The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

Huntley, Selene J. and Mahlberg, Michaela and Wiegand, Viola and van Gennip, Yves and Yang, Hui and Dean, Rachel S. and Brennan, Marnie L. (2018) Analysing the opinions of UK veterinarians on practice-based research using corpus linguistic and mathematical methods. Preventive Veterinary Medicine, 150 . pp. 60-69. ISSN 1873-1716

Access from the University of Nottingham repository:
http://eprints.nottingham.ac.uk/48917/1/1-s2.0-S0167587717300223-main.pdf

Copyright and reuse:

The Nottingham ePrints service makes this work by researchers of the University of Nottingham available open access under the following conditions.

This article is made available under the Creative Commons Attribution Non-commercial No Derivatives licence and may be reused according to the conditions of the licence. For more details see: http://creativecommons.org/licenses/by-nc-nd/2.5/

A note on versions:

The version presented here may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the repository url above for details on accessing the published version and note that access may require a subscription.

For more information, please contact eprints@nottingham.ac.uk

Preventive Veterinary Medicine 150 (2018) 60–69

Contents lists available at ScienceDirect

Preventive Veterinary Medicine

journal homepage: www.elsevier.com/locate/prevetmed

Analysing the opinions of UK veterinarians on practice-based research using corpus linguistic and mathematical methods

Selene J. Huntley[a], Michaela Mahlberg[b,1], Viola Wiegand[b,1], Yves van Gennip[c], Hui Yang[d], Rachel S. Dean[a], Marnie L. Brennan[a,*]

[a] Centre for Evidence-based Veterinary Medicine (CEVM), School of Veterinary Medicine and Science, University of Nottingham, Sutton Bonington Campus, Leicestershire, LE12 5RD, UK
[b] School of English,University Park, Nottingham, NG7 2RD, UK
[c] School of Mathematical Sciences, University of Nottingham,University Park, Nottingham, NG7 2RD, UK
[d] School of Computer Sciences, University of Nottingham, Jubilee Campus, Nottingham, NG8 1BB, UK

ARTICLE INFO

Keywords:
Evidence-based veterinary medicine
Practice-based research
Veterinarian
Veterinary surgeon
Topic modelling
Corpus linguistic analysis
Corpus linguistics
Survey
Questionnaire

ABSTRACT

The use of corpus linguistic techniques and other related mathematical analyses have rarely, if ever, been applied to qualitative data collected from the veterinary field. The aim of this study was to explore the use of a combination of corpus linguistic analyses and mathematical methods to investigate a free-text questionnaire dataset collected from 3796 UK veterinarians on evidence-based veterinary medicine, specifically, attitudes towards practice-based research (PBR) and improving the veterinary knowledge base.

The corpus methods of key word, concordance and collocate analyses were used to identify patterns of meanings within the free text responses. Key words were determined by comparing the questionnaire data with a wordlist from the British National Corpus (representing general English text) using cross-tabs and log-likelihood comparisons to identify words that occur significantly more frequently in the questionnaire data. Concordance and collocation analyses were used to account for the contextual patterns in which such key words occurred, involving qualitative analysis and Mutual Information Analysis (MI3). Additionally, a mathematical topic modelling approach was used as a comparative analysis; words within the free text responses were grouped into topics based on their weight or importance within each response to find starting points for analysis of textual patterns.

Results generated from using both qualitative and quantitative techniques identified that the perceived advantages of taking part in PBR centred on the themes of improving knowledge of both individuals and of the veterinary profession as a whole (illustrated by patterns around the words *learning, improving, contributing*). Time constraints (*lack of time, time issues, time commitments*) were the main concern of respondents in relation to taking part in PBR. Opinions of what vets could do to improve the veterinary knowledge base focussed on the collecting and sharing of information (*record, report*), particularly recording and discussing clinical cases (*interesting cases*), and undertaking relevant continuing professional development activities. The approach employed here demonstrated how corpus linguistics and mathematical methods can help to both identify and contextualise relevant linguistic patterns in the questionnaire responses. The results of the study inform those seeking to coordinate PBR initiatives about the motivators of veterinarians to participate in such initiatives and what concerns need to be addressed. The approach used in this study demonstrates a novel way of analysing textual data in veterinary research.

1. Introduction

The use of veterinary practice data can inform the veterinary profession on the prevalence and risk factors for animal diseases and aids disease surveillance (Radford et al., 2011; O'Neill et al., 2013; Robinson et al., 2014). Historically, the majority of veterinary clinical research has originated from referral centres or universities (O'Neill, 2015). Caseloads from specialist veterinary teaching hospitals may be biased

* Corresponding author.
E-mail address: marnie.brennan@nottingham.ac.uk (M.L. Brennan).
[1] Present address: Department of English Language and Applied Linguistics, School of English, Drama, American and Canadian Studies, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom.

https://doi.org/10.1016/j.prevetmed.2017.11.020
Received 6 January 2017; Received in revised form 20 November 2017; Accepted 22 November 2017
0167-5877/ © 2017 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/BY-NC-ND/4.0/).

towards conditions that are less common than those seen in the general population (Fleming et al., 2011). Recent advances in the reporting of data from first opinion practice with initiatives led by the Royal Veterinary College (VetCompass), University of Liverpool (SAVSNET), and Centre for Evidence-based Veterinary Medicine, University of Nottingham (O'Neill, 2015), seek to provide relevant research that is representative of the veterinary caseload in the UK. This is important for the integration of the principles of evidence-based veterinary medicine into veterinary practice as more relevant research is available. To build on and continue the long term success of these initiatives whilst ensuring that they are beneficial to all those involved, it is important to engage with veterinarians working in first opinion practice. In particular, to understand the benefits perceived by veterinarians by getting involved in this type of research and the factors motivating and impeding participation. Pre-emptive steps may then be taken to avoid scenarios of participatory bias, low rates of engagement, reduced compliance or study drop-out.

Gathering individuals' thoughts on participating in initiatives can be achieved by collecting data via questionnaires (Bowling, 2009). In particular, asking open questions is an effective way of extracting opinions. Open questioning has the advantage of allowing respondents to express their thoughts, ideas and emotions in their own words (White et al., 2005). This results in a more diverse array of responses than closed questions, where respondents are provided with a list of pre-coded response options (Reja et al., 2003). However, the analysis of open responses is more challenging because of the lack of pre-coding and the need to employ more qualitative methodologies. In veterinary studies, there is an expanding range of qualitative methods to study opinions and evaluations such as narrative research (Page-Jones and Abbey, 2015), thematic analysis (Mateus et al., 2014) and video tape analysis (Coe et al., 2009; Roshier and McBride, 2012). These methods usually rely on a researcher, or pair of researchers assessing participants' responses individually, and then drawing on the results of this analysis to identify key themes arising from the data (Braun and Clarke, 2006). Although of great value in terms of in-depth exploration of specific opinions, this can be an extensive process, and is not generally feasible for analysing data from a large number of respondents because of the requirement for extensive coding that typically has to be done manually (Reja et al., 2003; Bryman, 2012). It has been previously suggested that a combination of both quantitative and qualitative methods should be used to negate some of these issues (Upjohn et al., 2013).

