* **2018**, *21*, 130–140. [[CrossRef](#)]
10. Fiedler, J. Sport Horse Welfare and Social Licence to Operate: Informing a Social Licence to Operate Communication Framework: Attitudes to Sport Horse Welfare. Ph.D. Thesis, Central Queensland University, Queensland, Australia, 2020.
11. Heleski, C.; Stowe, C.; Fiedler, J.; Peterson, M.; Brady, C.; Wickens, C.; MacLeod, J. Thoroughbred Racehorse Welfare through the Lens of 'Social License to Operate—With an Emphasis on a U.S. Perspective. *Sustainability* **2020**, *12*, 1706. [[CrossRef](#)]
12. Fiedler, J.; McGreevy, P. Reconciling Horse Welfare, Worker Safety, and Public Expectations: Horse Event Incident Management Systems in Australia. *Animals* **2016**, *6*, 16. [[CrossRef](#)]
13. Von Borstel, U.U.; Duncan, I.J.H.; Shoveller, A.K.; Merkies, K.; Keeling, L.J.; Millman, S.T. Impact of riding in a coercively obtained Rollkur posture on welfare and fear of performance horses. *Appl. Anim. Behav. Sci.* **2009**, *116*, 228–236. [[CrossRef](#)]
14. Wilson, B.; Jones, B.; McGreevy, P. Longitudinal Trends in the Frequency of Medium and Fast Race Winning Times in Australian Harness Racing: Relationships with Rules Moderating Whip use. *PLoS ONE* **2018**, *13*, e0184091. [[CrossRef](#)] [[PubMed](#)]
15. Weller, D.; Franklin, S.; Shea, G.; White, P.; Fenner, K.; Wilson, B.; Wilkins, C.; McGreevy, P. The Reported Use of Nosebands in Racing and Equestrian Pursuits. *Animals* **2020**, *10*, 776. [[CrossRef](#)] [[PubMed](#)]
16. Thomson, P.C.; Hayek, A.R.; Jones, B.; Evans, D.L.; McGreevy, P.D. Number, causes and destinations of horses leaving the Australian Thoroughbred and Standardbred racing industries. *Aust. Vet. J.* **2014**, *92*, 303–311. [[CrossRef](#)]
17. Fowler, V.; Kennedy, M.; Marlin, D. A comparison of the Monty Roberts technique with a conventional UK technique for initial training of riding horses. *Anthrozoos* **2015**, *25*, 301–321. [[CrossRef](#)]
18. Momozawa, Y.; Ono, T.; Sato, F.; Kikusui, T.; Takeuchi, Y.; Mori, Y.; Kusunose, R. Assessment of equine temperament by a questionnaire survey to caretakers and evaluation of its reliability by simultaneous behavior test. *Appl. Anim. Behav. Sci.* **2003**, *84*, 127–138. [[CrossRef](#)]
19. Baker, J.; Leyland, A. Histological survey of tumours of the horse, with particular reference to those of the skin. *Vet. Rec.* **1975**. [[CrossRef](#)]
20. Parker, M.; Goodwin, D.; Redhead, E.S. Survey of breeders' management of horses in Europe, North America and Australia: Comparison of factors associated with the development of abnormal behaviour. *Appl. Anim. Behav. Sci.* **2008**, *114*, 206–215. [[CrossRef](#)]
21. McGreevy, P.; French, C.; Nicol, C. The prevalence of abnormal behaviours in dressage, eventing and endurance horses in relation to stabling. *Vet. Rec.* **1995**, *137*, 36–37. [[CrossRef](#)]

22. Fenner, K.; Freire, R.; McLean, A.; McGreevy, P. Behavioral, demographic and management influences on equine responses to negative reinforcement. *J. Vet. Behav. Clin. Appl. Res.* **2019**, *29*, 11–17. [[CrossRef](#)]
23. Padalino, B.; Raidal, S.; Hall, E.; Knight, P.; Celi, P.; Jeffcott, L.; Muscatello, G. Risk factors in equine transport-related health problems: A survey of the Australian equine industry. *Equine Vet. J.* **2017**, *49*, 507–511. [[CrossRef](#)] [[PubMed](#)]
24. Albright, J.; Mohammed, H.; Heleski, C.; Wickens, C.; Houpt, K. Crib-biting in US horses: Breed predispositions and owner perceptions of aetiology. *Equine Vet. J.* **2009**, *41*, 455–458. [[CrossRef](#)] [[PubMed](#)]
25. Henriksson, J.; Sauveroché, M.; Roth, L. Effects of size and personality on social learning and human-directed behaviour in horses (*Equus caballus*). *Anim. Cogn.* **2019**, *22*, 1001–1011. [[CrossRef](#)] [[PubMed](#)]
26. Widi, T.; Rakasiwi, G.; Nugroho, T.; Widyas, N. Personality assessment of different horse breeds trained for military purposes. In Proceedings of the IOP Conference Series: Earth and Environmental Science, Bali, Indonesia, 20–21 October 2018; pp. 1–7.
