Biases in Bias Elicitation

Giancarlo Manzi[*]   Martin Forster[†]

July 3, 2018

Abstract

We consider the biases that can arise in bias elicitation when expert assessors make random errors. After presenting a general framework of the phenomenon, we illustrate it for two examples: the case of omitting variables bias and that of the bias arising in adjusting relative risks. Results show that, even when assessors' elicitations of bias have desirable properties, the nonlinear nature of biases can lead to elicitations of bias that are, themselves, biased. We show the corrections which can be made to remove this bias and discuss the implications for the applied literature which employs these methods.

Keywords: bias assessment; expert elicitation; elicitation scales; omitted variable bias; relative risk

---

[*]Corresponding Author - Department of Economics, Management and Quantitative Methods, Università degli Studi di Milano - Via Conservatorio 7, 20122, Milan, Italy. E-mail: giancarlo.manzi@unimi.it.

[†]Department of Economics and Related Studies, University of York - York, United Kingdom. E-mail: mf8@york.ac.uk.

1   Introduction

There is growing interest in the use of experts to make elicitations about suspected biases in biomedical research (Thompson et al., 2011; Turner et al., 2009; Wilks et al., 2011). Such work sits within a broader literature which considers the role of expert opinion for dealing with uncertainty in statistical research (Oakley and O'Hagan, 2007; Greenland, 2005; Wolpert and Mengersen, 2004; Spiegelhalter and Best, 2003; Smith et al., 1995; Eddy et al., 1992; Zapata-Vásquez et al., 2014). According to Garthwaite et al. (2005) (p.698) expert opinion can add significantly more information to a study than can better data analysis or higher quality data sets.

Yet there is acknowledgement that the use of experts can present its own problems. Turner et al. (2009), in an analysis of a health technology appraisal in antenatal care, note that elicitation can be time-consuming, challenging and requires knowledgeable and motivated assessors who have, ideally, been recruited from a range of disciplines. Ioannidis (2011), commenting on bias adjustment for meta-analyses of observational studies, notes that some biases might be difficult to elicit with any degree of accuracy and that the sheer volume of potential biases can make it difficult for any expert to assess them rigorously. Kynn (2008) argues that much recent statistical research using probability elicitations has lagged behind psychological research. She presents a series of recommendations to improve elicitations, including the need to frame elicitation questions appropriately, to decompose the elicitation process into manageable tasks, to check for coherency in the elicitations and, if possible, to repeat the elicitation process at a later date to check the self-consistency of experts.

The work of Turner et al. (2009), in particular, has had a strong impact in the recent bias-adjusted meta-analysis and expert opinion modeling literature. Bias-adjusted meta-analysis has been extensively used after their proposal (Thompson

et al., 2011; Turner et al., 2012; Wilks et al., 2011, 2012; Schnell-Inderst et al., 2017), and has nurtured theoretical discussion (Welton et al., 2009), software development (Bujkiewicz et al., 2011) and practical extensions (Kaizar, 2011). Checklists for good practice in meta-analysis and meta-regression have also been proposed (Ades et al., 2013; Dias et al., 2013; Doi et al., 2013; Doi, 2014; Lash et al., 2014; Nasser et al., 2012; Verde and Ohmann, 2015).

This note considers the biases that can arise in bias elicitation when expert assessors make random errors in their elicitations. In this framework the term "bias" denotes a multidimensional problem and it is very application-specific. For example, in Turner et al. (2009), the authors described a set of biases arising for the external and internal validity in meta-analysis of clinical research. In Eddy et al. (1992) the authors presented a collection of possible bias functions associated to different pieces of evidence. Verde and Ohmann (2015) reviewed a large collection of papers dealing with bias modeling in clinical applications. In our work we are interested in the mathematical constituent parts of the bias subject to some assumptions, not in the various forms the bias could be detected in specific examples. We will present a general framework where the bias in bias elicitations could be evaluated and two important examples where we check for the bias which can arise when a group of assessors is asked to perform an evaluation of the bias in some estimates. We show that, even when the elicitation process is carried out by assessors of high quality - we define these as assessors who make unbiased, independent, elicitations of bias - bias elicitation can, itself, be biased. The result occurs when the bias term to be elicited is a nonlinear function of the random errors made by elicitors. The value of the bias may be approximated in a straightforward manner using a Taylor series polynomial of degree two.

The literature on expert opinion appears to exclude the possibility of eliciting a point estimate of an uncertain quantity from an expert with reference to lower and upper bounds, being the essence of elicitation the acknowledgement that the expert

is not (and cannot be) certain about this quantity. However, in many practical cases experts are asked very direct questions about something in a survey questionnaire-like format, like the following: 'Being an expert, how do you judge this econometric model? Very bad, bad, good or very good?'. Or, when the elicitor is asked to provide ratings choosing among some levels of a Likert scale, the uncertainty in the answers may be modeled in the framework of a cognitive decision process which leads an expert to express a rating about a certain quantity by the contemporary use of a *feeling* and *uncertainty* approach. This is the case of the CUB models (CUB stands for Combination of Uniform and shifted Binomial), which detects two latent variables driving individuals' responses on a rating scale, i.e., the *feeling*, accounting for logical and rational thinking, and for subjective individual's perceptions and evaluations for the objective being evaluated, and the *uncertainty*, accounting for inherent indecision present in the whole decision process (D'Elia and Piccolo, 2005; Manisera and Zuccolotto, 2014). Therefore they are particularly connected with the broad area of opinion survey and satisfaction research. In these cases, the knowledge of the properties of the point estimates can be useful.

We illustrate the phenomenon using two examples: elicitation for the classical 'omitted variables' problem in least squares regression, and the use of elicitation scales to assess bias in relative risk for studies used in a recent National Institute for Clinical Excellence technology appraisal in antenatal care (Turner et al., 2009). Proofs show the adjustments that must be made to remove the bias in the bias elicitation. We conclude by discussing the broader implications of our results for elicitations for other sources of bias that may be encountered in statistical research. Simulation uses Matlab R2015a and level curve figures are plotted using Maple 2017 and Wolfram Mathematica 11.

2 Biased elicitation of bias: general settings

Suppose a group of M expert assessors, indexed $l = 1, \ldots, M$ and operating independently of each other, is presented with an elicitation exercise about the assessment of the presence of bias in point estimates and their variances.

We assume that there exists complete accordance among experts about the 'identity' of the components of the bias in a given model. A facilitator could help ensure this, making a proposal about identifying potential estimates for which the possibility of bias is real, which could be discussed and finally agreed with the experts.