Corpus linguistic methods can be used to achieve a combination of quantitative and qualitative analysis to support the description of meanings in the analysis of textual data, e.g. in studies of subject-specific terminology and more general phraseology. Specialist corpus analysis software can be used to generate frequency data, or more tailor-made data processing techniques can be employed to investigate various corpora (large collections of electronic texts) (Mahlberg, 2014). In contrast to entirely computational approaches, corpus methods allow the data to be displayed in such a way that an analyst can link quantitative information with qualitative insights (Scott and Tribble, 2006). Corpus linguistics has become a mainstream approach in linguistics and is also being increasingly applied in other disciplines (Skelton et al., 2002). In particular, the way in which corpus methods are used to analyse specific discourses provides useful bridges between linguistics and other disciplines (Mahlberg, 2014). These methods have been used previously to support research in the medical sciences (Skelton et al., 2002; Seale et al., 2006) and health communication (Harvey, 2014). Corpus methods have also been used to study questionnaire data in other fields (Millar and Hunston, 2015). In veterinary medicine, linguistic methods have been used very little; to the authors' knowledge, a single conference abstract exists reporting the use of corpus linguistics methods (Ding and Riloff, 2015).

An example of a corpus linguistic method used to get an overview of datasets of texts is key word analysis. Key words are words that occur statistically significantly more frequently in a corpus under analysis compared to a reference corpus. They are measured, for instance, with the help of a log-likelihood comparison (Scott and Tribble, 2006). Key words are typically proper nouns, content words (as opposed to function words such as *and, of, the*) and words that might be stylistically relevant. A set of key words in themselves is not very meaningful. But key words are useful starting points for more qualitative analyses. Such analyses can move on to grouping key words together according to broader theoretical criteria (McEnery, 2009), by inductively finding groups (Fischer-Starcke, 2009) or a combination of both (Mahlberg and McIntyre, 2011). To provide a detailed account of the meanings and usage patterns of individual key words, a common subsequent analytical step is to study the textual contexts in which the words occur (Bondi and Scott, 2010). This type of analysis is usefully done with the help of concordances. A concordance is a display format that enables the researcher to view how words are used in a specific set of texts. For instance, a concordance for the word *time* in our data shows how the respondents of the questionnaire talk about *time constraints*. Such patterns of words occurring together, which are also called 'collocations' (Sinclair, 1991) provide insights into meanings of individual words, but also insights into specific discourse (as represented by specific data sets). While *time constraints* is a pattern in our questionnaire data, in a large corpus of general English, for instance, other frequent patterns of *time* include *time and time again* or *every time* (Mahlberg and Stockwell, 2016).

Another method that can be used is topic modelling (Blei et al., 2003; Kuang et al., 2015). In contrast to the key word comparison with an external reference corpus, topic modelling achieves groupings of words based on corpus-internal criteria. Each topic consists of a distribution over all the words in the corpus, in which each word is assigned a weight indicating its importance in that particular topic (Blei et al., 2003). Typically only the top words contribute significantly to a topic. Similarly to lists of key words, words grouped together in topics need further interpretation, requiring additional qualitative approaches (Kuang et al., 2015).

The aim of the current study was to use corpus linguistic methods and mathematical methods to analyse textual data collected via open questions in a survey of UK veterinarians. Our objective was to use these methods to identify the barriers and motivators of veterinarians participating in practice-based research (PBR). Additionally, exploration of responses from veterinarians on what they felt they could do to improve the veterinary knowledge base was also carried out. The aim of this paper was to present a novel methodological approach to demonstrate how corpus methods could be used to support traditional qualitative research approaches in veterinary research in a useful way.

The main hypothesis we work with here is that if we have a sufficiently large set of textual data that has been collected to gain insights into veterinarians' views on a specific topic, corpus methods can help us identify linguistic patterns that reflect shared and repeatedly expressed concerns. The questionnaire data was originally collected to use free-text responses to gain insights into practice-based research. The specific argument we put forward in this paper is that such free-text can be interpreted by identifying key meanings expressed in the form of linguistic patterns. It is important to note that we are not aiming to present an entirely automatic text mining approach. Our contribution lies in illustrating how qualitative and quantitative techniques can be usefully combined to significantly develop the study of practice-based research more widely.

2. Materials and methods

2.1. Veterinary questionnaire

The data analysed in this study was generated from a survey of the UK veterinary profession, with a questionnaire distributed to all members on the Royal College of Veterinary Surgeons (RCVS) list

(sampling frame). The target population was all veterinarians within the UK (both those practicing within veterinary clinics, and those who worked in other sectors of the veterinary profession). In order to work in the UK as a veterinary surgeon, individuals are required to be registered on this RCVS list, therefore being on this list enabled eligibility for the study. Respondents were sent a questionnaire through the post, and could either reply by post or complete the survey online to optimise response rates, along with the use of incentives (pens, chocolates, prize draws), using wording on the envelope to make it clear the survey was for research purposes, and by using coloured paper (Edwards et al., 2002). There were four main sections of the questionnaire which focused on gathering demographic information, the common conditions encountered by veterinarians in practice, the awareness of the term 'evidence-based veterinary medicine' and the sources of information or evidence accessed and used by veterinarians. Prior to sending the questionnaire, pre-testing and piloting phases were carried out to further improve reliability and validity of the results (Dillman et al., 2009); two reminders were sent to improve response rates, one six weeks after the initial mailing and the second four weeks after the first reminder. A post-hoc analysis was performed to assess the sample size achieved using Raosoft (www.raosoft.com/samplesize.html). Paper-based questionnaires were designed using TeleForm V.10.5.2 (Verity Inc. 2010). Online software provided by Cvent (2011 Cvent Inc.) was used to construct an almost identical online version of the paper questionnaire. For participants that completed the questionnaire on paper, a content capture system (TeleForm V.10.5.2; Verity Inc. 2010) was used to scan the completed questionnaires into a Microsoft Access V.14.06 (2010 Microsoft Corporation) database. Responses from open answer questions were transcribed verbatim by a data technician. Punctuation and spelling errors were not corrected and were included in the database. Ten percent of questionnaires were checked after data entry to highlight the error rate (minimal errors were identified during post-scanning verification). For responses collected via an online questionnaire using Cvent (2011 Cvent Inc.) software, data were stored in a Microsoft Excel V.14.0.6 (2010 Microsoft Corporation) spreadsheet containing the same fields as the Microsoft Access database and were transferred into the Microsoft Access database containing the paper questionnaire responses once collection had ceased. Reporting of results from other sections of the questionnaire can be seen in Nielsen et al. (2014), Nielsen et al. (2015) and Huntley et al. (2017).

The focus for the current study was a section of the survey which asked respondents about practice-based research (PBR) and improvement of the veterinary knowledge base, with three open questions posed. These were: **Question 1:** 'What do you think would be the main advantages of participating in practice-based research?'; **Question 2:** 'What would be your main concerns about conducting research in your practice?'; **Question 3:** 'What do you think you could do as a clinician to improve the veterinary knowledge base?'. For use with the corpus software, three separate text files were created in Notepad 6.3 (2013 Microsoft Corporation) and this data was then transferred to three WordSmith Tools (Scott, 2012a) files ready for corpus analysis. The data had been previously explored using traditional qualitative methods in a pilot study (Caulfield, 2013), but was performed by different authors in the current study (with the exception of MB).