27. Dashper, K.; Fenner, K.; Hyde, M.; Probyn-Rapsey, M.; Casper, G.; Henshall, C.; McGreevy, P. The anthropomorphic application of human-based gender stereotypes to animals. *Anthrozoos* **2018**, *31*, 673–684. [[CrossRef](#)]
28. Visser, E.; Van Wijk-Jansen, E. Diversity in horse enthusiasts with respect to horse welfare: An explorative study. *J. Vet. Behav. Clin. Appl. Res.* **2012**, *7*, 295–304. [[CrossRef](#)]
29. Luna, D.; Vasquez, R.; Rojas, M.; Tadich, T. Welfare Status of Working Horses and Owners' Perceptions of Their Animals. *Animals* **2017**, *7*, 56. [[CrossRef](#)]
30. McGreevy, P.; Oddie, C.; Burton, F.; McLean, A. The horse–human dyad: Can we align horse training and handling activities with the equid social ethogram? *Vet. J.* **2009**, *181*, 12–18. [[CrossRef](#)]
31. Bethlehem, J. Selection Bias in Web Surveys. *Int. Stat. Rev.* **2010**, *78*, 161–188. [[CrossRef](#)]
32. Fergusson, D.; Boden, J. Exploring issues arising from survey non-response. *Addiction* **2015**, *110*, 1513–1514. [[CrossRef](#)]
33. Smith, V.; Devane, D.; Begley, C.; Clarke, M. Methodology in Conducting a Systematic Review of Systematic Reviews of Healthcare Interventions. *BMC Med. Res. Methodol.* **2011**, *11*, 15. [[CrossRef](#)]
34. Hill, E.; McGreevy, P.; Caspar, G.; White, P.; McLean, A. Apparatus use in popular equestrian disciplines in Australia. *J. Vet. Behav. Clin. Appl. Res.* **2015**, *10*, 147–152. [[CrossRef](#)]
35. Aune, A.; Fenner, K.; Wilson, B.; Cameron, E.; McLean, A.; McGreevy, P. Reported Behavioural Differences between Geldings and Mares Challenge Sex-Driven Stereotypes in Ridden Equine Behaviour. *Animals* **2020**, *10*, 414. [[CrossRef](#)] [[PubMed](#)]
36. Althubaiti, A. Information bias in health research: Definition, pitfalls, and adjustment methods. *J. Multidiscip. Healthc.* **2016**, *9*, 211. [[CrossRef](#)] [[PubMed](#)]
37. Lloyd, A.S.; Martin, J.E.; Bornett-Gauci, H.L.I.; Wilkinson, R.G. Horse personality: Variation between breeds. *Appl. Anim. Behav. Sci.* **2008**, *112*, 369–383. [[CrossRef](#)]
38. Hockenhull, J.; Creighton, E. Management practices associated with owner-reported stable-related and handling behaviour problems in UK leisure horses. *Appl. Anim. Behav. Sci.* **2014**, *155*, 49. [[CrossRef](#)]
39. Hockenhull, J.; Creighton, E. The day-to-day management of UK leisure horses and the prevalence of owner-reported stable-related and handling behaviour problems. *Anim. Welf.* **2015**, *24*, 29–36. [[CrossRef](#)]
40. Hockenhull, J.; Creighton, E. The use of equipment and training practices and the prevalence of owner-reported ridden behaviour problems in UK leisure horses. *Equine Vet. J.* **2013**, *45*, 15–19. [[CrossRef](#)]
41. Hockenhull, J.; Creighton, E. Pre-feeding behaviour in UK leisure horses and associated feeding routine risk factors. *Anim. Welf.* **2014**, *23*, 297–308. [[CrossRef](#)]
42. Hockenhull, J.; Creighton, E. Unwanted oral investigative behaviour in horses: A note on the relationship between mugging behaviour, hand-feeding titbits and clicker training. *Appl. Anim. Behav. Sci.* **2010**, *127*, 104–107. [[CrossRef](#)]
