



## Quantifying veterinarians' beliefs on disease control and exploring the effect of new evidence: A Bayesian approach

H. M. Higgins,\*<sup>1</sup> J. N. Huxley,† W. Wapenaar,† and M. J. Green†

\*Faculty of Health and Medical Sciences, School of Veterinary Medicine, Guildford, University of Surrey, GU2 7TE, United Kingdom

†University of Nottingham, School of Veterinary Medicine and Science, Sutton Bonington Campus, Leicestershire, LE12 5RD, United Kingdom

### ABSTRACT

The clinical beliefs (expectations and demands) of veterinarians regarding herd-level strategies to control mastitis, lameness, and Johne's disease were quantified in a numerical format; 94 veterinarians working in England (UK) were randomly selected and, during interviews, a statistical technique called probabilistic elicitation was used to capture their clinical expectations as probability distributions. The results revealed that markedly different clinical expectations existed for all 3 diseases, and many pairs of veterinarians had expectations with nonoverlapping 95% Bayesian credible intervals. For example, for a 3-yr lameness intervention, the most pessimistic veterinarian was centered at an 11% population mean reduction in lameness prevalence (95% credible interval: 0–21%); the most enthusiastic veterinarian was centered at a 58% reduction (95% credible interval: 38–78%). This suggests that a major change in beliefs would be required to achieve clinical agreement. Veterinarians' clinical expectations were used as priors in Bayesian models where they were combined with synthetic data (from randomized clinical trials of different sizes) to explore the effect of new evidence on current clinical opinion. The mathematical models make predictions based on the assumption that veterinarians will update their beliefs logically. For example, for the lameness intervention, a 200-farm clinical trial that estimated a 30% mean reduction in lameness prevalence was predicted to be reasonably convincing to the most pessimist veterinarian; that is, in light of this data, they were predicted to believe there would be a 0.92 probability of exceeding the median clinical demand of this sample of veterinarians, which was a 20% mean reduction in lameness. Currently, controversy exists over the extent to which veterinarians update their beliefs logically, and further research on this is needed. This study has demonstrated that probabilistic elicitation and a Bayesian framework are useful for evaluating the diversity and strength of veterinarians' clinical

beliefs. The wide variations observed have implications for designing future projects. Although many factors influence disease control, nonetheless the heterogeneity in beliefs also raises concern over the extent to which a broadly consistent approach is currently being achieved; it supports the argument for more randomized clinical trials and for national programs to control nonstatutory endemic diseases.

**Key words:** evidence-based medicine, Bayesian updating, clinical belief, probabilistic elicitation

### INTRODUCTION

In the United Kingdom, most major endemic diseases of cattle are not under statutory control and hence the health of the national herd is critically reliant on farmers and their advisors. The clinical beliefs held by veterinarians working in private practice are thus important because they underpin the advice provided to farmers, the decisions that are made, and the way in which new research evidence is interpreted and applied. If new research is to have any effect it must be capable of changing veterinarians' current beliefs, and ultimately those of the farmers, otherwise it risks being of limited value and will not deliver a cost beneficial return on the investment necessary to conduct the work.

If the unknown parameter of interest, say  $\theta$ , is defined as the effect of a specified treatment (or disease intervention), then a veterinarian's clinical expectations regarding  $\theta$  pertain to the possible values they believe  $\theta$  will take, based on their current clinical experiences and knowledge (i.e., their beliefs, which includes information and influences from all possible sources). Within a Bayesian statistical framework, clinical expectations can be expressed in a numerical format as a probability distribution, using a statistical technique called probabilistic elicitation (O'Hagan et al., 2006). Once expectations are captured as probability distributions, it is possible to quantitatively study the variation in clinical expectations, and to explore, using Bayesian statistical models, the response of veterinarians with different clinical expectations to new information (Parmar et al., 1994). For a single continuous unknown parameter,  $\theta$ , Bayes theorem can be written as

Received May 31, 2013.

Accepted February 12, 2014.

<sup>1</sup>Corresponding author: [h.higgins@surrey.ac.uk](mailto:h.higgins@surrey.ac.uk)

$$\pi(\theta|x) \propto \pi(x|\theta) \pi(\theta), \quad [1]$$

where, in the context of this paper,  $\pi(\theta)$  is the veterinarian's current clinical expectation for  $\theta$ , expressed as a probability distribution;  $\pi(x|\theta)$  is the likelihood function (based on new synthetic data,  $x$ ); and  $\pi(\theta|x)$  is the posterior probability distribution that encapsulates everything that is now known about  $\theta$ , having updated the veterinarian's current clinical expectation with the new information contained in the data. In Bayesian terminology,  $\pi(\theta)$  is referred to as the prior. Equation [1] states that the extent of any logical change in belief depends on both the current clinical expectation  $\pi(\theta)$  and the strength of the new evidence  $\pi(x|\theta)$ ; thus, Bayes theorem quantifies (probabilistically) how new data should, logically, influence pre-existing expectations about an uncertain parameter.

Clinical expectations have been probabilistically elicited in human medicine and to a lesser extent in veterinary medicine (O'Hagan et al., 2006). In addition, at the same time as eliciting doctors' clinical expectations, some studies have elicited doctors' "clinical demands,"  $\theta^{min}$ . A new treatment almost invariably carries economic and other costs (e.g., adverse reactions, toxicity, or drug resistance). As a result, doctors usually demand that a "minimum clinically worthwhile effect" (i.e.,  $\theta^{min}$ ) must be achieved in order for them to consider the treatment worthwhile implementing when compared against alternative approaches. Eliciting a clinical demand requires a doctor to weigh all the potential advantages and disadvantages of the treatment and assign a single value (or range of values) to  $\theta^{min}$  (Parmar et al., 1994). To our knowledge, clinical demands have not been elicited from veterinarians working in farm animal medicine.

In a Bayesian analysis (Equation [1]),  $\theta^{min}$  can be juxtaposed on the posterior distribution  $\pi(\theta|x)$ , and the null hypothesis (no treatment effect) is rejected in favor of the alternative hypothesis ( $\theta > \theta^{min}$ ) when the posterior distribution assigns a "sufficiently large" probability to  $\theta$  taking values greater than  $\theta^{min}$  (Spiegelhalter et al., 2004). This probability, denoted here by  $\lambda$ , is a subjective choice but "sufficiently large" is often taken to be  $\lambda \geq 0.95$ . The choice for  $\lambda$  affects when the trial is stopped and the larger the value of  $\lambda$ , the greater the probability that the minimum clinically worthwhile effect will be realized.

In a retrospective Bayesian analysis, a recent study used a range of theoretical clinical expectations (priors) aimed at representing different opinions (from "severely skeptical" to "very enthusiastic") to model the effect of a clinical mastitis trial (Green et al., 2009). The study explored how the results of the trial would be interpreted by veterinarians with theoretically constructed

different initial beliefs and predicted that important differences in opinion would exist in light of the evidence. However, to the authors' knowledge, no studies have actually quantified veterinarians' clinical expectations or demands concerning herd-level interventions to reduce the endemic diseases of cattle. This omission is striking because, in analogy with human medicine (Spiegelhalter et al., 2004), a potentially important obstacle to the widespread implementation of veterinary interventions on farms may relate to the variability of veterinarians' beliefs concerning the effectiveness of disease control measures. If veterinarians currently have very different clinical expectations, then other approaches, such as more convincing research evidence, will be required to alter these beliefs. Although this is not to say that it is only research evidence that influences what veterinarians believe. Our hypothesis was that demonstrable variation exists for veterinarians' clinical expectations and demands regarding mastitis, lameness, and Johne's disease.

The first objective of this study was to report a probabilistic elicitation that aimed to quantify the current variation in veterinarians' clinical expectations and demands with respect to herd-level interventions to reduce mastitis, lameness, and Johne's disease in dairy cattle. The second objective was to explore, within a Bayesian statistical framework, the implications of the heterogeneity in beliefs. This was done by using a mathematical model to make predications regarding how new evidence will alter what veterinarians currently believe. The model makes the assumption that veterinarians will update their beliefs logically.

## MATERIALS AND METHODS

### Recruitment of Veterinarians

It was of interest to compare the beliefs of veterinarians with and without postgraduate qualifications. However, without some targeted selection, it was possible that no veterinarians holding postgraduate qualifications would be included in the sample. Therefore, to be eligible for selection, veterinarians must have been currently providing healthcare to dairy cattle in England during their normal working hours and be employed by a veterinary practice that contained at least one veterinarian possessing a postgraduate cattle qualification; that is, the European College of Bovine Health Management Diploma, the Royal College of Veterinary Surgeons Certificate or Diploma in Cattle Health and Production, or the University of Liverpool Diploma in Bovine Reproduction. None of the veterinarians were exclusively involved in providing emergency care or consultancy.