2.2. Corpus analysis

2.2.1. Keywords comparison

All analysis was undertaken in the corpus software WordSmith Tools (Scott, 2012a). The first step was a key word analysis. The corpus of answers for each question was compared with the British National Corpus (BNC) wordlist prepared for use in WordSmith Tools (Scott, 2012b). The BNC is a 100 million word corpus of general English and is frequently used in corpus linguistics studies as a reference corpus. This comparison gave us words that occurred significantly more frequently within the three data sets than in the BNC reference corpus. These 'key words' characterised each of the questions' responses in terms of their 'aboutness'. For each word appearing in the three corpora relating to our three questions, a keyness value was computed by cross-tabulating the frequency of the words appearing here with the frequency of the word in the BNC reference corpus. The log-likelihood comparison method was used as per the standard options in WordSmith Tools (Scott, 2016a). Key words are generally used as a starting point for further analyses and their main purpose is to select a useful set of preliminary words. Depending on the size of the corpus under analysis and the purpose of the study, a full key word list might be used, with a *p*-value used to determine a cut-off level below which key words would be considered less important. Often only the top key words are investigated further. For each of the three questions in the current study, the top ten key words with substantial keyness values were considered for further detailed analysis.

The next phase of the analysis involved the exploration of the patterns relating to the main key words by interpreting examples of individual words in context (Bondi and Scott, 2010). Concordance and collocation analysis were used to study the contexts surrounding the key words. A concordance displays all lines of text which feature the key word at the centre and a specified amount of text on either side (Sinclair, 1991). This display format allows the researcher to identify recurring patterns of words within the vertical display of the concordance lines, to highlight the context within which the key word is being used. The word in the centre of the concordance is typically referred to as the 'node' and words in its context that occur repeatedly are its 'collocates'. To describe collocations in more detail, statistical tests can be used to test the words appearing within a specified span around the node (e.g. a span of L5, R5 means investigating five words to the left and right of the node) (Evert, 2008). In this study the statistical test employed was the Mutual Information analysis (MI3) in WordSmith Tools (Scott, 2012b). An MI3 score is calculated by dividing the observed frequency of the co-occurring word by the expected frequency, and 'cubing' the result (McEnery et al., 2006). Using this statistic highlights words that collocate significantly frequently with the node. We can then further group key words together to describe sets of meanings. This can be done through deductive or inductive methods, or a combination of both to identify commonalities across all responses relating to key words (Fischer-Starcke, 2009; McEnery, 2009; Bondi and Scott, 2010).

2.2.2. Topic modelling

Topic modelling (Blei et al., 2003; Arora et al., 2012) uses the occurrence count for each word in a corpus per individual answer (a 'bag-of-words' description) to output a description of the corpus in terms of topics (themes). Each topic is described by a distribution of words. In contrast to the key word methodology, topics are generated without comparison to a reference corpus. Prior to processing, a number of characters (e.g. accents, symbols, punctuation etc.) and 'stop words' such as *I* and *as* were removed as these words do not generally differentiate between topics (Ma et al., 2013). To pre-process the data, the characters '@' and '#' were replaced with white spaces, all letters were converted to lower case and most characters that were not standard letters were removed from the beginnings and ends of words. The top fifty most common words in the corpus were then removed before running the algorithm to further target the analysis (Lai et al., 2014). For a single corpus (i.e. data from one survey question), it was assumed that m words were left after the pre-processing step and the number of answers was n. An m by n matrix of word counts was then created, with each of the n columns corresponding to a single answer and containing an occurrence count for each of the m words in the corpus. For the topic modelling, nonnegative matrix factorization (Kim and Park, 2008a, 2008b) was implemented in MATLAB (Kim and Park, 2010). Given the nonnegative m by n word count matrix A, and a pre-selected number of

Table 1
Number of respondents, and words in responses to the three open questions about practice-based research and the veterinary knowledge base in the questionnaire sent to UK veterinarians in 2010–2011.

Open questions	Number of respondents	**Number of key words**	**Total number of words**
Question 1: What do you think would be the main advantages of participating in practice-based research?	3279	137	46,753
Question 2: What would be your main concerns about conducting research in your practice?	3455	81	32,377
Question 3: What do you think you could do as a clinician to improve the veterinary knowledge base?	3237	125	40,645
Total			**119,775**

Table 2
Top ten key words identified within responses to the question 'What do you think would be the advantages of participating in practice-based research?' and the corresponding frequency and keyness values.

Keyword rank	Key word	Keyness value	Frequency
1	knowledge	6187.11	817
2	practice	5248.49	743
3	research	3734.33	635
4	clinical	3161.43	353
5	veterinary	2697.09	251
6	cases	2484.92	425
7	treatments	2133.72	208
8	more	1746.38	749
9	improve	1655.92	242
10	treatment	1536.13	268

topics k, nonnegative matrix factorization tries to find a m by k matrix W and a k by n matrix H such that the product WH approximates A and both W and H have nonnegative entries. The algorithm accomplishes this by minimising a weighted sum of the squared Frobenius norms of A-WH, W and H. Since W has nonnegative entries, it can be interpreted as a word-topic matrix, for example each topic is represented by a column of W whose entries represent the relevance of each word for that topic. Similarly, H is interpreted as a topic-document matrix. For each of the three questions, analysis with a preselected number of five ($k = 5$), and ten ($k = 10$), topics was carried out and of both of these, the number that represented the data in the most appropriate way based on the themes that arose from the outputs was selected. Per topic, the ten most important words were listed, including their relative importance in the topic (given in brackets, and computed by normalising the corresponding entry in W by sum of the entries of the corresponding column of W). These words were compared with the key words retrieved in the corpus analysis to add further contextualisation.

This project received ethical approval from the ethics committee at the School of Veterinary Medicine and Science at The University of Nottingham.

3. Results

Of the 14,532 questionnaires that were distributed, 5407 were returned. Five hundred and sixty five questionnaires received back were marked return to sender, had been sent to retired or deceased veterinarians, or were returned blank, resulting in 4842 responses (33%; CI 32% to 35%) eligible to be used for analysis. Posthoc analysis revealed a sample size of 3227 was required to demonstrate an adequate response (confidence level 99%, 2% margin of error, 50% response distribution, population size of 14,532) to the questionnaire.

Of these 4842 respondents, 3796 respondents answered at least one of the three questions. The number of respondents and total size of the corpus (the total number of words written by respondents) from the three questions is provided in Table 1. Unlike the responses generated from questions 1 and 3, responses generated from question 2 amounted to less than 40,000 words. The number of key words found was generally proportional to the corpora sizes found for each question.

3.1. Question 1: What do you think would be the main advantages of participating in practice-based research?

The key word analysis highlighted the terms *knowledge* and *practice* as occurring frequently in our corpus (Table 2). For the word *knowledge*, concordance and collocation analysis resulted in themes of maintaining (keeping up to date), improving or contributing to knowledge, either personally or for the greater veterinary community. This was illustrated in patterns of *adding to* or *advancing knowledge* during concordance analysis (Table 3).