Suppose that in a given model the presence of bias is suspected for a certain vector of estimators $\hat{\boldsymbol{\theta}}$ of a certain vector of parameters $\boldsymbol{\theta}$. The bias can be explicated in a functional form $f(\mathbf{d})$ and therefore the elicitors can express their opinion on the constituent parts $\mathbf{d}$ of this function, so that:

$$f(\mathbf{d}) = E\left[\hat{\boldsymbol{\theta}}\right] - \boldsymbol{\theta} \tag{1}$$

The experts should be able to elicit explicitly the values of the elements of $\mathbf{d}$ using, for example, a natural scale. This indeed could be a hard task for them. Again a facilitator could help by proposing levels in which the bias(es) could be elicited (using Likert scales, for example), and then helping the experts to formulate their guesses of point estimates on $\mathbf{d}$. However, our main concern here is not on the practical issues of carrying out an elicitation exercise, but on the theoretical properties *with strict assumptions on the elicitation process* which are outlined below.

Predicate the analysis on the following assumptions:

1 that experts correctly agree on the same form of the true model;

2 that the experts are presented with clear constituent parts $\mathbf{d}$, following the advice of Kynn (2008), and each expert makes an elicitation about these

constituents parts;

3 that assessors' elicitations are unbiased (that is, their elicitations are random variables with expected values equal to the true values of the omitted parameters);

4 that all elicitations have common variance (both within and between elicitors);

5 that the pooled elicitations of bias are obtained by averaging the assessors' elicitations using the arithmetic mean.

Assumptions [3] and [4] above are particularly important for our task, and can be summarized as follows for an elicitor indexed with e and for a given constituent part i:

$$d_i^e = \bar{d} + \epsilon_d \quad \text{and} \tag{2}$$

$$\tag{3}$$

where $\epsilon_d \sim g(0, \sigma_{\epsilon_d}^2)$, $\sigma_{\epsilon_d}^2 > 0$, and g is a density function.

Under these assumptions we can formulate the following general proposition about the elicitation for the bias in the estimates.

Proposition 2.1. *Under the assumptions made about the assessors' elicitations, the overall elicitation of the bias term in Eq. (1) will, itself, be biased, and its standard error will also result biased.*

We will prove this proposition and the related problem for the bias in the variance of the estimators for two important examples and give some hints about the corrections which can be made to remove the bias.

3 Biased elicitation of bias: important examples

3.1 Elicitation for omitted variables bias in least squares regression

While elicitation for the bias of risk-based estimates like the relative risk or the log odds ratio have been extensively reported in the expert opinion literature, especially in epidemiological research (see, for example, Schnell-Inderst et al. (2017) for a recently performed bias-adjusted meta-analysis of random control trials and observational studies reporting relative risks), to our knowledge elicitation for omitted variable biases has never been done in practice. Nevertheless, elicitation of expert opinion on omitted variable bias is such an important exercise that it merits theoretical consideration. In particular, what we address here is the case when important confounders are unmeasured, a situation very close to that described in McCandless et al. (2012) in a Bayesian context. In this section we consider this case.

The process explaining the bias arising from omitting predictors in a least square regression model could be briefly summarized as follows. Assume that the statistician (erroneously) believes that the true model is:

$$\mathbf{y} = \tilde{\mathbf{X}}\tilde{\boldsymbol{\beta}} + \mathbf{v}. \tag{4}$$

$\mathbf{y}$ is a $N \times 1$ vector of observations on the dependent variable, $\tilde{\mathbf{X}}$ a $N \times (K - L)$ matrix of observations on the regressors, and $\mathbf{v}$ the usual vector of errors whose generic component is distributed $v \sim N(0, \sigma_v^2)$. If the true model is instead

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \tag{5}$$

model (4) considered by the statistician has a number L of variables $X_{K-L+1}, \ldots, X_K$ from $\mathbf{X}$ which have been incorrectly omitted from the true model. In (5) the generic

component of the error term is distributed $u \sim N(0, \sigma_u^2)$.

Now define $\mathbf{z} = [x_{K-L+1}, \ldots, x_K]'$ as the $L \times 1$ vector containing the regressors from Eq. (5) that are incorrectly omitted in Eq. (4). It is the case that the generic component of the error term in (4) is such that $v = \mathbf{z}'\boldsymbol{\alpha} + u$, where $\boldsymbol{\alpha}$ is an $L \times 1$ vector of parameters. As is well known, as long as the variables in $\mathbf{z}$ are correlated with those in $\tilde{\mathbf{x}}$ and have non-zero correlation with $\mathbf{y}$ in Eq. (5), the least squares estimator $\hat{\tilde{\boldsymbol{\beta}}}$ from Eq. (4) will be biased, but more efficient, than that from Eq. (5), as follows:

$$E\left[\hat{\tilde{\boldsymbol{\beta}}}\right] = \boldsymbol{\beta}^* + \mathbf{b}\boldsymbol{\beta}^\dagger \tag{6}$$

$$\mathrm{var}\left(\hat{\tilde{\boldsymbol{\beta}}}\right) = \sigma_u^2 (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}. \tag{7}$$

In (6), $\boldsymbol{\beta}^* = [\beta_1, \ldots, \beta_{K-L}]'$ (the true parameters for the first $K - L$ variables in Eq. (5)), $\mathbf{b}$ is a $(K - L) \times L$ matrix containing the appropriate regression coefficients from the auxiliary regressions, the regressions of the excluded variables on all of the included variables[1], and $\boldsymbol{\beta}^\dagger = [\beta_{K-L+1}, \ldots, \beta_K]'$, the true parameters for the final L variables in Eq. (5):

$$y = \tilde{\mathbf{X}}\boldsymbol{\beta}^* + \mathbf{Z}\boldsymbol{\beta}^\dagger + u, \tag{9}$$

where $\mathbf{Z}$ is the stacked matrix of the observations of the omitted variables.

---

[1]The auxiliary regressions are:

$$
\begin{aligned}
x_{K-L+1} &= a_{K-L+1} + b_{K-L+1,1}x_1 + \ldots + b_{K,K-L}x_{K-L} + w_{K-L+1} \\
\ldots &= \ldots \\
x_K &= a_K + b_{K,1}x_1 + \ldots + b_{K,K-L}x_{K-L} + w_K, \tag{8}
\end{aligned}
$$

where the ws are assumed to be zero mean and constant variance random variables. The matrix of coefficients is therefore:

$$
\mathbf{b} = \begin{bmatrix} b_{K-L+1,1} & \ldots & b_{K,1} \\ \ldots & \ldots & \ldots \\ b_{K-L+1,K-L} & \ldots & b_{K,K-L} \end{bmatrix}.
$$

The bias term for the parameter vector is, therefore:

$$E\left[\hat{\tilde{\boldsymbol{\beta}}}\right] - E\left[\hat{\boldsymbol{\beta}}^*\right] \;=\; \mathbf{b}\boldsymbol{\beta}^\dagger. \tag{10}$$

And the difference in efficiency of the two estimators is (Greene, 2007):

$$\mathrm{var}\left(\hat{\tilde{\boldsymbol{\beta}}}\right) - \mathrm{var}\left(\hat{\boldsymbol{\beta}}^*\right) = \sigma_u^2(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1} - \sigma_u^2\left[\tilde{\mathbf{X}}'\tilde{\mathbf{X}} - \tilde{\mathbf{X}}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\tilde{\mathbf{X}}\right]^{-1}. \tag{11}$$

