Elicitation of Opinion about Multinomial Models

Paul Garthwaite

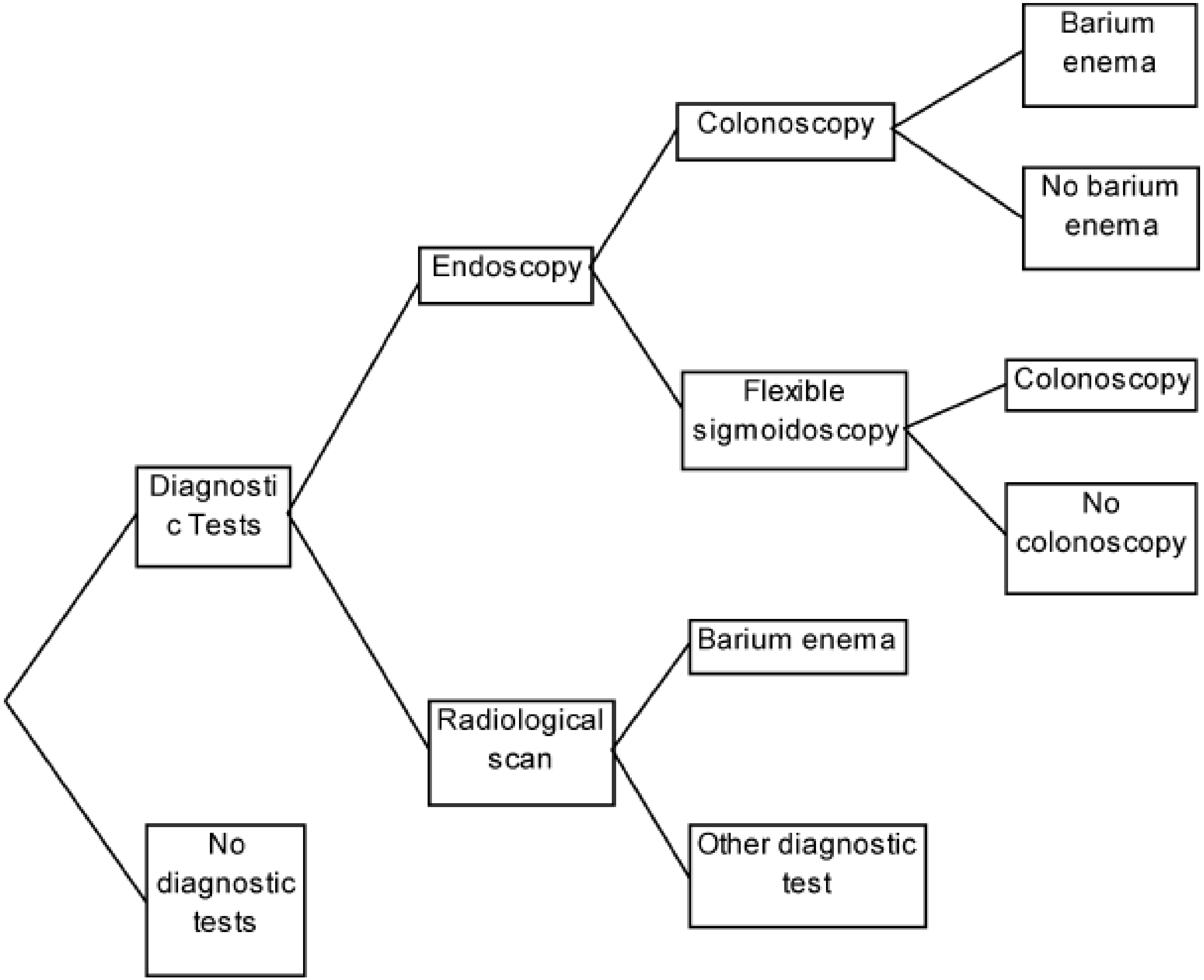
Open University, UK

(Joint work with Fadlalla Elfadaly)

Initial motivation:
Filling in gaps in a treatment pathway model

- The UK National Health Service (NHS) initiated a study to estimate the benefits of current bowel cancer services in England and examine costs and benefits of alternative developments in service provision.
- A treatment pathway model was developed that gave the possible sequences of presentation, diagnosis, treatment, and outcomes that could be followed by a patient with suspected colorectal cancer.
- Model parameters had to be specified that gave the probabilities or probability distributions governing the path taken at each branch of the pathway model.
- The majority of information required for the study could be quantified from available data sources.
- For some quantities, however, information was only available in the background knowledge and experience of experts.

Pathway for diagnostic tests



In the pathway model, some of the nodes we considered led to four alternatives. However, our software used logistic regression to model expert opinion so we separated the node into two nodes and had two alternatives from each. The pairings made sense so that seemed OK. But clearly a method was needed for situations with more than two alternatives – we needed a method of quantifying opinion about multinomial models.

In the multinomial model we have a number of categories.

Each observation will be in exactly one category and expert opinion must:

- provide an estimate of the probability of each category
- quantify the accuracy of the estimates.

Perhaps we might also want the expert to quantify the correlation between his or her assessments.

There may also be covariates in the model.

With multivariate problems it seems essential to model the expert’s opinion by a parametric distribution.

Then the task of quantifying the expert’s opinion reduces to choosing parameter values that approximate to her opinion.

We will model opinion as

- A Dirichlet distribution
- A Connor-Mosimann distribution
- A Gaussian copula prior
- A multinomial logistic prior.

The Dirichlet distribution is the conjugate prior distribution for a multinomial model, so that using it is easy.

Sampling model: $\mathbf{X} = (X_1, \ldots, X_k)$ is multinomially distributed with k categories and probabilities $\mathbf{p} = (p_1, \ldots, p_k)$ and

$$f(x_1, x_2, \ldots, x_k) = \frac{n!}{x_1! x_2! \ldots x_k!} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k}.$$

Dirichlet prior:

$$\pi(p_1, p_2, \ldots, p_k) = \frac{\Gamma(N)}{\Gamma(a_1)\Gamma(a_2)\ldots\Gamma(a_k)} p_1^{a_1-1} p_2^{a_2-1} \ldots p_k^{a_k-1},$$

where $N = \sum a_i$ and $a_i > 0$.

Methods of assessing Dirichlt priors have been proposed.

Dickey et al (1983) give one method of eliciting its parameters:

Estimates of $p_1, \ldots p_k$ are elicited and reconciled to sum to 1. Hypothetical data is given to the expert who then revises her assessments. Small revision gives a large value for N while large revision gives a small N.

Under the Dirichlet prior, the marginal distribution of each p_i is a beta distribution.

This is exploited in other methods of assessing the Dirichlet parameters (e.g. Chaloner and Duncan (1987)).

SHELF (Sheffield Elicitation Framework)

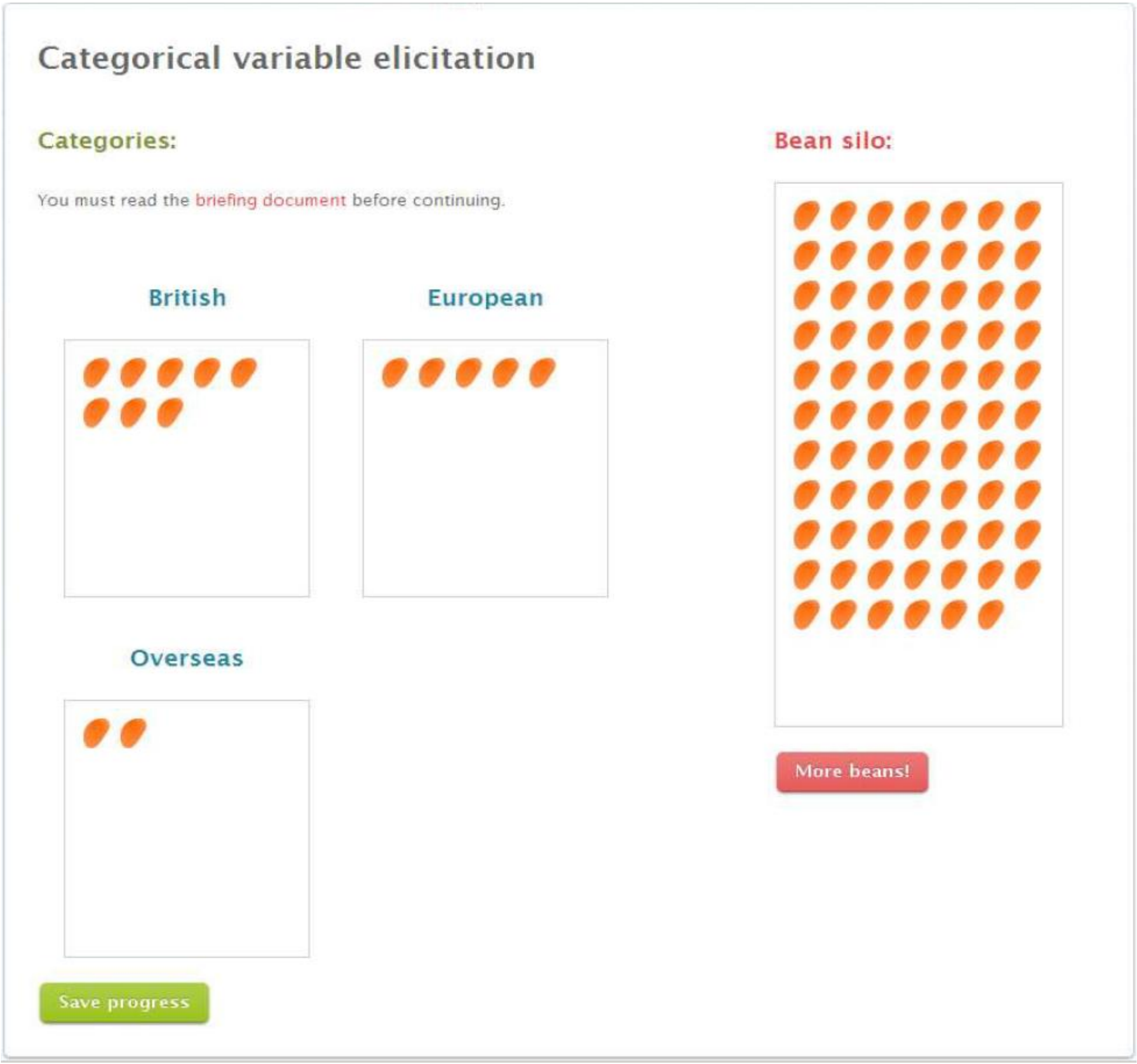
O'Hagan and Oakley have software to carry out elicitation of probability distributions (aimed particularly at quantifying uncertainty from a group of experts).

