On Quantifying Expert Opinion about Multinomial Models that Contain Covariates

Fadlalla G. Elfadaly

The Open University, UK & Cairo University, Egypt

and

Paul H. Garthwaite

The Open University, UK

Abstract: This paper addresses the task of forming a prior distribution to represent expert opinion about a multinomial model that contains covariates. The task has not previously been addressed. We suppose the sampling model is a multinomial logistic regression and represent expert opinion about the regression coefficients by a multivariate normal distribution. This logistic-normal model gives a flexible prior distribution that can capture a broad variety of expert opinion. The challenge is to (i) find meaningful assessment tasks that an expert can perform and which should yield appropriate information to determine the values of parameters in the prior distribution, and (ii) develop theory for determining the parameter values from the assessments. A method is proposed that meets this challenge.

The method is implemented in interactive user-friendly software that is freely available. It provides a graphical interface that the expert uses to assess quartiles of sets of proportions and the method determines a mean vector and a positive-definite covariance matrix to represent the expert's opinions. The chosen assessment tasks yield parameter values that satisfy the usual laws of probability without the expert being aware of the constraints this imposes. Special attention is given to feedback that encourages the expert to consider his/her opinions from a different perspective. The method is illustrated in an example that shows its viability and usefulness.

Keywords and phrases: Elicitation method, Interactive graphical software, Logistic normal prior, Multinomial logit model, Multinomial logistic model, Prior distribution.

1. Introduction

The purpose of an elicitation method is to help an expert quantify his or her opinions in a mathematically useful form. The expert performs meaningful assessment tasks and a probability distribution is formed from the elicited assessments. In a Bayesian analysis this distribution would be the prior distribution, but the distribution might also be used to express uncertainty about parameters in a decision analysis, or to communicate an expert's opinion succinctly. Different elicitation methods are required for different situations depending, in particular, on the sampling model and the model used to represent opinion.

Here we consider the important case where the sampling model is a multinomial model with covariates.

Multinomial models arise when there is a set of complementary and mutually exclusive categories and each observation falls into one of these categories. In the simplest case, the probability of falling into any specified category is the same for each observation and observations are independent of each other. Then observations follow a multinomial distribution with, say, probability p_i that an observation falls in the ith category. A number of papers have addressed the task of quantifying an expert's opinions about the p_i so as to form a prior distribution for this situation. Early work focused on representing expert opinion by a Dirichlet distribution, which is the natural conjugate prior distribution for multinomial sampling; reviews of this work may be found in Garthwaite *et al.* (2005) and O'Hagan *et al.* (2006). Modelling opinion by a Dirichlet distribution has continued to attract attention (Zapata-Vázquez *et al.*, 2014; Evans *et al.*, 2017), but methods for eliciting more flexible prior distributions have also been proposed. Elfadaly and Garthwaite (2013) give a method for quantifying expert opinion as both a Dirichlet distribution and a Connor-Mosimann distribution, and give an example to illustrate the greater flexibility of the latter. More flexible models have also been obtained through the use of copulas. Elfadaly and Garthwaite (2017) use a Gaussian copula prior distribution to model expert opinion and Wilson (2017) uses a D-vine copula. With the latter copula, specification of the multivariate prior distribution can be separated into assessments about univariate marginal distributions and unconditional bivariate copulas, thus simplifying the assessment tasks that the expert performs (Wilson, 2017). Recent applications in which experts have assessed prior distributions for a multinomial model include Németh *et al.* (2017), who quantify the opinions of three experts about treatments for schizophrenia, representing their opinions by Dirichlet distributions, and Wilson *et al.* (2017), who elicited expert opinion about disease progression of untreated melanoma, representing opinions by Connor-Mosimann distributions.

The task of quantifying opinion about covariates is of importance and has merited significant amounts of time from experts. In clinical epidemiology, for instance, Miettinen *et al.* (2008) elicited opinions of 22 experts from three separate panels, who filled out questionnaires with hypothetical patient profiles for the diagnosis of pneumonia. Bergstraesser *et al.* (2014) used case vignettes and fed assessments from 37 experts into proportional-odds logistic regression so as to evaluate a paediatric palliative screening scale. The aims in these works was to obtain point estimates of regression coefficients for a binomial response, but their approaches could be extended to a multinomial response, and their methods could form the initial step in constructing a prior distribution. The work illustrates the growing use that is being made of carefully quantified expert opinion.

Previous work has not addressed the task of eliciting a prior distribution for the common situation where covariates influence the values of the p_i (the multinomial probabilities), other than in the special case where there are only two categories and the sampling model reduces to logistic regression

(Bedrick *et al.*, 1996; Garthwaite *et al.*, 2013). Here we develop an elicitation method for quantifying opinion about a multinomial model when there are covariates and any number of categories. We take a multinomial logistic regression as the sampling model, which is the most common form of multinomial model with covariates. A linear regression links the covariates to functions of the multinomial probabilities (Agresti, 2002).

While the Dirichlet distribution is the standard prior distribution for a multinomial sampling model when there are no covariates, using the logistic normal distribution as the prior distribution has also been considered (see, for example, O'Hagan and Forster (2004), Sections 12.14 to 12.19). However, ways of choosing its parameters to model an expert's opinions have not been proposed and in this paper we first develop an elicitation method to rectify this deficiency. It does not seem to have been suggested before, but it is easy to expand the scope of the logistic normal prior so that it is a suitable prior distribution for a multinomial logistic regression model. We extend our elicitation method so it elicits the parameters of the resulting logistic normal prior.

Eliciting parameters of multivariate distributions is not, in general, an easy task, especially if variates are not independent (O'Hagan *et al.*, 2006). The assessed subjective distribution must satisfy the usual laws of probability and, with multinomial models, these include additional requirements because the probabilities of the different categories must sum to one. The elicitation method proposed here uses assessment tasks and a task structure that leads to coherent assessments without the expert having to be conscious of coherence constraints. The method has been implemented in interactive graphical user-friendly software that aims to help the expert quantify his/her opinion effectively and as painlessly as possible. (We will use he/his rather than she/her because the expert in the example we give is male.)

In Section 2 we consider multinomial models that do not contain covariates; we define the logistic normal prior model and consider its assumptions. The required assessments and their use to elicit the prior and obtain feedback is given in Section 3. In Section 4 we consider the case where the sampling model does contain covariates and extend our elicitation method to model opinion about a logistic normal distribution. An application in which the elicitation method was used is given in Section 5 and some concluding comments are made in Section 6.

2. Elicitation for a multinomial sampling model

The flexibility of a logistic normal distribution makes it an attractive means of representing expert opinion about a multinomial sampling model. Here we quantify opinion as an *additive* logistic normal prior distribution. Different forms of multivariate logistic transformations are given in the literature, but this form is the most widely used [see, for example, Aitchison (1986)].

We suppose there are k categories and that each observation in a sample belongs to exactly one category. Under a multinomial sampling model, observations are independent of each other and the probability that an observation belongs to any specified category is the same for each observation. Let p_i denote the probability that an observations is in the ith category ($i = 1, \ldots, k$) and put $\mathbf{p} = (p_1, \ldots, p_k)$. Also, let $\boldsymbol{Y}_{k/1} = (Y_2, \ldots, Y_k)'$, where the first category is suppressed. The additive logistic transformation from $\boldsymbol{Y}_{k/1}$ to $\mathbf{p}$ is defined by

$$p_1 = \frac{1}{1 + \sum_{j=2}^{k} \exp(Y_j)}, \qquad p_i = \frac{\exp(Y_i)}{1 + \sum_{j=2}^{k} \exp(Y_j)}, \quad i = 2, 3, \ldots, k. \tag{1}$$

It follows that $\sum_{i=1}^{k} p_i = 1$, which will be referred to as the *unit sum constraint*. The inverse transformation is

$$Y_i = \log(p_i/p_1), \quad i = 2, 3, \ldots, k, \tag{2}$$

and we let

$$r_i = p_i/p_1, \quad i = 2, 3, \ldots, k \tag{3}$$

denote the probability ratios. The vector $\mathbf{p}$ is said to have a logistic normal distribution if

$$\boldsymbol{Y}_{k/1} \sim \mathrm{MVN}(\boldsymbol{\mu}_{k/1}, \boldsymbol{\Sigma}_{k/1}) \tag{4}$$

and $\boldsymbol{\Sigma}_{k/1}$ is positive-definite. We assume the prior distribution takes this form.

In the above formulae, p_1 seems to be treated differently to the other p_i variables. However, the additive logistic normal distribution is permutation invariant. That is, whatever the ordering of the elements of the vector $\mathbf{p}$, the density function given above is invariant. For a theoretical proof of this property see Aitchison (1986). Hence, any order of the elements of $\mathbf{p}$ can be considered. With the representation given in equations (1)–(4), p_1 is referred to as the fill-up variable and the first category as the fill-up category. Its choice is arbitrary, in principle, though we believe that our assessment tasks are easier for the expert if categories are ordered so that the most common category is the first category. (In multinomial logistic regression, the standard choice for the fill-up category is the first or last category.)

O'Hagan and Forster (2004, Sections 12.15 to 12.18) give the prior-posterior analysis for a $(k-1) \times 1$ vector Φ whose components are a set of log contrasts. Their results can be applied to the prior distribution that we elicit by equating Φ to $\boldsymbol{Y}_{k/1}$ and taking the distribution of $\boldsymbol{Y}_{k/1}$ as the prior distribution of Φ.

3. Elicitation method without covariates

An expert should be asked only meaningful questions, so he will not be asked directly about the $Y_i = \log(r_i) = \log(p_i/p_1)$, as the Y_i are not easily interpretable. However, his assessments must provide information about the Y_i. Our approach is to ask him to focus on two categories at a time, one of which is the fill-up category. Thus, for the ith category we ask the expert to consider all the items that fall in the ith category or the first category. The expert is then asked about the proportion of those items that fall in the ith category relative to the proportion that falls in the first category. He makes his assessments on a bar chart using a graphical interface (see Fig. 1) and may think about this proportion as the volume in the lower (blue) part of a box relative to the volume in the upper (orange) part of the box. (Labels naming colours have been added to figures in this paper to assist understanding with black-and-white copies.) That is, he focuses on r_i. Alternatively, if he finds it easier he can consider $p_i/(p_1+p_i)$, the proportion of the complete bar that is blue. The expert assesses medians and quartiles (medians are being assessed in Fig. 1) and in both cases we treat his assessments as values of r_i.

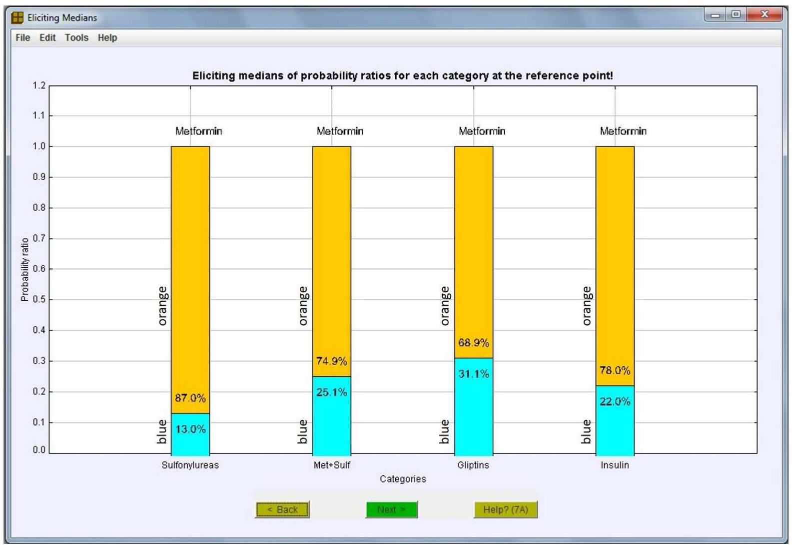

FIG 1. *Assessing medians of probability ratios*

The task of assessing ratios of proportions (such as r_i) is similar to the tasks of assessing ratios of probabilities, odds ratios and likelihood ratios. Assessing point estimates of these ratios has been used in elicitation methods for many years. The assessment of likelihood ratios forms the basis of Edward's probability information processing system (Edwards, 1962) and several associated experiments have found it better to elicit likelihood ratios rather than likelihoods (Edwards *et al.*, 1968; Kaplan and Newman, 1966; Schum *et al.*, 1966). Assessment of ratios of probabilities is advocated in Por and Budescu (2017), where references to a number of earlier studies may be found. Both point and interval estimates have been assessed for odds ratios in medical applications, notably in the context of trial design for a rare disease (Hampson *et al.*, 2015), sensitivity analysis (Shardell *et al.*, 2010), and missing data that is not missing at random (Shepherd *et al.*, 2008).

As the log transformation from r_i to Y_i is strictly monotonic increasing, quantile assessments of r_i provide the corresponding quantiles of Y_i. We will make substantial use of this approach and have found that it gives a viable means of eliciting opinion about Y_i. Where possible, feedback is also given to the expert in the form of unconditional marginal quantiles of each probability p_i $(i = 1, \ldots, k)$, so as to improve the quality of the elicitation method (*c.f.* Section 3.3).