A 2-stage cluster design stratified by geographic location was used. First, 20 practices were selected with probability proportional to the number of veterinarians they contained. Subsequently, if the practice had >5 eligible members, then 5 were selected using the random number generator function in the software program R version 2.15.2 (R Core Team, 2012). In practices that contained  $\leq 5$  eligible participants, all veterinarians were recruited. This sampling strategy meant that every veterinarian had approximately the same chance of selection, independent of the size of the practice in which they worked (Kalton, 1987). The Royal College of Veterinary Surgeons online database (<http://www.rcvs.org.uk/>) provided a list of veterinary practices. Face-to-face, individual interviews were conducted by H.M.H. between June 8 and September 1, 2011. Veterinarians were provided with an inconvenience allowance of £100/h (pro rata). A standard script was used for consistency, and basic descriptive information concerning the veterinarians was gathered.

### Clinical Expectations

We examined 3 epistemic parameters of interest ( $\theta_m$ ,  $\theta_l$ ,  $\theta_j$ ), each of which related to the population mean effect of an intervention strategy aimed at reducing an endemic disease. The subscript refers to the endemic disease, such that  $\theta_m$  was defined as the population mean effect of a 1-yr mastitis intervention,  $\theta_l$  a 3-yr lameness intervention, and  $\theta_j$  an 8-yr Johne's disease intervention. In addition,  $\theta_d$  is used to refer more generally to the population mean effect of a disease intervention ( $d = m, l, j$ ). For each disease, what was meant by "an intervention strategy" was described to each participant in detail (see Appendix) and can be summarized as follows. For all 3 diseases, an initial assessment of any available farm data was conducted, in conjunction with a farm visit, to assess the risks and achieve a working diagnosis. For mastitis and Johne's disease, the initial assessment was additionally taken to include some strategic laboratory testing, and for lameness, some locomotion scoring. As a result, a list of farm-specific recommendations was made, based on current best evidence where possible, prioritized, and discussed with the farmer. During quarterly farm revisits, the farm data and risks were reassessed, allowing the control measures to be modified in a timely manner, and encouragement and support were provided to implement the control measures. Note, that in the UK, uncomplicated cases of mastitis and lameness are often diagnosed and treated initially by farm personnel, following a standard protocol agreed with their veterinarian.

Different time periods were chosen for the interventions and these reflect the fact that the time taken to

make clinically observable reductions in disease levels varies with the disease. For the task, it was assumed that all dairy farms in England had been recently visited by their local veterinarian to discuss the disease in question. The population under consideration was all the farmers who, following this consultation, subsequently decided to proceed with the disease intervention.

The disease outcomes were as follows. For the mastitis intervention, it was assumed that in the year before and in the year after the intervention took place, the number of clinical cases of mastitis per 100 cows was reliably measured for each farm in the population, to calculate the population mean mastitis incidence before and after the intervention; subclinical mastitis was not considered. For lameness, it was assumed that for each farm in the population, the entire dairy herd was reliably scored for lameness at the start of the intervention and then again at the end, to determine the population mean prevalence of lameness; lame was defined as observably lame in one or more limbs, with any severity of limp. For Johne's disease, it was assumed that for each farm in the population all the adult cattle ( $\geq 24$  mo old) were blood sampled before and after the intervention and the percentage of blood-test-positive adults was used as an approximation for the population mean prevalence of Johne's disease; it is acknowledged that this is an estimate because of the limited sensitivity of the test. These outcomes were chosen because veterinarians were likely to be familiar with them. The population mean outcome before and after the intervention are denoted  $\psi_d^0$  and  $\psi_d^1$ , respectively. As before, subscript  $d$  refers to the endemic disease ( $d = m, l, j$ ).

The scale used was the difference in the population mean outcome (before and after the intervention) divided by the population mean outcome before the intervention, expressed as a percentage:

$$\theta_d = \frac{\psi_d^0 - \psi_d^1}{\psi_d^0} \times 100, \quad \psi_d^0 > 0. \quad [2]$$

Hence,  $\theta_d \in [-\infty, +100]$  for  $d = m, l, j$ . On this scale, positive percentages reflect reductions in disease and negative percentages reflect increases in disease. Thus, for example,  $\theta_l = 20\%$  is a 20% reduction in lameness prevalence, averaged across all farms, with a 3-yr lameness intervention. Veterinarians' expectations for  $\theta_d$  were elicited using a method that enabled probability distributions to be fitted to their answers (see next section).

The 3 parameters ( $\theta_d$ ) were elicited conditional on the assumption that without an intervention delivered with the help of a veterinarian, the change in population mean over the time period would be zero for each

disease. In addition, despite the fact that the population parameters related to farmers who had volunteered to participate with the intervention, it was possible that some farmers may not actually implement any of the recommendations either at all or to a sufficient standard, such that their intervention had no chance of having any effect at all. Therefore, veterinarians were asked to provide a single estimate for the percentage of farmers who would fall into this category of zero implementation for each disease. Veterinarians were told to exclude these farmers from the population under consideration when they made their estimates for  $\theta_d$ ; however, it was also stated that even with this exclusion, there was likely to remain a wide diversity in the scale and effectiveness of farmer implementation.

In summary, veterinarians' beliefs were sought for the expected (future) outcome of the specified disease intervention, on average, across all farms who implemented at least some of the recommendations ( $\theta_d$ ). For example, for the 8-yr Johne's program,  $\theta_j$  documents veterinarians' beliefs regarding the final outcome of this intervention, given that it commenced today.

### **Probabilistic Elicitation Method**

The fixed-interval method was used, such that for each parameter ( $\theta_d$ ), 5 values were elicited to allow a probability distribution to be fitted: the minimum and maximum values, median, and lower and upper quartiles. The minimum and maximum values for the parameter together constituted a plausible range, such that the veterinarian considered it "exceedingly unlikely" that the actual answer would fall outside this range. The order in which these values are elicited is important to avoid an anchoring and adjustment heuristic; thus, the plausible range was elicited first so that further judgments were made relative to this range. If, for example, the median were elicited first, then there would be a tendency for this to act as an anchor, and the maximum and minimum values may not be adjusted far enough away from the median (O'Hagan et al., 2006).

Current best practice for probabilistic elicitation was used, whereby a face-to-face interview between the first author and each veterinarian was conducted at the veterinarian's own place of work and lasted between 70 and 100 min. The first part of the elicitation involved a 15- to 25-min training exercise, whereby the scale upon which veterinarians needed to give their answers was explained, as well as the meaning of the 5 judgments they needed to make, using a fictitious example of a veterinarian who was convinced that an unspecified disease intervention would certainly make the disease problem worse (Parmar et al., 1994). This example was

used on the grounds that the majority of veterinarians would believe that the disease interventions in question would typically reduce disease, and therefore the example explained how they should give their answers but avoided biasing them toward a certain disease reduction. It is also worth noting that during their interviews the veterinarians were asked for their current clinical expectations and in particular, they were not shown any data pertaining to any of the parameters,  $\theta_d$ .

The Sheffield Elicitation Framework is a freely available package of guidance documents, templates, and software specifically designed for carrying out probabilistic elicitation (SHELF; <http://www.tonyohagan.co.uk/shelf/>). Probability distributions were fitted to the veterinarians' 5 judgments using computer code provided with this package (SHELF version 1.01) and run in the software program R, version 2.15.2 (R Core Team, 2012). The probability distributions were fitted during the interviews and graphically presented to each veterinarian. Furthermore, some of the inferred, but not directly elicited, probabilities were cited to them; this enabled veterinarians to visualize their answers and gave them the opportunity to revise their 5 judgments appropriately until they felt the fitted distributions were a fair reflection of their beliefs.

### **Clinical Demands**

For each disease, veterinarians were asked to weigh all the benefits of implementing the intervention on many thousands of farms, against all the costs (see Appendix). The minimum population mean reduction they would want the intervention to deliver, to consider the intervention clinically worthwhile implementing on a large scale, is subsequently referred to as their "clinical demands" ( $\theta_d^{min}$ ,  $d = m, l, j$ ). Veterinarians were asked to either give a single value or, if they preferred, a range of values.

In addition, for each disease intervention, veterinarians were asked to give a value for  $\lambda$ :  $\lambda_d$ ,  $d = m, l, j$ . That is, they were asked for the probability (as a percentage chance) that the actual mean reduction in disease would be greater than what they had stated was a minimally clinically worthwhile effect. This value for  $\lambda$  was placed in the context of the veterinarian needing to be "convinced that the disease intervention is clinically useful and worthwhile implementing on many farms," based on the evidence produced by a clinical trial alone.