The univariate distributions it uses to model opinion are:

Normal, Student *t*, scaled beta, gamma, log-normal and log Student-*t*.

An extension quantifies opinion about a multinomial distribution by first eliciting marginal (beta) probabilities for each category, and then reconciling them to form a Dirichlet distribution.

Offers a choice of assessment tasks for quantifying probabilities: quartiles, tertiles, fixed interval, roulette.

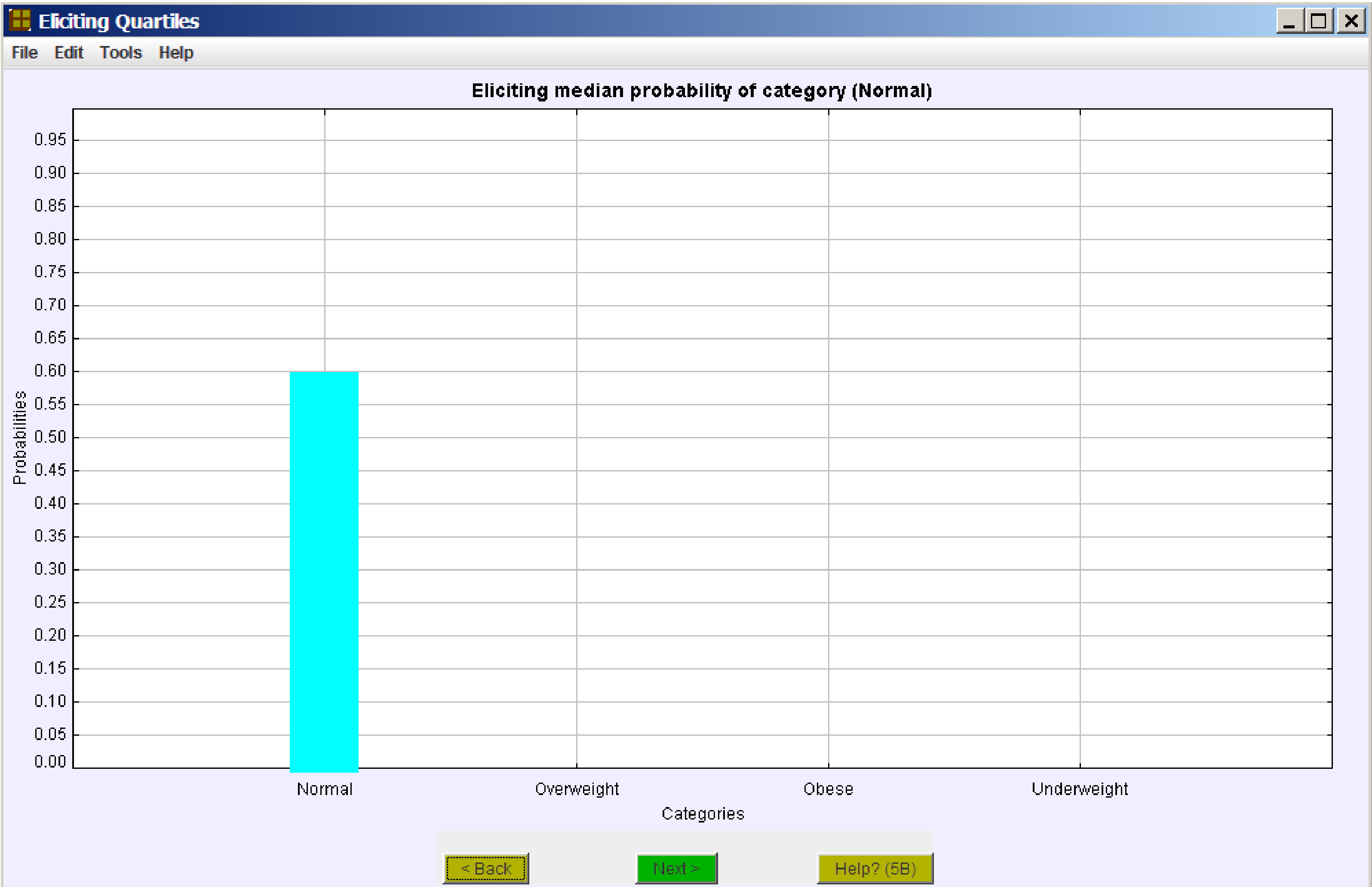

Categorical variable elicitation
Categories:
You must read the briefing document before continuing.
British
European
Overseas
Save progress
Bean silo:
More beans!

Elicitation methods we have developed form a package called *PEGS* (Probability Elicitation using Graphical Software).
An example will be used to describe its method of Quantifying opinion as a Dirichlet distribution.