The top collocates of *knowledge* included words describing the acquirement of further knowledge, or the improvement of knowledge, such as *improving, increasing, contributing,* and *improve* (Table 4).

When repeated phrases in a concordance were investigated, the most common cluster (i.e. a verbatim repetition of a sequence of words) around the term *knowledge* was *veterinary knowledge base*. The context showed that the notion of advancing knowledge was applied to the overall veterinary knowledge base (i.e. that of the profession) and was expressed with verb forms like *improving, expanding* and *contributing* (Appendix A).

Table 3
Fifteen illustrative examples of 817 occurrences of *knowledge* (sorted by the L1 concordance position) within responses to the question: 'What do you think would be the main advantages of participating in practice-based research?'.

1	niversities etc *adding* to a	**knowledge**	base and improving animal h
2	ice and possibly gain *added*	**knowledge**	Gain further insight into c
3	observed Gaining *additional*	**knowledge**	with a view to improving se
4	ractice conditions. *Advance*	**knowledge**	and learning in the process
5	otherwise not last *advance*	**knowledge**	benefit long term from info
6	finding answers. *Advanceing*	**knowledge**	Keeping up to date with bes
7	cally stimulating *Advancing*	**knowledge**	and understanding of clinic
8	t's work schedule *Advancing*	**knowledge**.	Applying a practical aspec
9	REST IN CASES AND *ADVANCING*	**KNOWLEDGE**"	Contributing to the knowle
10	e of satisfaction *advancing*	**knowledge**.	Encourages awareness of su
11	was successful in *advancing*	**knowledge**	etc" change of targeting re
12	job satisfaction. *Advancing*	**knowledge**	nire accurate representatio
13	client compliance *advancing*	**knowledge**	- personal and for practice
14	ssion as a whole. *advancing*	**knowledge**.	Subsidized bloodtests etc
15	er involvement in *advancing*	**knowledge**	the results may be useful a

Table 4
Top 15 collocates of *knowledge* (with a span of L5-R5 position, ranked by MI3 score) within responses to the question 'What do you think would be the advantages of participating in practice-based research?'.

Collocate rank	Collocate	MI3 score	Frequency
1	to	20.74	401
2	of	20.60	366
3	base	20.42	174
4	and	19.97	284
5	improving	18.90	113
6	the	18.85	225
7	veterinary	18.83	127
8	in	18.37	187
9	contributing	17.86	80
10	increasing	17.29	60
11	improve	17.25	87
12	own	17.24	64
13	my	17.14	72
14	increased	17.11	71
15	further	16.97	56

Topics arising from the topic modelling approach implied similar themes of improvement of knowledge (e.g. *learning*) for either personal gain or for the benefit of the profession (Table 5). The most frequent collocate to the left of the word *learning* was *more* and by exploring this further in concordance lines, it was evident that *more* referred to *cases*, a *subject* or *disease* (Table 5). Other themes that arose were of gaining experience (topic 2), improved job satisfaction (topic 3), improvement in animal welfare (topic 5) and provision of relevant information (topic 6). The remaining topics are included in Appendix B.

3.2. Question 2: What would be your main concerns about conducting research in your practice?

The majority of respondents (2583/3455; 74.8%) to this question mentioned the word *time* at least once in their response. Amongst the top ten key words that were observed were *time*, *constraints*, *client*, *paperwork*, *compliance*, *consuming* and *lack* (Table 6). Time issues and constraints were clearly the largest concerns for respondents, reflected by their large keyness scores.

Time constraints and *time commitment* were often statements on their own and not part of a larger sentence, and many of the statements regarding time were extremely similar between respondents (Table 7). While the phrase was sometimes part of a longer sentence (as in line 4 in Table 7), it was often used as a single 'bullet point'. *Constraints* and, similarly, *lack* were among the top collocates of *time* (Table 8). *Paperwork* collocated as requiring additional time or as part of another point (Appendix C).

Similarly, words arising from topic modelling were around *time issues* (topic 1), *time commitment* (topic 2) and *ethical considerations* (topic 3; Table 9). Concordance lines from further contextualisation in WordSmith Tools were also strongly focused around the themes of time commitment and constraints, already feeling overworked and client compliance. Themes of data quality and validity, compliance of colleagues and ethical considerations also arose. The remaining topics are included in Appendix D.

Table 5
Top ten words with sample concordance lines for each of the 5 sample topics displayed (out of ten total topics) within responses to the question 'What do you think would be the advantages of participating in practice-based research?'. The numbers represent the weight or importance of a word in relation to the other words appearing in the text.

Topic	Top ten words per topic with sample concordance lines
	practitioners (0.0764), best (0.0460), **learning (0.0362),** protocols (0.0215), available (0.0213), outcomes (0.0086), realistic (0.0085), way (0.0078), relevance (0.0078), think (0.0078)
1	Concordance lines of *learning* & its most frequent R1 collocate *more* (n=15): 1 first opinion cases used **learning** *more* about a subject/ hav 2 nt response and prognosis **Learning** *more* about disease epidem 3 ith changing theories etc **learning** *more* about the cases invo
2	**experience (0.0656),** access (0.0323), st (0.0226), practices (0.0193), getting (0.0170), sharing (0.0169), field (0.0111), protocols (0.0108), areas (0.0096), gaining (0.0079) Concordance lines of *experience* & its most frequent L1 collocates: 1 are my personal *clinical* **experience** with all the others if a 2 equine medicine *Gaining* **experience** widening my norizion Abl 3 ledge getting *first hand* **experience** of new developments. Dev
3	**job (0.1165)**, outcomes (0.0453), interesting (0.0296), time (0.0140), subject (0.0094), confidence (0.0079), hopefully (0.0069), edge (0.0068), learn (0.0067), cutting (0.0067) Concordance lines of *job* & its most frequent R1 collocate *satisfaction* (n=58): 1 ore research data. increased **job** *satisfaction* for practition 2 client satisfaction improved **job** *satisfaction*" Better ground 3 clients and staff" Increase **job** *satisfaction*. Improve ones
5	**animal (0.0813)**, welfare (0.0378), client (0.0314), care (0.0312), contribute (0.0163), health (0.0142), increasing (0.0123), patient (0.0112), able (0.0109), diseases (0.0078) Concordance lines of *animal* & its most frequent R1 collocate *welfare* (n=20): 1 do and outcomes. improving **animal** *welfare* Encouraging me to 2 als and clients to improve **animal** *welfare*. To reduce the nee 3 f-satisfaction Benefits to **animal** *welfare* Makes you consider
6	**provide (0.1140)**, approach (0.0254), accurate (0.0185), effective (0.0153), diseases (0.0150), different (0.0139), useful (0.0135), studies (0.0124), certain (0.0110), used (0.0085) Concordance lines of *provide* & its most frequent R1 collocate *more* (n=8): 1 ituation and give scope to **provide** *more* accurate and forcefu 2 ary research being able to **provide** *more* data on prevalence o 3 ractice based research can **provide** *more* valid information re

Table 6
Top ten key words identified within responses to the question 'What would be your main concerns about conducting research in your practice?' and the corresponding frequency and keyness values.