3.1. Eliciting the mean vector $\boldsymbol{\mu}_{k/1}$ using median assessments

For $i = 2, \ldots, k$, let m_i^* denote the median of $r_i = p_i/p_1$ and let m_i denote the median of Y_i. (With this notation, deletion of an asterisk 'transforms' a median of r_i to a median of Y_i.) We use $C(\,.\,)$ to denote the centre (median) of the quantity in braces.

Here we relate $\boldsymbol{\mu}_{k/1} = (\mu_2, \ldots, \mu_k)'$ to the m_i^*. Using the symmetry of the distribution of Y_i $(i = 2, \ldots, k)$, together with equations (2)–(4), we have that $\mu_i = E(Y_i) = C(Y_i) = C[\log(r_i)]$. Hence, the components of $\boldsymbol{\mu}_{k/1}$ are given by

$$\mu_i = \log(m_i^*) \quad \text{for } i = 2, 3, \cdots, k. \tag{5}$$

Using this approach, $\boldsymbol{\mu}_{k/1}$ is assessed in a simple direct way. The expert assesses unconditional medians, m_i^*, of r_i for each category. These assessments do not need to be constrained at all. However, based on these assessments, unconditional marginal medians of p_i that sum to 1 are approximated and presented as feedback to the expert who is invited to revise them until they form an acceptable representation of his opinion. More than one iteration is seldom needed. Our approach to approximate the marginal medians of p_i $(i = 1, \ldots, k)$ is discussed in Section 3.3.

The expert uses interactive graphics to give his median assessments, m_i^*, of the probability ratio r_i. The screen-shot given earlier in Fig. 1 illustrates the process. The expert has assessed his median of the probability ratio for each of the four categories relative to the first category. He did this by clicking the mouse-pointer on each (blue/orange) stacked bar to divide it in two according to his median assessment for the ratio of p_i, represented by the lower (blue) area of each stacked bar, to p_1, which is represented by the upper (orange) area of each bar. The first category forms the lower part of each bar and so it does not have its own bar. Fig. 1 and all the subsequent figures in this paper are from an example reported in Section 5, in which an expert quantified his opinion about the potential cost of a new drug in Malta.

3.2. Eliciting the variance matrix $\boldsymbol{\Sigma}_{k/1}$ using conditional assessments

Kadane *et al.* (1980) give a method of eliciting the "spread" matrix of a vector-variate that has a multivariate-t distribution. We adopt their method with minor modifications and with the degrees of freedom (df) set equal to infinity, so that the multivariate-t distribution becomes an MVN distribution and the spread matrix becomes a variance-covariance matrix. Kadane *et al.* (1980) use $S(.)$ to denote the spread of a t distribution but here $S(.)$ will be the variance of a normal distribution.

Our procedure for determining $\mathbf{\Sigma}_{k/1} = S(\boldsymbol{Y}_{k/1})$ requires the following quantities.

1. $C(Y_i)$ for $i = 2, \ldots, k$.
2. $S(Y_2)$ and $S(Y_{i+1} \,|\, Y_2 = y_2^\diamond, \ldots, Y_i = y_i^\diamond)$ for $i = 2, \ldots, k$.
3. $C(Y_j \,|\, Y_2 = y_2^\diamond, \ldots, Y_{i-1} = y_{i-1}^\diamond, Y_i = y_i^0)$ for $i = 2, \ldots, k-1$; $j = i+1, \ldots, k$, where $y_2^\diamond, \ldots, y_{i-1}^\diamond$ and y_i^0 are chosen as detailed below.

The purpose of Section 3.1 was to elicit $\boldsymbol{\mu}_{k/1}$, whose ith component is $C(Y_i)$. Hence the quantities in 1. have already been assessed.

3.2.1. Assessing $S(Y_2)$ *and* $S(Y_{i+1} \,|\, Y_2 = y_2^\diamond, \ldots, Y_i = y_i^\diamond)$

To obtain $S(Y_2)$ the expert is asked to assess the lower and upper quartiles for the probability ratio $r_2 = p_2/p_1$, which we denote by L_2^* and U_2^*. Then the upper and lower quartiles of Y_2 are

$$L_2 = \log(L_2^*) \quad \text{and} \quad U_2 = \log(U_2^*), \tag{6}$$

since the log transformation from r_2 to Y_2 is monotonic strictly increasing. As the interquartile range of a standard normal distribution is 1.349, we put

$$S(Y_2) = \left[\frac{U_2 - L_2}{1.349}\right]^2. \tag{7}$$

So as to define the assessment tasks involving conditional spreads, values must be specified for $y_2^\diamond, \ldots, y_{k-1}^\diamond$. Kadane *et al.* (1980) set them equal to the expert's assessment of upper quartiles. We prefer to give them values from the middle of the expert's subjective distribution, when possible, because of the unit sum constraint. Otherwise, conditioning can imply that some categories have a tiny probability of occurring, even though the expert initially thought their occurrence is quite likely. For this reason we choose each $y_i^\diamond$ so that the corresponding condition on r_i is $r_i = m_i^*$. Hence, $y_i^\diamond = m_i$ for $i = 2, \ldots, k-1$. To shorten notation, we replace $S(Y_{i+1} \,|\, Y_2 = m_2, \ldots, Y_i = m_i)$ by $S(Y_{i+1} \,|\, Y_2, \ldots, Y_i)$ as, for any MVN distribution, the values actually specified for $Y_2, \ldots, Y_i$ do not affect $S(Y_{i+1} \,|\, Y_2 \ldots, Y_i)$.

To obtain $S(Y_{i+1} \,|\, Y_2, \ldots, Y_i)$ the expert assesses quartiles of r_{i+1} conditional on $r_j = m_j^*$ for $j = 2, \ldots, i$. Let L_{i+1}^* and U_{i+1}^* denote the expert's lower and upper quartile assessments. The log transformation from r_{i+1} to Y_{i+1} is monotonic strictly increasing. Also, $r_2, \ldots, r_i$ determine $Y_2, \ldots, Y_i$. Hence,

$$L_{i+1} = \log(L_{i+1}^*) \quad \text{and} \quad U_{i+1} = \log(U_{i+1}^*) \tag{8}$$

are the lower and upper quartiles of $Y_{i+1} \,|\, Y_2, \ldots, Y_i$. We put,

$$S(Y_{i+1} \,|\, Y_2, \ldots, Y_i) = \left[\frac{U_{i+1} - L_{i+1}}{1.349}\right]^2, \quad \text{for } i = 2, \ldots, k-1. \tag{9}$$

The interactive graphical interface is used to elicit the required quartile assessments. *Help* messages describe the meaning of quartiles to the expert and explain how to use the method of bisection to assess them. The expert first clicks the computer's mouse on the left-most stacked bar to assess L_2^* and U_2^*. Then, for each remaining r_{i+1} $(i = 2, \ldots, k-1)$, he assesses quartiles L_{i+1}^* and U_{i+1}^* given that $r_2 = m_2^*, \ldots, r_i = m_i^*$. This is illustrated in Fig. 2, where the expert has assessed the two quartiles of r_4, represented by the two short (dark blue) horizontal lines on the third (blue/orange) bar, conditional on r_2 and r_3 having the median values given by the two leftmost (purple/orange) bars. To assist the expert, the software presents an interactive graph showing the pdf curve of the lognormal distribution of $(r_{i+1}|r_2 = m_2^*, \ldots, r_i = m_i^*)$, for $i = 2, \cdots, k-1$; see Fig. 2. The expert can change his assessed conditional quartiles of r_{i+1} until the conditional pdf curve forms an acceptable representation of his opinion.

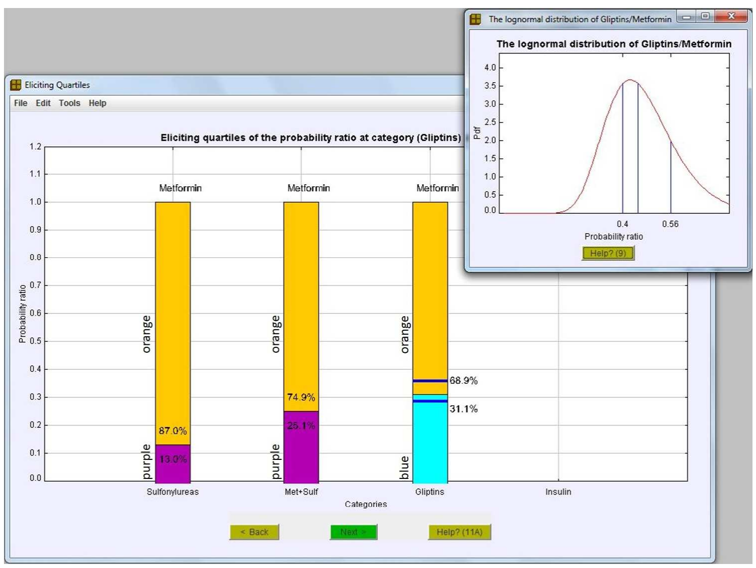

FIG 2. *Assessing conditional quartiles of probability ratios*

3.2.2. *Assessing* $C(Y_j \,|\, Y_2 = y_2^\diamond, \ldots, Y_{i-1} = y_{i-1}^\diamond, Y_i = y_i^0)$

The value we give to y_i^0 must cause the expert to revise his opinions, so it must differ from m_i by a reasonable amount. In particular y_i^0 must not equal $y_i^\diamond = m_i$, as in that case the conditional centre would exactly equal m_i. We set $y_i^0 = L_i$ for $i = 2, \ldots, k-1$. The condition $Y_2 = L_2$ is equivalent to $r_2 = L_2^*$ and, the condition $Y_i = L_i$ is equivalent to $r_i = L_i^*$. The expert has earlier assessed $L_2^*, \ldots, L_{k-1}^*$, giving $L_2, \ldots, L_{k-1}$ from equations (6) and (8). Here the expert first assesses the median of $(r_j \,|\, r_2 = L_2^*)$ for $j = 3, \ldots, k$, and we let m_{2j}^* denote the assessments. He next assesses the median of $(r_j \,|\, r_2 = m_2^*, \ldots, r_{i-1} =$

$m^*_{i-1}, r_i = L^*_i$) for $i = 3, \ldots, k-1$, $j = i+1, \ldots, k$ and we let m^*_{ij} denote the assessments.

Let $C(Y_j\,|y^0_2)$ denote $C(Y_j\,|Y_2 = y^0_2)$ and let $C(Y_j \,|\, y^\diamond_2, \ldots, y^\diamond_{i-1}, y^0_i)$ denote $C(Y_j \,|\, Y_2 = y^\diamond_2, \ldots, Y_{i-1} = y^\diamond_{i-1}, Y_i = y^0_i)$. We have

$$C(Y_j \,|\, y^0_2) = \log(m^*_{2j}) \quad \text{for } j = 3, \ldots, k \tag{10}$$

and

$$C(Y_j\,|y^\diamond_2, \ldots, y^\diamond_{i-1}, y^0_i) = \log(m^*_{ij}) \quad \text{for } i = 3.\ldots, k-1;\ j = i+1, \ldots, k. \tag{11}$$

To implement this part of the elicitation process, the interactive program displays a bar-chart with a pair of stacked bars for each category (except for the first category). The left-hand stacked bars show the expert's unconditional median assessments; the right-hand bars relate to the conditional assessments, some giving conditions and the remainder showing the expert's conditional assessments. The expert is first asked to assume that L^*_2 is the actual value of r_2. This condition is specified by the right-hand stacked bar in the first pair of bars. Taking this information into account, he clicks the mouse on the right-hand bar in each remaining pair to re-assesses his medians of $r_3, \ldots, r_k$, giving $m^*_{22}, \ldots, m^*_{2k}$. These conditional median assessments, as with the conditional median assessments m^*_i $(i = 2, \ldots, k)$, are not constrained and may take any positive value.

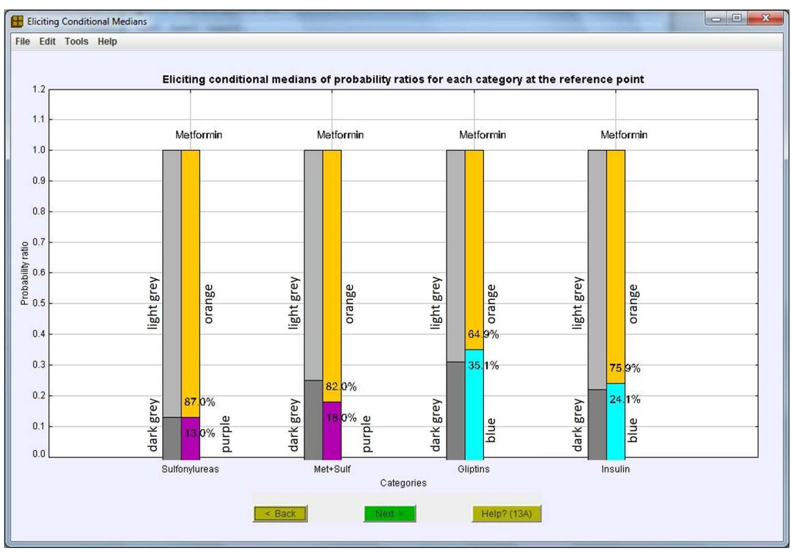

FIG 3. *Assessing conditional medians*

Then, in each successive step i (for $i = 3, \cdots, k-1$), the expert is asked to suppose that the true values of $r_2, \ldots, r_{i-1}$ and r_i are $m^*_2, \ldots, m^*_{i-1}$ and

L_i^*, respectively. In Fig. 3, these conditions are shown by the right-hand (purple/orange) stacked bars in the first two categories. Given this information, the expert is asked to revise his earlier unconditional assessments of the medians of $r_{i+1}, \ldots, r_k$. His new medians are m_{ij}^* $(j = i+1, \ldots, k)$ and in Fig. 3 these are the right-hand (blue/orange) bars in the last two categories. The left-hand (dark/light grey) bars in Fig. 3 are his unconditional median assessments and they are displayed to aid the expert.