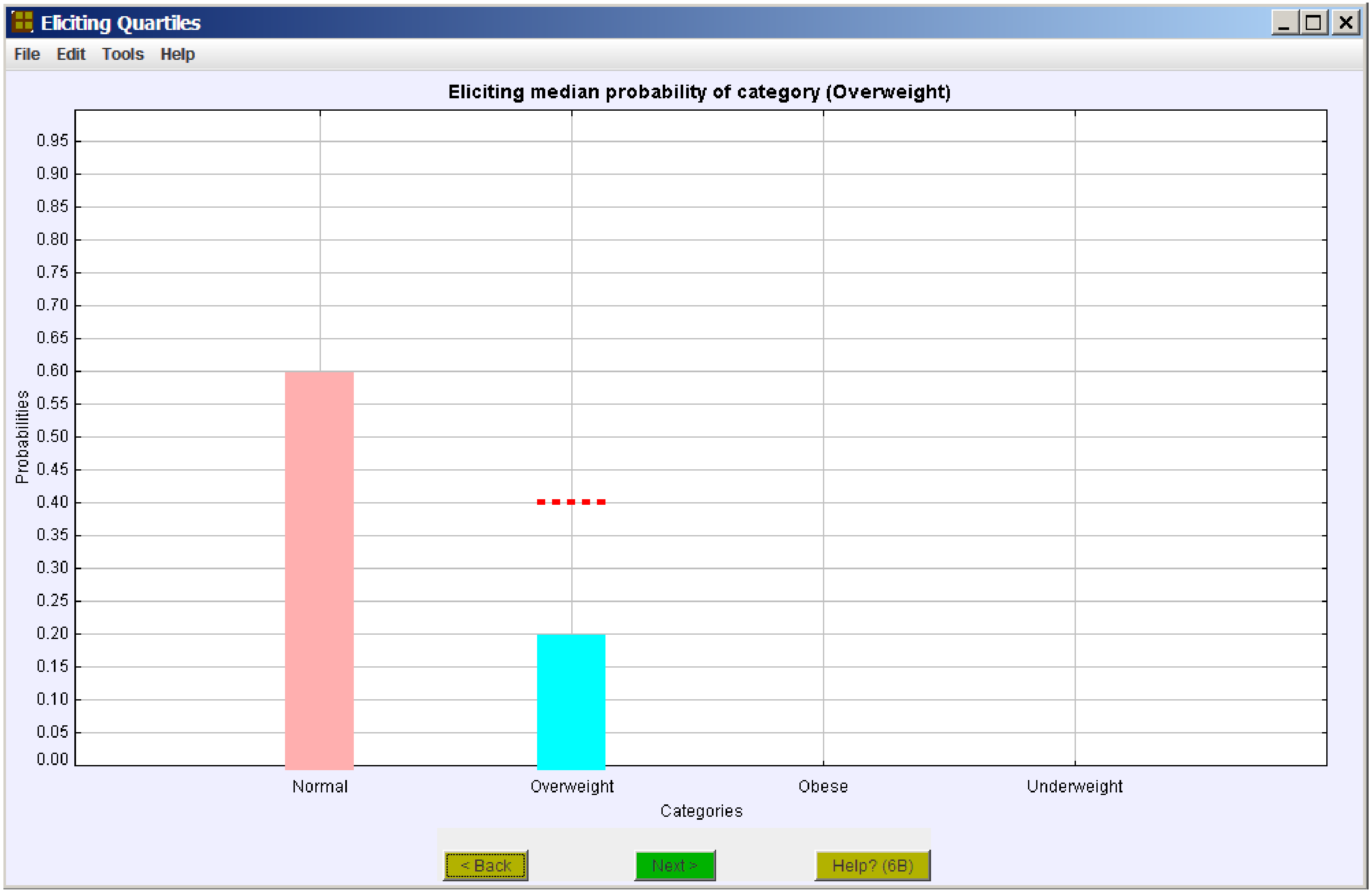
Example: Misclassification rates of BMI

- A person in Malta gives their height and weight in a questionnaire and their calculated BMI is in the *normal* range.
- Their true BMI is in one of the four categories: *normal; overweight; obese; underweight.*
- We want to question an expert to assess the probabilities that the person's true BMI is in each of these categories.

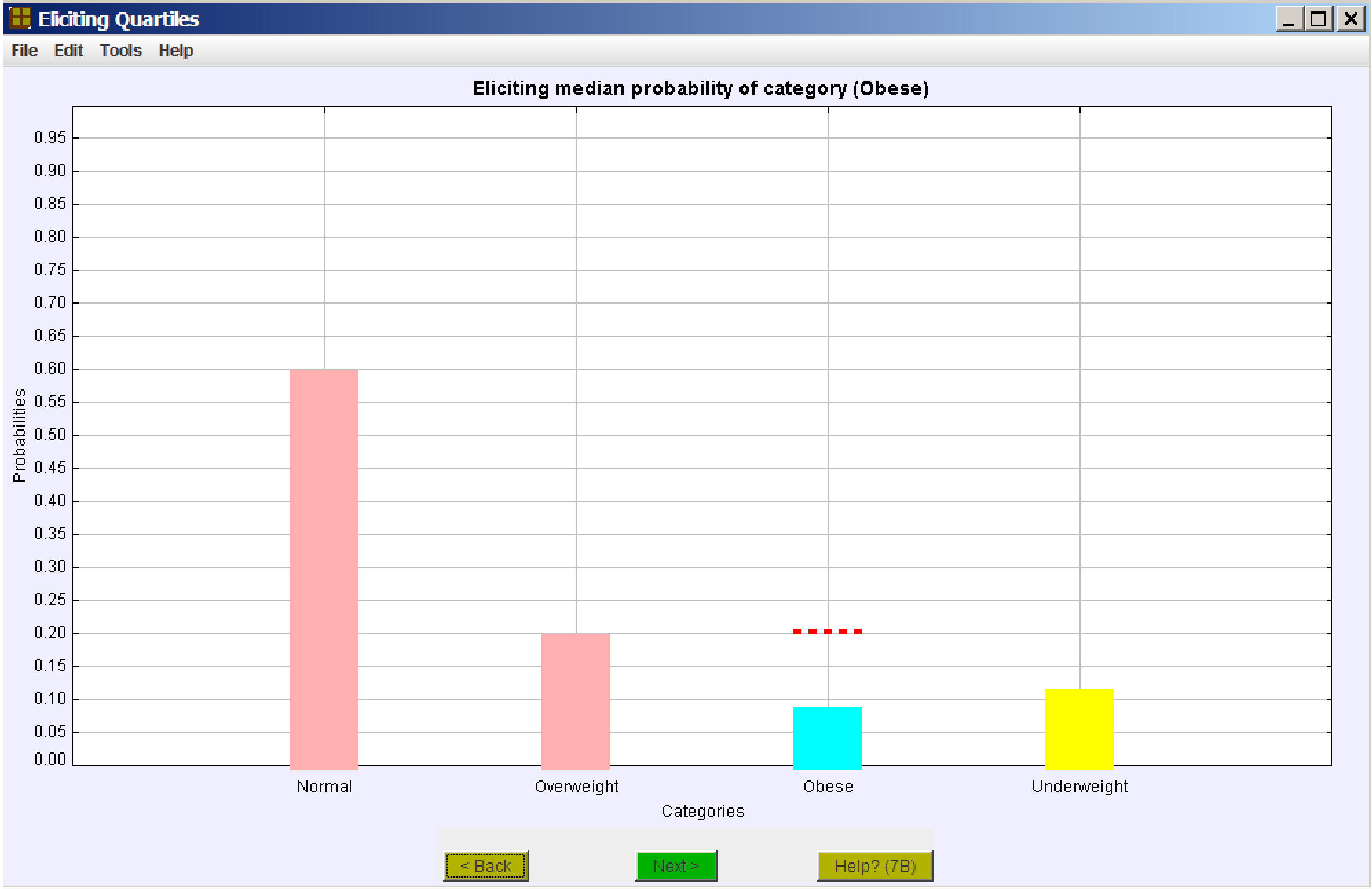
Median assessment of p for the first category



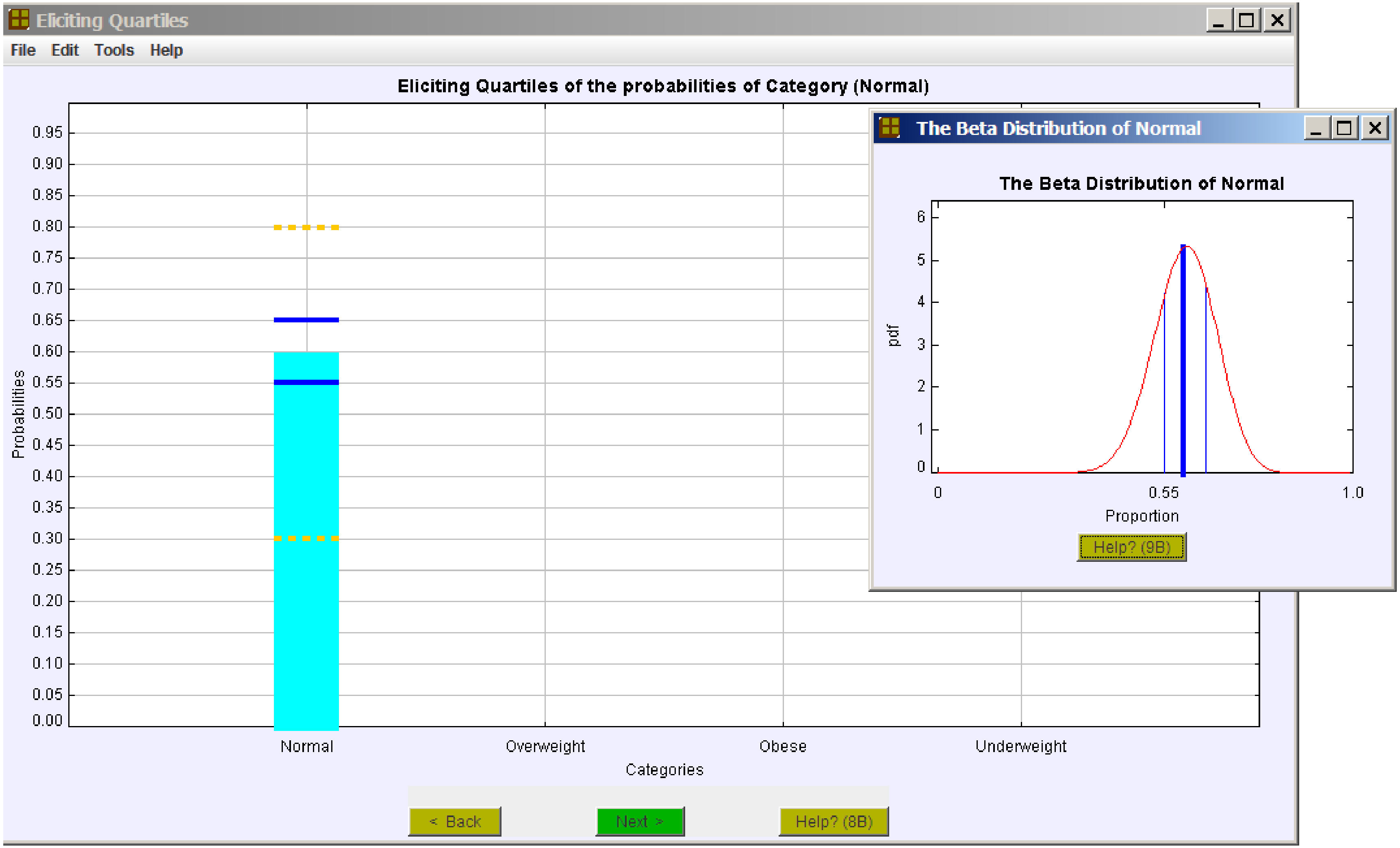
Assessing the probability of *overweight*, given that the red bar is correct.