Key word rank	Key word	Keyness value	Frequency
1	time	17051.74	2780
2	constraints	3947.97	372
3	client	3446.48	395
4	clients	2416.10	288
5	paperwork	2399.73	201
6	compliance	2305.40	222
7	consuming	2178.08	187
8	practice	1766.50	295
9	research	1693.49	321
10	lack	1482.43	227

3.3. Question 3: What do you think you could do as a clinician to improve the veterinary knowledge base?

The top key words within responses to question 3 included *clinical*, *cases* and *CPD* (CPD denotes Continuing Professional Development; Table 10). Collocates to the key word *cases* indicated that respondents viewed the recording and sharing of their own cases as an important activity in improving the knowledge base (*report, record, seen, write, discuss*; Table 11). In particular, *interesting* and *unusual* cases were highlighted in this regard. Sample occurrences of *interesting cases* in context (Table 12), demonstrated how the recording and publishing of such cases was described as desirable in instructions to self like *keep a file of abnormal/interesting cases* (line 1) or *publish more interesting cases* (line 14).

CPD was also a highly ranking key word, which frequently collocated with *more*. The context of this collocation pair showed respondents' intentions to *attend more CPD* (Appendix E).

Among the topics identified in topic modelling, high ranking words included the abbreviated word *info*, the word *discussion* and the word *samples* (Table 13). When analysed contextually in WordSmith Tools, the noun *info* tended to occur in phrases like *info on (…)* where the final slot was a noun like *cases*, *results* or *outcomes* (topic 1). The word *discussion* collocates with *groups*, *online* and *forums* (topic 2) and *samples* co-occurred with the word *submit* (topic 7). Themes arising from concordance lines included feeding back information on caseload and outcomes, submitting samples, increased attendance at CPD meetings, sharing knowledge via internet discussion groups and reading more. The remaining topics are included in Appendix F.

4. Discussion

The results presented here provide unique insight into the attitudes of clinical veterinarians towards practice-based research (PBR) and the perceived role of the clinician in advancing the veterinary knowledge base, using corpus linguistic and mathematical methodologies applied in a novel way. From the language and phraseology used, it appears veterinarians are aware of the benefits to themselves and the profession by being involved in research that could potentially further the relevance of the knowledge base to commonly encountered cases in first opinion veterinary practice.

The main benefits perceived by veterinarians in taking part in PBR appeared to be improvement of both personal knowledge and that of the profession. This is generally in agreement with a survey of UK veterinarians about CPD where intrinsic factors were the main motivators for most of the CPD undertaken. Examples cited were "keeping up to date" and "for the love of learning" (Dale et al., 2013). Veterinarians also had clear ideas about how they could contribute to the veterinary knowledge base. In general veterinarians consider their role in PBR to be the providers of data, either by the provision of case data, biological samples, or by sharing their own knowledge amongst veterinary colleagues. The increased opportunities for the use of electronic patient record data (O'Neill, 2015) has made the first of these easier to achieve than previously possible.

The main disadvantage of taking part in PBR stated was time constraints. This is perhaps not surprising as veterinarians are already pushed for time during their average working week; the number of hours worked has been reported as one of the greatest contributors to stress in the veterinary profession (Bartram et al., 2009) with almost half of consults exceeding an allotted 10 min consult time in one UK study (Robinson et al., 2014). As *time* is the most frequent noun in general English (Mahlberg, 2005), it is expected to appear commonly in corpora. However, the fact that it appears as a key word in the responses to question 2 compared to the BNC list shows that time constraints are of particular concern in this context. Client factors and paperwork were also frequently mentioned as a disadvantage in participation in PBR. Barriers such as client expectations and administrative and clerical tasks have been mentioned as some of the main stressors reported by veterinarians (Bartram et al., 2009). Client participation and owner compliance was an often mentioned concern in our survey. In human medicine, a study of participation in human epidemiological research reported that people were generally prepared to participate in epidemiological studies (Slegers et al., 2015). However, in veterinary research, barriers to patient involvement may be more complex than those in human medicine since it is the owner of an animal making the decision on behalf of a third party (the animal). Nevertheless, benefits to animals, recommendation for enrolment by veterinarians and trust in organisations were found to be important factors in owner participation in clinical trials of cats (Gruen et al., 2014). It may be that animal owner understanding of what PBR entails, or the communication from veterinarians about what PBR is, could be another reason for vets highlighting it as a concern; further work would need to be undertaken to explore this in more detail.

The barriers and motivators identified here can be utilised by researchers wanting to conduct research within a practice environment with veterinarians. It is apparent that veterinarians have a desire to get involved for a number of personal and professional reasons, so motivation exists. However, barriers such as the time commitments required and paperwork expectations need to be considered and preferably minimised by researchers looking to successfully work with veterinarians within the practice environment.

Table 7
Top ten illustrative examples of 341 occurrences of *time (*sorted by the R1 concordance position) within responses to the question 'What would be your main concerns about conducting research in your practice?'.

1	research time. Accuracy	**time constraints**.	Limited access to advic
2	collect data accurately	**time constraints**	lack of reliability eg w
3	keeping problems. also	**time constraints**	time consuming. unprofit
4	duct any research. also	**time constraints**	would be a problem Time
5	rd keeping and analysis	**time constraints**.	Owner compliance (if re
6	retrosopective analysis	**Time constraints**	time constraints time an
7	ews Confidentiality and	**time constraints**	Time sometimes my job is
8	ted number of cases and	**time constraints**	time. Never enough to do
9	arch. time Economic and	**time constraints**	validity of the results
10	ons. Cost. time and £££	**Time constraints**,	owner compliance time "

Table 8
Top 15 collocates of *time* (ranked by the MI3 statistic) within responses to the question 'What would be your main concerns about conducting research in your practice?'.

Collocate rank	Collocate	**MI3 score**	**Frequency**
1	to	23.52	1095
2	of	23.08	900
3	constraints	22.76	774
4	and	22.65	749
5	the	21.33	549
6	client	21.58	515
7	it	21.24	458
8	consuming	20.55	361
9	in	20.07	352
10	lack	20.16	316
11	involved	19.89	288
12	paperwork	19.87	286
13	not	19.29	276
14	compliance	19.62	262
15	do	19.48	250

In addition to the insights gained about PBR, this study has illustrated an innovative approach to veterinary research as our findings were generated via the novel application of corpus linguistic and mathematical methods. The key word methods helped to select for more detailed analysis and both concordance and collocation analysis contributed to investigating key words in their context (Scott, 2016b). The utilisation of a number of different techniques enabled the use of both qualitative and quantitative approaches in a concurrent fashion which increases the value of the findings. The 'marriage' of approaches in this way is novel in veterinary sciences where a combination of both quantitative and qualitative methodologies has infrequently being utilised in research. When utilised, separate qualitative (e.g. interviews, focus groups) and quantitative (e.g. questionnaires using closed questions) research phases tend to be employed to investigate factors identified in the initial stages of the work across a wider population (Scantlebury et al., 2014; Nöremark et al., 2016). Being able to use a combination of both concurrently provides new methodological avenues for researchers. While key words provide a list of words that are retrieved by comparison with a reference corpus, topic modelling highlights topics as a result of intra-corpus comparison. By combining both approaches the grouping and selecting of key words is supported by the information from the topics. As with any qualitative approach, there is an element of interpretation involved when assessing the corpus data. The employment of corpus linguistics and topic modelling to interpret the questionnaire data here is no exception to this. A traditionally employed qualitative approach uses techniques such as individual and group interviews, focus groups and ethnographic studies which employ a number of different analyses, including thematic analysis, which rely on individual researchers' interpretation of each

Table 10
Top ten key words identified within responses to the question 'What do you think you could do as a clinician to improve the veterinary knowledge base?' and the corresponding frequency and keyness values.