The purpose of eliciting  $\lambda$  was to explore any differences between veterinarians in terms of the strength of evidence needed to be convincing to them. However, this is not to say that it is only evidence from clinical trials that influences how convinced veterinarians

are regarding the outcome of a disease intervention in reality. Many factors (e.g., marketing campaigns, clinical experience, farmer experience, colleagues) can have an effect. It was beyond the scope of this project to quantify and mathematically model the influence of all possible factors. However, because it is hoped that new evidence has some influence on veterinarians' confidence in the usefulness of disease interventions, it is of interest to know how veterinarians differ in terms of the strength of evidence they would like to be presented with; by eliciting  $\lambda$ , this is what our study has focused on.

The study was approved by the Research and Ethics committee, School of Veterinary Medicine and Science, University of Nottingham.

### Data Analysis

Due to the scale involved; that is,  $\theta_d \in [-\infty, +100]$ , a suitable choice was to fit normal (Gaussian) probability density functions to the 5 elicited values for  $\theta_m$ ,  $\theta_l$ , and  $\theta_j$ , for each veterinarian. The normal family of probability density functions is characterized by 2 parameters: the mean and the variance. In keeping with Bayesian terminology, these are referred to as "hyperparameters" to distinguish them from the elicited (population) parameters. The R code provided in SHELF was used to find the best fitting hyperparameters by minimizing the sum of the squared differences between the fitted cumulative distribution and the elicited cumulative distribution; this uses numerical optimization based on the simplex algorithm (Nelder and Mead, 1965). In addition for each disease, the 5 raw elicited values were each arithmetically averaged (across the veterinarians), and a normal distribution fitted to these 5 averaged numbers; for each disease, this distribution will be referred to as the "clinical expectations for the average veterinarian." It should be noted that this is a theoretical concept because it is unlikely to represent the actual belief of any individual person.

**Visualizing the Diversity in Clinical Opinion for  $\theta_d$ .** To illustrate the diversity in clinical expectations among the sample of veterinarians for each disease, 3 different individuals were identified: (a) the most ambivalent, (b) the most pessimistic, and (c) the most enthusiastic. Veterinarian (a) was identified based on the widest plausible range for  $\theta_d$ , veterinarians (b) and (c) were identified based first on the lowest and highest median belief for  $\theta_d$ , respectively, and then if tied results existed (e.g., two or more veterinarians had the same highest median belief) the veterinarian was selected that had the narrowest plausible range. In other words, "enthusiasm/pessimism" was based on center of location, and if more than one veterinarian was

centered in the same enthusiastic or pessimistic place, then the most confident veterinarian was selected.

**Modeling the Implications of the Heterogeneity in Clinical Beliefs.** The clinical expectations of the most enthusiastic and pessimistic veterinarians were used as priors in a Bayesian analysis to predict how their current clinical expectations would be updated in the light of new data, as follows. The prior belonging to each veterinarian was combined using Bayes theorem (Equation [1]) with synthetic data originating from a randomized clinical trial to produce a posterior distribution for each veterinarian. This posterior distribution can be considered to represent what each veterinarian would believe if they were shown the data and updated their current clinical expectations logically in accordance with Bayes theorem.

For the Bayesian analysis, it was necessary to specify a likelihood function (i.e., to probabilistically specify the new information about  $\theta_d$  arising from an experiment). For each disease, a clinical trial was assumed to have been conducted and designed such that farms were randomly assigned to 1 of 2 groups (control and intervention) with  $n$  farms in each group. The control farms received no intervention. For all farms enrolled in the trial, in the year before and after the intervention the clinical outcome was measured (as previously described), and from this information the sample mean change in disease was calculated for the control and intervention farms (in analogy with Equation [2]) and denoted  $\bar{y}_c$  and  $\bar{y}_d$ , respectively. Assuming independent random observations, then by virtue of the central limit theorem,

$$\bar{y}_d \approx N\left(\theta_d, \frac{\sigma_d^2}{n}\right) \text{ and } \bar{y}_c \approx N\left(\theta_c, \frac{\sigma_d^2}{n}\right),$$

where  $\theta_d$  is the population mean change in disease for farms with an intervention in place,  $\theta_c$  is the same parameter but for the control farms, and  $\sigma_d^2$  is the population variance for the disease and is assumed to be the same for both control and intervention farms. Subscript  $d$  is used to refer generally to the diseases:  $d = m, l, j$  for mastitis, lameness, and Johne's disease, respectively.

The parameter of interest was the difference in the population mean change in disease:  $\theta_d - \theta_c$ . A point estimate for this difference from the trial data can be obtained using  $y = \bar{y}_d - \bar{y}_c$ , and the normal likelihood is given by

$$y \approx N\left(\theta_d - \theta_c, \frac{\sigma_d^2}{n} + \frac{\sigma_d^2}{n}\right);$$

$$\text{that is; } y \approx N\left(\theta_d, \frac{2\sigma_d^2}{n}\right), \quad [3]$$

because, in keeping with the condition used for eliciting the veterinarians' beliefs,  $\theta_c$  is assumed to be zero for each disease. Veterinarians' clinical expectations were elicited for  $\theta_d$  given  $\theta_c = 0$ , and because normal distributions were fitted to their beliefs, the priors took the form

$$p(\theta_{di}) \propto \exp\left(-\frac{1}{2} \frac{(\theta_{di} - \varphi_{di})^2}{\tau_{di}^2}\right),$$

where  $i$  denotes the veterinarian and takes values  $i = 1, \dots, 94$  (94 veterinarians were interviewed), and  $\varphi_{di}$  and  $\tau_{di}$  are the fitted hyperparameters (mean and standard deviation) for each veterinarian. With the normal likelihood specified by Equation [3]; that is,

$$y | \theta_d \propto \exp\left(-\frac{1}{2} \frac{(y - \theta_d)^2 n}{2\sigma_d^2}\right),$$

the Bayesian analysis (i.e., Equation [1]) is conjugate, and it can be shown (Spiegelhalter et al., 2004) that the posterior distribution also takes the shape of a normal distribution, as follows:

$$\theta_{di} | y \sim N\left(\frac{\frac{\varphi_{di}}{\tau_{di}^2} + \frac{yn}{2\sigma_d^2}}{\frac{1}{\tau_{di}^2} + \frac{n}{2\sigma_d^2}}, \frac{1}{\frac{1}{\tau_{di}^2} + \frac{n}{2\sigma_d^2}}\right). \quad [4]$$

In combination with the elicited values for  $\theta_d^{min}$  and  $\lambda_d$ , Equation [4] can be used to explore the implications of the heterogeneity in veterinarians' clinical expectations because it provides the formula to calculate the posterior distribution (i.e., the predicted belief) for each veterinarian and for each disease.

In Equation [4],  $\varphi_{di}$  and  $\tau_{di}$  were derived from the raw elicitation data (obtained during the interviews), and the remaining variables ( $y, n, \sigma_d^2$ ) require specification. The value chosen for  $y$  (i.e., the difference in the sample means between the control and intervention farms in the trial) was the median value (i.e., across the 94 participating veterinarians) of the median clinical expectation for  $\theta_d$ . For each disease, the population variance  $\sigma_d^2$  is unknown and therefore sensitivity of the results to 2 choices were explored:  $\sigma_d = 10\%$  and  $25\%$  for  $d = m, l, j$ . As  $n$  is the number of farms in each arm of the trial,

$2n$  is the total size of the clinical trial and was given values of 20, 50, 100, and 200 farms. For clinical demands, because  $\theta_d^{min}$  was elicited from each veterinarian, the median value (across all veterinarians) was used; note that because a few veterinarians elected to give a range of values, the lower value of the range was taken to calculate the median clinical demand. As described, this Bayesian model explores how large a clinical trial would be required to convince the most pessimistic veterinarian that the median clinical demand of this sample of veterinarians will be exceeded. It also explores, theoretically, how large a clinical trial would be required to bring the most pessimistic and most enthusiastic veterinarian into general agreement; that is, assuming veterinarians revise their beliefs logically in keeping with Bayes theorem. As explained, the model makes some assumptions regarding the data produced by a new clinical trial.

## RESULTS

### Response Rates and Description of Participants

Not all the veterinarians in one initially selected veterinary practice agreed to participate; another practice was selected from the same region and consented. From the 20 consenting practices, 94 of the 96 selected veterinarians were interviewed (2 were unavailable during the data collection period). Of the 20 practices, 7 were located in the North, 2 in the Midlands, and 11 in the South of England.