3.1.1 Bias elicitation

Bias elicitation for $\hat{\tilde{\boldsymbol{\beta}}}$

Under these assumptions, Proposition (2.1) assumes the following form:

Proposition 3.1. *Given the true model in (5) and the assumptions made about the assessors' elicitations, the overall elicitation of the bias term in Eq. (10) will, itself, be biased, unless there exists zero correlation between the errors in the assessors' elicitations*

Proof. Let the lth expert's elicitation of the bias associated with the parameters in $\mathbf{b}$ and $\boldsymbol{\beta}^\dagger$ in Eq. (10) be denoted by the superscript (e, l) and the bias-elicited matrices for the lth expert be $\mathbf{b}^{(e,l)}$ and $\boldsymbol{\beta}^{\dagger(e,l)}$, as follows:

$$\mathbf{b}^{(e,l)} \;=\; \begin{bmatrix} b_{K-L+1,1}^{(e,l)} & b_{K-L+2,1}^{(e,l)} & \cdots & b_{K,1}^{(e,l)} \\ b_{K-L+1,2}^{(e,l)} & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ b_{K-L+1,K-L}^{(e,l)} & b_{K-L+2,K-L}^{(e,l)} & \cdots & b_{K,K-L}^{(e,l)} \end{bmatrix} \tag{12}$$

and

$$\boldsymbol{\beta}^{\dagger(e,l)} \;=\; \begin{bmatrix} \beta_{K-L+1}^{(e,l)} \\ \beta_{K-L+2}^{(e,l)} \\ \cdots \\ \beta_K^{(e,l)} \end{bmatrix}. \tag{13}$$

Given the assumptions about the elicitations, $b_{j,k}^{(e,l)} = b_{j,k} + \epsilon_{j,k}$ and $\beta_j^{(e,l)} = \beta_j + \epsilon_j$, where $j = K-L+1, \ldots, K$ and $k = 1, \ldots, K-L$, where all ϵs are zero-mean random variables with common variance. Substituting these expressions into Eqs. (12) and (13), obtaining the product $\mathbf{b}^{(e,l)} \boldsymbol{\beta}^{\dagger(e,l)}$ and applying the expectation operator gives:

$$
E\left[\mathbf{b}^{(e,l)} \boldsymbol{\beta}^{\dagger(e,l)}\right] = \mathbf{b}\boldsymbol{\beta}^{\dagger} + \begin{bmatrix} \sum_{j=K-L+1}^{K} \mathrm{cov}(\epsilon_{j,1}, \epsilon_j) \\ \sum_{j=K-L+1}^{K} \mathrm{cov}(\epsilon_{j,2}, \epsilon_j) \\ \ldots \\ \sum_{j=K-L+1}^{K} \mathrm{cov}(\epsilon_{j,K-L}, \epsilon_j) \end{bmatrix}. \tag{14}
$$

The expectation of the matrix product of the individual assessors' elicitations therefore equals the true bias in Eq. (10), plus a bias term, which will be non-zero unless there exists zero correlation between the errors in the assessors' elicitations. $\qquad\square$

Bias elicitation for $var\left(\hat{\tilde{\beta}}\right)$

For ease of exposition, we consider a version of Eq. (5) which contains an intercept term and only two regressors, X and Z, one of which is incorrectly omitted. The true model is therefore:

$$
y = \alpha + \beta_x x + \beta_z z + u, \tag{15}
$$

and we assume that the incorrectly specified regression omits Z. For simplicity, we assume that σ_x^2, the variance of variable X, σ_w^2, the variance of variable Z, and σ_u^2, the variance of the error term u, are known. As in the general case, since we assume that $cov(X, Z) \neq 0$, the relationship between variables X and Z is supposed to be as follows

$$
Z = a + b_{xz} X + w. \tag{16}
$$

10

The assumptions about elicitations are, in this case:

$$b_{xz}^{(e,l)} = b_{xz} + \epsilon_{xz}, \tag{17}$$

and

$$\beta_z^{(e,l)} = \beta_z + \epsilon_z, \tag{18}$$

with $var(\epsilon_{xz}) = \sigma_{\epsilon_{xz}}^2$ and $var(\epsilon_z) = \sigma_{\epsilon_z}^2$, respectively, and covariance between ϵ_{xz} and ϵ_z given by $\sigma_{\epsilon_z \epsilon_{xz}}$.

Proposition 3.2. *Under the dgp given by Eq. (15), the elicitation of the bias associated with the variance of $\hat{\beta}_x$ in a regression which omits Z:*

cannot be separated from elicitation of the bias for the point estimate, and

will, itself, be biased, even if there exists zero correlation between the errors in the assessors' elicitations.

Proof. For the two variable case, Eq. (11) simplifies to:

$$\mathrm{var}\left(\hat{\boldsymbol{\beta}}^*\right) = \mathrm{var}\left(\hat{\tilde{\boldsymbol{\beta}}}\right)\phi,$$

where:

$$\phi = \frac{1}{\left(1 - \left(b_{xz}\frac{\sigma_z}{\sigma_x}\right)^2\right)} > 1 \tag{19}$$

is the multiplicative bias term. Since we have assumed that σ_x and σ_z are known, the assessors may use their elicitations $b_{zx}^{(e,l)}$ that would have been used for the elicitation of bias for the point estimator, and substitute them into (19).

Substitute $b_{xz}^{(e,l)} = b_{xz} + \epsilon_{xz}$ into (19) (where all ϵ_{xz}s are zero-mean random variables with common variance $\sigma_{\epsilon_{xz}}^2$), rearrange and apply the expectation operator to obtain:

$$E[\phi^{(e,l)}] = E\left[\frac{\sigma_x^2}{\sigma_x^2 - (b_{xz} + \epsilon_{xz})^2\sigma_z^2}\right].$$

Use a Taylor series polynomial of degree two to approximate the term on the right hand side:

$$E[\phi^{(e,l)}] \approx \phi + \sigma^2_{\epsilon_{xz}}\left(\frac{\sigma^2_z\sigma^2_x\left(3\sigma^2_x(b_{xz})^2 + \sigma^2_z\right)}{(\sigma^2_z - \sigma^2_x(b_{xz})^2)^3}\right). \tag{20}$$

Hence:

the appearance of the term b_{xz} from the elicitation for the parameter estimator means that elicitation for the variance cannot be separated from elicitation for the parameter estimator itself;

the expectation of the elicited adjustment required to the variance of the estimator in Eq. (15) is, itself, biased.