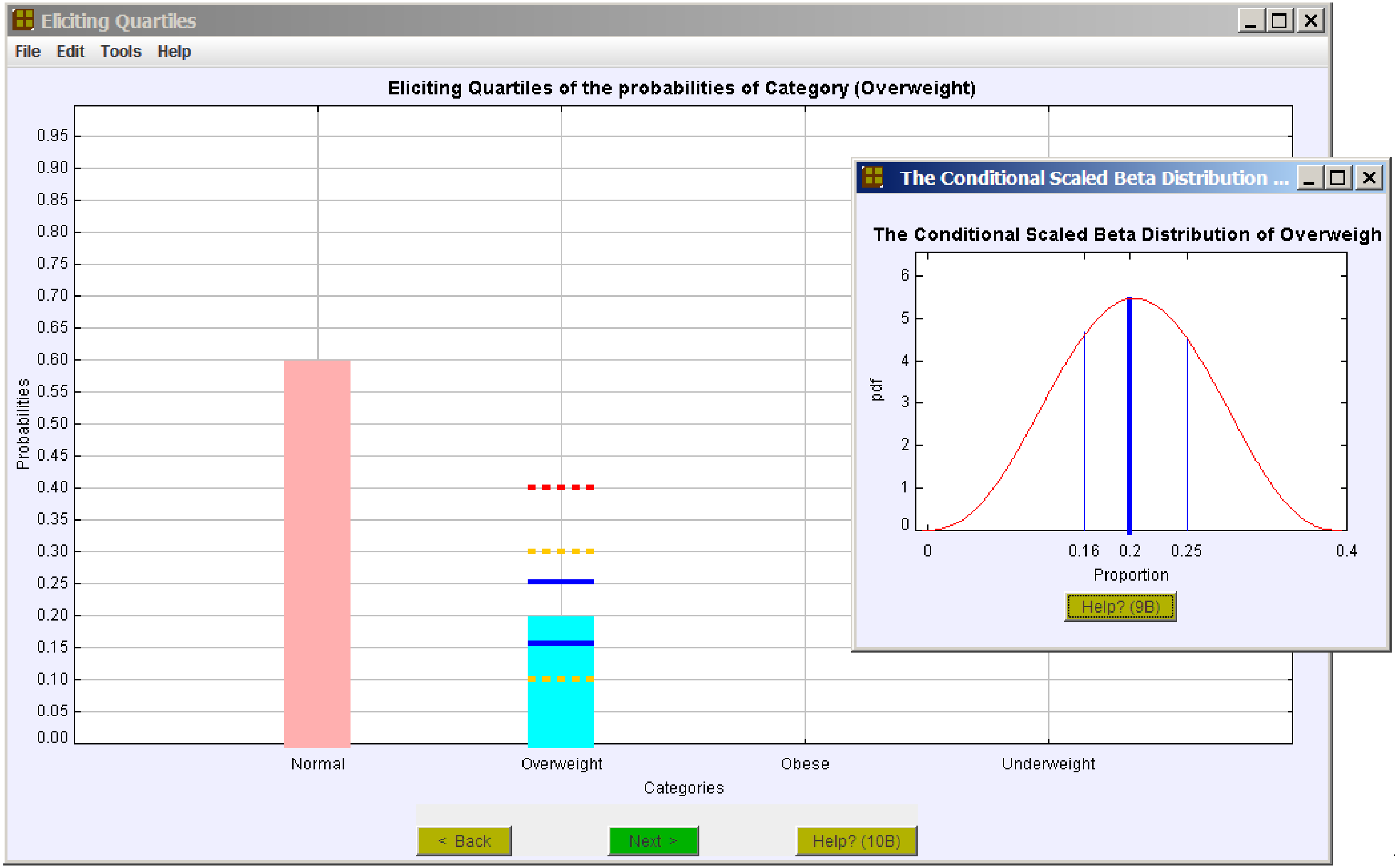
Assessing the probability for obese, conditional on the red categories are correct. (The probability for the yellow category follows automatically.)



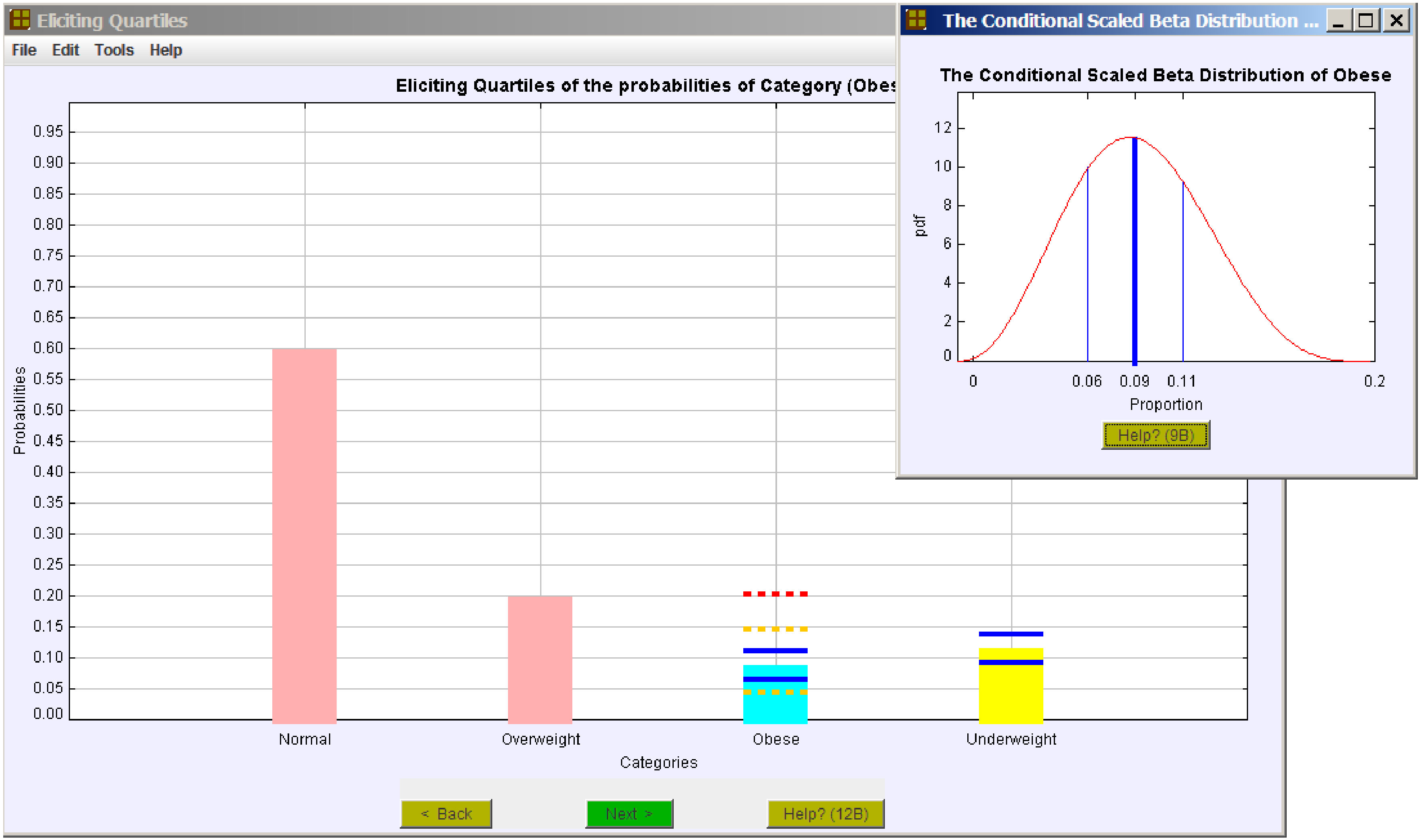
The short blue lines are the expert's lower and upper quartile assessments for the first category. The insert shows the probability density function for the first probability.



Blue lines are the expert's quartile assessments for the probability of *overweight*, conditional that 0.60 is the probability of *normal*.



Quartile assessments for *obese* (also giving those for *underweight*.)