3.2.3. Determining $\boldsymbol{\Sigma}_{k/1}$

Conformably partition $\boldsymbol{\Sigma}_{k/1} = \text{Var}(\mathbf{Y}_{k/1})$ as

$$\boldsymbol{\Sigma}_{k/1} = \begin{bmatrix} \boldsymbol{\Sigma}_{(k-1)/1} & \boldsymbol{\zeta} \\ \boldsymbol{\zeta}' & \sigma_k^2 \end{bmatrix}, \tag{12}$$

where $\boldsymbol{\Sigma}_{(k-1)/1} = \text{Var}(Y_2, \ldots, Y_{k-1})$. A slight variant of the method of Kadane *et al.* (1980) is used to estimate all the elements of $\boldsymbol{\Sigma}_{k/1}$. Differences between our method and that of Kadane *et al.* (1980) arise from the choices of $y_2^{\diamond}, \ldots, y_{k-1}^{\diamond}$. The required spreads and centres are obtained from the elicited assessments, using equations (7) and (9)–(11). Our procedure gives the desired property that $\boldsymbol{\Sigma}_{k/1}$ is certain to be positive-definite. Details are given in the appendix.

3.3. Feedback using marginal quartiles

After initial hyperparameter values of the logistic normal distribution have been assessed, the expert is given feedback about his prior distribution through a bar-chart. This displays the unconditional median and unconditional quartiles of p_j $(j = 1, \ldots, k)$ that are implied by his prior distribution. This form of feedback has two clear benefits.

(i) It encourages the expert to examine his assessed opinions from a different perspective. The feedback requires the expert to consider the probabilities p_j, rather than the probability ratios $r_j = p_j/p_1$.
(ii) Conditional assessment tasks are generally harder to perform than unconditional tasks. Other than for the probability ratio r_2, all assessments of upper and lower quartiles have been conditional assessments for the remaining probability ratios. The feedback asks the expert to consider *un*conditional assessments.

An example of the bar-chart is shown in Fig. 4, where the left-hand (grey) bar in each pair and the associated short (black) horizontal lines show the median and quartile values given as feedback. When the feedback is displayed, the expert is invited to change the median and/or quartiles of any categories he wants by adjusting the right-hand (blue) bars and the short (dark blue) horizontal lines.

To add this feedback to the software, a method is required for estimating unconditional quartiles from the elicited hyperparameters $\boldsymbol{\mu}_{k/1}$ and $\boldsymbol{\Sigma}_{k/1}$. Unfortunately, a closed-form method for estimating the unconditional moments, or quartiles, of the logistic normal distribution does not exist. This is only a minor inconvenience though – we simply generate a large sample of, say, 100000 random observations of $\boldsymbol{Y}_{k/1}$ from MVN$(\boldsymbol{\mu}_{k/1}, \boldsymbol{\Sigma}_{k/1})$. Each observation gives a value of $(p_1, \ldots, p_k)$, so the generated observations yield a random sample of each p_i $(i = 1, \ldots, k)$, from which quantiles of each p_i's marginal distribution can be calculated. Further detail is given in the appendix.

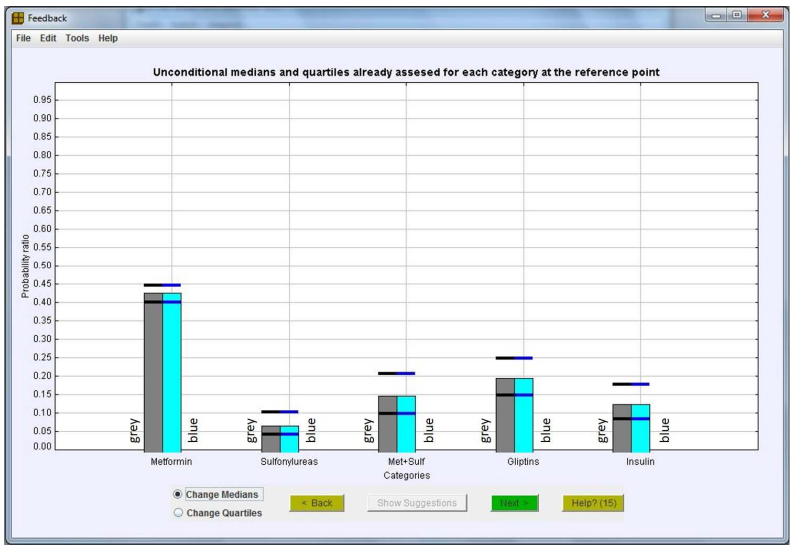

FIG 4. *Feedback giving marginal medians and quartiles of the* p_j

If the expert revises the quantiles of the p_i, we also require a method of determining $\boldsymbol{\mu}_{k/1}$ and $\boldsymbol{\Sigma}_{k/1}$ from the new quantiles. This is a trickier problem and the procedure we construct is described in the appendix. An important feature of the procedure is that the different categories are treated symmetrically. In particular, no category has the distinction of being classified as the 'fill-up category'. To achieve this we define the $k \times (k-1)$ matrix $\mathbf{H}$ by

$$\mathbf{H} = \begin{bmatrix} -1/k & -1/k & -1/k & \ldots & -1/k \\ (k-1)/k & -1/k & -1/k & \ldots & -1/k \\ -1/k & (k-1)/k & -1/k & \ldots & -1/k \\ \vdots & & & & \\ -1/k & -1/k & -1/k & \ldots & (k-1)/k \end{bmatrix}.$$

That is, each element of the top row of $\mathbf{H}$ is $-1/k$, while its other $k-1$ rows form a square matrix whose diagonal elements are $(k-1)/k$ and whose non-diagonal elements are $-1/k$, We transform from $\boldsymbol{Y}_{k/1}$ to the $k \times 1$ vector $\boldsymbol{T} = (T_1, \ldots, T_k)'$

by putting $\boldsymbol{T} = \mathbf{H}\,\boldsymbol{Y}_{k/1}$. Then

$$T_i = \log(p_i) - \frac{1}{k}\sum_{j=1}^{k}\log(p_j) \qquad \text{for } i = 1, \ldots, k. \tag{13}$$

The distribution of $\boldsymbol{T}$ is a singular normal distribution ($\sum_{j=1}^{k} T_i = 0$). We let $\boldsymbol{\Gamma} = (\gamma_1, \ldots, \gamma_k)' = \mathbf{H}\boldsymbol{\mu}_{k/1}$ denote its mean and write its variance matrix, $\mathbf{H}\boldsymbol{\Sigma}_{k/1}\mathbf{H}'$, as $\mathbf{A}\boldsymbol{\Omega}\mathbf{A}$, where $\mathbf{A}$ is a diagonal matrix with ith diagonal element $a_i = \{\mathrm{Var}(T_i)\}^{1/2}$ and $\boldsymbol{\Omega}$ is the correlation matrix of $\boldsymbol{T}$. Aitchison (1986) refers to $\mathbf{A}\boldsymbol{\Omega}\mathbf{A}$ as the *centred logratio covariance matrix*.

In modifying the prior distribution to reflect an expert's re-assessments of quantiles of $p_1, \ldots, p_k$, we treat $\boldsymbol{\Omega}$ as fixed and only revise the estimates of $\mathbf{A}$ and $\boldsymbol{\Gamma}$. It is because we want to take the correlation matrix as fixed that we must transform to variables that treat the different categories symmetrically. From equation (13), revision of the quantiles of p_i primarily affect γ_i and a_i.

If the expert revises quantiles of $p_1, \ldots, p_k$, we determine $\boldsymbol{\Gamma}$ and $\mathbf{A}$ (using the method described in the appendix), re-estimate $\boldsymbol{\mu}_{k/1}$ and $\boldsymbol{\Sigma}_{k/1}$, and generate a large sample of observations from $\mathrm{MVN}(\boldsymbol{\mu}_{k/1}, \boldsymbol{\Sigma}_{k/1})$. From the sample, we calculate quantiles of the p_i that mimic any asymmetry in the expert's previous quantile assessments. The quantiles are unlikely to exactly match the expert's previous assessments, as the expert gave more assessments than the number of parameters, so his assessments will seldom satisfy the requirements for statistical coherence. The expert is invited to revise the suggested quantiles and the cycle is repeated until the suggested quantiles form an acceptable representation of his assessments.

On completion of the elicitation, the software outputs the elicited hyperparameters of the logistic normal prior distribution, $\boldsymbol{\mu}_{k/1}$ and $\boldsymbol{\Sigma}_{k/1}$, in a suitable format for further Bayesian analysis.

4. Elicitation method with covariates

We now turn to the situation that motivated the work in this paper, where the multinomial sampling model contains one or more covariates that influence the membership probabilities of different categories. We suppose the sampling model is a multinomial logistic regression, which is a generalisation of logistic regression. In Section 4.1 we define the multinomial logistic model and also the form of our prior distribution for its parameters, which is a multivariate normal distribution.

Our method of quantifying opinion to obtain the parameters of the prior distribution is given in Sections 4.2 and 4.3. The method makes extensive use of the elicitation procedure developed in Section 3, repeatedly using the procedure to quantify opinion when the covariates take specified values, for a variety of specified values.

4.1. Sampling and prior models

If a covariate is a factor with i levels, we assume that it has been re-expressed as $i-1$ dummy 0/1 variables: these all equal 0 for the first level of the factor and a different one of them equals 1 for each other factor level. We suppose the dummy variables and continuous covariates give m variables in total. Let $\mathbf{x} = (x_1, \ldots, x_m)'$ be the vector that they form and let $p_i(\mathbf{x})$ denote the probability that an observation with covariate values $\mathbf{x}$ falls in the ith category $(i = 1, \ldots, k)$. The sampling model is obtained by setting Y_i equal to $\alpha_i + \mathbf{x}'\boldsymbol{\beta}_i$ in equation (1), where α_i and $\boldsymbol{\beta}_i = (\beta_{1,i}, \ldots, \beta_{m,i})'$ are the constant and vector of regression coefficients for the ith category $(i = 2, \ldots, k)$. This gives

$$p_i(\mathbf{x}) = \begin{cases} \dfrac{1}{1 + \sum_{j=2}^{k} \exp(\alpha_j + \mathbf{x}'\boldsymbol{\beta}_j)}, & i = 1 \\ \\ \dfrac{\exp(\alpha_i + \mathbf{x}'\boldsymbol{\beta}_i)}{1 + \sum_{j=2}^{k} \exp(\alpha_j + \mathbf{x}'\boldsymbol{\beta}_j)}, & i = 2, \cdots, k, \end{cases} \tag{14}$$

which defines the multinomial logistic (logit) model. To simplify the notation, we henceforth denote $p_i(\mathbf{x})$ by p_i. As defined earlier, $\boldsymbol{Y}_{k/1} = (Y_2, \ldots, Y_k)'$, where $Y_i = \log(p_i/p_1)$ for $i = 2, \ldots, k$. While the prior distribution for the α_i and $\boldsymbol{\beta}_i$ $(i = 2, \ldots, k)$ is a multivariate normal distribution, the prior model for $(p_1, \ldots, p_k)$ is referred to as a logistic normal distribution.

It is convenient to rearrange the regression coefficients into a matrix, say $\mathbf{B}$, of the form

$$\mathbf{B} = \left[\begin{pmatrix} \alpha_2 \\ \boldsymbol{\beta}_2 \end{pmatrix}, \quad \cdots, \quad \begin{pmatrix} \alpha_k \\ \boldsymbol{\beta}_k \end{pmatrix} \right], \tag{15}$$

because we will be focusing on one row of $\mathbf{B}$ at a time. We define the new set of vectors $\boldsymbol{\alpha}$, $\boldsymbol{\beta}_{(r)}$, for $r = 1, 2, \cdots, m$, as the rows of $\mathbf{B}$, so that

$$\boldsymbol{\alpha} = (\alpha_2, \quad \alpha_3, \quad \cdots, \quad \alpha_k)', \tag{16}$$

$$\boldsymbol{\beta}_{(r)} = (\beta_{r,2}, \quad \beta_{r,3}, \quad \cdots, \quad \beta_{r,k})'. \tag{17}$$

We assume that the prior distribution for $(\boldsymbol{\alpha}', \boldsymbol{\beta}'_{(1)}, \ldots, \boldsymbol{\beta}'_{(m)})'$ is a multivariate normal distribution, MVN$(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The task is to elicit $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

To make the elicitation problem manageable we shall assume that, given the value of $\boldsymbol{\alpha}$, the vectors $\boldsymbol{\beta}_{(r)}$ and $\boldsymbol{\beta}_{(s)}$ are *a priori* conditionally independent for all r and s $(r \neq s)$. Thus, $\boldsymbol{\Sigma}_{|\alpha} = S(\boldsymbol{\beta}'_{(1)}, \ldots, \boldsymbol{\beta}'_{(m)} \,|\, \boldsymbol{\alpha})$ is a block-diagonal matrix:

$$\boldsymbol{\Sigma}_{|\alpha} = \begin{pmatrix} \boldsymbol{\Sigma}_{\beta,1|\alpha} & 0 & \ldots & 0 \\ 0 & \boldsymbol{\Sigma}_{\beta,2|\alpha} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \boldsymbol{\Sigma}_{\beta,m|\alpha} \end{pmatrix}, \tag{18}$$

where $\boldsymbol{\Sigma}_{\beta,r|\alpha} = S(\boldsymbol{\beta}_{(r)}|\boldsymbol{\alpha})$. We do not make further independence assumptions for the following reasons.