□

3.1.2 Simulation

To illustrate the results, consider a scenario in which ten assessors are asked to make elicitations for omitted variable bias for a parameter estimate and its variance in the two-regressor scenario of Eq. (15), using the methods and assumptions described above. The parameter values we choose for the simulation are summarized in Table 1 for five scenarios. Simulation 1 is the reference simulation. Simulations 2 and 4 follow a pattern where $\sigma^2_{\epsilon_z}$, $\sigma^2_{\epsilon_{xz}}$ and $\sigma^2_{\epsilon_z\epsilon_{xz}}$ have an increasing trend with respect to their values in simulation 1. Simulations 3 and 5 follow a pattern where $\sigma^2_{\epsilon_z}$, $\sigma^2_{\epsilon_{xz}}$ and $\sigma^2_{\epsilon_z\epsilon_{xz}}$ have an increasing trend with respect to their values in simulation 1. Note that the non-zero covariance implies that the assessors' elicitations are not independent. We run the elicitation exercise $100,000$ times and calculate the elicitation biases that are given in Eqs. (14) and (20).

- Elicitation for bias in $\hat{\tilde{\boldsymbol{\beta}}}_{\boldsymbol{x}}$. Given the result in Eq. (14), we would expect the bias in the elicitations to equal the covariance between ϵ_z and ϵ_{xz}. From Table 2, in simulation 1 the covariance equals 0.16667. The average of the bias in

Table 1: Simulation input parameter values (100,000 runs)

Input parameter	Simulation				
	1	2	3	4	5
Standard deviation of ϵ_z defined in (18)	1	0.5	2	0.25	4
Standard deviation of ϵ_{xz} defined in (17)	0.7	0.35	1.4	0.175	2.8
Value of β_z defined in (15)	10	10	10	10	10
Value of b_{xz} defined in (16)	1	1	1	1	1
Standard deviation of u defined in (15)	2	2	2	2	2
Standard deviation of X defined in (15)	1	1	1	1	1
Standard deviation of Z defined in (15)	5	5	5	5	5
Covariance between ϵ_z and ϵ_{xz} as in (17) and (18)	1/6	1/12	1/3	1/24	2/3

Table 2: Simulation output estimates (100,000 runs)

Output estimate	Simulation				
	1	2	3	4	5
Avg. elicitation for bias in $\hat{\tilde{\beta}}_x$	0.169	0.083	0.332	0.042	0.671
(% diff. from true bias)	(1.2%)	(0.2%)	(-0.5%)	(-0.1%)	(0.6%)
Avg. elicitation for bias in $\text{var}(\hat{\beta}_x)$	0.025	0.006	0.106	0.002	0.574
(% diff. from true bias)	(0.8%)	(6.8%)	(11.0%)	(1.3%)	(-28.9%)

the elicitations across the 100,000 simulations is 0.16861, a difference of 1.2%. Percentage differences from the true bias for the other simulations are very similar and well below 1%.

- Elicitation for bias in $\text{var}(\hat{\beta}_x)$. The true bias associated with the elicitations for the adjustment required to the variance in (20) is 0.02501. The average of the bias across the simulations is 0.02521, a difference of 0.8%. Percentage differences from the true bias for simulations 2 and 4 tend to be greater than that of simulation 1, whereas percentage differences for simulations 3 and 5 have a slight tendency to inflate in absolute value.

3.1.3 Impact of the bias in the bias elicitations

It could be useful to show an example on the impact the bias in the bias elicitations might have on the results of a regression model affected by omitted variable bias. To this purpose we perform a simulation based on what was presented in Žliobaite

and Custers (2016) for the case of an important predictor (in a regression model explaining the salary level) which might have been omitted due to its sensitive intrinsic meaning (in some countries this omission could be mandatory due to anti-discriminating laws which forbid the use of characteristics like race, immigrant status, unemployment status, etc.).

Suppose a situation in a simplified society where the monthly salary is assumed to be dependent on years of education and the unemployment status in recent years. The underlying supposed true model could be, for example:

$$salary = 1000 + 100 \times education + 50 \times unemployment, \tag{21}$$

where *salary* is the monthly salary, *education* is the number of years spent to achieve the highest individual qualification and *unemployment* is the number of months of unemployment in the last five years.

Now, suppose that the variables involved are collected through a small survey questionnaire, whose results are presented in Table 3, but researchers are forced to omit the responses obtained from the question about unemployment. This would result in a model where *unemployment* is omitted:

$$salary = \beta_0 + \beta_1 \times education + \epsilon.$$

After running a standard OLS estimation on the data in Table 3, the following model is obtained:

$$salary = 549.7 + 138.3 \times education, \tag{22}$$

so that with this model people with no education get 549.7 euros instead of 1000 euros if we had used the full model with the predictor *unemployment*. The fitted model is therefore biased because people with low education are disadvantaged more

14

Table 3: Questionnaire data

Education	Unemployment in the last 5 years (in months)	Monthly salary (in Euros)
1	10	600
2	9	750
3	8.5	875
4	4	1200
10	2	1900
1	5	850
6	4	1400
7	4	1500
9	3	1750
10	0	2000

than necessary, whereas people with high education are favored more than deserved. This bias generates incorrect policies and recommendations. In this situation model (22) implements model (4) and the auxiliary regression (8) is:

$$unemployment = 9.01 - 0.77 \times education. \tag{23}$$

Now, suppose that four assessors are asked to formulate their guesses about the constituent parts of the omitted variable bias, i.e. the parameters $\mathbf{b}$ and $\boldsymbol{\beta}^{\dagger}$ in Eq. (10), which in this example corresponds to -0.77 (model (23)) and 50 (model (21)). We simulate their evaluations by using the assumptions about the elicitations, i.e. the assessors elicit the true parameters with errors ϵs distributed as zero-mean random variables with common variance. For these ϵs we generate normal values with zero mean and common variance. We start with the variance equal to 0.2 and increase it until 200, with some intermediate values chosen in order to better define the "behavior" of the assessors' elicitations. In this way we obtain simulated values for $\mathbf{b}^{e,l}$ and $\boldsymbol{\beta}^{\dagger(e,l)}$, with $e = 1, \ldots, 4$ and $l = 1$ (since we have four assessors, one parameter in the auxiliary regression (23) and one omitted variable) at different levels of common variance. Finally we take the product $\mathbf{b}^{e,l}\boldsymbol{\beta}^{\dagger(e,l)}$ for each assessor and average over them. These means are then compared with the true bias term given by Eq. (10) in terms of relative percentage difference. Results are shown in

Table 4: Simulated elicitations

ϵ	Elicitors				Averages	Absolute relative differences
	Elicitor 1	Elicitor 2	Elicitor 3	Elicitor 4		
$N(0, 0.02)$	-39.60	-23.37	-26.56	-31.48	-30.25	20.93%
$N(0, 0.05)$	-50.44	-45.28	-22.45	-29.79	-39.24	2.56%
$N(0, 1)$	-44.77	37.05	20.04	-5.59	1.68	104.39%
$N(0, 1.5)$	-106.18	-24.45	-65.36	-99.75	-73.94	93.23%
$N(0, 2)$	-103.41	-19.86	-121.07	114.13	-32.55	14.93%
$N(0, 2.5)$	2.85	-143.77	24.20	53.75	-15.74	58.87%
$N(0, 3)$	49.22	-83.80	172.93	19.71	39.51	203.27%
$N(0, 4)$	140.94	117.60	-25.41	-462.76	-57.41	50.03%
$N(0, 5)$	-190.39	-576.94	259.90	-52.25	-139.92	256.67%
$N(0, 10)$	-39.97	372.29	399.50	298.45	257.57	773.13%
$N(0, 20)$	-20.14	143.19	74.85	-1010.50	-203.15	430.90%
$N(0, 50)$	176.72	6341.21	2275.53	-2198.95	1648.60	4408.44%
$N(0, 100)$	-510.56	-4696.95	-1139.94	1090.15	-1314.33	3334.85%
$N(0, 200)$	298.82	4743.18	27877.16	22855.08	13943.56	36540.00%

Table 4.