Put

$$p_r^* = \frac{p_r}{1 - \sum_{i=1}^{r-1} p_i}$$

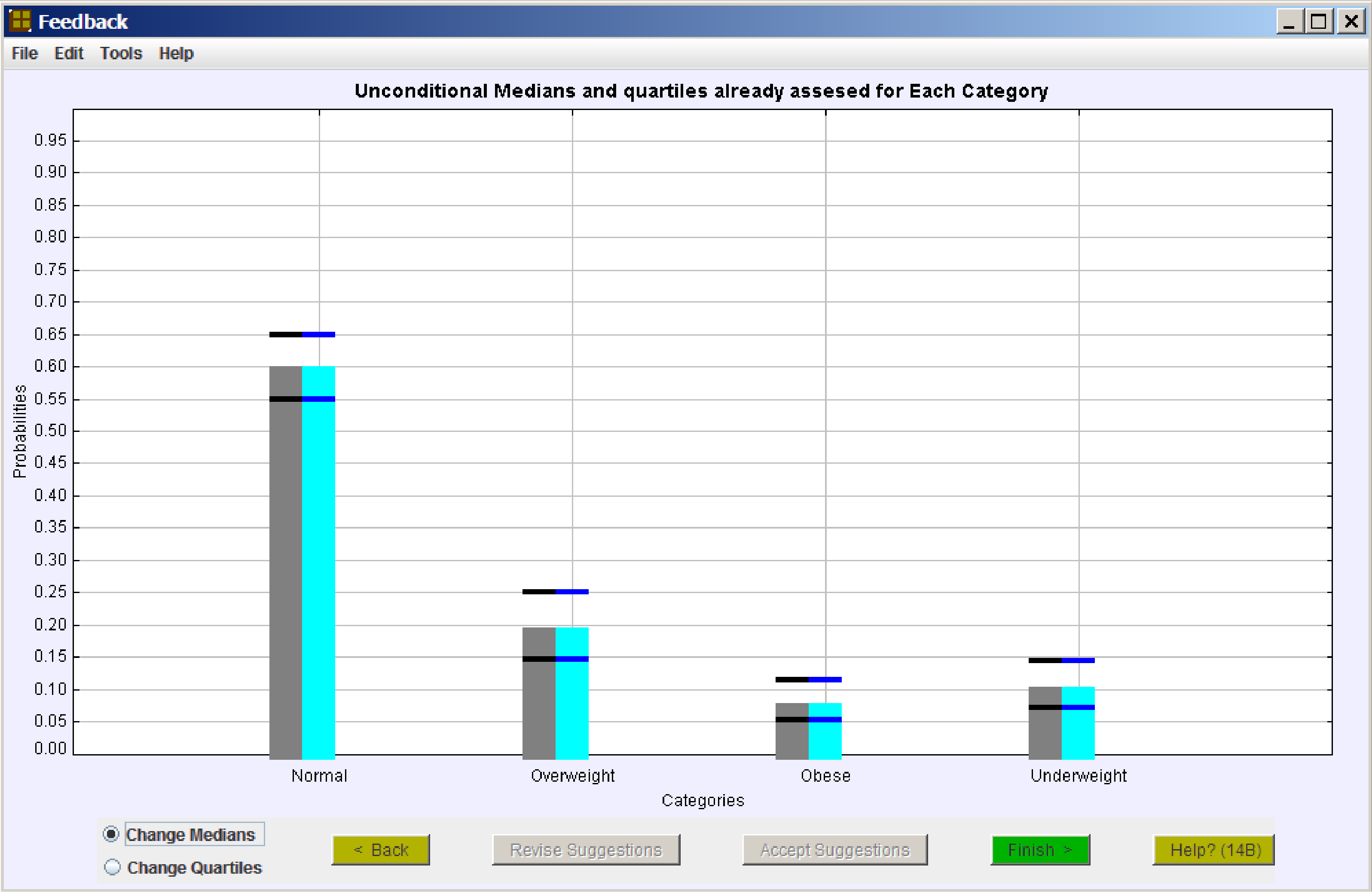
The assessments relate to $p_r^* \mid p_1, \ldots, p_{r-1}$ and

$$p_r^* \mid p_1, \ldots, p_{r-1} \sim \text{beta}(a_r, \sum_{i=r+1}^{k} a_i).$$

We have far more assessments than we need, so we use a form of reconciliation to estimate the a_i's.

We then calculate marginal distributions of the p_i and give a feedback screen to the expert on which he can change the marginal quartiles.

Marginal distributions are shown to the expert as feedback.
The expert can make modifications if (s)he wishes.



A drawback of the Dirichlet prior is that it only has *k* parameters – the same as the number of parameters in the multinomial distribution.

The Connor-Mosimann distribution has almost twice as many parameters (2*k* – 2), so that it should be able to capture a broader range of expert opinion.

The distribution is

$$\pi(p_1,\ldots,p_k) = \prod_{i=1}^{k-1}\left[\frac{\Gamma(a_i+b_i)}{\Gamma(a_i)\Gamma(b_i)}\,p_i^{a_i-1}\left(\sum_{j=i}^{k}p_j\right)^{b_{i-1}-(a_i+b_i)}\right]p_k^{b_{k-1}-1}.$$

It is a conjugate prior distribution for a multinomial distribution.

The same assessments used for the Dirichlet distribution give the Connor-Mosimann distribution (less reconciliation is needed).

- The Dirichlet distribution fits the expert's assessments well, so fitting the more flexible Connor-Mossiman distribution essentially yields the same distribution.

	Assessed median	Standard Dirichlet		Connor-Mossinan	
		$E(p)$	S.D.(p)	$E(p)$	S.D.(p)
Normal	0.65	0.653	0.089	0.653	0.089
Overweight	0.20	0.201	0.076	0.201	0.078
Obese	0.09	0.087	0.053	0.087	0.053
Underweight	0.06	0.058	0.044	0.058	0.042

Nevertheless, the Connor-Mossiman distribution fits the expert's assessments better.

The table compares the quartiles of both elicited priors to the expert's assessments.

Prob.	Lower Quartile			Median			Upper Quartile		
	Exprt	Dirich	C-M	Exprt	Dirich	C-M	Exprt	Dirich	C-M
p_1	0.50	0.54	0.49	0.65	0.65	0.66	0.80	0.75	0.80
$p_2 \mid p_1 = 0.65$	0.17	0.14	0.16	0.20	0.21	0.21	0.25	0.27	0.25
$p_3 \mid p_1=0.65, p_2=0.2$	0.07	0.05	0.07	0.09	0.10	0.09	0.11	0.13	0.11
$p_4 \mid p_1=0.65, p_2=0.2$	0.04	0.02	0.04	0.06	0.06	0.06	0.08	0.10	0.08

The Connor-Mossiman quantiles are closer to the expert's assessments.

Gaussian copula prior distribution

The Connor-Mosimann prior offers a more flexibility means of modelling expert opinion than the Dirichlet prior, but perhaps not enough.

A Gaussian copula prior offers greater flexibility. A copula joins marginal distributions into a joint distribution that has those marginals.

A Gaussian copula is defined at the point $(x_1, \ldots, x_k)$ as

$$C[G_1(x_1), \ldots, G_k(x_k)] = \Phi_{k,\mathbf{R}}\left\{\phi^{-1}[G_1(x_1)], \ldots, \phi^{-1}[G_k(x_k)]\right\}$$

Here $\Phi_{k,\mathbf{R}}$ is the CDF of a *k*-variate normal distribution with zero means, unit variances, and a correlation matrix **R** that reflects the desired dependence structure.

$$C[G_1(x_1),\ldots,G_k(x_k)] = \Phi_{k,\mathbf{R}}\left\{\phi^{-1}[G_1(x_1)],\ldots,\phi^{-1}[G_k(x_k)]\right\}$$

We cannot put $Y_i = \phi^{-1}[G_i(p_i)]$ because then $\sum p_i = 1$ would prevent the Y_i from being normally distributed.

[Conditional distributions would not have an infinite range – given, say, the value of Y_1 implies a value of p_1 and the other p_i would necessarily be less than 1.]