Keyword rank	Keyword	Keyness value	Frequency
1	clinical	6826.86	657
2	cases	5637.66	773
3	CPD	4336.10	304
4	practice	3224.40	493
5	research	2655.36	475
6	studies	2568.01	393
7	trials	2337.15	252
8	data	2333.00	391
9	treatment	1754.16	286
10	outcomes	1730.03	172

Table 11
Top 15 collocates of *cases* (ranked by the MI3 statistic) within responses to the question 'What do you think you could do as a clinician to improve the veterinary knowledge base?'.

Collocate rank	Collocate	**MI3 score**	**Frequency**
1	of	20.68	328
2	interesting	19.60	123
3	to	19.47	280
4	report	19.33	152
5	and	19.24	235
6	unusual	19.18	110
7	clinical	18.27	158
8	in	18.14	185
9	on	17.62	122
10	up	17.55	89
11	more	17.18	124
12	record	17.04	79
13	seen	16.95	64
14	a	16.83	109
15	write	16.75	61

Table 9
Top ten words with sample concordance lines for each of the 3 sample topics displayed (out of 10 topics) within responses to the question 'What would be your main concerns about conducting research in your practice?'. The numbers represent the weight or importance of a word in relation to the other words appearing in the text.

Topic	Top ten words per topic with sample concordance lines from WordSmith
1	**commitment (0.1218),** takes (0.0255), properly (0.0236), lot (0.0194), additional (0.0188), samples (0.0175), filling (0.0121), employer (0.0101), run (0.0093), collected (0.0091) Concordance lines of *commitment* & its most frequent L1 collocate *time* (n=35): 1 pliance of farmers the *time* **commitment,** lack of continuation 2 not add up to evidence *Time* **commitment,** particulalry if a de 3 ming client compliance *Time* **commitment,** re paperwork etc. Ti
2	**issues (0.0749),** possible (0.0324), increased (0.0223), funding (0.0203), doing (0.0196), workload (0.0190), cooperation (0.0133), persuading (0.0128), welfare (0.0123), boss (0.0121) Concordance lines of *issues* & its most frequent L1 collocate *time* (n=10): 1 the practice principle." *Time* **issues** and owner consent Clien 2 ady heavy caseload/ free *time* **issues** client compliance. time 3 be fed back to practice *Time* **issues** and accuracy That your
3	**ethical (0.0955),** placebo (0.0584), use (0.0245), considerations (0.0214), limitations (0.0170), controlled (0.0154), trials (0.0152), blind (0.0138), ethics (0.0133), approval (0.0122) Concordance lines of *ethical* & its most frequent R1 collocate *considerations* (n=8): 1 rly Time and cost. *Ethical* **considerations.** Administrations w 2 out Cost and time *Ethical* **considerations** when choosing to u 3 esourse limitation *Ethical* **considerations.** Lack of standardi

Table 12
Top 15 illustrative examples from 73 occurrences of *interesting cases* within responses to the question 'What do you think you could do as a clinician to improve the veterinary knowledge base?'.

1	eep a file of abnormal/ **interesting cases** participate in informat
2	ovide information about **interesting cases** data collection. Resear
3	report more unusual and **interesting cases** and submit to a paper/
4	network and posting any **interesting cases** probably shut up and go
5	ts Little publicise any **interesting cases**/ results you have found
6	n and publish feed back **interesting cases** - to where? report on m
7	es write up and collate **interesting cases** presentations and outco
8	relevant cases document **interesting cases** report interesting case
9	eference works document **interesting cases** Attend CPD and take par
10	or CPD and research for **interesting cases**. Take time to conduct c
11	ithin our practice from **interesting cases** or cpd courses document
12	ectiveness Publish good **interesting cases** (if investigated proper
13	unusual cases log more **interesting cases** collect and record trea
14	this area publish more **interesting cases** i have seen. Improve my
15	le to present my 2 most **interesting cases**/ month to other vets pu

Table 13
Top ten words with sample concordance lines for each of the 3 sample topics (out of 10 topics) displayed within responses to the question 'What do you think you could do as a clinician to improve the veterinary knowledge base?'. The numbers represent the weight or importance of a word in relation to the other words appearing in the text.

Topic	Top ten words per topic with sample concordance lines from WordSmith
1	**info (0.0918),** audits (0.0494), meetings (0.0244), collate (0.0199), discuss (0.0184), analysis (0.0154), audit (0.0152), house (0.0143), complete (0.0104), exchange (0.0090) Concordance lines of *info* & its most frequent R1 collocate *on* (n=14): 1 adiographs, but I need more **info** *on* the use of these investi 2 aving access to statistical **info** *on* results from other pract 3 ars to carry out? collating **info** *on* clinical cases eg Previo
2	**discussion (0.0716)**, online (0.0596), forums (0.0442), groups (0.0408), internet (0.0215), discuss (0.0174), sharing (0.0161), experiences (0.0137), meetings (0.0136), attend (0.0133) Concordance lines of *discussion* & its most frequent R1 collocate groups (n=16): 1 es. Attending forums/ **discussion** *groups* more clinical audits 2 t communication as in **discussion** *groups* submit interesting c 3 g to vets involved in **discussion** *groups*. Further education -
7	**samples (0.0541)**, outcome (0.0243), collection (0.0236), collect (0.0204), submit (0.0201), particular (0.0195), blood (0.0192), projects (0.0149), owners (0.0126), referring (0.0119) Concordance lines of *samples* with frequent L1 collocates 1 e prepared to submit blood **samples**, histories etc Gather dat 2 etails in studies, collect **samples** record cases. discuss in 3 approach. take laboratory **samples** where possible and take r

response in an iterative fashion. In comparison, our method reduces subjectivity and at the same time makes it possible to consider larger amounts of data which have been reported disadvantages of the more traditional approaches (Bryman, 2012; Green and Thorogood, 2014). The study is also of value as an example of a transdisciplinary approach to address a problem where current methods typically used within a field are cumbersome and time consuming (Hadorn et al., 2008).

There were a number of limitations with this study. In terms of the application of corpus methods, respondents did not always use full sentences, which was reflected by the 'bullet-point' nature of some responses and a lack of 'function' words such as *by, they* and *you* when the data were analysed. The brevity of the responses poses a potential issue for the corpus linguistic analysis as well as the topic modelling, because co-occurrence counts of words may continue across responses by different participants. However, there are other studies similar to ours successfully utilising these analytical approaches for corpora of short questionnaire responses (Millar and Hunston, 2015; Nolte et al., 2018). For further in-depth understanding of the barriers to and motivators for PBR (and of ways to improve the knowledge base), other methods such as interviews or focus groups could complement our findings. However, such methods are not usually possible to conduct on as large a scale as this survey was. Previous pilot analysis of the data using more traditional methods of qualitative analysis, namely thematic analysis, elicited similar results to those found in this study (Caulfield, 2013). More extensive comparative analysis would need to be undertaken to be able to confidently report equivalence between these methods and other qualitative approaches.