As expected, the impact on the bias, as revealed by the absolute relative differences, increases exponentially with the variance of the ϵs, and therefore greater uncertainty equates to greater impact on bias. We also performed a similar simulation (whose results are not presented here) with the uniform distribution on the ϵ, leading to even more impact on bias.

Table 5: Hermann et al. study data (Hermann et al., 1984)

	Cases with sensitization	Cases with no sensitization	
Treated	2	527	529
Control	10	635	645
	12	1162	1274

3.2 Elicitation for bias in log relative risks using elicitation scales

Turner et al. (2009) consider bias elicitation for a range of biases in a series of studies in antenatal care. The following example presented in their paper could help in understanding a typical elicitation exercise in this context. Rhesus negative pregnant women are at risk of 'sensitization' when their foetus is Rhesus positive, most probably causing a condition that is known as haemolytic disease of the newborn which can, in severe cases, result in stillbirth, disability or neonatal death. One of the possible remedies is the offer of a dose of blood product called anti-D immunoglobulin after birth of a Rhesus positive baby. There is still not enough evidence to conclude that this treatment is successful in avoiding the insurgence of haemolytic disease of the newborn, since it occurs in 1 case in about 21,000. There are many studies conducted to test the effectiveness of this treatment, differing in intervention dose, randomization, outcome measurement etc. An example is the study of Hermann et al. (1984), whose key data are reported in Table 5. In the application of their bias elicitation proposal, Turner et al. (2009) considered this study because of the presence of important biases such as population bias (controls were recruited in an earlier period than women receiving the intervention; no inclusion-exclusion criteria were formulated; no confounders were considered), performance bias (lack of blinding likely), and attrition bias (the outcome were reported 8 months later with respect to when the women delivered Rhesus positive babies). In our study, the odds ratio is estimated to be 0.24 (95 % CI 0.05, 1.10) and the log odds ratio is estimated to be -1.42.

To illustrate biased elicitation of bias, we consider elicitation for one possible source of bias in log relative risk from one published study only, using the elicitation scale approach proposed by Turner et al. (2009). Table 6 may help to understand how the elicited bias affects the original estimate and the width of the original confidence interval in a single study. The original OR estimate, confidence interval and total sample size are taken from the Hermann et al. study (Hermann et al., 1984). The effect is kept the same all over the scenarios (OR=0.24), and only the number of treated and non-treated patients is varied. For the ease of the elicitation exercise and to express in a similar way results favoring, in turn, the intervention and the control arm, elicitors are given an elicitation scale where the left hand part of the scale runs from 0.1 (risk much lower in the intervention group) to 1 (no bias)) and the right hand part of the scale runs from 1 (no bias) to 0.1 (risk much lower in the control group)).[2] The elicited lower (a) and upper (b) bounds are also varied. As a result, different adjusted CI widths and OR are obtained. If $a = b$, the elicitor believes that the biases 'cancel'; they are as likely to favour the intervention group as they are the control. An interval for which $a \neq b$ states that the mean bias is not null and expresses a certain 'bias unbalance' favoring the treatment or the control arm. If an assessor thinks that there is no bias, the interval will be $a = 1; b = 1$. For example, an elicited interval of, say, $(0.9; 0.9)$ in the elicitation scale means that, in elicitor's mind, the bias could act symmetrically, favoring the intervention or the control with the same magnitude, but the average bias effect is null, and so the interval represents small uncertainty around a zero-valued average bias. In the calculations, this interval must be expressed as $(0.9; 1.1111)$, using $1/b$ instead of b. Therefore, in the table values of bias-adjusted OR are equal to the original OR when the assessor states a symmetric bias interval like $a = 0.9; 1/b = 1.111$. A negative adjustment is produced when elicited bias intervals are of the type $a < b$; a positive adjustment is produced when $a > b$. As for the width of the adjusted OR

---

[2]A full description may be found in Turner et al. (2009), pages 29-30.

Table 6: Effect of the elicited bias on the original estimates in Hermann et al. study (Hermann et al., 1984)

No. of cases in treatment arm	No. of cases in control arm	LB of elicited bias (a)	UB of elicited bias ($1/b$)	Width of the adjusted OR CI	Value of the adjusted OR
		Unadjusted results:			
2	10	1	1	3.045	0.24
4	20	1	1	2.159	0.24
8	40	1	1	1.536	0.24
16	80	1	1	1.098	0.24
		Adjusted results:			
2	10	0.9	1.111	3.256	0.24
4	20	0.9	1.111	2.370	0.24
8	40	0.9	1.111	1.746	0.24
16	80	0.9	1.111	1.309	0.24
2	10	0.7	1.429	3.758	0.24
4	20	0.7	1.429	2.873	0.24
8	40	0.7	1.429	2.249	0.24
16	80	0.7	1.429	1.811	0.24
2	10	0.7	1.111	3.507	0.21
4	20	0.7	1.111	2.621	0.21
8	40	0.7	1.111	1.998	0.21
16	80	0.7	1.111	1.560	0.21
2	10	0.9	1.429	3.507	0.27
4	20	0.9	1.429	2.621	0.27
8	40	0.9	1.429	1.998	0.27
16	80	0.9	1-429	1.560	0.27

CIs, they decrease as the number of cases both in the intervention and in the control arm increase. Considering symmetric elicited bias intervals, the larger they are, the larger the width of the adjusted OR CIs. At constant width, asymmetric elicited bias intervals in general produces larger adjusted OR CIs than those produced by symmetric elicited bias intervals.

In general and more formally, the study parameter estimate of log relative risk, $\hat{\theta}$, is adjusted by adding a value μ^e, a pooled estimate of bias, derived from assessors' elicitations μ_i^e, $i = 1, \ldots, M$, for M assessors (where, once again, we use the superscript 'e' to denote 'elicitation'). The standard error of $\hat{\theta}$, s^2, is adjusted by adding the pooled estimate of the bias for the standard error, $(\sigma^e)^2$.