(i) As in ordinary linear regression, changing the origins of the x-variables will change the correlation between α_i and $\boldsymbol{\beta}_i$ $(i = 1, \ldots, k)$. Hence, as there is no reason to believe that the x-variables each have a natural origin, it is unreasonable to assume that α_i and $\boldsymbol{\beta}_i$ are uncorrelated in the prior model.

(ii) If we only consider items that have some specified covariate values, $\mathbf{x}^*$ say, then these items follow a simple multinomial distribution with membership probabilities $p_1, \ldots, p_k$. Thus, if changing the value of the r^*th component of $\mathbf{x}$ increased one of these probabilities, then it must decrease one or more of the other probabilities, because of the unit sum constraint. It follows that the components of $\boldsymbol{\beta}_{(r^*)}$ cannot be independent of each other.

We finish this section with an overview of the elicitation procedure.

1. We elicit $E(\boldsymbol{\alpha}) = \boldsymbol{\mu}_\alpha$ and $E(\boldsymbol{\beta}_{(r)}) = \boldsymbol{\mu}_{\beta,r}$ for $r = 1, \ldots, m$. Then obtain $\boldsymbol{\mu}$ from

$$\boldsymbol{\mu} = (\boldsymbol{\mu}'_\alpha, \boldsymbol{\mu}'_{\beta,1}, \ldots, \boldsymbol{\mu}'_{\beta,m})'. \tag{19}$$

2. For $r = 1, \ldots, m$, define $\boldsymbol{\Sigma}_\alpha$, $\boldsymbol{\Sigma}_{\alpha,\beta,r}$ and $\boldsymbol{\Sigma}_{\beta,r}$ from

$$S\begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta}_{(r)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_\alpha & \boldsymbol{\Sigma}'_{\alpha,\beta,r} \\ \boldsymbol{\Sigma}_{\alpha,\beta,r} & \boldsymbol{\Sigma}_{\beta,r} \end{pmatrix}, \tag{20}$$

where the partitioning is conformal. We elicit $\boldsymbol{\Sigma}_\alpha$ and, for $r = 1, \ldots, m$, we also elicit $\boldsymbol{\Sigma}_{\alpha,\beta,r}$ and $\boldsymbol{\Sigma}_{\beta,r|\alpha}$. The latter determine $\boldsymbol{\Sigma}_{|\alpha}$ (equation (18)). We put $\boldsymbol{\Sigma}_{\alpha,\beta} = (\boldsymbol{\Sigma}'_{\alpha,\beta,1}, \ldots, \boldsymbol{\Sigma}'_{\alpha,\beta,m})'$ and

$$\boldsymbol{\Sigma}_\beta = \boldsymbol{\Sigma}_{|\alpha} + \boldsymbol{\Sigma}_{\alpha,\beta}\boldsymbol{\Sigma}_\alpha^{-1}\boldsymbol{\Sigma}'_{\alpha,\beta}. \tag{21}$$

Then we obtain $\boldsymbol{\Sigma}$ from

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_\alpha & \boldsymbol{\Sigma}'_{\alpha,\beta} \\ \boldsymbol{\Sigma}_{\alpha,\beta} & \boldsymbol{\Sigma}_\beta \end{pmatrix}. \tag{22}$$

In Section 4.2 we give our method of eliciting $\boldsymbol{\mu}_\alpha$, $\boldsymbol{\Sigma}_\alpha$, $\boldsymbol{\mu}_{\beta,r}$ (for $r = 1, \ldots, m$) and $\boldsymbol{\Sigma}_{|\alpha}$. In Section 4.3 we give our method of eliciting $\boldsymbol{\Sigma}_{\alpha,\beta,r}$ for $r = 1, \ldots, m$.

4.2. Assessment of $\boldsymbol{\mu_\alpha}$, $\boldsymbol{\Sigma_\alpha}$, $\boldsymbol{\mu_{\beta,r}}$ and $\boldsymbol{\Sigma_{|\alpha}}$

To make the assessment tasks easier for the expert, each covariate is given a reference value and, in the elicitation procedure, only one covariate at a time is varied. All other covariates are assumed to be at their reference values/levels. By doing this for each covariate in turn, the expert can concentrate on revising his assessments as a result of the change in just one variable. For notational convenience, we assume that the scale of each continuous covariate has been centred so that 0 is its reference value. For a factor, its first level is taken as its reference level, so that its dummy variables each equal 0 at the reference level. Let $\mathbf{x}_0$ be the $m \times 1$ vector of 0s. When $\mathbf{x} = \mathbf{x}_0$, all variables take their reference value (0), so we refer to $\mathbf{x}_0$ as the *reference point.* Also, for $r = 1, \ldots, m$, let $\mathbf{x}_r^*$

Fig 5. *Assessing medians for specified covariate values*

denote a $m \times 1$ vector whose elements are 0 apart from its rth element, which equals some value that we specify, x_r^* say. (For a factor level, x_r^* is set equal to 1.)

(I) To elicit $\boldsymbol{\mu}_\alpha$, the expert is asked to restrict attention to items whose covariates are all at their reference values/levels; i.e. to items for which $\mathbf{x} = \mathbf{x}_0$. Considering just this subpopulation, he performs the same assessment tasks as in Section 3, where $\boldsymbol{\mu}_{k/1} = E(\boldsymbol{Y}_{k/1})$ and $\boldsymbol{\Sigma}_{k/1} = S(\boldsymbol{Y}_{k/1})$ were assessed. Now, however, the assessments yield $E(\boldsymbol{Y}_{k/1} \,|\, \mathbf{x}_0)$ and $S(\boldsymbol{Y}_{k/1} \,|\, \mathbf{x}_0)$. We then set $\boldsymbol{\mu}_\alpha = E(\boldsymbol{Y}_{k/1} \,|\, \mathbf{x}_0)$ and $\boldsymbol{\Sigma}_\alpha = S(\boldsymbol{Y}_{k/1} \,|\, \mathbf{x}_0)$.

(II) To elicit $\boldsymbol{\mu}_{\beta,r}$, the subpopulation is restricted to items for which $\mathbf{x} = \mathbf{x}_r^*$. The expert repeats the assessment tasks of Section 3.1 and these yield $E(\boldsymbol{Y}_{k/1} \,|\, \mathbf{x}_r^*)$. Now, as x_r^* is the only non-zero element of $\mathbf{x}_r^*$, $E(\boldsymbol{Y}_{k/1} \,|\, \mathbf{x}_r^*) = \boldsymbol{\mu}_\alpha + x_r^* \boldsymbol{\mu}_{\beta,r}$. As $\boldsymbol{\mu}_\alpha$ is given by the assessments in (I) and x_r^* is a value we specify, this gives $\boldsymbol{\mu}_{\beta,r}$. During this stage, the software displays the previously assessed medians when all covariates were at their reference values/levels. These are presented as the left-hand (dark/light grey) bar of each pair of stacked bars in Fig. 5. The expert should primarily consider the difference between the left-hand and right-hand (blue/orange) stacked bars in assessing the latter. The reference value/level of each covariate is also displayed, as in the table in the upper-left of Fig. 5.

(III) To obtain $\boldsymbol{\Sigma}_{\beta,r|\alpha}$, the expert is asked to assume that his earlier median assessments of the probability ratios $r_2(\mathbf{x}_0), \ldots, r_k(\mathbf{x}_0)$ are the true values of these ratios. Under this assumption, which fixes the value of $\boldsymbol{\alpha}$, the elicitation procedure in Section 3.2 is repeated with the population

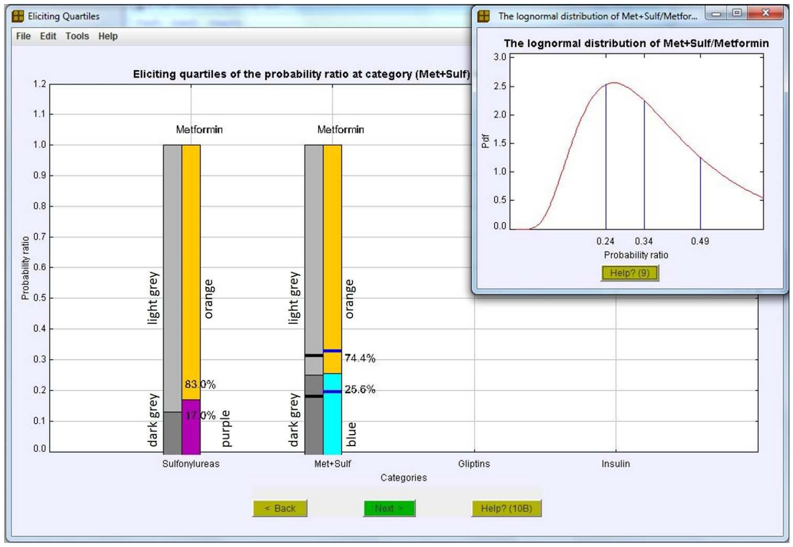

FIG 6. *Assessing conditional quartiles for specified covariate values*

restricted to items for which $\mathbf{x} = \mathbf{x}_r^*$, as in (II). Fig. 6 illustrates the graphics used for assessing conditional quartiles in this step. The right-hand (purple/orange) bar in the first pair of bars is the true value of r_2 for the subpopulation with covariate values $\mathbf{x} = \mathbf{x}_r^*$. (This purple/orange bar was his median assessment of r_2 for that subpopulation.) The left-hand (dark/light grey) bar of each pair show the expert's earlier median assessments of $r_2(\mathbf{x}_0), \ldots, r_k(\mathbf{x}_0)$ and the short (black) horizontal lines above and below the tops of the dark-grey boxes show the unconditional quartiles at the reference point. The right-hand (blue/orange) bars have the same meaning as in Section 3.2 (c.f. Fig. 2) but now the quartiles reflect uncertainty in the difference in height between the left-hand (dark/light grey) and right-hand (blue/orange) bars.

Fig. 7 illustrates the graphics used to assess conditional medians. The expert is asked to assume that, as in Fig. 6,

(i) The left-hand (dark/light grey) bar in each pair of bars (which were his median assessments at the reference point) are the true values of the proportions at the reference point.

(ii) The right-hand (purple/orange) bar in the first pair of bars is the true value of r_2 for the subpopulation with covariate values $\mathbf{x} = \mathbf{x}_r^*$.

He is also asked to assume that:

(iii) The narrow inner (red/yellow) bar in the second set of bars is the true value for r_3 for the subpopulation with covariate values $\mathbf{x} = \mathbf{x}_r^*$. This value differs from his median assessment of r_3 for that subpopulation, which corresponds to the outer (purple/orange) part of right-hand

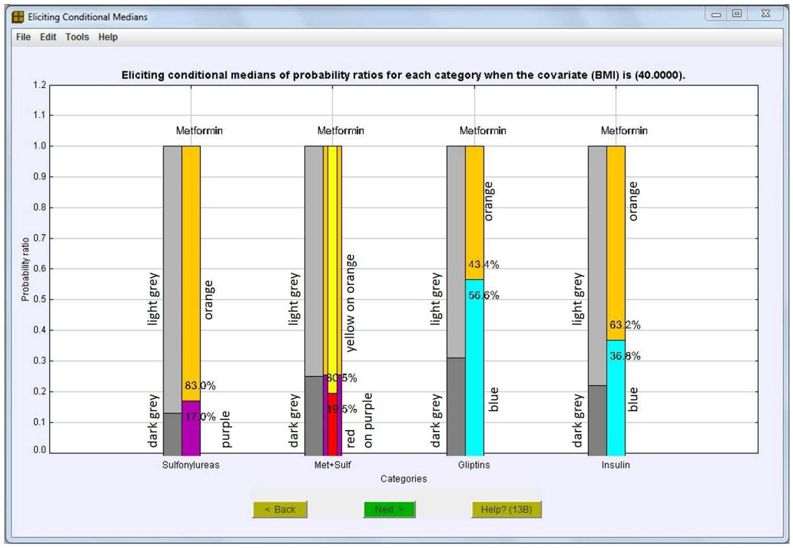

FIG 7. *Conditional medians at specified covariate values*

bar in that set. This assumption might alter the expert's opinion about the probability ratios for the remaining categories.