In terms of possible limitations relating to the collection of data via survey, information bias can occur when using questionnaire methods to collect data (Dohoo et al., 2010), particularly in relation to the use of ambiguous questions. It is possible in this instance that because 'practice-based research' was not defined, it could lead to a number of different understandings by respondents leading to 'local' interpretations specific to each individual (Weiss, 2001). However, it was this 'local' interpretation that was being sought in the study to further understand the barriers as perceived by practitioners, and pre-testing and piloting of the questionnaire was undertaken to minimise the ambiguity of the questions asked. Selection bias can also be a disadvantage when using a questionnaire based study (Dohoo et al., 2010). A census approach was used in this survey in an attempt to mitigate this type of bias, with all veterinary surgeons listed on the RCVS mailing list invited to participate. Those individuals with an interest, or dislike of evidence-based veterinary medicine, may have been more likely to respond to this study. However, techniques were employed to encourage all those invited to complete the survey, including multiple reminders and incentives, and adequate power was demonstrated. The general demographics of the respondents reflected a variety of different age groups and occupations, with the proportion of women and distribution of work places similar to that found in previous research (Robertson-Smith et al., 2010). Other limitations to conducting surveys and using

questionnaires for data collection have been addressed in previous publications (Nielsen et al., 2014; Nielsen et al., 2015).

5. Conclusions

From our study, it appears that members of the veterinary profession are generally positive about PBR and see clear benefits, both personally and at the professional level; some concerns exist, such as the amount of time required, which should be addressed prior to the commencement of research studies. Veterinary practitioners perceive themselves as contributors to the veterinary knowledge base by acting as suppliers of data and by sharing personal experiences with colleagues. Researchers running PBR initiatives should take into account these perceptions when explaining the requirements of research involvement to veterinarians.

Beyond these findings, our paper makes an original contribution to the use of novel methodologies in veterinary research by demonstrating the value of corpus linguistic and mathematical methods for the analysis of textual data. Within veterinary research, questionnaires are commonly used for research and therefore these approaches open new interdisciplinary avenues for the field. Future work should focus on the compilation of veterinary data sets that are specifically collected for corpus approaches.

Acknowledgements

The authors wish to thank the veterinarians who participated in the survey. The questionnaire study was supported by an unrestricted grant from Elanco Animal Health and The University of Nottingham; the corpus linguistic and topic modelling analyses were completed as part of the 'Mining the Meaning' project funded by The University of Nottingham.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.prevetmed.2017.11.020.