The data set included 36 partners and 58 assistants; sex was split 59 males to 35 females; and years qualified ranged from 0 to 37 yr, with a median of 7 yr. Of the 94 veterinarians, 19 (20%) held a postgraduate cattle qualification. Time spent working with dairy cows was as follows: 6 participants spent 0 to <25% of their time, 16 participants 25 to <50%, 27 participants 50 to <75%, and 45 participants 75 to <100%.

### Elicited Clinical Beliefs

Descriptive statistics concerning the raw elicited parameters  $\theta_d$  are provided in Table 1. For each disease and each veterinarian, their raw elicited plausible range and median values for  $\theta_d$  are presented in Figure 1, along with  $\theta_d^{min}$ .

In Figure 2, the fitted normal distributions are plotted for the most ambivalent, enthusiastic, pessimistic, and "average" veterinarian. For example, Figure 2B shows that the most pessimistic veterinarian concerning the 3-yr lameness intervention had a normal distribution centered at an 11% population mean reduction in lameness prevalence with a 95% credible interval of 0

**Table 1.** The elicited clinical expectations of 94 veterinarians for the population mean effect of a disease intervention<sup>1</sup>

Elicited value	Population mean reduction (%)		
	Mastitis incidence	Lameness prevalence	Johne's disease prevalence
Minimum	10 (5–15)	10 (5–20)	10 (5–30)
Lower quartile	24 (17–33)	24 (15–35)	33 (20–50)
Median	30 (20–40)	30 (20–40)	40 (30–60)
Upper quartile	36 (25–46)	39 (25–50)	52 (38–65)
Maximum	50 (40–60)	50 (40–65)	70 (50–80)

<sup>1</sup>Values shown are median (interquartile range).

to 21%, whereas the most enthusiastic veterinarian was centered at a 58% reduction with a 95% credible interval of 38 to 78%; thus the enthusiastic veterinarian had a more uncertain belief (wider credible interval) compared with the most pessimistic veterinarian. Figures 1 and 2 together show the considerable heterogeneity in veterinarians' clinical expectations, both in terms of central location and clinical confidence. The diversity in opinion was greater for the Johne's disease intervention compared with the other diseases, as highlighted by the marked polarization in beliefs displayed by the most enthusiastic and pessimistic veterinarians (Figure 2C).

Across the 94 veterinarians, the median clinical demand was a 20% population mean reduction in incidence for a 1-yr mastitis intervention with an interquartile range of 10 to 25%, a 20% reduction in mean prevalence for a 3-yr lameness intervention (interquartile range 10–30%), and a 35% reduction in mean prevalence for an 8-yr Johne's disease intervention (interquartile range 25–50%). Comparing these figures with Table 1, it can be seen that the median clinical demand was below the median clinical expectation for all 3 diseases.

Figure 3 shows how each veterinarian's clinical demand compared with their median (best guess) clinical expectation; that is, whether each of the crosses (Figure 1) was below, above, or in the same place as the corresponding dot. It reveals that for both the mastitis and lameness interventions, most veterinarians ( $\geq 79\%$ ) had clinical demands that were lower than their median expectations. In contrast, for Johne's disease, only 51% of veterinarians had clinical demands that were lower than their median expectation; that is, there was more pessimism in terms of the effect that veterinarians clinically wanted and what they expected to happen for Johne's disease compared with mastitis and lameness.

The summary statistics for veterinarians' values for  $\lambda_d$ ; that is, the probability they would require (as a percentage chance) that the mean reduction would be at least their own estimate for  $\theta_d^{min}$ , were similar for each disease intervention and are as follows. Median

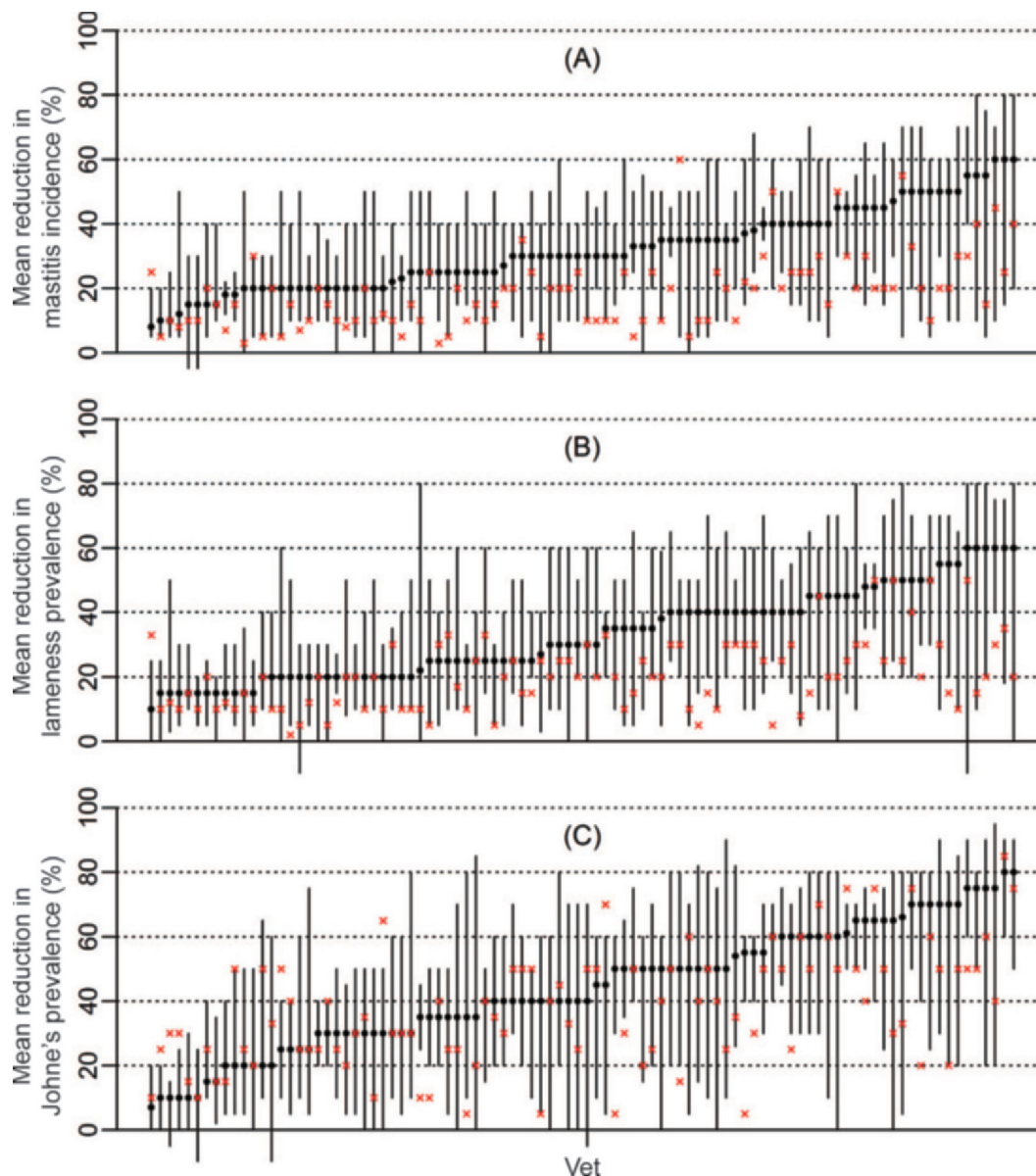
values:  $\lambda_m = \lambda_l = \lambda_j = 80\%$ ; lower quartile  $\lambda_m = \lambda_l = \lambda_j = 75\%$ ; upper quartile:  $\lambda_m = 88\%$ ,  $\lambda_l = \lambda_j = 90\%$ .

The distributions and summary statistics for veterinarians' estimates of the percentage of farmers who they believed would implement none of the disease interventions to any effective extent, despite initially agreeing to participate, are provided in Figure 4.

### Modeling the Implications of the Heterogeneity in Beliefs

Figure 5 predicts (using Equation [4]) how the veterinarian with the most pessimistic clinical expectation and the veterinarian with the most enthusiastic clinical expectation, regarding the lameness intervention, would update their expectations in light of new (synthetic) clinical trial data of different strengths. In Figure 5, the 4 plots (A to D) show the effects of clinical trials of different sizes, involving 20, 50, 100, and 200 farms in total. For each clinical trial, it was assumed that the data provided a point estimate,  $y$ , equal to a 30% mean reduction in lameness prevalence in favor of the intervention farms (over the control farms), and that the population variance,  $\sigma_d$ , was 25%. Hence, the only difference between the clinical trials is that they are increasing in size, and because the likelihood is encapsulating the information arising from the clinical trial, it can be seen that the evidence from the trial becomes stronger (i.e., narrower and taller) from plot (5A) through to plot (5D) as the trial increases in size, but remains centered at 30%.