Given this information, the expert is asked to revise his median assessments for the remaining probability ratios. His revised conditional median assessments are represented by the right-hand (blue/orange) bars in the last two pairs of bars in Fig. 7. This time the assessments yield $S(\boldsymbol{Y}_{k/1} \,|\, \mathbf{x} = \mathbf{x}_r^*, \boldsymbol{\alpha}) = S(x_r^* \boldsymbol{\beta}_{(r)} \,|\, \boldsymbol{\alpha})$. We put

$$\boldsymbol{\Sigma}_{\beta,r|\alpha} = S(\boldsymbol{Y}_{k/1} \,|\, \mathbf{x} = \mathbf{x}_r^*, \boldsymbol{\alpha}) \,/\, (x_r^*)^2. \tag{23}$$

The process is repeated for $r = 1, \ldots, m$, giving $\boldsymbol{\Sigma}_{\beta,1|\alpha}, \ldots, \boldsymbol{\Sigma}_{\beta,m|\alpha}$, and $\boldsymbol{\Sigma}_{|\alpha}$ is determined from equation (18).

The above procedure requires a number of assessments if there are several covariates. As a short-cut, for any $\mathbf{x}_r^*$ (but not for $\mathbf{x}_0$) the expert may choose to skip to the feedback screen instead of performing steps (II) and (III). The feedback screen displays a pair of bars for each category (c.f. Fig. 8). The left-hand (grey) bars give the unconditional median probabilities when all the covariates are at their reference values and short dark-blue horizontal lines in the left-hand bars are his unconditional lower and upper quartiles. The medians have been calculated from the expert's earlier assessments and he is told to assume they are the true probabilities when $\mathbf{x} = \mathbf{x}_0$. The expert assesses medians and quartiles for membership probabilities that reflect his opinions when the values of the covariates are $\mathbf{x}_r^*$ (rather than $\mathbf{x}_0$). His assessments form the right-hand (sky-blue) bars and short (dark-blue) horizontal lines. In making these assessments the expert should focus on the difference between the left-hand and right-hand

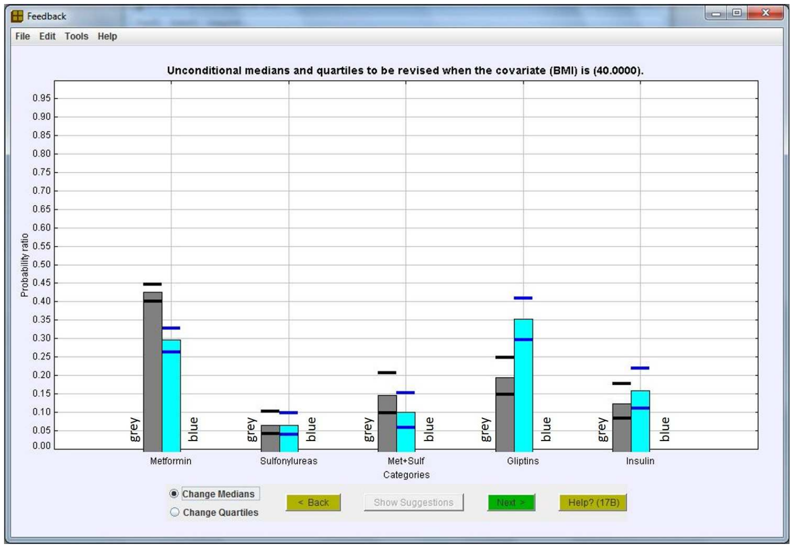

FIG 8. *The short-cut option using feedback*

bars in a pair, as his assessments should reflect the change in probabilities as the covariate values change from $\mathbf{x}_0$ to $\mathbf{x}_r^*$.

Skipping steps (II) and (III) shortens the elicitation procedure significantly, but requires the assumption that prior knowledge gives the same correlation structure to changes in membership probabilities as to the membership probabilities themselves.

4.3. Assessment of $\boldsymbol{\Sigma}_{\alpha,\beta,r}$

Section 4.2 yields the parameters $E(\boldsymbol{\alpha}) = \boldsymbol{\mu}_\alpha$, $E(\boldsymbol{\beta}_{(r)}) = \boldsymbol{\mu}_{\beta,r}$ and $\text{Var}(\boldsymbol{\alpha}) = \boldsymbol{\Sigma}_\alpha$. Now

$$E(\boldsymbol{\beta}_{(r)}\,|\boldsymbol{\alpha} = \boldsymbol{\alpha}^*) = \boldsymbol{\mu}_{\beta,r} + \boldsymbol{\Sigma}_{\alpha,\beta,r}\boldsymbol{\Sigma}_\alpha^{-1}(\boldsymbol{\alpha}^* - \boldsymbol{\mu}_\alpha). \tag{24}$$

The strategy for obtaining $\boldsymbol{\Sigma}_{\alpha,\beta,r}$ is to choose $k-1$ different values for $\boldsymbol{\alpha}^*$, say $\boldsymbol{\alpha}_2^*, \ldots, \boldsymbol{\alpha}_k^*$ and, for each $\boldsymbol{\alpha}_i^*$, obtain assessments from the expert that determine $E(\boldsymbol{\beta}_{(r)}|\boldsymbol{\alpha} = \boldsymbol{\alpha}_i^*)$. For $i = 2, \ldots, k$, define $\mathbf{w}_{r,i} = E(\boldsymbol{\beta}_{(r)}|\boldsymbol{\alpha} = \boldsymbol{\alpha}_i^*) - \boldsymbol{\mu}_{\beta,r}$ and $\boldsymbol{\psi}_i = \boldsymbol{\alpha}_i^* - \boldsymbol{\mu}_\alpha$. Put

$$\mathbf{W}_r = (\mathbf{w}_{r,2}, \ldots, \mathbf{w}_{r,k}) \quad \text{and} \quad \boldsymbol{\Psi} = (\boldsymbol{\psi}_2, \ldots, \boldsymbol{\psi}_k). \tag{25}$$

Then, from (24),

$$\boldsymbol{\Sigma}_{\alpha,\beta,r} = \mathbf{W}_r\boldsymbol{\Psi}^{-1}\boldsymbol{\Sigma}_\alpha. \tag{26}$$

Details of our implementation are as follows. Let $\boldsymbol{\mu}_\alpha = (\mu_{\alpha,2}, \ldots, \mu_{\alpha,k})'$ and let ξ_i^2 be the $(i-1,\, i-1)$ diagonal element of $\boldsymbol{\Sigma}_\alpha$ $(i = 2, \ldots, k)$. Then each component of $\boldsymbol{\alpha}_i^*$ is set equal to the corresponding component of $\boldsymbol{\mu}_\alpha$, apart from

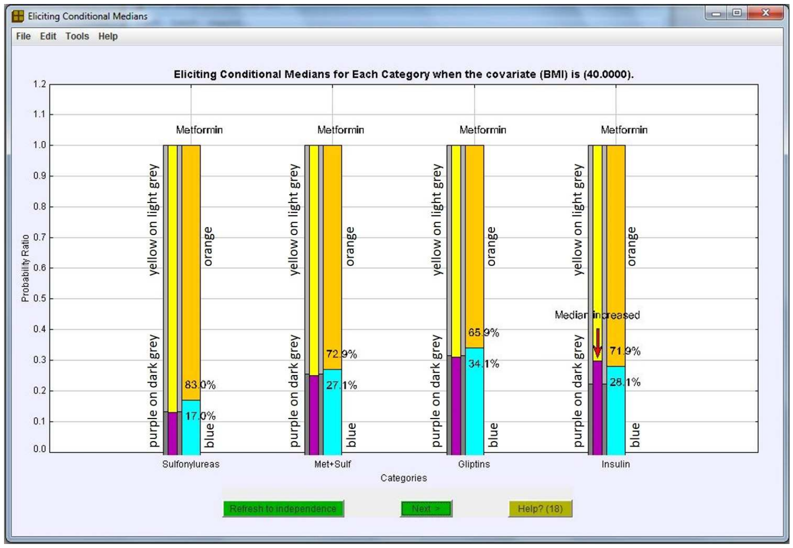

FIG 9. *Assessing correlations between α and β*

the ith component, which is set equal to $\mu_{\alpha,i+1} + 0.674\xi_{i+1}$. (This is the upper quartile of the prior distribution of the ith component of $\boldsymbol{\alpha}$.) Then $\boldsymbol{\Psi}$ is a diagonal matrix with diagonal elements $0.674\xi_2, \ldots, 0.674\xi_k$, so it is positive-definite and $\boldsymbol{\Psi}^{-1}$ is well-defined. The probability ratios $r_2(\mathbf{x}_0), \ldots, r_k(\mathbf{x}_0)$ are calculated under the assumption that $\boldsymbol{\alpha} = \boldsymbol{\alpha}_i^*$. That is, we put $r_j(\mathbf{x}_0) = \exp(\alpha_{i,j}^*)$, where $\alpha_{i,j}^*$ is the jth component of $\boldsymbol{\alpha}_i^*$. The expert is asked to assume that these are the true probability ratios for items whose covariate values are at their reference values. With the population restricted to items for which $\mathbf{x} = \mathbf{x}_r^*$, he then repeats the assessment tasks of Section 3.1. These give $E(\boldsymbol{Y}_{k/1}|\mathbf{x}_r^*, \boldsymbol{\alpha} = \boldsymbol{\alpha}_i^*)$ and we obtain

$$E(\boldsymbol{\beta}_{(r)}|\,\boldsymbol{\alpha} = \boldsymbol{\alpha}_i^*) = \{E(\boldsymbol{Y}_{k/1}|\,\mathbf{x}_r^*, \boldsymbol{\alpha} = \boldsymbol{\alpha}_i^*) - \boldsymbol{\alpha}_i^*\}/x_r. \tag{27}$$

Fig. 9 illustrates the assessment tasks that the expert performs. The left-hand (dark/light grey) stacked bar in each pair shows the median assessment of the probability ratio at the reference point. Within each of these bars is a narrower (purple/yellow) bar. The expert is asked to assume that these narrower bars give the true probability ratios for items with covariates at their reference value. These 'true' ratios correspond to the condition $\boldsymbol{\alpha} = \boldsymbol{\alpha}_i^*$. The right-hand (blue/orange) bars are the expert's conditional median assessments of probability ratios (given $\boldsymbol{\alpha} = \boldsymbol{\alpha}_i^*$) for items with covariate values $\mathbf{x} = \mathbf{x}_r^*$.

To help the expert with this assessment task, we suggest median values for the conditional probability ratios that the expert may accept or change. The suggested values are calculated under the assumption that $\boldsymbol{\beta}_{(r)}$ and $\boldsymbol{\alpha}$ are independent *a priori*. The value of $\boldsymbol{\mu}_{\beta,r} = E(\boldsymbol{\beta}_{(r)})$ has already been assessed, say with components $\mu_{\beta,r,j}$ ($j = 2, \ldots, k$). The median values that we suggest are

$r_j(\mathbf{x}_r^*) = \exp(\alpha_{i,j}^* + x_r^* \, \mu_{\beta,r,j})$ $(j = 2, \ldots, k)$. The expert may accept these values as a reasonable representation, or assess any that he chooses.

5. Example: Potential cost of a new drug

In Malta, the standard treatments for diabetes could be classified into four groups, as Metformin, Sulfonylureas, Metformin plus Sulfonylureas, and Insulin. A fifth new treatment, Gliptins, is to be introduced. However, this will increase the overall cost to Malta of treating diabetes, as the new treatment is more expensive per patient. Increased health costs are a concern to Malta's Ministry of Finance so, to estimate the future cost, an expert quantified his opinion about the proportions of patients in Malta who would be on each treatment. The expert, Dr Neville Calleja, is Director of Health Information and Research in Malta's Ministry for Health.

The expert thought that body-mass index (BMI) and co-morbidity would affect which treatment a diabetic patient was likely to take, where co-morbidity is a binary yes/no variable, with 'yes' meaning the patient also suffers from hypertension or has a history of heart disease. Thus BMI and co-morbidity are the covariates in this example. The categories in our multinomial model are the five diabetes treatments Metformin (Mtf), Sulfonylureas (Sul), Metformin plus Sulfonylureas (M+S), Gliptins (Glip) and Insulin (Ins) with membership probabilities $p_1, \ldots, p_5$, respectively. The task is to quantify the expert's opinions about the membership probabilities for these five categories and how the probabilities are affected by the covariates. We take Mtf as the fill-up category. The following describes the assessments that the expert made.

5.1. Assessments at the reference point

As a reference point ($\mathbf{x}_0$), the expert chose patients who had a BMI of 22.5 and no co-morbidity diseases, and he was asked to consider the subpopulation of patients with these characteristics. Unconditional median assessments of the probability ratios for each of the last four categories were first elicited. The expert gave 0.13, 0.25, 0.31 and 0.22 as his medians. Fig. 1 shows these assessments. He then assessed the two quartiles of the probability ratio for the second category as 0.08 and 0.18. Next, he was asked to assume that the probability ratio for the second category is exactly 0.13; given this information he gave his two conditional quartiles for the third category's probability ratio as 0.18 and 0.31. Conditional on 0.13 and 0.25 being the true probability ratios for the second and third categories, he gave 0.29 and 0.36 as his two quartiles for the fourth category – these conditional assessments are shown in the screen-shot in Fig. 2. Finally, conditional on the probability ratios for the second, third and fourth categories being 0.13, 0.25 and 0.31, respectively, he gave 0.15 and 0.25 as his two quartiles for the probability ratio for the fifth category. Table 1 summarizes this set of assessments.