43. Hartmann, E.; Boe, K.; Christensen, J.; Hyyppä, S.; Jansson, G.; Ladewig, J.; Mejdell, C.; Norling, Y.; Rundgren, M.; Sarkijarvi, S.; et al. A Nordic Survey of Management Practices and Owners' Attitudes towards Keeping Horses in groups. *J. Anim. Sci.* **2015**, *93*, 4564–4574. [[CrossRef](#)]
44. Gronqvist, G.; Rogers, C.; Gee, E. The Management of Horses during Fireworks in New Zealand. *Animals* **2016**, *6*, 20. [[CrossRef](#)]

45. Hartmann, E.; Boe, K.; Jorgensen, G.; Mejdell, C.; Dahlborn, K. Management of Horses with Focus on Blanketing and Clipping Practices Reported by Members of the Swedish and Norwegian Equestrian community. *J. Anim. Sci.* **2017**, *95*, 1104–1117. [[CrossRef](#)] [[PubMed](#)]
46. Hockenfull, J.; Creighton, E. The strengths of statistical techniques in identifying patterns underlying apparently random behavioral problems in horses. *J. Vet. Behav. Clin. Appl. Res.* **2012**, *7*, 305–310. [[CrossRef](#)]
47. Ross, S.; Murray, J.; Roberts, V. Prevalence of headshaking within the equine population in the UK. *Equine Vet. J.* **2018**, *50*, 73–78. [[CrossRef](#)] [[PubMed](#)]
48. Hartmann, E.; Christensen, J.; McGreevy, P. Dominance and Leadership: Useful Concepts in Human–Horse Interactions? *J. Equine Vet. Sci.* **2017**, *52*, 1–9. [[CrossRef](#)]
49. Lloyd, A.S.; Martin, J.E.; Bornett-Gauci, H.L.I.; Wilkinson, R.G. Evaluation of a novel method of horse personality assessment: Rater-agreement and links to behaviour. *Appl. Anim. Behav. Sci.* **2007**, *105*, 205–222. [[CrossRef](#)]
50. McGreevy, P.D. The advent of equitation science. *Vet. J.* **2007**, *174*, 492–500. [[CrossRef](#)]
51. Godlonton, S.; Hernandez, M.; Murphy, M. Anchoring Bias in Recall Data: Evidence from Central America. *Am. J. Agric. Econ.* **2018**, *100*. [[CrossRef](#)]
52. Seitzinger, P.; Tataryn, J.; Osgood, N.; Waldner, C. Foodborne Outbreak Investigation: Effect of Recall Inaccuracies on Food Histories. *J. Food Prot.* **2019**, *82*. [[CrossRef](#)]
53. Liu, J.; Qu, B.; Hu, B.; Jing, N.; Wang, D. The quality of life of men who have sex with men in china: Reliability and validity testing of the SF-36 questionnaire. *PLoS ONE* **2013**, *8*, e83362. [[CrossRef](#)]
54. Duffy, D.; Diniz de Moura, R.; Serpell, J. Development and evaluation of the Fe-BARQ: A new survey instrument for measuring behavior in domestic cats (*Felis s. catus*). *Behav. Process.* **2017**, *141*, 329–341. [[CrossRef](#)] [[PubMed](#)]
55. Behr, D. Translating Answers to Open-ended Survey Questions in Cross-cultural Research: A Case Study on the Interplay between Translation, Coding, and Analysis. *Field Methods* **2015**, *27*, 284–299. [[CrossRef](#)]
56. Engaging Volunteers: Guide to Engaging Volunteers in Citizen Science Projects. Available online: <https://www.tcv.org.uk/wp-content/uploads/2014/11/EngagingVolunteersCitizenScience.pdf> (accessed on 11 June 2020).
57. Fenner, K.; Dashper, K.; Wilkins, C.; Serpell, J.; McLean, A.; Wilson, B.; McGreevy, P. Building Bridges between Theory and Practice: How Citizen Science Can Bring Equine Researchers and Practitioners Together. *Animals* **2020**, *10*, 1644. [[CrossRef](#)] [[PubMed](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).