On each plot in Figure 5, normal distributions have been fitted to the elicited values belonging to the most pessimistic and enthusiastic veterinarians and reflect their current clinical expectations (priors), as they stated during their interviews; that is, the "prior enthusiast" and "prior pessimist" curves. The plots show how the priors are updated when combined with the likelihood (i.e., the new data) using Bayes theorem. Hence the "updated enthusiast" and "updated pessimist" curves are the posterior distributions and represent what the



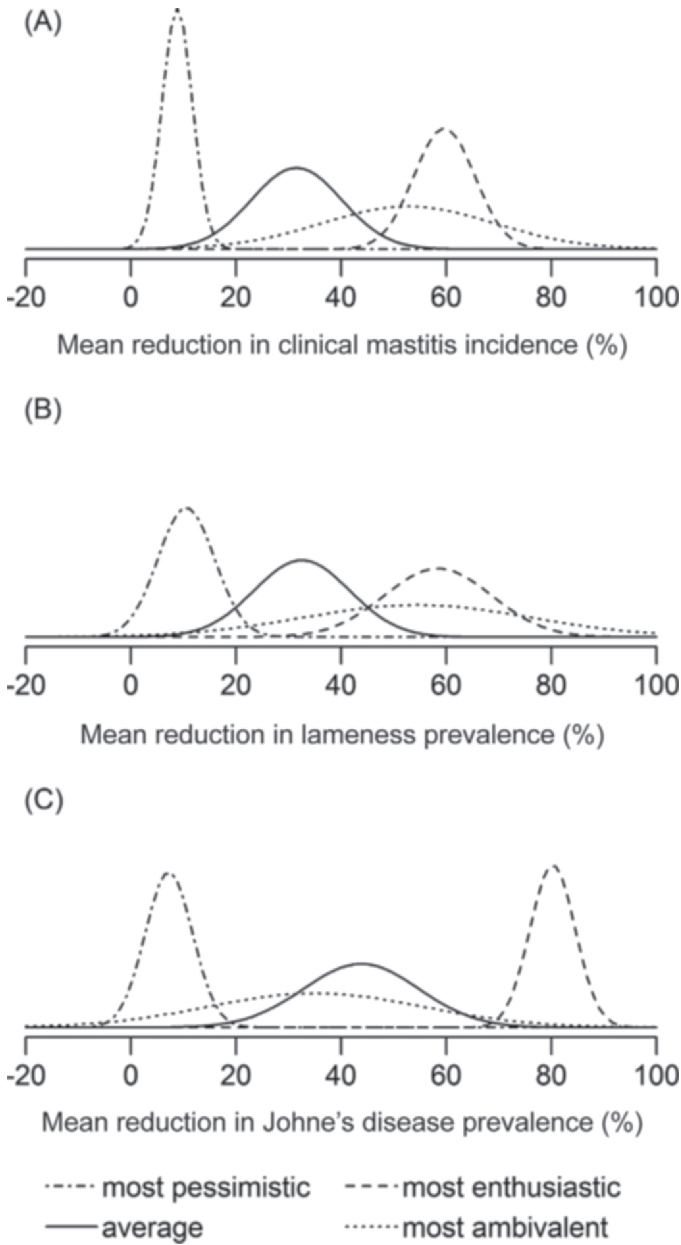
**Figure 1.** The clinical expectations of 94 veterinarians concerning the population mean reduction in disease achievable with (A) a 1-yr mastitis intervention, (B) a 3-yr lameness intervention, and (C) an 8-yr Johne’s disease intervention; median beliefs (dots) and plausible range (lines). Crosses (x) denote the minimum clinically worthwhile reduction stipulated by each veterinarian (i.e., the clinical demand). For each disease, veterinarians were ordered along the x-axis by their median belief (dots). Color version available in the online PDF.

most pessimistic and enthusiastic veterinarian are predicted to believe if they had been shown the clinical trial data, and updated their opinions coherently in accordance with Bayes theorem. To emphasize, the veterinarians were not shown any data during their interviews, and the posterior distributions shown in Figure 5 are predictions arising from the Bayesian models.

Across the 94 veterinarians, the median clinical demand for the 3-yr lameness intervention was a 20% mean reduction in prevalence, shown in Figure 5 by the

straight vertical line. Plot 5A shows that in light of a clinical trial involving 20 farms, the most pessimistic veterinarian is predicted to believe that there will only be an 0.11 probability (11% chance) that this clinical demand will be exceeded (area under the “updated pessimist” curve to the right of the vertical 20% line, as a proportion of the total area under the curve). As the size of the clinical trial increased, the most pessimistic veterinarian was predicted to become more convinced that this goal would be realized, such that with a trial

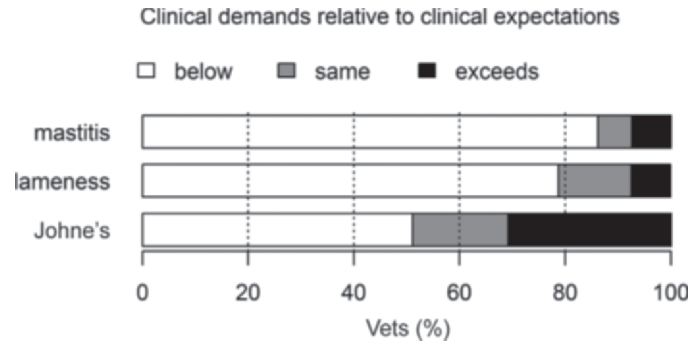




**Figure 2.** Normal probability distributions fitted to the elicited clinical expectations ( $\theta_d$ ) belonging to the most ambivalent, enthusiastic, pessimistic, and “average” veterinarians: (A) a 1-yr mastitis intervention, (B) a 3-yr lameness intervention, and (C) an 8-yr Johne’s disease intervention.

involving 200 farms, they were predicted to believe that there was a 0.92 probability that this clinical demand would be exceeded (plot 5D).

The null hypothesis is usually rejected in favor of the alternative hypothesis when the posterior distribution assigns a “sufficiently large” probability to  $\theta$  taking values greater than  $\theta^{min}$ ; because this probability ( $\lambda$ ) is often taken to be  $\geq 0.95$ , then, on this basis, a 200-farm clinical trial is not strong enough evidence to convince



**Figure 3.** Each veterinarian’s clinical demand,  $\theta_d^{min}$ , compared with their median (best guess) for the expected clinical outcome  $\theta_d$ .

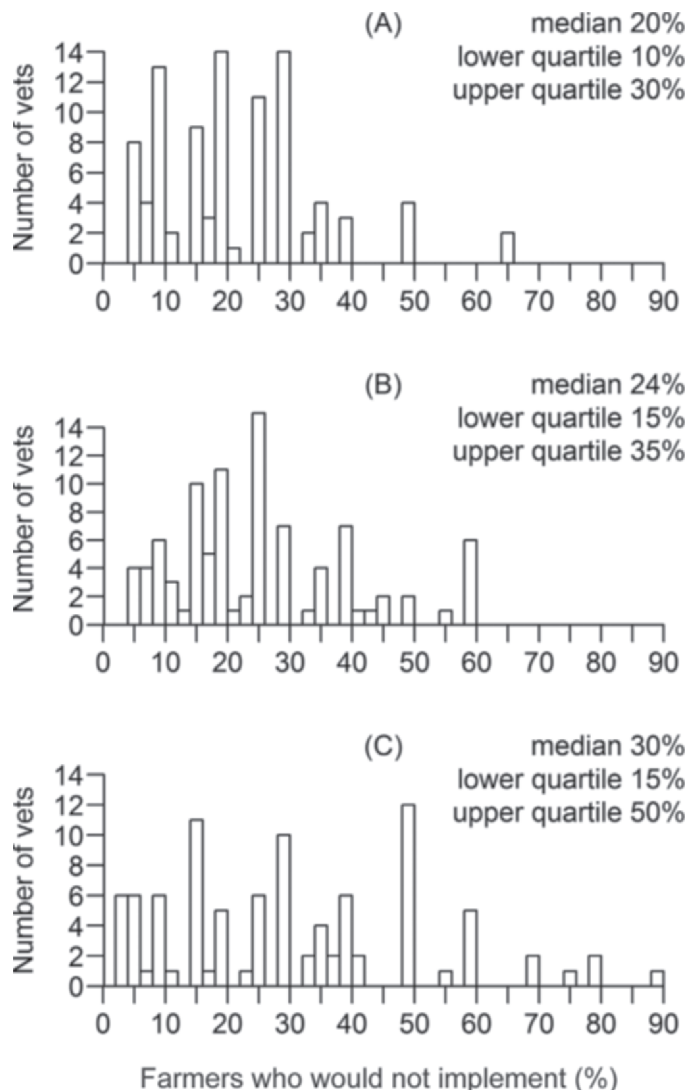
a “reasonable skeptic” (i.e., the most pessimistic veterinarian in this sample) that the intervention is clinically worthwhile. However, for the lameness intervention, the majority of veterinarians interviewed provided a value for  $\lambda_i$  that was  $< 0.90$  (upper quartile  $\lambda_i = 90\%$ ), and therefore on this basis, a clinical trial of 200 farms may be expected to be sufficiently convincing to the majority of veterinarians who participated in this study.