TABLE 1
Median and conditional quartile assessments at the reference point

	$\frac{p_2}{p_1+p_2}$	$\frac{p_3}{p_1+p_3}$	$\frac{p_4}{p_1+p_4}$	$\frac{p_5}{p_1+p_5}$
Upper quartile	0.18	0.31	0.36	0.25
Median	0.13	0.25	0.31	0.22
Lower quartile	0.08	0.18	0.29	0.15

TABLE 2
Conditional median assessments at the reference point[1]

	$\frac{p_2}{p_1+p_2}$	$\frac{p_3}{p_1+p_3}$	$\frac{p_4}{p_1+p_4}$	$\frac{p_5}{p_1+p_5}$
Given $p_2/(p_1+p_2)$	$\underline{0.08}$	0.27	0.33	0.22
Given $p_2/(p_1+p_2)$ and $p_3/(p_1+p_3)$	$\underline{0.13}$	$\underline{0.18}$	0.35	0.24
Given $p_2/(p_1+p_2)$, $p_3/(p_1+p_3)$ and $p_4/(p_1+p_4)$	$\underline{0.13}$	$\underline{0.25}$	$\underline{0.29}$	0.25

[1] Underlined values are the conditions; assessments are not underlined.

Conditional median assessments were elicited next. Conditional on the second category having a probability ratio ($p_2/(p_1+p_2)$) of 0.08, the expert gave 0.27, 0.33 and 0.22 as his conditional median assessments for the probability ratios of the remaining three categories. Then conditional on the second and third categories having probability ratios ($p_2/(p_1+p_2)$ and $p_3/(p_1+p_3)$) of 0.13 and 0.18, respectively, he assessed his conditional medians for the probability ratios of the last two categories as 0.35 and 0.24 − Fig. 3 shows these assessments. Conditional on the second, third and fourth categories having probability ratios ($p_2/(p_1+p_2)$, $p_3/(p_1+p_3)$ and $p_4/(p_1+p_4)$) of 0.13, 0.25 and 0.29, respectively, the expert assessed 0.25 as his conditional median for the probability ratio of the last category. These sets of conditional median assessments are summarized in Table 2.

The expert was then shown the feedback graph in Fig. 4, which displays the unconditional median and quartile values that were calculated by the software based on his assessments at the reference point. He accepted the suggested unconditional medians and quartiles as a reasonable representation of his opinions. This completed the assessments that determine $\boldsymbol{\mu}_\alpha$ and $\boldsymbol{\Sigma}_\alpha$.

5.2. Assessments at different covariate values

During the rest of this elicitation process, for each covariate in turn, the expert was asked to consider people who had a specified value for one covariate (a value not equal to its reference value), while the other covariate was at its reference value/level.

5.2.1. Assessments at BMI = 40

The expert gave 40 as a value of BMI for which he could make reasonable assessments and which differed markedly from the reference BMI value of 22.5. To quantify his opinions about $\boldsymbol{\mu}_{\beta,1}$ and $\boldsymbol{\Sigma}_{\beta,1\,|\alpha}$, the expert made modifications to the feedback screen that had resulted from his assessments at the reference point. He was asked to suppose that his median assessments at the reference point were correct and to assess medians and quartiles for a person who had a BMI of 40 instead of the reference value of 22.5 (but who still had no co-morbidity). This step is illustrated in Fig. 8 – his new median assessments were 0.29, 0.06, 0.10, 0.35 and 0.16. The computer then suggested values for the unconditional quartiles and the expert found them an acceptable representation of his opinions.

To obtain $\mathbf{\Sigma}_{\alpha,\beta,1}$, the expert again assessed probability ratios for just the subpopulation with BMI=40 and the other covariate is at its reference level (i.e. no co-morbidity). These assessments were conditional assessments, where the conditions specified the probability ratios at the reference point $(\alpha_2, \ldots, \alpha_4)$. Most of the specified values equalled his assessed medians at the reference point, but for one category at a time (starting with the second category) a slightly larger value was specified. This information could change the expert opinions about the subpopulation with BMI=40, and he was invited to revise his median assessments of the probability ratios for that subpopulation. The expert did not want to make any changes when the second category had the larger value, and similarly for the third category. However, when the fourth or fifth category had a larger 'true' probability ratio at the reference point than his median assessment, he revised some of his opinions and gave the following as his median probability ratios for the subpopulation with BMI=40.0 and no co-morbidity:

	$\frac{p_2}{p_1+p_2}$	$\frac{p_3}{p_1+p_3}$	$\frac{p_4}{p_1+p_4}$	$\frac{p_5}{p_1+p_5}$
α_4 increased	0.23	0.37	0.56	0.36
α_5 increased	0.15	0.27	0.34	0.28

Fig. 9 is a screen shot of the assessments when a value larger than the expert's median assessment was specified for α_5 (the probability ratio at the reference point for the fifth category).

5.2.2. Assessments at Co-morbidity = yes

The assessment tasks for the subpopulation with BMI=40.0 and co-morbidity at its reference level (i.e. no co-morbidity) were repeated for the subpopulation that *had* co-morbidity and had BMI at the reference level of 22.5. Conditional on the probability ratios at the reference point equalling the expert's median assessments, the expert gave 0.14, 0.11, 0.10, 0.40 and 0.20 as his medians for this new subpopulation. All other values suggested by the computer were accepted by the expert as a fair reflection of his opinions, except when conditioning specified that the fourth category had a larger value at the reference point than his median assessment. This larger value affected the expert's opinions about some of the probability ratios of this new subpopulation – he set the conditional medians for the fourth and fifth categories to 0.64 and 0.46.

5.3. Results

The output from the software gives the hyperparameters ($\boldsymbol{\mu}$ and $\mathbf{\Sigma}$) of a multivariate normal distribution that corresponds to the expert's assessments. These

TABLE 3
The elicited mean vector $\boldsymbol{\mu}$

Constant term ($\boldsymbol{\mu}'_{\alpha}$)				BMI ($\boldsymbol{\mu}'_{\beta,1}$)				Co-morbidity ($\boldsymbol{\mu}'_{\beta,2}$)			
Sul	M+S	Glip	Ins	Sul	M+S	Glip	Ins	Sul	M+S	Glip	Ins
-1.90	-1.10	-0.80	-1.26	0.01	0.00	0.05	0.04	1.66	0.76	1.87	1.62

TABLE 4
The elicited variance matrix $\boldsymbol{\Sigma}$

Constant term				BMI				Co-morbidity			
Sul	M+S	Glip	Ins	Sul	M+S	Glip	Ins	Sul	M+S	Glip	Ins
0.51	-0.09	-0.08	0.00	0.00	0.00	0.00	0.00	0.01	-0.01	0.00	0.00
-0.09	0.32	-0.12	-0.08	0.00	0.00	0.00	0.00	0.01	-0.01	0.00	0.00
-0.08	-0.12	0.13	-0.05	0.02	0.02	-0.03	0.00	0.13	0.32	-0.29	-0.37
0.00	-0.08	-0.05	0.38	0.00	0.03	-0.02	-0.03	0.01	-0.01	0.00	0.00
0.00	0.00	0.02	0.00	0.02	0.02	-0.02	-0.01	0.09	0.20	-0.18	-0.23
0.00	0.00	0.02	0.03	0.02	0.03	-0.03	-0.01	0.11	0.23	-0.21	-0.27
0.00	0.00	-0.03	-0.02	-0.02	-0.03	0.03	0.01	-0.12	-0.27	0.25	0.31
0.00	0.00	0.00	-0.03	-0.01	-0.01	0.01	0.01	-0.04	-0.08	0.08	0.10
0.01	0.01	0.13	0.01	0.09	0.11	-0.12	-0.04	1.03	1.03	-1.02	-1.20
-0.01	-0.01	0.32	-0.01	0.20	0.23	-0.27	-0.08	1.03	2.71	-2.26	-2.79
0.00	0.00	-0.29	0.00	-0.18	-0.21	0.25	0.08	-1.02	-2.26	2.08	2.54
0.00	0.00	-0.37	0.00	-0.23	-0.27	0.31	0.10	-1.20	-2.79	2.54	3.58

hyperparameters are given in Tables 3 and 4. The output also presents the correlations corresponding to the off-diagonal elements of $\boldsymbol{\Sigma}$. Figure 10 shows the correlations of the hyperparameters given by $\boldsymbol{\Sigma}$ in a graphical display developed by Murdoch and Chow (1996).

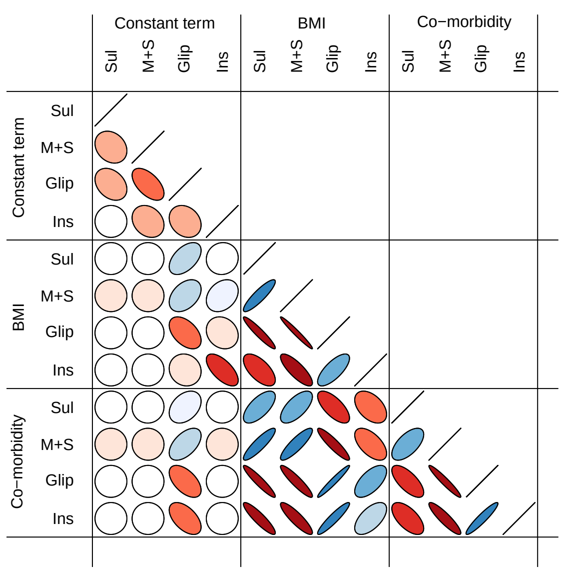

FIG 10. *Correlations between hyperparameters given by the prior distribution*

Table 5
BMI and co-morbidity status of 217 diabetics

BMI	Co-morbidity		Total
	Yes	No	
<20.0	4	3	7
20.0 < 22.5	3	0	3
22.5 < 25.0	20	14	34
25.0 < 30.0	56	38	94
30.0 < 40.0	50	22	72
⩾ 40	6	1	7
Total	139	78	217

The reason for quantifying the expert's opinion was to estimate the cost of treating diabetes in Malta when Gliptins can be prescribed through its national health service. To this end, the expert's opinions were combined with data from a survey that had been conducted in Malta as part of the European Health Interview Survey, 2008 (Ministry for Social Policy, 2008). The survey was conducted on a sample of 5,500 individuals aged 15 years or over. It was drawn from a population register for the Maltese islands and was stratified by age, gender and locality. The survey gave complete data on the required covariates (BMI and co-morbidity status) for 217 diabetics, so for each of these individuals we have the vector $\mathbf{x}$ in equation (14). A summary of these data is given in Table 5. The other information that was needed were the average monthly costs per patient of each treatment. To the nearest five euros, these were Metformin: 65; Sulfonylureas: 40; Metformin plus Sulfonylureas: 85; Gliptins: 230 and Insulin: 30. (The Metformin plus Sulfonylureas treatment comes as a single medication, which reduces its cost relative to the combined costs of its two constituents.)

A vector was randomly generated from the multivariate normal distribution that represented the expert's opinions. This vector gives the values of α_j and $\boldsymbol{\beta}_j$, say α_j^* and $\boldsymbol{\beta}_j^*$ $(j = 2, \ldots, 5)$. Applying equation (14), we determined $p_1(\mathbf{x}), \ldots, p_5(\mathbf{x})$ for each of the diabetics with covariate data. These probabilities were multiplied by the costs of each treatment to give the expected cost for that person, and hence the average cost per patient for the people surveyed could be determined. The number of people with diabetes in Malta is known reasonably accurately (approximately 20,000 adult diabetics), so scaling up gave the cost of treating diabetes in Malta when $\alpha_j = \alpha_j^*$ and $\boldsymbol{\beta}_j = \boldsymbol{\beta}_j^*$ for $j = 2, \ldots, 5$.

The calculation was repeated for 10000 randomly selected vectors from the expert's multivariate normal distribution, each giving an estimate of the cost of treating diabetes in Malta. Their average was 2.38 million euros and, based on 250^{th} smallest and 250^{th} largest estimates, a 95% credible interval for the cost is (1.29, 3.77). R code that performed these calculations is given in supplementary material.

6. Concluding comments

Prior distributions are both a weakness and a strength of Bayesian statistics. They are a weakness when expert opinion is not to be included in the statistical analysis (expert opinion can introduce bias as well as knowledge), as it can be difficult to find a prior distribution that genuinely conveys no information. However, prior distributions are a strength when expert opinion is a resource to be exploited since, in principle, an informative prior distribution then provides an optimal mechanism for using expert opinion.

Different sampling models require different forms of prior distribution and, until now, no prior distribution had been proposed for a multinomial sampling model that contains covariates. However, in this context a logistic normal distribution has commonly been used as the sampling model, making it a natural choice as a prior model for representing expert opinion. The distribution contains a large number of parameters and we made reasonable independence assumptions to reduce their number and make the elicitation problem manageable. Nevertheless, quantifying expert opinion as a logistic normal distribution would be well nigh impossible without the use of interactive computing, but the use of interactive graphics leads to a viable elicitation method.