References

Arora, S., Ge, R., Moitra, A., 2012. Learning topic models–going beyond SVIn, 2012D. In, 2012 IEEE 53rd. FOCS) 1–10.
Bartram, D.J., Yadegarfar, G., Baldwin, D.S., 2009. Psychosocial working conditions and work-related stressors among UK veterinary surgeons. Occup. Med. 59, 334–341.
Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent dirichlet allocation. J. Mach. Learn. Res. 3, 993–1022.
Bondi, M., Scott, M., 2010. Keyness in Texts. John Benjamins, Amsterdam.
Bowling, A., 2009. Research Methods in Health. Open University Press, Maidenhead.
Braun, V., Clarke, V., 2006. Using thematic analysis in psychology. Qual. Res. Psychol. 3, 77–101.
Bryman, A., 2012. Social Research Methods. Oxford University Press, USA.
Caulfield, S.J.E., 2013. Exploring the Motivators and Barriers of Vets Undertaking Practice-based Research. University of Nottingham Research dissertation.
Coe, J.B., Adams, C.L., Bonnett, B.N., 2009. Prevalence and nature of cost discussions during clinical appointments in companion animal practice. J. Am. Vet. Med. Assoc. 234, 1418–1424.
Dale, V., Pierce, S., May, S., 2013. Motivating factors and perceived barriers to participating in continuing professional development: a national survey of veterinary surgeons. Vet. Rec. 173, 247.
Dillman, D.A., Smyth, J.D., Christian, L.M., 2009. Mail and Internet Surveys: The Tailored Design Method. John Wiley & Sons, USA.
Ding, H., Riloff, E., 2015. Extracting information about medication use from veterinary discussions. In: Conf. of the N. Amer. Chapt. Assoc. Comp. Ling. ? Human Language Technologies. Denver, Colorado, USA.
Dohoo, I., Martin, W., Stryhn, H., 2010. Veterinary Epidemiologic Research. AVC, Incorporated, Charlottetown, Canada.
Edwards, P., Roberts, I., Clarke, M., DiGuiseppi, C., Pratap, S., Wentz, R., Kwan, I., 2002. Increasing response rates to postal questionnaires: systematic review. Br. Med. J. 324, 1183.
Evert, S., 2008. Corpora and collocations. In: Lüdeling, A., Kytö, M. (Eds.), Corpus Linguistics: An International Handbook. Walter de Gruyter, Berlin.
Fischer-Starcke, B., 2009. Keywords and frequent phrases of Jane Austen's Pride and Prejudice A corpus-stylistic analysis. Int. J. Corpus Linguist. 14, 492–523.
Fleming, J., Creevy, K., Promislow, D., 2011. Mortality in north american dogs from 1984 to 2004: an investigation into age-, size-, and breed-related causes of death. J. Vet. Intern. Med. 25, 187–198.
Green, J., Thorogood, N. (Eds.), 2014. Qualitative Methods for Health Research, 3rd Edition. SAGE publications Limited, London.
Gruen, M.E., Jiamachello, K.N., Thomson, A., Lascelles, B.D.X., 2014. Clinical trials involving cats What factors affect owner participation? J. Feline Med. Surg. 16, 727–735.
Hirsch Hadorn, G., Hoffmann-Reim, H., Biber-Klemm, S., Grossenbacher, W., Joye, D., Pohl, C., Wiesmann, U., Zemp, E., 2008. Handbook of Transdisciplinary Research. Springer, Netherlands.
Harvey, K., 2014. Investigating Adolescent Health Communication: A Corpus Linguistics Approach. Bloomsbury Publishing, London.
Huntley, S.J., Dean, R.S., Brennan, M.L., 2017. The awareness of the international veterinary profession of evidence-based veterinary medicine and preferred methods of training. Vet. Sci. 4. http://dx.doi.org/10.3390/vetsci4010015.
Kim, H., Park, H., 2008a. Nonnegative matrix factorization based on alternating nonnegativity constrained least squares and active set method. SIAM J. Matrix Anal. Appl. 30, 713–730.
Kim, J., Park, H., 2008b. Toward faster nonnegative matrix factorization: a new algorithm and comparisons, 2008. In: Eighth IEEE International Conference on Data Mining. Pisa, Italy. pp. 353–362.
Kim, J., Park, H., 2010. Nonnegative Matrix Factorization by Alternating Nonnegativity Constrained Least Squares Using Block Principal Pivoting/active Set Method (MATLAB Code). School of Computational Science and Engineering, Georgia Institute of Technology.
Kuang, D., Choo, J., Park, H., 2015. Nonnegative matrix factorization for interaction topic modelling and document clustering. In: Celebi, M. (Ed.), Partitional Clustering Algorithms. Springer, Switzerland, pp. 215–243.
Lai, E., Moyer, D., Yuan, B., Fox, E., Hunter, B., Bertozzi, A.L., Brantingham, J., 2014. Topic Time Series Analysis of Microblogs. DTIC Document.
Ma, H., Zhao, W., Shi, Z., 2013. A nonnegative matrix factorization framework for semi-supervised document clustering with dual constraints. Knowl. Inform. Syst. 36, 629–651.
Mahlberg, M., McIntyre, D., 2011. A case for corpus stylistics: ian fleming's casino royale. Engl. Text Constr. 4, 204–227.
Mahlberg, M., Stockwell, P., et al., 2016. Point and CLiC: teaching literature with corpusstylistic tools. In: Burke (Ed.), Scientific Approaches to Literature in Learning Environments. John Benjamins, Amsterdam, pp. 251–267.
Mahlberg, M., 2005. English General Nouns: A Corpus Theoretical Approach. John Benjamins Publishing, Amsterdam.
Mahlberg, M., 2014. Corpus linguistics and discourse analysis. Pragmat. Discourse: Handb. Pragmat. 3, 215–238.
Mateus, A.L., Brodbelt, D.C., Barber, N., Stärk, K.D., 2014. Qualitative study of factors associated with antimicrobial usage in seven small animal veterinary practices in the UK. Prev. Vet. Med. 117, 68–78.
McEnery, T., Xiao, R., Yukio, T., 2006. Corpus-based Language Studies: An Advanced Resource Book. Taylor & Francis, Oxford.
McEnery, T., 2009. Keywords and moral panics: Mary Whitehouse and media censorship. What's in Word-list. Ashgate, Farnham.
Millar, N., Hunston, S., 2015. Adjectives, communities, and taxonomies of evaluative meaning. Funct. Lang. 22, 297–331.
Nöremark, M., Lewerin, S.S., Ernholm, L., Frössling, J., 2016. Swedish farmersö opinions about biosecurity and their intention to make professionals use clean protective clothing when entering the stable. Front. Vet. Sci. 3. http://dx.doi.org/10.3389/fvets.2016.00046.
Nielsen, T., Dean, R., Robinson, N., Massey, A., Brennan, M., 2014. Survey of the UK veterinary profession: common species and conditions nominated by veterinarians in practice. Vet. Rec vetrec-2013-101745.
Nielsen, T., Dean, R., Massey, A., Brennan, M., 2015. Survey of the UK veterinary profession 2: sources of information used by veterinarians. Vet. Rec. 177, 257 vol 177, pg 172, 2015.
Nolte, I., Ancarno, C., Jones, R., 2018. Using Corpus Methods to Explore a Survey Oninter-religious Relations in Yorubaland. Corpora., southwest Nigeria.
O'Neill, D., Elliott, J., Church, D., McGreevy, P., Thomson, P., Brodbelt, D., 2013. Chronic kidney disease in dogs in UK veterinary practices: prevalence, risk factors, and survival. J. Vet. Intern. Med. 27, 814–821.
O'Neill, D., 2015. Building practice-based research in the UK. Vet. Rec. 176, 461–462.
Page-Jones, S., Abbey, G., 2015. Career identity in the veterinary profession. Vet. Rec. 176, 433.
Radford, A., Noble, P., Coyne, K., Gaskell, R., Jones, P., Bryan, J., Setzkorn, C., Tierney, Á., Dawson, S., 2011. Antibacterial prescribing patterns in small animal veterinary practice identified via SAVSNET: the small animal veterinary surveillance network. Vet. Rec. 169, 310.
Reja, U., Manfreda, K.L., Hlebec, V., Vehovar, V., 2003. Open-ended vs. close-ended questions in web questionnaires. Dev. Appl. Stat. 19, 159–177.
Robertson-Smith, G., Robinson, D., Hicks, B., Khambhaita, P., Hayday, S., 2010. The 2010 RCVS survey of the UK veterinary and veterinary nursing professions, institute for employment studies. Brighton.
Robinson, N., Dean, R., Cobb, M., Brennan, M., 2014. Short communication: consultation length in first opinion small animal practice. Vet. Rec. 175, 486.
Roshier, A., McBride, E., 2012. Canine behaviour problems: discussions between veterinarians and dog owners during annual booster consultations. Vet. Rec vetrec-2012-101125.
Scantlebury, C.E., Perkins, E., Pinchbeck, G.L., Archer, D.C., Christley, R.M., 2014. Could it be colic? Horse-owner decision making and practices in response to equine colic.

BMC Vet. Res. 10. http://dx.doi.org/10.1186/1746-6148-10-S1-S1.
Scott, M., Tribble, C., 2006. Textual Patterns: Key Words and Corpus Analysis in Language Education. John Benjamins, Amsterdam.
Scott, M., 2012a. WordSmith Tools Lexical Analysis Software, Stroud.
Scott, M., 2012b. Downloading BNC or ANC Word Lists. http://lexically.net/downloads/BNC_wordlists/downloading%20BNC.htm (Accessed 19.11.17).
Scott, M., 2016a. How Key Words Are Calculated. http://www.lexically.net/downloads/version5/HTML/?keywords_calculate_info.htm (Accessed 19.11.17).
Scott, M., 2016b. What Is Concord and What's It For? http://www.lexically.net/downloads/version5/HTML/?concord_basics.htm (Accessed 19.11.17).
Seale, C., Ziebland, S., Charteris-Black, J., 2006. Gender, cancer experience and internet use: a comparative keyword analysis of interviews and online cancer support groups. Soc. Sci. Med. 62, 2577–2590.
Sinclair, J., 1991. Corpus, Concordance, Collocation. Oxford University Press, Oxford.
Skelton, J.R., Wearn, A.M., Hobbs, F.R., 2002. 'I'and 'we': a concordancing analysis of how doctors and patients use first person pronouns in primary care consultations. Family Pract. 19, 484–488.
Slegers, C., Zion, D., Glass, D., Kelsall, H., Fritschi, L., Brown, N., Loff, B., 2015. Why do people participate in epidemiological research? J. Bioeth. Inq. 12, 227–237.
Upjohn, M., Attwood, G., Lerotholi, T., Pfeiffer, D., Verheyen, K., 2013. Quantitative versus qualitative approaches: a comparison of two research methods applied to identification of key health issues for working horses in Lesotho. Prev. Vet. Med. 108, 313–320.
Weiss, M.G., 2001. Cultural epidemiology: an introduction and overview. Anthropol. Med. 8, 5–29.
White, P.C., Jennings, N.V., Renwick, A.R., Barker, N.H., 2005. Review: questionnaires in ecology: a review of past use and recommendations for best practice. J. Appl. Ecol. 42, 421–430.