Figure 5 also shows how the most enthusiastic veterinarian is predicted to have their enthusiasm damped by the trial results. The most enthusiastic veterinarian was expecting a mean reduction in prevalence centered in the region of 60% (“prior enthusiast” curve), and for example, plot 5C shows that this is predicted to be reduced to a more confident belief that is centered around a 35% reduction (“updated enthusiast” curve) with a 100-farm trial.

Figure 5C also shows that in the light of the evidence originating from a 100-farm trial, the most pessimistic and enthusiastic veterinarians are predicted to have updated beliefs that are still almost entirely in disagreement (updated curves barely overlapping).

In plots A to D (Figure 5), the population standard deviation  $\sigma_i$  was assumed to be 25%. For a given size of clinical trial, taking  $\sigma_i$  to be 10%, results in a stronger likelihood (because there is less variation of farms around the mean value), and therefore affects the analysis. Plots E to H (Figure 5) show an identical analysis to plots A to D (Figure 5), but with  $\sigma_i$  assumed to be 10%. It shows that because the evidence from the trial is stronger, a 50-farm trial is predicted to convince (with a 0.99 probability) the prior pessimist that a 20% mean reduction in lameness prevalence will be realized.

A similar analysis to that presented in Figure 5 was run for the mastitis and Johne’s disease interventions; for the mastitis intervention, the median clinical demand was 20% (the same as for the lameness intervention) and for Johne’s disease, it was 35%. For the mastitis intervention, the synthetic trial data were



**Figure 4.** Veterinarians' estimates for the percentage of dairy farmers who, having agreed to undertake a structured disease intervention and had an initial visit, subsequently fail to implement any of the required changes to any successful extent: (A) a 1-yr mastitis intervention, (B) a 3-yr lameness intervention, and (C) an 8-yr Johne's disease intervention.

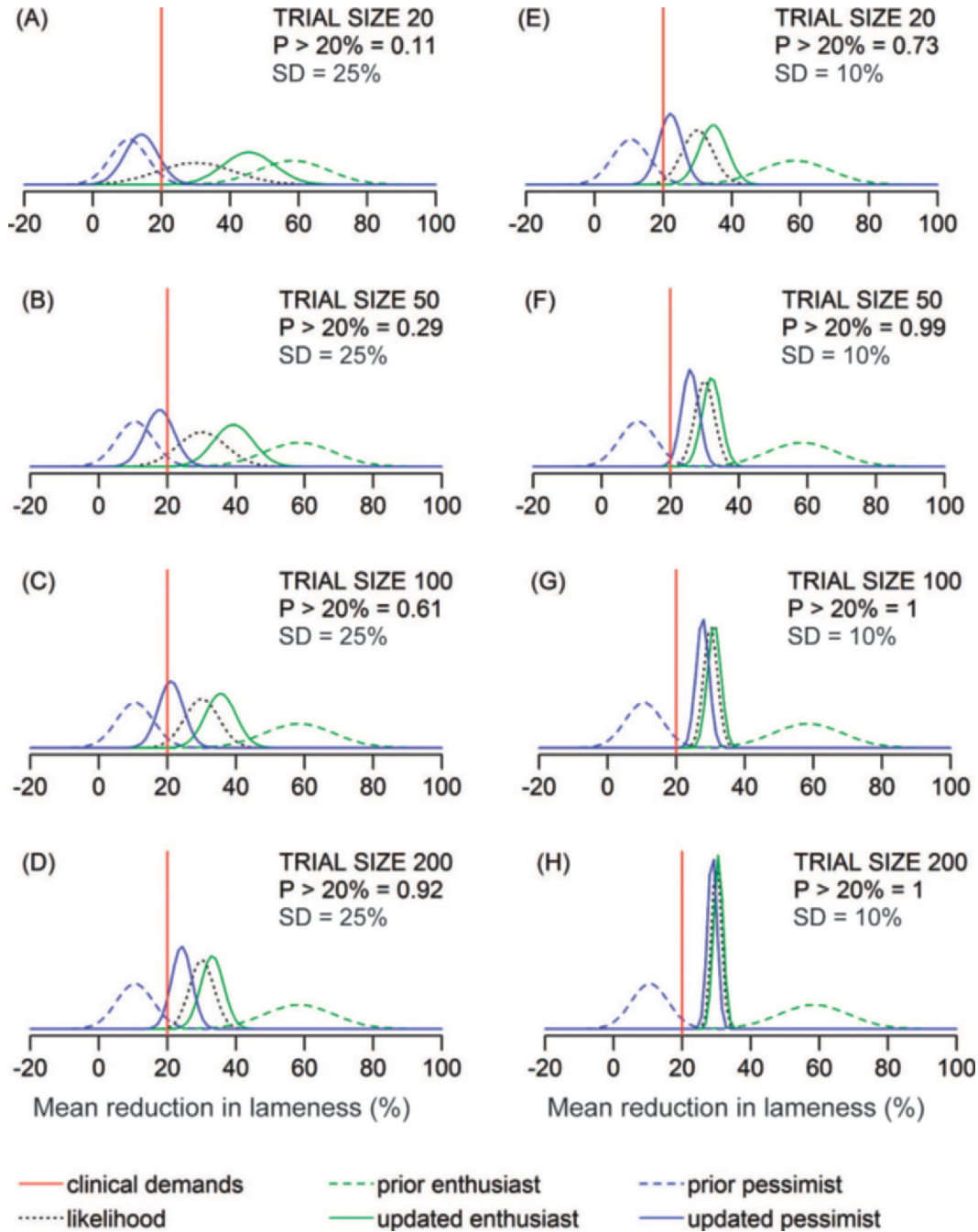
set to provide a point estimate of a 30% reduction in clinical mastitis incidence, and for Johne's disease, a 40% reduction in prevalence (these figures were the median expected reduction in disease elicited across the 94 veterinarians, Table 1). The analyses were run for population standard deviations of 10% and 25% and, for each disease, the predicted probability of exceeding the median clinical demand in light of the clinical trial data was calculated for the most pessimistic veterinarian. The results are presented in Table 2, where the lameness results from Figure 5 have also been included to allow comparison.

Figure 2 shows that the most pessimistic veterinarian for the mastitis intervention (plot A) was more confidently pessimistic (narrower and taller distribution) compared with the most pessimistic veterinarian for the lameness intervention (plot B); consequently, as Table 2 reveals, this veterinarian needs stronger evidence, according to Bayes theorem, to be convinced that the clinical demand will be exceeded. Thus, the most pessimistic veterinarian concerning mastitis intervention is predicted to be entirely unconvinced, even by a trial involving 200 farms (if the population standard deviation is indeed 25%) because they are predicted to give only a 12% chance that a 20% reduction in mastitis will be achieved. This is demonstrably different from the most pessimistic veterinarian concerning the lameness intervention, who would be substantially convinced by a 200-farm trial (0.92 probability), with the same population variance assumption.

**DISCUSSION**

The results revealed that markedly different clinical opinions currently exist; hence, for all 3 disease interventions, a major change in clinical expectations would be required to bring the most enthusiastic and most pessimistic veterinarians into clinical agreement. It may not be desirable for veterinarians to be completely united in their clinical beliefs, but broad agreement for the major endemic diseases is important both for the credibility of the profession and to provide a unified approach to disease control. The wide heterogeneity observed raises concerns over the extent to which a broadly consistent approach to controlling these endemic diseases is currently being achieved. This in turn supports the argument for the implementation of national disease control programs for nonstatutory endemic diseases. This is because national disease control programs help to provide a coordinated and consistent approach to disease control. As the variability between veterinarians increases, the potential gain derived from implementing a national disease control program also increases.