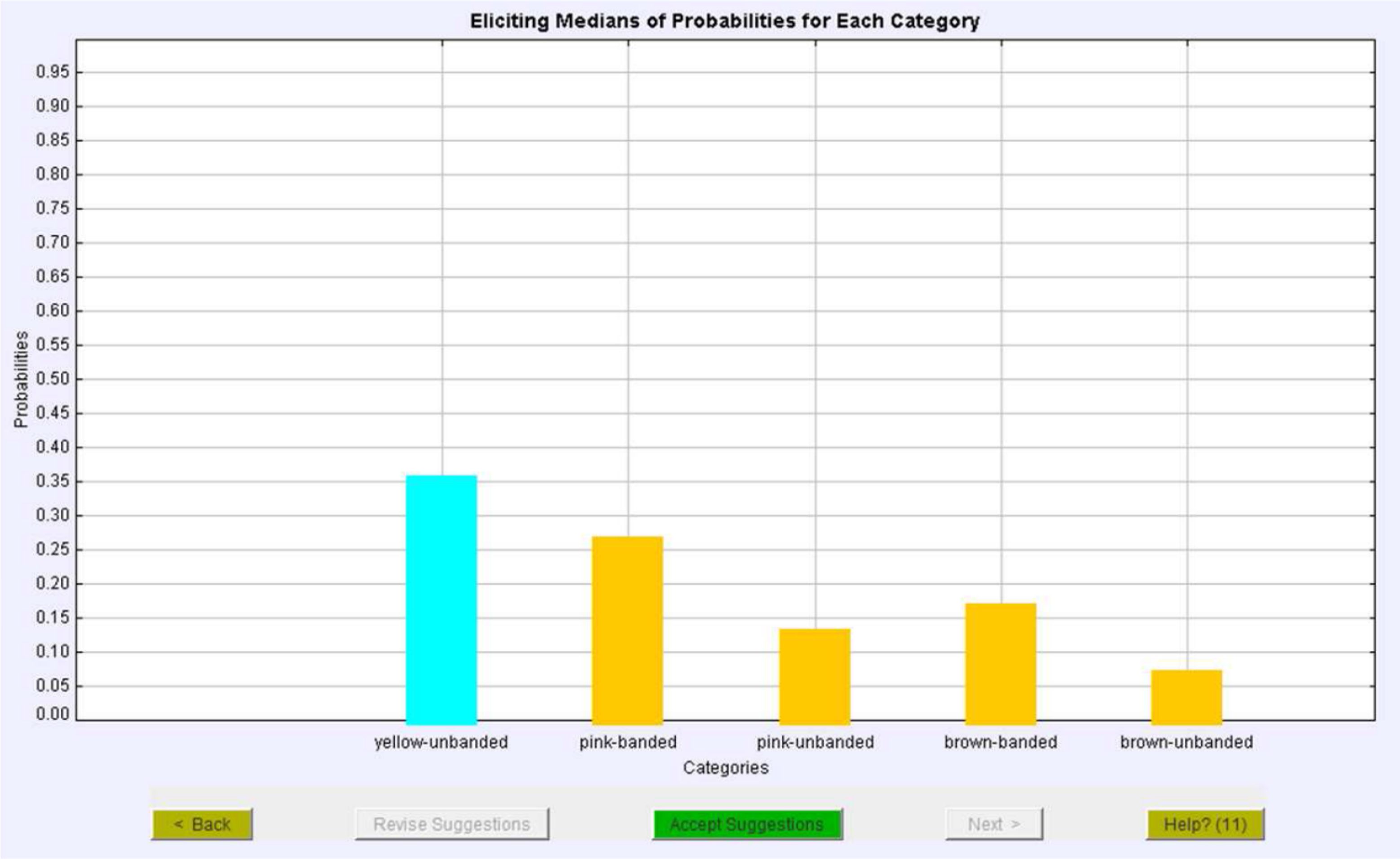
Instead we define new variables $Z_1,\ldots,Z_k$:

$$Z_1 = p_1, \quad Z_i = \frac{p_i}{1-\sum_{j=1}^{i-1} p_j} \text{ for } i = 2,\ldots,k-1, \quad Z_k = 1.$$

Each Z_i has a marginal beta distribution and we put

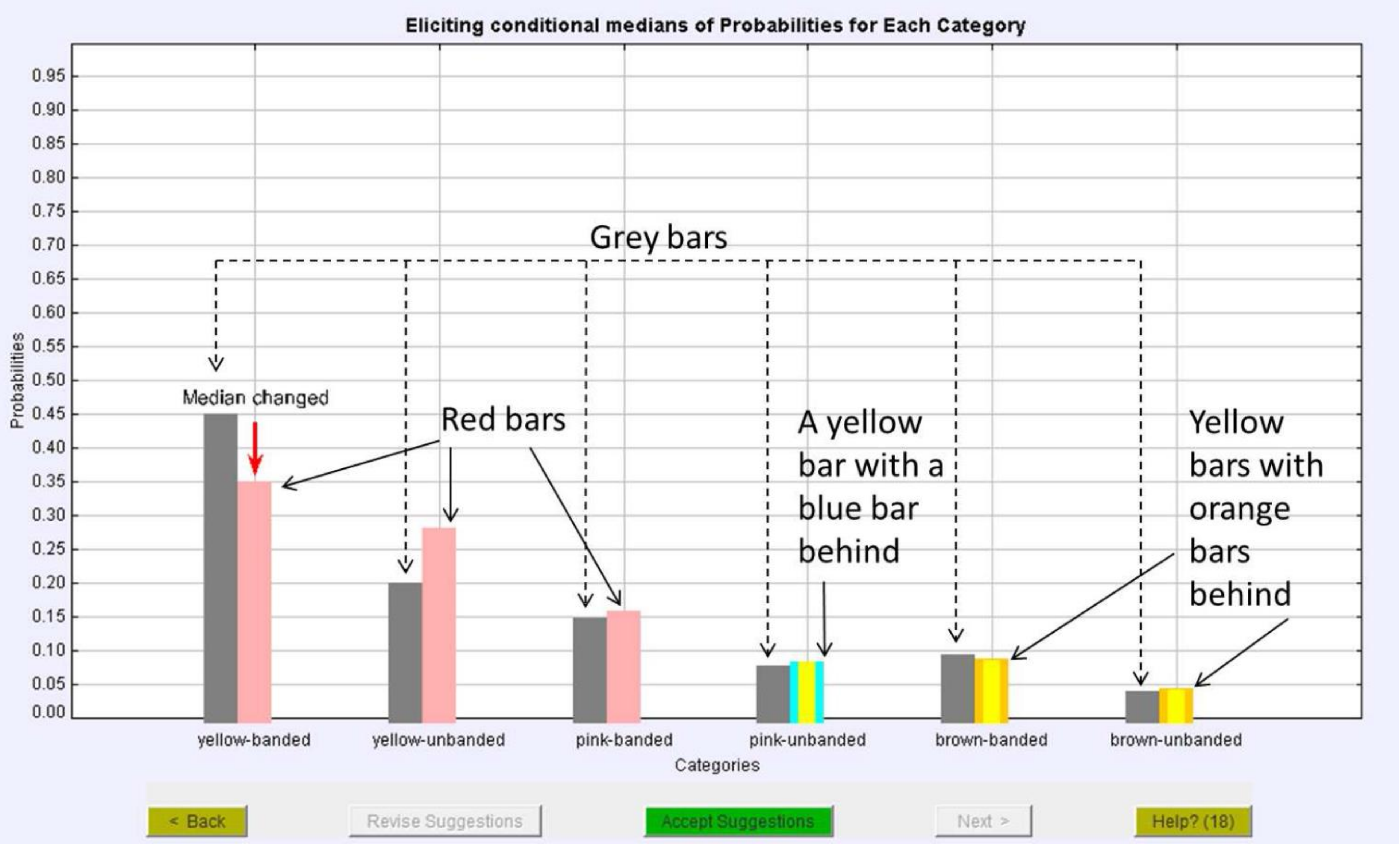
$$Y_i = \phi^{-1}[G_i(Z_i)] \text{ for } i = 1,\ldots,k.$$

Median assessment for the second category. The expert is told: "Assessments are only required for the categories that are labelled and have blue or orange boxes. Assume that you know an item falls in one of these categories.



- Our method for determining the correlation matrix is an adaptation of a method developed Kadane, Dickey et al (1980).
- Their method can be used in many contexts (they developed it for linear regression) and determines the variance-covariance matrix of an MVN or multivariate-*t* distribution.
- It requires a sequence of conditional assessments to quantify the expert's opinions about the relationships between variables.

The grey bars are the expert's unconditional median assessments. The expert is asked to assume that the proportions in the first three boxes are given by the pink boxes and gives assessments for the other boxes.



Multinomial models that contain covariates

- We still have *k* categories, but now membership probabilities depend upon covariates.
- In medical contexts for instance, age and gender will typically affect probabilities.
- After the multinomial model, the best-known sampling model for proportions is the logistic normal distribution, in which proportions are transformed to variables that (by assumption) follow an MVN distribution (Aitchison, 1986).
- This has been considered as a prior (e.g. O'Hagan and Forster (2004)) but methods of eliciting the prior have not previously been proposed.
- The logistic normal model can be extended to include covariates (Aitchison, 1986), yielding the multinomial logistic model (also called the multinomial logit model).

The sampling model.

$$Y_1 = \log\left\{p_1 / (1 - p_{1)}\right\}; \quad Y_i = \log(p_i / p_1) \text{ for } i = 2, \ldots, k$$

Covariates: $Y_i = \alpha_i + \mathbf{x}'\beta_i$

where α_i and $\beta_i = (\beta_{1,i}, \ldots, \beta_{m,i})'$ are the constant and regression coefficients for the *i*th category (i=2,...,k).

This gives

$$p_i(\mathbf{x}) = \begin{cases} \dfrac{1}{1 + \sum_{j=2}^{k} \exp(\alpha_j + \mathbf{x}'\beta_j)}, & i = 1 \\ \dfrac{\exp(\alpha_i + \mathbf{x}'\beta_i)}{1 + \sum_{j=2}^{k} \exp(\alpha_j + \mathbf{x}'\beta_j)}, & i = 2, \ldots, k. \end{cases}$$

Arrange the regression coefficients as a matrix:

$$\mathbf{B} = \left[\begin{pmatrix} \alpha_2 \\ \beta_2 \end{pmatrix}, \ldots, \begin{pmatrix} \alpha_k \\ \beta_k \end{pmatrix} \right].$$

We focus on one *row* of **B** at a time, defining new vectors:

$\alpha = (\alpha_2, \ldots, \alpha_k)'$, $\beta_{(r)} = (\beta_{r,2}, \ldots, \beta_{r,k})'$ for $r = 1, 2, \ldots, m$.