Assessor i's elicitations for the two bias parameters, μ_i^e and σ_i^e, are given by:

$$\mu_i^e = \frac{\log(a_i^e) + \log(1/b_i^e)}{2} = \frac{1}{2}\log\left(\frac{a_i^e}{b_i^e}\right) \quad \text{and} \tag{24}$$

$$\sigma_i^e = \frac{\log(1/b_i^e) - \log a_i^e}{2} = \frac{1}{2}\log\left(\frac{1}{a_i^e b_i^e}\right), \tag{25}$$

where a_i^e and b_i^e are elicitor i's chosen upper and lower ranges on an elicitation scale for the degree of bias in the intervention group. Assume that the true values of the lower and upper ranges that should be selected are $\bar{a}$ and $\bar{b}$ and that assessors, on

average, get their elicitations of these two values correct, but with random error with constant variance. Then the elicitations of the range end-points in Eqs. (24) and (25) are given by the random variables:

$$a_i^e = \bar{a} + \epsilon_a \quad \text{and} \tag{26}$$

$$b_i^e = \bar{b} + \epsilon_b, \tag{27}$$

where $\epsilon_a \sim f(0, \sigma_{\epsilon_a}^2)$ and $\epsilon_b \sim g(0, \sigma_{\epsilon_b}^2)$, $\sigma_{\epsilon_a}^2 > 0, \sigma_{\epsilon_b}^2 > 0$, f and g being two density functions. Two examples of possible situations are shown in Figure 1, with f being a $N \sim (0, 0.1)$ (case (a)), and a $Unif \sim (-0.1725, 0.1725)$ (case (b) - having therefore mean 0 and variance 0.1).

We assume that the pooled elicitations are obtained by averaging the μ_is and σ_is, that is, $\mu^e = \sum_{i=1}^M \mu_i^e / M$ and $\sigma^e = \sum_{i=1}^M \sigma_i^e / M^3$.

Figure 1 represents an ideal situation for the impact of the bias in bias elicitation, since both the chosen densities are centered on the true bias and the impact is negligible. If we slightly shift either the location or the shape parameters of the densities, the impact on the bias dramatically changes, as shown in Figure 2 for the case of a $Unif \sim (0, 1)$.

Proposition 3.3. *Under the bias elicitation process and assumptions described above, the pooled elicitations of the bias terms μ^e and σ^e will, themselves, be biased.*

Proof. Substitute (26) and (27) into Eqs. (24) and (25) and calculate approximations to the expected values $E[\mu^e]$ and $E[\sigma^e]$, using a Talyor series polynomial of degree two:

$$E[\mu^e] \approx \frac{1}{2} \log\left(\frac{\bar{a}}{\bar{b}}\right) - \frac{\sigma_{\epsilon_a}^2}{4\bar{a}^2} - \frac{\sigma_{\epsilon_b}^2}{4\bar{b}^2}, \tag{28}$$

$$E[\sigma^e] \approx \frac{1}{2} \log\left(\frac{1}{\bar{a}\bar{b}}\right) + \frac{\sigma_{\epsilon_a}^2}{4\bar{a}^2} - \frac{\sigma_{\epsilon_b}^2}{4\bar{b}^2}. \tag{29}$$

---

[3]Other pooling methods - for example, using the median, as in Turner et al. (2009) - could be considered. However, in the case of the median $Me^e(\mu_1^e, \ldots, \mu_1^e)$, the final result is the same since $\mu^e = Me^e(\mu_1^e, \ldots, \mu_1^e) = \frac{1}{2} \log \frac{\bar{a}+\epsilon_a}{\bar{b}+\epsilon_b}$.
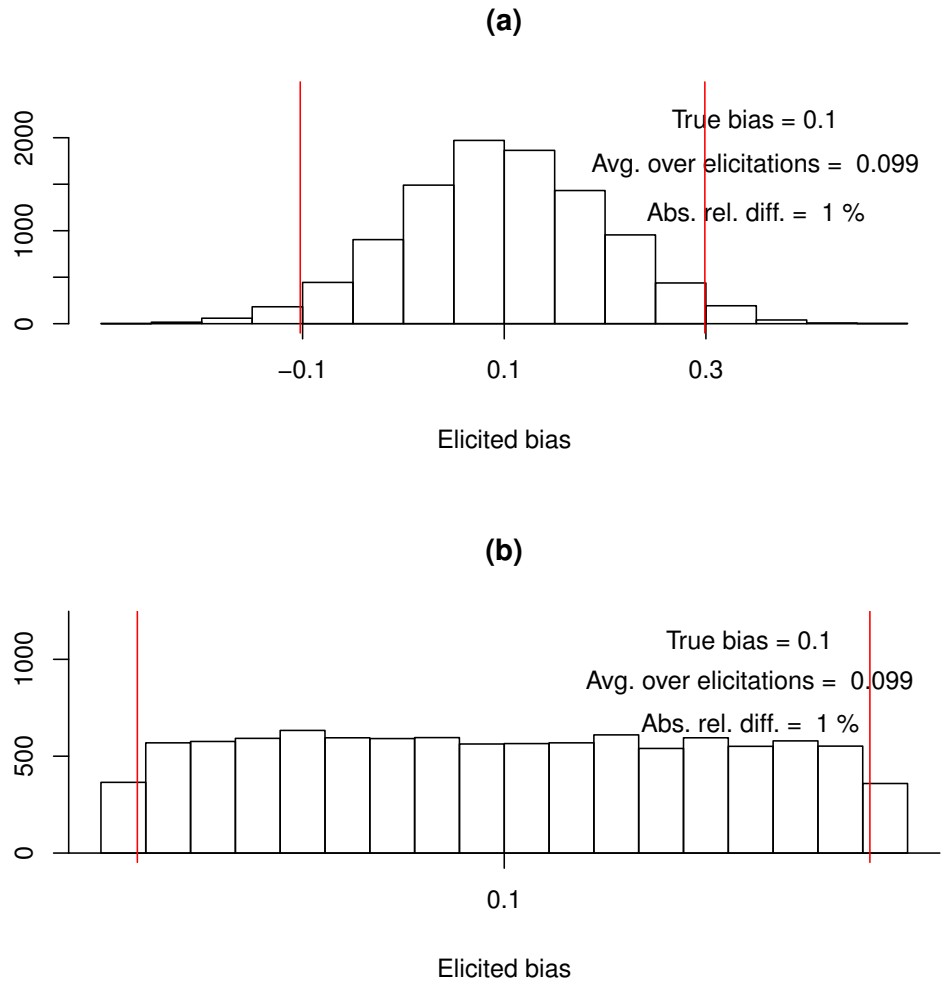
Figure 1: Two examples of random elicitation of the bias when the true bias is 0.1. (a): $N \sim (0, 0.1)$; (b): $Unif(-0.1725, 0.1725)$. Red vertical bars represent the 25th and the 75th percentiles
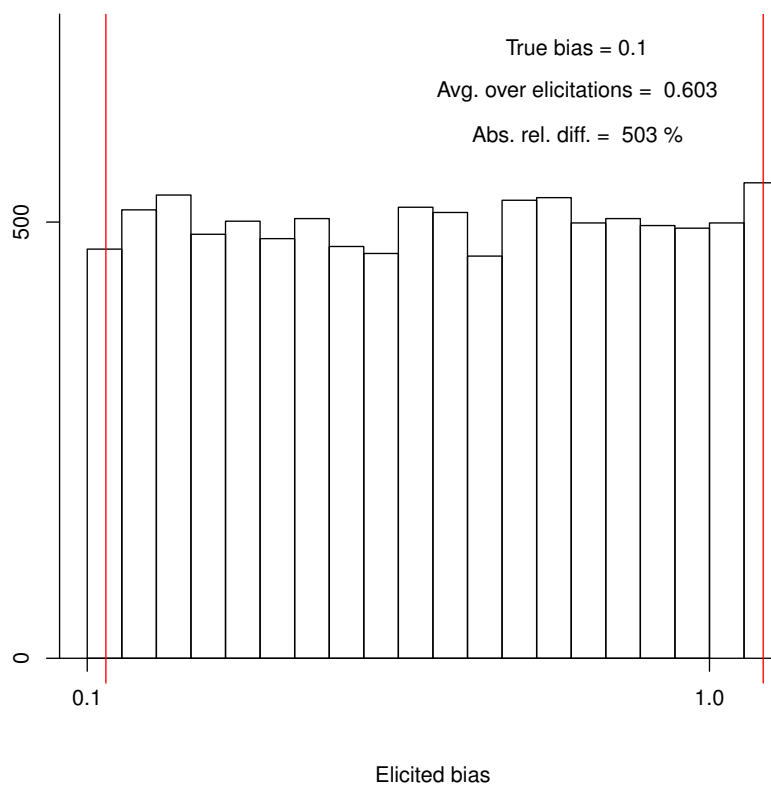
Figure 2: Simulated elicitation of the bias when the true bias is 0.1 and a $Unif(0,1)$ is chosen for the density function f