The elicitation method requires a relatively large number of specifications from the expert even in relatively few dimensions. In practice, we believe the method will be of most use when the number of categories is between three and eight (logistic regression seems preferable if there are only two categories), though using it for up to about 12 categories might be possible with a highly motivated expert (if, say, the expert needed the prior distribution to progress their own work). As well as the number of categories, the complexity of the elicitation task is also dependent on the number of covariates and, in the case of covariates that are factors, the number of factor levels. We envisage the method being used with up to four covariates, with the number of continuous variables plus the total number of factor levels being fewer than ten. A modest-sized problem will typically require an elicitation session of one-and-a-half to two hours, and a large problem will generally take about three hours.

In principle, the assessed prior distribution should be invariant to the choice of fill-up category but, in practice, the assessed prior distribution will vary if the expert repeated the elicitation process but with a different category chosen as the fill-up category. We have not explored how to use the different prior distributions that repeating the elicitation process would yield, but there are obvious possibilities and potential benefits. The expert could be shown figures that highlighted similarities and differences between important features of the prior distribution, such as the median and quartiles of the marginal probability of each category (similar to the feedback screen shown in Figure 4). The expert might then choose the prior distribution that he thought was a better representation of his opinions, or use the elicitation method again to try to reconcile any large differences. Another possibility is to combine the prior distributions through some form of mathematical or algorithmic aggregation rule. This has similarities to mathematically combining the opinions of multiple experts, but is a little simpler than this latter task. When combining the opinions of multiple experts, questions arise as to how much weight to give the different experts (Cooke and Goossens, 2008) and the degree of overlap between different experts' knowledge (Winkler, 1981; French, 2011). These questions do not arise when combining the prior distributions of a single expert – it is reasonable to give each prior the same weight and there is complete overlap of background knowledge. Reviews of methods of aggregating expert judgement may be found, for example, in (O'Hagan *et al.*, 2006, Chapter 9), French (2011) and Werner *et al.* (2017).

In the example in Section 5, the data only give the BMI and co-morbidity of a sample of diabetics, and do not give their current (or future) medication. Thus the data cannot be combined with a prior distribution to form a posterior distribution, as that would require information about their future medication. For the same reason, the question of whether there is conflict between the prior distribution and observed data does not arise. However, a subjectively assessed prior distribution may contain bias as well as information so, when possible, it can be prudent to compare an assessed prior distribution with data. This may be done using Box's p-value, or methods related to it, which compare prior predictive distributions of the data or summary statistics with observed values (a brief summary of such methods may be found in Held and Sauter (2017)). Prior-data conflict can also be examined through calibration-in-the-large, where the observed counts in categories are compared with predicted values at a population-based level, or slope calibration, where the predicted effect of covariates is compared with sample estimates. Description and application of these techniques in multinomial logistic regression is given in Van Hoorde *et al.* (2014) and Van Calster *et al.* (2017). These papers also suggest frequentist methods for recalibrating models derived from one setting so that they can be applied to a second setting – perhaps a different geographical location or a more recent time frame.

To guard against or mitigate prior-data conflict, hyperparameters can be added to a prior distribution that make the prior distribution more robust. As in Held *et al.* (2012), an additional unknown scalar hyperparameter λ might be introduced as a scalar multiplier of the subjective prior covariance matrix. Thus, the assessed covariance matrix $\mathbf{\Sigma}$ would be replaced by $\lambda\mathbf{\Sigma}$. The hyperparameter λ must be given a prior distribution; Held *et al.* (2012) assume that λ follows a chi-squared distribution with one degree of freedom, although a diffuse prior distribution would be an obvious alternative choice. More than one hyperparameter might also be introduced to the subjective prior, as illustrated by Held and Sauter (2017) and Al-Awadhi and Garthwaite (2006). These papers both concern logistic regression models and multiply the subjective variance of the constant term by one hyperparameter and the subjective covariance matrix of the other regression coefficients by a second hyperparameter. Al-Awadhi and Garthwaite (2006) also consider the biases that might influence an expert's assessments and suggest multiplying the diagonal elements of the covariance matrix by one parameter and its off-diagonal elements by a different hyperparameter. This approach could be applied to a covariance matrix assessed through the elicitation method proposed here.

In most of the elicitation tasks that the expert performs, assessments are required of a probability ratio: the proportion of items that fall in a specified category relative to the proportion that fall in the fill-up category. It may seem simpler to ask the expert about the actual proportion of items that fall in the category (i.e. ask about p_i rather than p_i/p_1). Indeed, that was the assessment task we tried to use in a prototype version of the elicitation method developed here. Unfortunately, it led to assessment tasks in which the expert was asked to assume that p_j took a specified value and his opinion about the distribution of $Y_i \mid p_j$ was elicited. This caused problems, because the value of p_j does not determine the value of Y_j, so while the distribution of $Y_i \mid Y_j$ would be a normal distribution, $Y_i \mid p_j$ does not follow a normal distribution, making mathematics intractable. The problem does not arise when a value of $r_j = p_j/p_1$ is specified, as r_j does determine Y_j (we have $Y_j = \log r_j$), so that $Y_i \mid r_j$ follows a normal distribution.

The people intending to use the prior distribution that is assessed by an expert will typically include one or more people with good knowledge of statistics. If the expert's background in statistics is limited, then a facilitator who has good knowledge of statistics should guide the expert through the elicitation process. Features from the Sheffield elicitation framework (SHELF) might also form part of the elicitation process if a transparent record is required. SHELF contains a protocol for capturing information about an elicitation exercise including an expert's backgrounds and potential conflicts of interest and any reasoning or key sources of information that underpin the expert's judgements (Gosling, 2018). The elicitation method has been implemented in interactive software that is freely available on the web at `http://statistics.open.ac.uk/elicitation`. All screens include *help* buttons to assist the expert and facilitator. There is also a *User Guide* that includes some training material to give the expert practice at quantifying his/her opinion. The output available from the software includes files that can be run in WinBUGS (Lunn *et al.*, 2000), OpenBUGS (Lunn *et al.*, 2009) or rJAGS (Plummer, 2015), making it straightforward to combine the prior distribution with data. The output also gives a file that documents the assessments elicited at each step of the process and, using a second file that is also provided, the software can sequentially reproduce screenshots of the elicitation process. As the example in Section 5 indicates, the elicitation software provides a practical means of quantifying expert opinion about a multinomial model that includes covariates, and the assessed prior distribution may be readily exploited to yield useful inferences.

Acknowledgement

The work reported here was supported by a studentship from the Open University. We are very grateful to Dr Neville Calleja whose opinion is quantified in the example. We are grateful to two reviewers and the editor for constructive comments that led to clear improvements in this paper and the work it reports.

Appendix: Mathematical detail

Obtaining $\boldsymbol{\Sigma}_{k/1}$

Following Kadane *et al.* (1980), we sequentially determine matrices $\mathbf{U}_2, \ldots, \mathbf{U}_k$ where $\mathbf{U}_2 = S(Y_2) > 0$ and

$$\mathbf{U}_{i+1} = \begin{bmatrix} \mathbf{U}_i & \mathbf{U}_i \mathbf{g}_{i+1} \\ \mathbf{g}'_{i+1}\mathbf{U}_i & S(Y_{i+1}) \end{bmatrix}.$$

After obtaining $\mathbf{U}_i$, we determine $\mathbf{g}_{i+1}$ from the equation $\mathbf{g}_{i+1} = \mathbf{M}_{i+1}^{-1}\mathbf{h}_{i+1}$, where

$$\mathbf{h}_{i+1} = \begin{bmatrix} C(Y_{i+1} \,|\, y_2^0) - C(Y_{i+1}) \\ C(Y_{i+1} \,|\, y_2^\diamond, y_3^0) - C(Y_{i+1}) \\ \vdots \\ C(Y_{i+1} \,|\, y_2^\diamond, \ldots, y_{i-1}^\diamond, y_i^0) - C(Y_{i+1}) \end{bmatrix}$$

and

$$\mathbf{M}_{i+1} = \begin{bmatrix} y_2^0 - C(Y_2) & C(Y_3 \,|\, y_2^0) - C(Y_3) & C(Y_4 \,|\, y_2^0) - C(Y_4) & \ldots & C(Y_i \,|\, y_2^0) - C(Y_i) \\ 0 & y_3^0 - C(Y_3) & C(Y_4 \,|\, y_2^\diamond, y_3^0) - C(Y_4) & \ldots & C(Y_i \,|\, y_2^\diamond, y_3^0) - C(Y_i) \\ 0 & 0 & y_4^0 - C(Y_4) & \ldots & C(Y_i \,|\, y_2^\diamond, y_3^\diamond, y_4^0) - C(Y_i) \\ \vdots & \vdots & \ddots & \ddots & \\ 0 & 0 & \ldots & 0 & y_i^0 - C(Y_i) \end{bmatrix}.$$

Our choice of $y_2^\diamond, \ldots, y_{k-1}^\diamond$ results in a slightly different definition of $\mathbf{M}_{i+1}$ to that of Kadane *et al.* (1980). However, as $\mathbf{M}_{i+1}$ is an upper-triangular matrix with non-zero diagonal elements, it is invertible and easy to invert. To fully determine $\mathbf{U}_{i+1}$ we define

$$S(Y_{i+1}) = S(Y_{i+1} \,|\, Y_2, \ldots, Y_i) + \mathbf{g}'_{i+1}\mathbf{U}_i\mathbf{g}_{i+1}.$$

The process of determining $\mathbf{U}_{i+1}$ from $\mathbf{U}_i$ is repeated until $\mathbf{U}_k$ has been obtained. As $S(Y_{i+1} \,|\, Y_2, \ldots, Y_i) > 0$ for $i = 2, \ldots, k-1$, it follows that $\mathbf{U}_k$ is a positive-definite matrix (see Kadane *et al.*, 1980). We then put $\boldsymbol{\Sigma}_{k/1} = \mathbf{U}_k$.

Feedback: Obtaining unconditional quartiles from $\boldsymbol{\mu}_{k/1}$ and $\boldsymbol{\Sigma}_{k/1}$

After $\boldsymbol{\mu}_{k/1}$ and $\boldsymbol{\Sigma}_{k/1}$ have been determined from the expert's assessments, we generate a sample of 100000 observations from MVN$(\boldsymbol{\mu}_{k/1}, \boldsymbol{\Sigma}_{k/1})$. Each observation is a value of $(Y_2, \ldots, Y_k)$, and this is converted into an observation of $(p_1, p_2, \ldots, p_k)$ by using equation (1). This yields 100000 observations of each p_i and from these we calculate the sample median and sample quartiles of p_i $(i = 1, \ldots, k)$.

1. If this is the first time a feedback screen is being presented, then we simply show the medians and sample quartiles on the feedback screen (as in Fig. 4).

2. Suppose a feedback screen has previously been presented and that we are currently choosing a median and quartiles to present for the ith category. Let q_1, q_2 and q_3 denote the lower quartile, median and upper quartile of p_i on the previous feedback screen, and let q_1^*, q_2^* and q_3^* denote the corresponding quartiles given by the sample of 100000 observations. If q_1^{**}, q_2^{**} and q_3^{**} denote the quartiles to be shown on the current feedback screen, we choose these values so that

$$q_2^{**} = q_2^* \tag{28}$$

$$\frac{\text{logit}(q_2^{**}) - \text{logit}(q_1^{**})}{\text{logit}(q_3^{**}) - \text{logit}(q_2^{**})} = \frac{\text{logit}(q_2) - \text{logit}(q_1)}{\text{logit}(q_3) - \text{logit}(q_2)} \tag{29}$$

and

$$\log(q_3^{**}) - \log(q_1^{**}) = \log(q_3^*) - \log(q_1^*). \tag{30}$$

Equation (28) implies that the median shown to the expert is the median given by the random sample. Equation (29) means that we mimic previous asymmetry in quartiles about the median when values are considered on the logit scale. (A benefit of the logit scale is that its range is unbounded.) Equation (30) implies that, on the log scale, the interquartile range presented to the expert matches the interquartile range given by the random sample. While q_1^{**}, q_2^{**} and q_3^{**} give the same values for the mean and variance of Y as those given by q_1^*, q_2^* and q_3^*, we believe that q_1^{**}, q_2^{**} and q_3^{**} are more likely to be acceptable to the expert as representative of his opinion. Values of q_1^{**} and q_3^{**} that satisfy (29) and (30) are found by a simple numeric search.

Feedback: Revising $\boldsymbol{\mu}_{k/1}$ and $\boldsymbol{\Sigma}_{k/1}$ following re-assessment of medians or quartiles

We obtain $\boldsymbol{\mu}_{k/1}$ and $\boldsymbol{\Sigma}_{k/1}$ from quantiles of the p_i by estimating the parameters of the distribution of $\boldsymbol{T} = (T_1, \ldots, T_k)'$. As noted in Section 3.3, $\boldsymbol{T}$ follows an MVN distribution with mean $\boldsymbol{\Gamma} = (\gamma_1, \ldots, \gamma_k)' = \mathbf{H}\boldsymbol{\mu}_{k/1}$ and variance $\mathbf{A\Omega A} = \mathbf{H}\boldsymbol{\Sigma}_{k/1}\mathbf{H}'$, where $\mathbf{H}$ is known and $\boldsymbol{\Omega}$ is fixed by the expert's earlier assessments. We must estimate $\boldsymbol{\Gamma}$ and $\mathbf{A} = \text{diag}(a_1, \ldots, a_k)$.