Prior distribution: $(\alpha', \beta_{(1)}', \ldots, \beta_{(m)}')' \sim \text{MVN}(\mu, \Sigma)$
with $\Sigma_{|\alpha}$ block-diagonal.

$$\Sigma_{|\alpha} = \begin{pmatrix} \Sigma_{\beta,1|\alpha} & 0 & \cdots & 0 \\ 0 & \Sigma_{\beta,2|\alpha} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \Sigma_{\beta,m|\alpha} \end{pmatrix}.$$

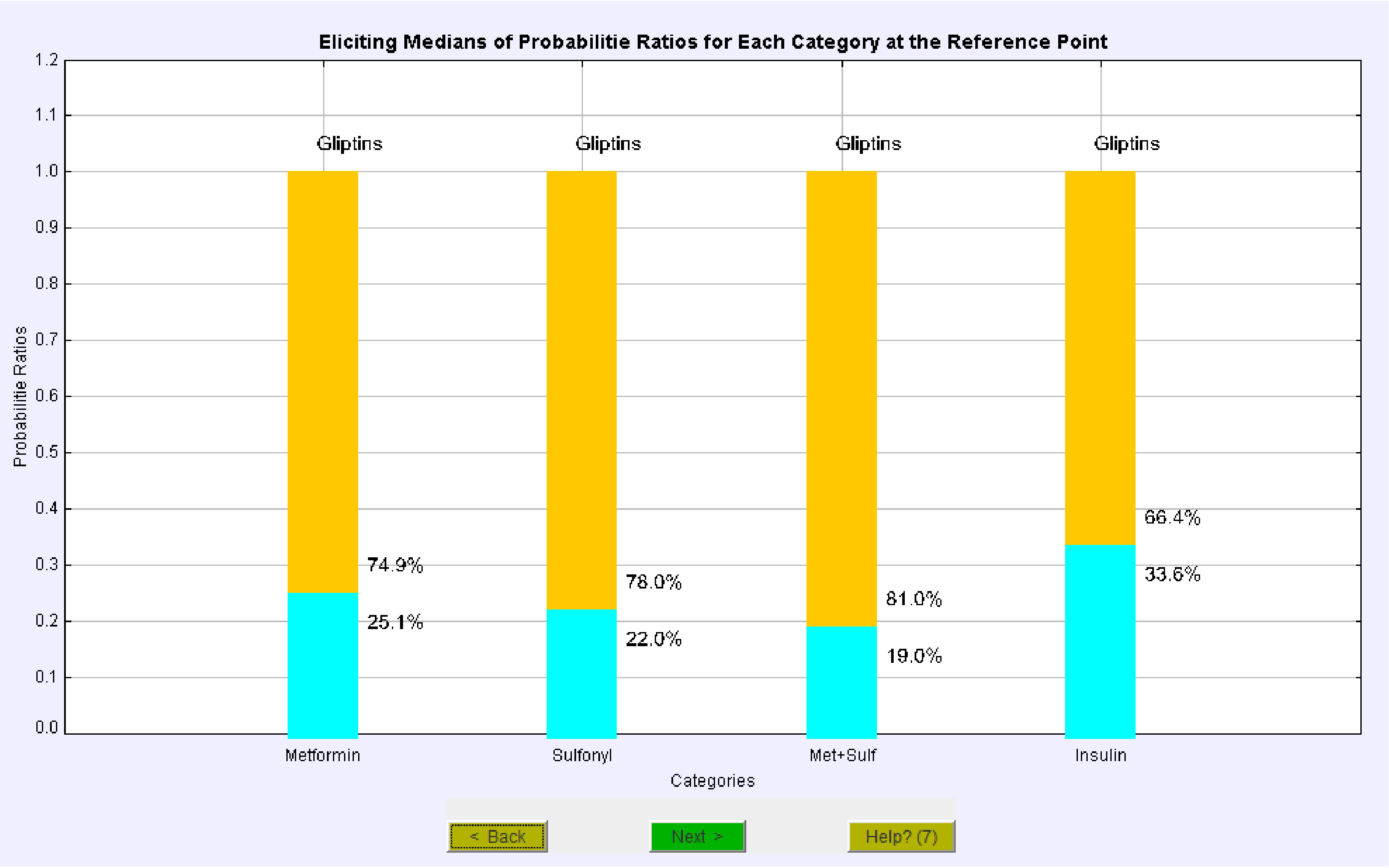
We fix on one set of covariate values and elicit assessments for those values.

The first set of values gives the mean and variance of the vector α.

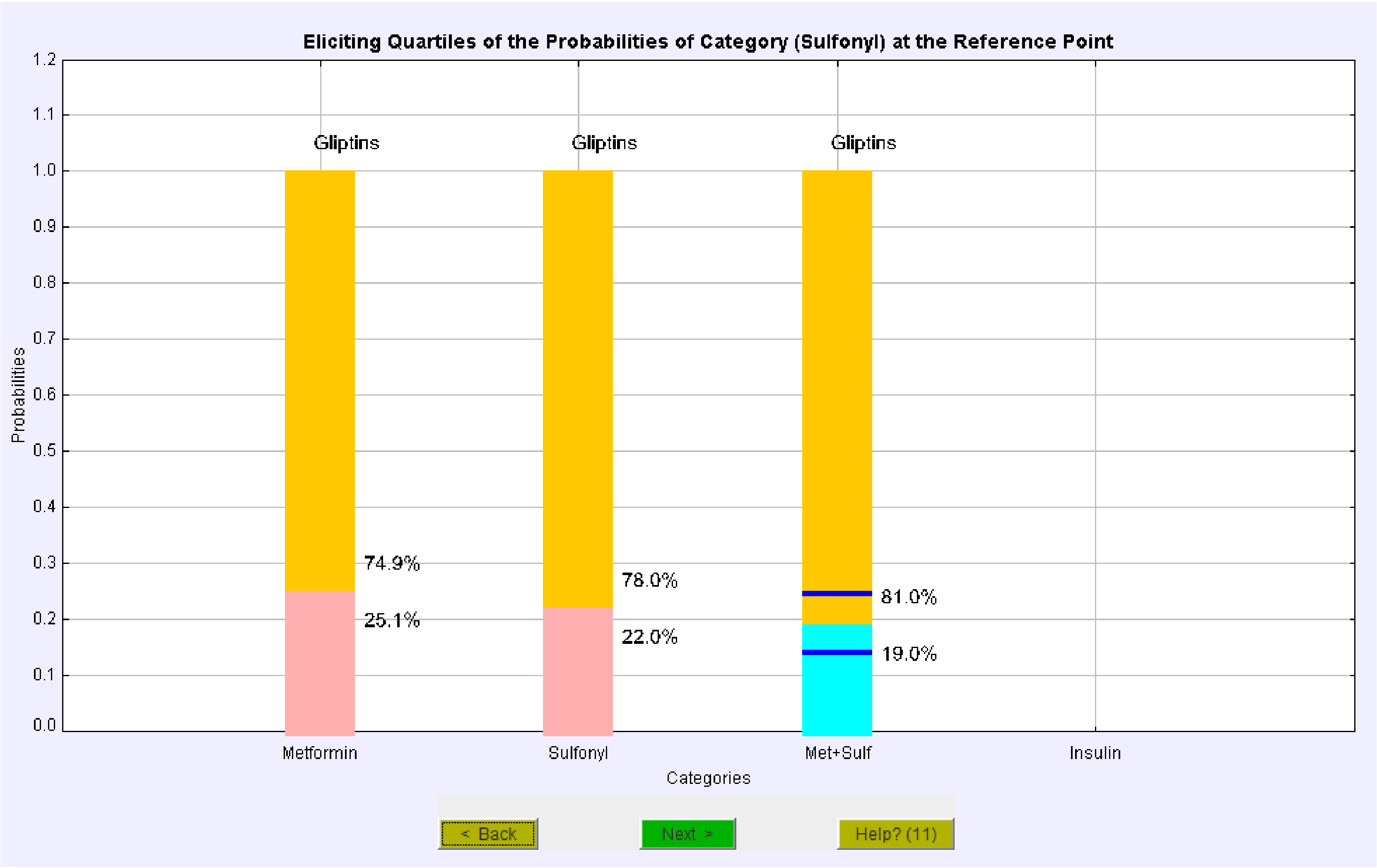
One of the covariates, age say, is given a new value. Assessments for the new set of covariate values give the mean and variance of the regression coefficients for age (one regression coefficient for each category).

This is repeated for each continuous covariate and factor level in turn.