Hence the pooled bias elicitations μ^e and σ^e are themselves biased. $\square$

Eq. (28) shows that the bias in the bias elicitation - the final two terms in the equation - can be positive, zero, or negative. Eq. (25) shows that the bias is strictly positive. Figure 3 plots the bias in bias elicitation term from Eq. (28) using level curves. Figure 3 shows how this bias in bias elicitation changes as $\bar{a}$ and $\bar{b}$ - the true lower and upper limits on the elicitation scale - change (for the purposes of illustration, we assume that $\sigma^2_{\epsilon_a} = \sigma^2_{\epsilon_a} = 1$). The figure shows that, when $\bar{a} = \bar{b}$, the bias in the bias elicitation equals zero, because our assumptions on $\sigma^2_{\epsilon_a}$ and $\sigma^2_{\epsilon_b}$ mean that the two bias terms in Eq. (28) cancel. Further, we know from Eq. (24) that the true bias equals zero in this situation. Hence, if $\bar{a} = \bar{b}$, which corresponds to there being no difference between intervention and control in terms of: 1. the degree of bias favouring each and 2. the amount of bias adjustment which is required to the elicited bias term, the overall adjustment of $\hat{\theta}$ for elicited bias and its bias equals zero. However, as Figure 3 shows, the more $\bar{a}$ and $\bar{b}$ differ, the greater the bias in the bias elicitation: to the left of the line $\bar{a} = \bar{b}$, the bias term for the elicitation of bias is negative and decreasing; to the right of the line it is positive and increasing. Hence, deviations of either $\bar{a}$ or $\bar{b}$ from the line $\bar{a} = \bar{b}$ will lead to the bias adjustment being biased. The level curves in Figure 3 show that the severity of the impact of a marginal change in either $\bar{a}$ or $\bar{b}$ is higher the closer one moves to the point (1/10,1/10); for studies with large biases, a relatively small difference between $\bar{a}$ and $\bar{b}$ will lead to a greater absolute bias in the bias adjustment term. This is due to the strictly concave nature of the bias function for a^e and the strictly convex nature of the bias function for b^e. Figure 4 plots the level curves for the bias term for σ. The appendix gives more on the intuition behind these results.
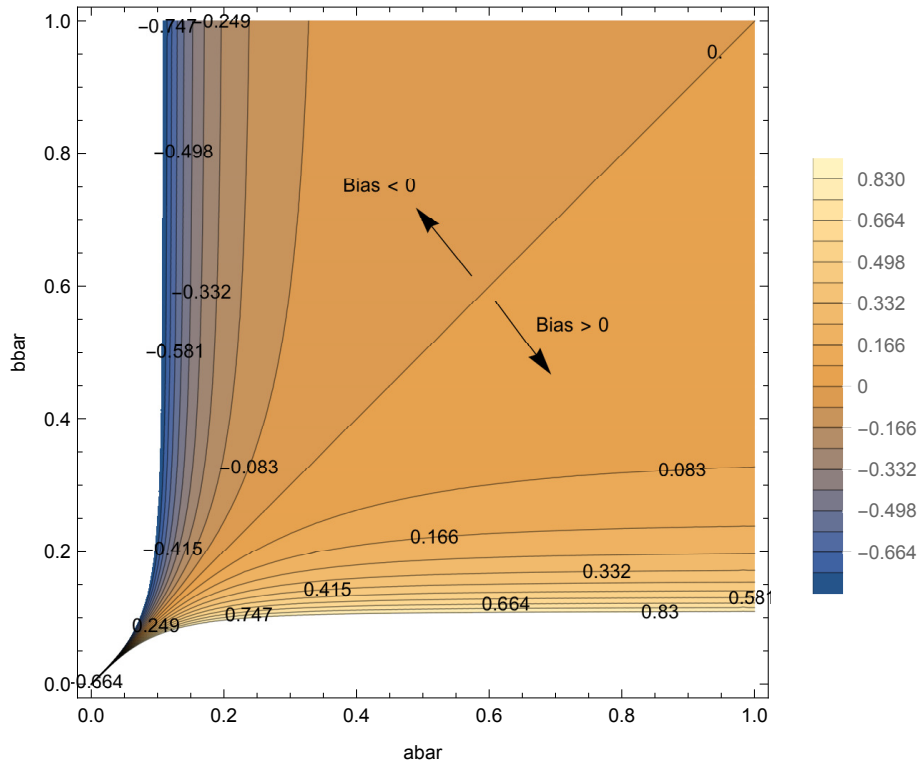
Figure 3: Level curves for the bias in bias elicitation term for μ in Eq. (28): $-\sigma_{\epsilon_a}^2/(4\bar{a}^2) + \sigma_{\epsilon_b}^2/(4\bar{b}^2)$, assuming $\sigma_{\epsilon_a}^2 = \sigma_{\epsilon_b}^2 = 0.01$. Numbers inside the plot represent value levels for bias
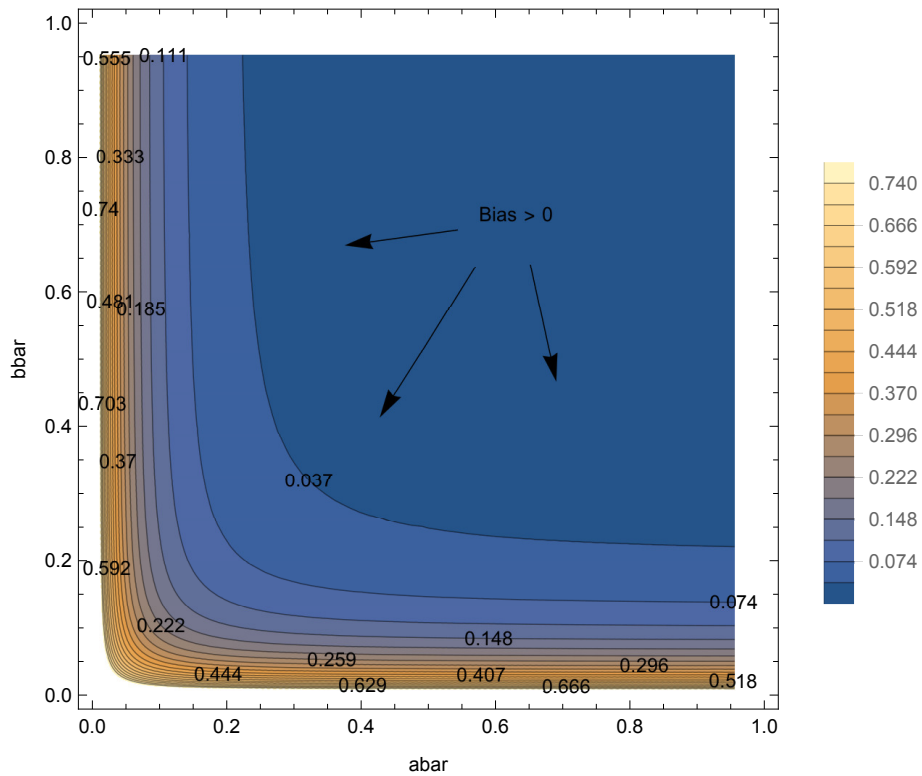
Figure 4: Level curves for the bias in bias elicitation term for σ in Eq. (29): $\sigma^2_{\epsilon_a}/(4\bar{a}^2) + \sigma^2_{\epsilon_b}/(4\bar{b}^2)$, assuming $\sigma^2_{\epsilon_a} = \sigma^2_{\epsilon_b} = 0.01$. Numbers inside the plot represent value levels for bias