Explanation for the observed heterogeneity in clinical expectations is likely to be multifactorial. In this regard, the paucity of literature quantifying the effects of these interventions is worth noting. To our knowledge, only 1 randomized clinical trial has estimated the effect of a clinical mastitis intervention of the type described here (Green et al., 2007); this 1-yr intervention study found a 22% reduction in the proportion of cows affected with clinical mastitis on the intervention farms compared with the control farms. There appears to be an absolute lack of randomized clinical trials estimating the effects of lameness and Johne's disease interventions. More



**Figure 5.** A Bayesian analysis to predict how the most pessimistic and enthusiastic veterinarians regarding the 3-yr lameness intervention, would update their clinical opinions in light of new synthetic trial data of different sizes. For plots A to D, the population standard deviation (SD) was assumed to be 25%; for plots E to H, SD = 10%. In each plot, the trial data (likelihood) was assumed to provide a point estimate equal to a 30% mean reduction in lameness prevalence. “Updated enthusiast/pessimist” is the posterior distribution obtained by combining the relevant prior belief with the likelihood using Bayes theorem. “Clinical demands” (= 20%) is the median “minimum clinically worthwhile reduction” elicited across the 94 veterinarians. The probability ( $P$ ) that the “updated pessimist” assigns to this value being exceeded is given (area under the curve to the right of the straight vertical line). Color version available in the online PDF.

**Table 2.** The predicted probability of exceeding the median clinical demand (20% mastitis, 20% lameness, 35% Johne’s disease) for the most pessimistic veterinarian, when their prior clinical expectation is combined with synthetic trial data of different sizes using Bayes theorem; the trial data provided a point estimate of a 30% reduction in clinical outcome for mastitis and lameness, and 40% for the Johne’s disease intervention

Disease intervention	Population SD (%)	Size of randomized clinical trial (no. of farms)			
		20	50	100	200
Mastitis	10	0.02	0.43	0.97	1.00
	25	<0.001	0.001	0.01	0.12
Lameness	10	0.73	0.99	1.00	1.00
	25	0.11	0.29	0.61	0.92
Johne’s disease	10	<0.001	0.037	0.41	0.94
	25	<0.001	<0.001	<0.001	0.003

clinical trials to quantify the effects of intervention strategies to reduce the major endemic diseases of dairy cattle are required because there is a lack of knowledge in this area. In our data set, we did not find any statistically significant differences in veterinarians’ beliefs by the following factors: time spent working with dairy cows, possession of postgraduate qualifications or not, job status (assistant vs. partner), or sex of respondent. However, this is not to say that these factors are not associated with different opinions.

Notably, clinical expectations were elicited conditional on an assumption that the population mean change in disease would be zero in the absence of a disease intervention; it is necessary to know what veterinarians were basing their clinical expectations for a disease intervention against. It is acknowledged that some may believe this to be a larger assumption to make for the Johne’s intervention compared with the mastitis and lameness interventions. We made this assumption for the purpose of simplicity and in the absence of any data.

The results also showed considerable variation in veterinarians’ estimates for the percentage of farmers who would not effectively implement any of the intervention to any extent, despite initially agreeing to do so. The median estimates were nontrivial percentages: 20, 24, and 30% for the mastitis, lameness, and Johne’s disease interventions, respectively. This may, in part, reflect veterinarians’ practical experiences with maintaining farmer engagement with a disease intervention. Understanding ways to facilitate and support farmers to make the changes needed to implement a disease intervention is important.

“Clinical demands” in a herd health context could be interpreted in 4 different ways. First, the reduction the veterinarian would want to consider the intervention clinically worthwhile. Second, what the veterinarian thinks the farmer would want to hear to be convinced enough to implement the intervention. Third, what the farmer would actually demand to be prepared to implement the intervention, with its associated time,

financial, and labor investments. Fourth, the reduction the veterinarian needs to consider the intervention clinically worthwhile solely from the perspective of their patient; that is, the cow. In a sense, the latter is analogous to the scenarios for which clinical demands have been elicited and reported in human medicine, where the doctor is faced with deciding what is best for their patient, given the treatment will be freely available and the patient will almost certainly take it.

In this study, it was stipulated that clinical demands related to the first point only. However, it may be difficult to compartmentalize these issues. Thus, part of the variation in veterinarians’ clinical demands may have related to how they chose to weigh the vested interests of all stakeholders; for example, veterinarians who estimated a very small clinical demand may have given greatest weight to the interests of the cow, or to the environment, whereas those who gave the highest clinical demands may have given more weight to the financial cost-effectiveness for the farmer. Another possible explanation for the variation in clinical demands may concern differences between veterinarians in terms of their beliefs about the economic benefits of controlling the different diseases. This may, in turn, reflect differences in undergraduate teaching between universities.

It is important to note that the Bayesian analysis presented here has modeled the effect of new clinical trial data on veterinarians’ current opinions, assuming they would update their beliefs logically in accordance with Bayes theorem. This means that as the size of the trial increases, the mathematical model predicts that veterinarians will become more convinced because the evidence from larger trials is stronger. However, controversy currently exists in the literature over the extent to which people are actually able to logically value, interpret, and revise their opinions when presented with new information (Tversky and Kahneman, 1974; Kynn, 2008; Westover et al., 2011). Recent findings suggest that people can update their beliefs in a Bayesian manner, but how the task is framed is important (Giger-

enzer and Edwards, 2003; Gigerenzer and Hoffrage, 2007). In reality, it may be challenging for veterinarians to differentiate between evidence of varying strength and quality. Other factors, such as the perceived reputation of the scientists conducting the work, the final concluding remarks of the paper, or the point estimate alone (not in combination with the confidence or credible interval) may be more influential. More research is needed to establish how logically veterinarians update their beliefs about clinical parameters, using information published in veterinary journals. This is important to ascertain, because investing money in larger clinical trials to convince skeptical clinicians will not provide a return unless they revise their beliefs in at least a somewhat logical manner. And, of course, due consideration should be given to all the other factors that influence veterinary beliefs and other approaches employed as necessary. This could include communication strategies, as well as helping veterinarians to interpret new information appropriately, and facilitating and improving veterinary education (both undergraduate and postgraduate).

It is also worth noting that, in human medicine,  $\lambda$  has not been elicited from doctors, even though doctors' clinical demands have been elicited. A choice for  $\lambda$  that is less than 0.95 may not be generally acceptable to the scientific community, because it may not be regarded as robust enough evidence. However, it is a subjective choice and clinically useful information may go unreported if only clinical trials achieving  $\lambda \geq 0.95$  are published.

A potential limitation of this study is the repeatability of the results at a later date in the absence of any genuine change in veterinarians' beliefs. There is no literature reporting the repeatability of the method. However, during the interviews, extra clarification was given if required. This reduced misunderstanding, which is one reason why different results may be obtained on different dates.

#### ACKNOWLEDGMENTS

Our thanks go to all the veterinarians who participated in this study. H. M. Higgins was funded by a Wellcome Trust Research Fellowship [087797/Z/08/Z] and hosted by the University of Nottingham, UK.

#### REFERENCES

- Gigerenzer, G., and A. Edwards. 2003. Simple tools for understanding risks: From innumeracy to insight. *BMJ* 327:741–744.
- Gigerenzer, G., and U. Hoffrage. 2007. The role of representation in Bayesian reasoning: Correcting common misconceptions. *Behav. Brain Sci.* 30:264–267.
- Green, M. J., W. J. Browne, L. E. Green, A. J. Bradley, K. A. Leach, J. E. Breen, and G. F. Medley. 2009. Bayesian analysis of a mastitis control plan to investigate the influence of veterinary prior beliefs on clinical interpretation. *Prev. Vet. Med.* 91:209–217.
- Green, M. J., K. A. Leach, J. E. Breen, L. E. Green, and A. J. Bradley. 2007. National intervention study of mastitis control in dairy herds in England and Wales. *Vet. Rec.* 160:287–293.
- Kalton, G. 1987. *Introduction to Survey Sampling*. Sage Publications, London, UK.
- Kynn, M. 2008. The 'heuristics and biases' bias in expert elicitation. *J. R. Stat. Soc. A Stat. Soc.* 171:239–264.
- Nelder, J. A., and R. Mead. 1965. A simplex algorithm for function minimization. *Comput. J.* 7:308–313.
- O'Hagan, A., C. E. Buck, A. Daneshkhan, J. R. Eiser, P. H. Garthwaite, D. J. Jenkinson, J. E. Oakley, and T. Rakow. 2006. *Uncertain Judgements: Eliciting Experts' Probabilities*. Wiley, Chichester, UK.
- Parmar, M. K. B., D. J. Spiegelhalter, and L. S. Freedman. 1994. The CHART trials: Bayesian design and monitoring in practice. *Stat. Med.* 13:1297–1312.
- R Core Team. 2012. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
- Spiegelhalter, D. J., K. R. Abrams, and J. P. Myles. 2004. *Bayesian Approaches to Clinical Trials and Healthcare Evaluation*. Wiley, Chichester, UK.
- Tversky, A., and D. Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185:1124–1131.
- Westover, M. B., K. D. Westover, and M. T. Bianchi. 2011. Significance testing as perverse probabilistic reasoning. *BMC Med.* 9:20.