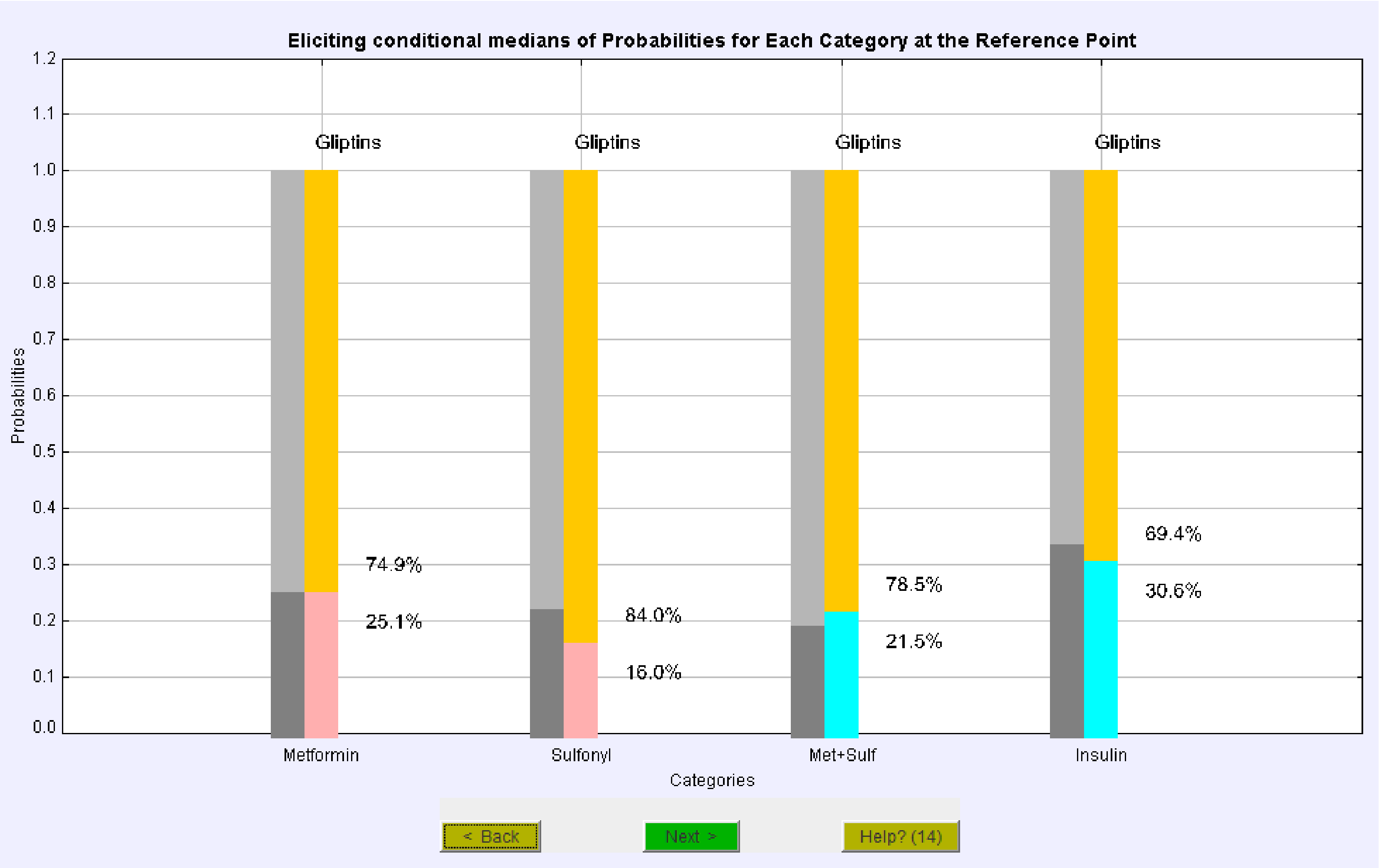
The first category is special. The expert gives point estimates of each category's probability relative to the first category.



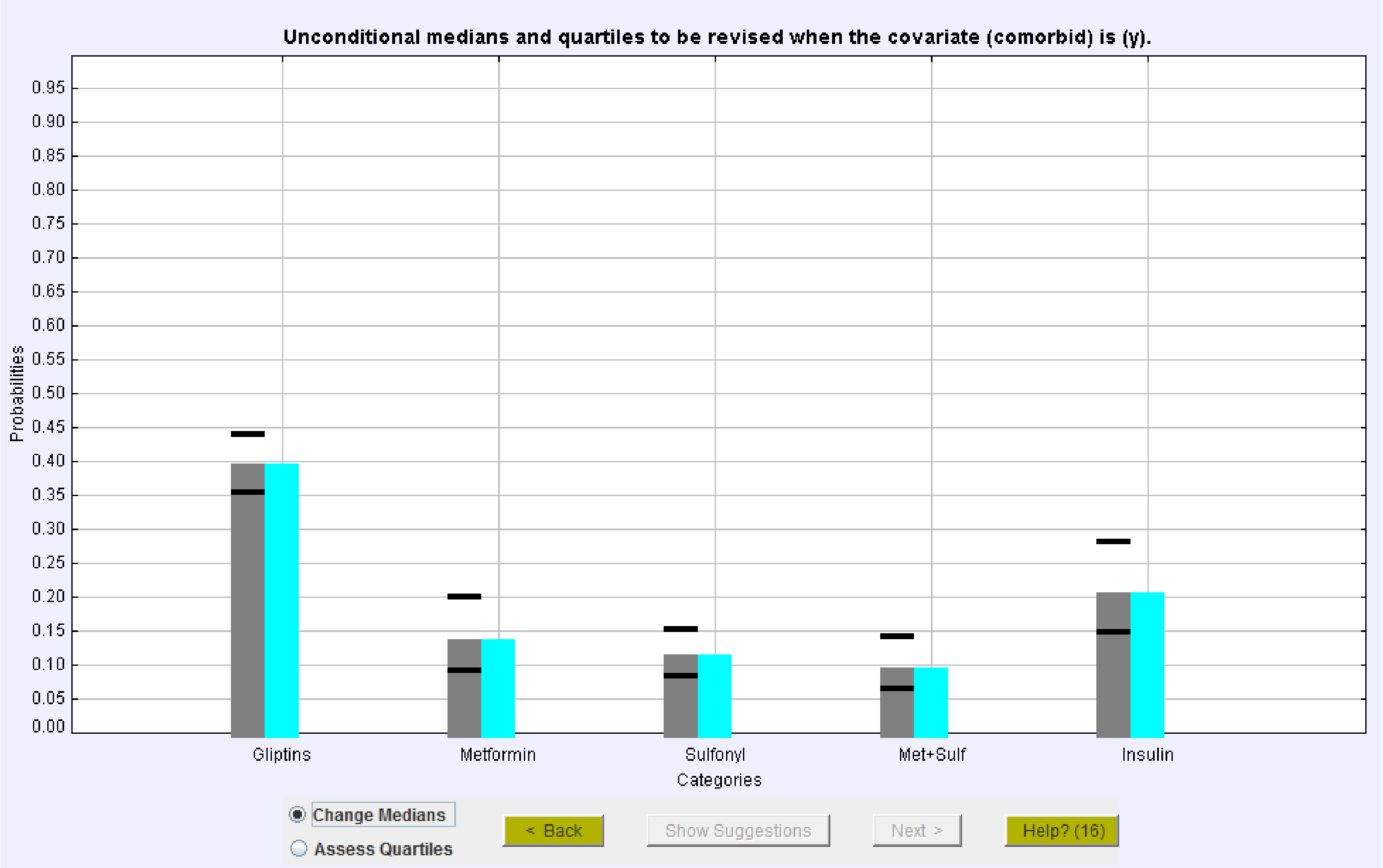
Conditional quartiles are assessed. The expert is asked to treat the proportions she assessed for the preceding categories as correct.



To assess dependencies, the expert gave the values in blue, conditional on the values in pink, only one of which differs from her original assessments (the dark grey boxes).



This feedback screen gives marginal medians and quartiles for category. For subsequent settings of the covariates, the expert may modify this screen to quantify her opinions.



PEGS (Probability Elicitation Graphical Software)

http://statistics.open.ac.uk/elicitation

Multinomial distribution.

Separate programs (and a single combined program) elicit:

- Dirichlet and Connor-Mossiman priors.
- Dirichlet and Gaussian copula priors.
- MVN prior for multinomial logistic model.

Piecewise-linear GLMs

Program that elicits an MVN prior also quantifies opinion about:

- The error variance in a normal linear model.
- The scale parameter in a gamma GLM.

These are also available in separate stand-alone programs.

References

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

Chaloner, K and Duncam, G T. (1987). Some properties of the Dirichlet-multinomial distribution and its use in prior elicitation. *Communications in Statistics – Theory and Methods*, 16, 511-523.

Dickey, J M, Jiang, J M and Kadane, J B. (1983). Bayesiam methods for multinomial sampling with noninformatively missing data. *Research Report 6/18 - #15*. State University of New York at Albany. Department of Mathematics and Statistics.

Garthwaite, P H, Chilcott, J B, Jenkinson, D J and Tappenden, P. (2008). Use of expert knowledge in evaluating costs and benefits of alternative service provision: A case study. *Int. J. Technol. Assess. Health Care*, 24, 350-357.

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

Elfadaly, F G and Garthwaite, P H. Eliciting Dirichlet and Gaussian copula prior distributions for multinomial models. Submitted.

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.

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

Thank You