To draw the feedback screen on which the expert made his re-assessments, a sample of 100000 observations was generated from the estimated distribution of $\boldsymbol{T}$. Let $\mathbf{A}^* = \text{diag}(a_1^*, \ldots, a_k^*)$ denote the estimate of $\mathbf{A}$ that was used for generating these observations. The sample yields 100000 values for each p_i, from which we estimate the lower quartile, median and upper quartile of each p_i's marginal distribution, denoting the estimates for p_i by q_{i1}^*, q_{i2}^* and q_{i3}^*, respectively ($i = 1, \ldots, k$). For each i, we also calculate the sample mean of $\log(p_i)$, denoting its value by τ_i^*. The expert's re-assessment of the lower quartile, median and upper quartile of p_i are denoted by q_{i1}, q_{i2} and q_{i3} ($i = 1, \ldots, k$).

To obtain a revised estimate of $\boldsymbol{\Gamma}$, we require the expert's mean estimate of $\log(p_i)$ for $i = 1, \ldots, k$. We cannot ask the expert to directly assess its value, as the logarithm of a probability is not something to which people can relate. However, the transformation from p_i to $\log(p_i)$ is a monotonic transformation and the expert has assessed q_{i2} as the median of p_i. Hence, we have $\log(q_{i2})$ as the expert's median of $\log(p_i)$. Medians and means are closely related and, moreover, $\log(q^*_{i2})$ and τ^*_i [the median and mean of $\log(p_i)$ in the generated sample] provide information about their relationship in the ith category. For $i = 1, \dots, k$, we take

$$\tau_i = \frac{\tau^*_i}{\log(q^*_{i2})} \log(q_{i2}) \tag{31}$$

as the expert's mean of $\log(p_i)$. From equation (13), $T_i = \log(p_i) - \frac{1}{k}\sum_{j=1}^{k} \log(p_j)$ so, for $i = 1, \dots, k$, we have

$$\gamma_i = E(T_i) = \tau_i - \frac{1}{k}\sum_{j=1}^{k} \tau_j.$$

The γ_i automatically satisfy the constraint, $\sum_{i=1}^{k} \gamma_i = 0$.

To obtain a revised estimate of a_i, we assume that, for minor changes in the expert's opinions, the standard deviation of $[\log(p_i) - \frac{1}{k}\sum_{j=1}^{k} \log(p_j)]$ is approximately proportional to the standard deviation of $\log(p_i)$. As $\mathrm{Var}(T_i) = a_i^2$, it follows that a_i is approximately proportional to the interquartile of the marginal distribution of $\log(p_i)$. The constant of proportionality can be estimated from the fact that $\log(q^*_{i3}) - \log(q^*_{i1})$ was the sample interquartile range of $\log(p_i)$ when the sample was generated with $\mathbf{A}$ set equal to $\mathrm{diag}(a^*_1, \dots, a^*_k)$. The expert's revised interquartile range for $\log(p_i)$ is $\log(q_{i3}) - \log(q_{i1})$, so we take

$$\frac{\log(q_{i3}) - \log(q_{i1})}{\log(q^*_{i3}) - \log(q^*_{i1})} a^*_i$$

as the revised value of a_i for $i = 1, \dots, k$.

To obtain $\boldsymbol{\mu}_{k/1}$ and $\boldsymbol{\Sigma}_{k/1}$ from $\boldsymbol{\Gamma}$ and $\mathbf{A\Omega A}$, define the $(k-1) \times k$ matrix $\widetilde{\mathbf{H}}$ by

$$\widetilde{\mathbf{H}} = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 \\ -1 & 0 & 1 & \dots & 0 \\ \vdots & & & \ddots & \\ -1 & 0 & 0 & \dots & 1 \end{bmatrix}.$$

That is, each element of the first column of $\widetilde{\mathbf{H}}$ is -1, while its other $k-1$ columns form $\mathbf{I}_{(k-1)}$, the $(k-1) \times (k-1)$ identity matrix. Matrix multiplication shows the $\widetilde{\mathbf{H}}\mathbf{H} = \mathbf{I}_{(k-1)}$, so we put $\boldsymbol{\mu}_{k/1} = \widetilde{\mathbf{H}}\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}_{k/1} = \widetilde{\mathbf{H}}\mathbf{A\Omega A}\widetilde{\mathbf{H}}'$.

Our approach for estimating $\boldsymbol{\mu}_{k/1}$ and $\boldsymbol{\Sigma}_{k/1}$ from the expert's quantile reassessments is somewhat *ad hoc*. However, the mechanism for determining quantiles of the p_i from $\boldsymbol{\mu}_{k/1}$ and $\boldsymbol{\Sigma}_{k/1}$ is exact (as long as the sample size is large). We repeatedly update the feedback screen until the expert finds the suggested quantiles of the p_i are an adequate representation of his opinions. At that point we adopt the estimates of $\boldsymbol{\mu}_{k/1}$ and $\boldsymbol{\Sigma}_{k/1}$ that generated the acceptable quantiles – we do not recalculate $\boldsymbol{\mu}_{k/1}$ and $\boldsymbol{\Sigma}_{k/1}$, so the *ad hoc* nature of that calculation does not impact directly on their estimates.

References

Agresti, A. (2002). *Categorical Data Analysis*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., New Jersey, second edition.

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. Chapman and Hall, London.

Al-Awadhi, S. A. and Garthwaite, P. H. (2006). Quantifying expert opinion for modelling fauna habitat distributions. *Computational Statistics*, **21**, 121–140.

Bedrick, E. J., Christensen, R., and Johnson, W. (1996). A new perspective on priors for generalized linear models. *Journal of the American Statistical Association*, **91**, 1450–1460.

Bergstraesser, E., Paul, M., Rufibach, K., Hain, R. D., and Held, L. (2014). The paediatric palliative screening scale: Further validity testing. *Palliative Medicine*, **28**, 530–533.

Cooke, R. M. and Goossens, L. L. H. J. (2008). TU Delft expert judgement database. *Reliability Engineering and System Safety*, **93**, 657–674.

Edwards, W. (1962). Dynamic decision theory and probabilistic information processings. *Human Factors*, **4**, 59–74.

Edwards, W., Phillips, L. D., Hays, W. L., and Goodman, B. C. (1968). Probabilistic information processing systems: Design and evaluation. *IEEE Transactions on Systems Science and Cybernetics*, **SSC-4**, 248–265.

Elfadaly, F. G. and Garthwaite, P. H. (2013). Eliciting Dirichlet and Connor–Mosimann prior distributions for multinomial models. *Test*, **22**, 628–646.

Elfadaly, F. G. and Garthwaite, P. H. (2017). Eliciting Dirichlet and Gaussian copula prior distributions for multinomial models. *Statistics and Computing*, **27**, 449–467.

Evans, M., Guttman, I., and Li, P. (2017). Prior elicitation, assessment and inference with a Dirichlet prior. *Entropy*, **19:564**, doi:10.3390/e19100564.

French, S. (2011). Aggregating expert judgement. *Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales. Serie A. Matematicas*, **105**, 181–206.

Garthwaite, P. H., Kadane, J. B., and O'Hagan, A. (2005). Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association*, **100**, 680–701.

Garthwaite, P. H., Al-Awadhi, S. A., Elfadaly, F. G., and Jenkinson, D. J. (2013). Prior distribution elicitation for generalized linear and piecewise-linear models. *Journal of Applied Statistics*, **40**, 59–75.

Gosling, J. P. (2018). SHELF: the sheffield elicitation framework. In L. C. Dias, A. Morton, and J. Quigley, editors, *Elicitation: The Science and Art of Structuring Judgement*. Springer, New York.

Hampson, L. V., Whitehead, J., Eleftheriou, D., Tudur-Smith, C., Jones, R., Jayne, D., Hickey, H., Beresford, M. W., Bracaglia, C., Caldas, A., Cimaz, R., Dehoorne, J., Dolezalova, P., Friswell, M., Jelusic, M., Marks, S. D., Martin, N., McMahon, A.-M., Peitz, J., van Royen-Kerkhof, A., Soylemezoglu, O., and Brogan, P. A. (2015). Elicitation of expert prior opinion: Application to the MYPAN trial in childhood polyarteritis nodosa. *PLoS ONE*, **10**, doi:10.1371/journal.pone.0120981.

Held, L. and Sauter, R. (2017). Adaptive prior weighting in generalized regression. *Biometrics*, **73**, 242–251.

Held, U., Bové, D. S., Steurer, J., and Held, L. (2012). Validating and updating a risk model for pneumonia – a case study. *BMC Medical Research Methodology*, **12:99**, 1–8.

Kadane, J. B., Dickey, J. M., Winkler, R., Smith, W., and Peters, S. (1980). Interactive elicitation of opinion for a normal linear model. *Journal of the American Statistical Association*, **75**, 845–854.

Kaplan, R. J. and Newman, J. R. (1966). Studies in probabilistic information processing. *IEEE Transactions on Human Factors in Electronics*, **HFE-7**, 49–63.

Lunn, D., Thomas, A., Best, N., and Spiegelhalter, D. (2000). WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, **10**, 325–337.

Lunn, D., Spiegelhalter, D., Thomas, A., and Best, N. (2009). The BUGS project: Evolution, critique and future directions (with discussion). *Statistics in Medicine*, **28**, 3049–3082.

Ministry for Social Policy (2008). *European Health Interview Survey 2008*. Department of Health Information and Research, Malta.

Miettinen, O. S., Flegel, K. M., and Steurer, J. (2008). Clinical diagnosis of pneumonia, typical of experts. *Journal of Evaluation in Clinical Practice*, **14**, 343–350.

Murdoch, D. J. and Chow, E. D. (1996). A graphical display of large correlation matrices. *The American Statistician*, **50**, 178–180.

Németh, B., Molnár, A., Akehurst, R., Horváth, M., Kóczián, K., Németh, G., Götze, A., and Vokó, Z. (2017). Quality-adjusted life year difference in patients with predominant negative symptoms of schizophrenia treated with cariprazine and risperidone. *Journal of Comparative Effectiveness Research*, **6**, 639–648.

O'Hagan, A. and Forster, J. (2004). B*ayesian Inference*, volume 2B of *Kendall's Advanced Theory of Statistics*. Arnold, London, second edition.

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E., and Rakow, T. (2006). *Uncertain Judgements: Eliciting Expert Probabilities*. John Wiley, Chichester.

Plummer, M. (2015). rjags: Bayesian graphical models using MCMC. R package version 3-15.

Por, H.-H. and Budescu, D. V. (2017). Eliciting subjective probabilities through pair-wise comparisons. *Journal of Behavioral Decision Making*, **30**, 181–196.

Schum, D. A., Goldstein, I. L., and Southard, J. F. (1966). Research on a simulated Bayesian information-processing system. *IEEE Transactions on Human Factors in Electronics*, **HFE-7**, 37–48.

Shardell, M., Hicks, G. E., Miller, R. R., and Magaziner, J. (2010). Semiparametric regression models for repeated measures of mortal cohorts with non-monotone missing outcomes and time-dependent covariates. *Statistics in Medicine*, **29**, 2282–2296.

Shepherd, B. E., Redman, M. W., and Ankerst, D. P. (2008). Does finasteride affect the severity of prostate cancer? A causal sensitivity analysis. *Journal of the American Statistical Association*, **103**, 1392–1404.

Van Calster, B., Van Hoorde, K., Vergouwe, Y., Bobdiwala, S., Condous, G., Kirk, E., Bourne, T., and Steyerberg, E. W. (2017). Validation and updating of risk models based on multinomial logistic regression. *Diagnostic and Prognostic Research*, **1:2**, 1–14.

Van Hoorde, K., Vergouwe, Y., Timmerman, D., Van Huffel, S., Steyerberg, E. W., and Van Calster, B. (2014). Assessing calibration of multinomial risk prediction models. *Statistics in Medicine*, **33**, 2585–2596.

Werner, C., Bedford, T., Cooke, R. M., Hanea, A. M., and Morales-Nápoles, O. (2017). Expert judgement for dependence in probabilistic modelling: A systematic literature review and future research directions. *European Journal of Operational Research*, **258**, 801–819.

Wilson, E. C. F., Usher-Smith, J. A., Emery, J., Corrie, P. G., and Walter, F. M. (2017). Expert elicitation of multinomial probabilities for decision-analytic modeling: an application to rates of disease progression in undiagnosed and untreated melanoma. *Value in Health*. dx.doi.org/10.1016/j.jval.2017.10.009.

Wilson, K. J. (2017). Specification of informative prior distributions for multinomial models using vine copulas. *Bayesian Analysis*. doi:10.1214/17-BA1068.

Winkler, R. L. (1981). Combining probability distributions from dependent information sources. *Management Science*, **27**, 479–488.

Zapata-Vázquez, R. E., O'Hagan, A., and Bastos, L. S. (2014). Eliciting expert judgements about a set of proportions. *Journal of Applied Statistics*, **41**, 1919–1933.