## APPENDIX

### Standard Elicitation Script

#### *Clinical Expectations: Mastitis Intervention.*

My first question is about a clinical mastitis intervention and how successful you think it might be when implemented on a large number of dairy farms.

Please imagine that all the dairy farms in England are visited by their own local veterinarian today, specifically to discuss their current clinical mastitis situations. As a result, we have identified a large number of farms, for which both the farmer and their local vet have agreed there is a clinical mastitis problem and they have decided that the farmer will go ahead with a structured clinical mastitis intervention with the help of their local vet over the next year.

By a structured clinical mastitis intervention over the next year, I mean the following:

Initially, the vet tries to identify the main cause (s) of the mastitis problem. The vet carries out some strategic laboratory testing, assesses any available milk recording data and any other relevant farm data that is available (e.g., clinical mastitis records). The vet then makes a specific visit to walk round the farm and assess the risks that may be contributing to the problem. The vet quantifies the risks using our current best evidence for clinical mastitis. As a result the vet produces a list of recommendations, prioritizes them and discusses them with the farmer. Over the next year, the vet revisits the farm at least every quarter, reassesses any data available and the on-farm risks, and modifies the recom-

mentations as necessary. These are discussed with the farmer, and the vet supports and encourages the farmer to implement the required changes. This approach, over a year, is what I mean by a “structured clinical mastitis intervention.” Note: the recommendations can be anything, including major farm reinvestment, but the vet is realistic and will not make suggestions that will outstrip the current labor and capital resources of this particular farm.

When we think about this large number of farms, despite the fact that the farmers have all decided to implement a mastitis intervention and they have all had a first initial visit by their vet, nevertheless some of them may not actually implement any of the recommendations either at all or sufficiently well over the year, such that their intervention had no chance of having any effect at all simply due to farmer compliance.

Question: What is your best estimate for the percentage of farms that, in reality, would fall into this category?

Now we are going to think about how successful the mastitis interventions will be on average, across all the farms. However, when doing this we are going to exclude the farms we have just talked about; that is, those who will do nothing or so little that their interventions have no chance of working at all.

[Scale:] We will be working in terms of the percentage reduction (or percentage increase) of the original mastitis problem. By this I mean that in the year before we start the mastitis intervention, we have measured (for each farm) the number of clinical cases of mastitis per 100 cows per year. And in the year after the intervention, we will again measure the number of clinical cases of mastitis per 100 cows per year. So for example, a 50% reduction in the farms original clinical mastitis problem means that a farm with 120 clinical cases per 100 cows occurring in the year before the intervention would have 60 clinical cases per 100 cows in the year after the intervention. Do you have any questions about this?

If we don't intervene, we will assume that the mastitis problems on the farms will not change *on average*.

*Range:* It is natural that you will have some uncertainty about the average effect of the mastitis intervention. Therefore, please can you give me a *plausible range* for the average effect, as a percentage increase or reduction of the original clinical mastitis problem? That is, tell me the worst (X) and the best (Y) possible answer for the average, such that you think it exceedingly unlikely that the actual answer could turn out to fall outside of this range?

Take time to think about your answer. Feel free to ask any questions or talk it through.

So, your plausible range for the average effect of a mastitis intervention is between X and Y. Please take time to consider if this properly reflects the amount of uncertainty you have in what the average effect of this mastitis intervention might be? You can revise your answer if you wish. Note that the more uncertainty you have, the wider your range should be.

*Median:* Given your range, please tell me your “best guess” (median value, M) for the average effect of the clinical mastitis intervention. This should be a value such that you believe it is just as likely that the true answer will turn out to fall somewhere between your worst possible value (which was X) and this number, and this number and your best possible value (which was Y). Please let me know if you have any queries about this.

*Lower and upper quartiles:* Please can you give me a value for your lower quartile for the average effect. This is a number such that you think it is equally likely that the actual answer will fall between your worst possible value (X) and this value, and this value and your best guess (which was M). Please let me know if you have any queries about this.

Please can you give me a value for your upper quartile for the average effect. This is a number such that you think it is equally likely that the actual answer will fall between your best guess (which was M) and this value, and this value and your best possible answer (which was Y). Please let me know if you have any queries about this.

*Fitted probability distribution:* [Researcher fits probability distribution to the 5 values (range, median, lower and upper quartile) and produces a graph using SHELF software. The graph is shown to veterinarian (with explanation). The fitted probabilities are compared with the elicited values. Researcher feeds back 2 implied probabilities (10th and 90th percentiles) from the fitted distribution].

Does the shape and distribution of this graph represent what you believe? (*yes/no*)

[If yes, proceed. If no, invite them to vary some values to fit others that are believed to be more pivotal. Re-fit a new distribution and feedback until the veterinarian is satisfied that the distribution is an accurate reflection of their belief.]

***Clinical Demands: Mastitis Intervention.*** So far, we've discussed what we *expect* might actually happen if we were to implement a clinical mastitis intervention on a large number of dairy farms, in terms of the average effect. For my next question, I just want a single answer and it has nothing to do with what we've said previously about what we are expecting to happen!

We are interested in what you *want* the average reduction in clinical mastitis to be, in order for you to feel that the intervention itself typically provides what you consider a clinically worthwhile and important reduction and therefore should be implemented on a large number of farms.

Rolling out an intervention onto thousands of farms has costs to the whole industry but also benefits. This question is about weighing up all the costs and all the benefits.

If I told you that the average effect of a mastitis intervention is an increase in the original disease problem, say 10% on average, we wouldn't consider it a good intervention! Similarly if I told you that the average effect of a mastitis intervention is zero change on average, we wouldn't consider it a good intervention!

For this 1-yr clinical mastitis intervention, please tell me, what is the minimum that you would want the average reduction in clinical mastitis to be in order for you to feel that the mastitis intervention itself is clinically useful when implemented on a large number of farms? This can just be a single answer (or you can give me a range if you prefer).

In reality, all researchers can do is *estimate* the average effect of a mastitis intervention when implemented on a large number of farms, by carrying out a clinical trial. They would implement the mastitis intervention on a small number of farms and compare the average reduction observed to a small number of control farms. Hence, researchers cannot prove to us with *absolute certainty* that the average reduction will definitely be greater than what you wanted. There will always be some uncertainty.

$\lambda$ : My next question is about how certain you personally would need to be (as percentage chance), that the average reduction is going to be at least what you said you wanted, in order that you personally would be convinced that the mastitis intervention is useful and worthwhile implementing on many farms, based on the results of a clinical trial alone.

By this I mean that if you considered the evidence that some researchers had produced, and as a result you are now sure enough that the average reduction is going to be at least what you wanted, then you would be convinced that the mastitis intervention is clinically useful. How sure would you personally need to be (as a percentage chance) in order for you to be convinced that the mastitis intervention is clinically useful and worthwhile implementing on many farms?

My next 2 questions are identical in format to this clinical mastitis intervention question, except that they relate to a lameness intervention and a Johne's disease intervention. Let's start with lameness first.

### **Lameness Intervention**

Repeat the above text, but substitute "lameness" for "clinical mastitis" and "3 years" for "1 year." For the description of the structured lameness intervention, the text is modified to include mobility (locomotion) scoring the milking cows. For the scale, substitute the following text:

We will be working in terms of the percentage reduction (or percentage increase) of the original lameness problem. So at the start of the lameness intervention, we mobility (locomotion) score the entire herd for lameness and hence we know, for every farm (reliably) what percentage of the herd is lame. Note: by "lame" we mean observably lame in one or more limbs, with any severity of limp. At the end of the 3-yr intervention, we will again mobility (locomotion) score the entire herd for lameness and hence we know, for every farm (reliably) what percentage of the herd is lame. So, for example, a 50% reduction in the farm's original lameness problem means that a farm with 40% of the herd lame has 20% of the herd lame after 3 yr.

### **Johne's Disease Intervention**

Repeat the above text, but substitute "Johne's disease" for "clinical mastitis" and "8 years" for "1 year." For the scale, substitute the following text:

We will be working in terms of the percentage reduction (or percentage increase) of the original Johne's problem. So, we blood sample all the adult cattle (over 24 mo old) and hence we will know, for every farm, what percentage of the adults are blood test positive for Johne's at the start. After the end of the 8-yr intervention, we will again blood test the adults and hence we know, for every farm, what percentage of the herd is blood test positive for Johne's at the end. So for example, a 50% reduction in the farms original Johne's problem after an 8-yr Johne's intervention means that a farm with 40% of adults testing positive for Johne's at the start has 20% of adults testing positive after 8 yr.