4 Discussion

Our results show that, even when high quality assessors are tasked with making elicitations for bias, the nonlinearities in biases can lead to biased elicitations of bias. For the case of omitted variables in least squares regression, Proposition 1 shows that the bias associated with the assessors' elicitations for point estimators will equal zero only if the elicitations have zero covariance. Proposition 2 shows that bias in bias elicitations for the variance of the point estimators is present even if the errors in the elicitations have zero covariance: it is sufficient that assessors' elicitations are random, that is, that they have non-zero variance (the term σ_ϵ^2 in Eq. (20)). In the case of bias elicitation for log relative risk in Turner et al. (2009), if it is assumed that assessors make unbiased elicitations of the lower and upper ranges on the elicitation scale, elicitations are biased.

According to Chavalarias and Ioannidis (2010), there are 235 potential biases in biomedical research. Since biases and elicitation methods are likely to differ across studies, such analysis could proceed on a case-by-case basis, or a group of researchers could catalogue the main approaches to bias elicitation and the biases therein. We believe that the technique of Taylor series polynomials presented here offers an accessible and elegant approach to approximating and interpreting these biases in bias elicitations.

Possible future developments of this work may include, in its omitted variable bias part, the analysis of observational studies where researchers usually tend to include a large number of variables into a regression model and is hard to explicitly detect omitted variables. This problem is strictly related to the choice of variable selection methods used to reduce the prediction error. In this case, elicitation (and quantification) of important variables from experts can be incorporated into the analysis. At the same time, another important aspect will be to analyze wrongly chosen predictors and the bias arising from this selection, in an "bias in a selection

bias elicitation" framework.

Another important application of the ideas presented in this paper could be the adjustment of baseline variables in Randomized Clinical Trials, where the variables of a model like (5) has to be specified into the protocol of investigation (European Medicines Agency, 2015). If a committee of experts decide to review the protocol and modify the variables in the model, then the ideas presented in this paper become important.

References

Ades, A.E. Caldwell, D.M. Reken, S. Welton, N.J. Sutton, A.J. Dias, S. (2013). Evidence synthesis for decision making 7: A reviewer's checklist. *Medical Decision Making* 33: 679–691.

Bujkiewicz, S. Jones, H.E. Lai, M.C.W. Cooper, N.J. Hawkins, N. Squires, H. Abrams, K.R. Spiegelhalter, D.J. Sutton, A.J. (2011) Development of a transparent interactive decision interrogator to facilitate the decision-making process in health care. *Value Health* 14(5): 768–776.

Chavalarias, D. Ioannidis, J.P. (2010). Science mapping analysis characterizes 235 biases in biomedical research. *Journal of Clinical Epidemiology* 63: 1205–1215.

D'Elia, A. Piccolo, D. (2005). A Mixture Model for Preference Data Analysis. *Computational Statistics & Data Analysis* 49(3): 917-934.

Dias, S. Sutton, A.J. Welton, N.J. Ades, A.E. (2013). Evidence synthesis for decision making 3: Heterogeneity - Subgroups, meta-regression, bias, and bias-adjustment. *Medical Decision Making* 33: 618–640.

Doi, S.A.R. (2014). Evidence synthesis for medical decision making and the appropriate use of quality scores. *Clinical Medicine & Research* 12(1-2): 40-46.

Doi, S.A.R. Barendregt, J-J. Onitilo, A.A. (2013). Methods for the bias adjustment of meta-analyses of published observational studies. *Journal of Evaluation in Clinical Practice* 19: 653-657.

Eddy, D.M. Hasselblad, V. Schachter, R. (1992).*Meta-analysis by the Confidence Profile Method: the Statistical Synthesis of Evidence.* San Diego (CA): Academic Press.

European Medicines Agency (2015).*Guideline on adjustment for baseline covariates in clinical trials.* London: European Union.

Garthwaite, P. H. Kadane, J. B. O'Hagan, A. (2005). Statistical Methods for Eliciting Probability Distributions. Journal of the American Statistical Association 100(470): 680–701.

Greene, W. (2007). *Econometric Analysis*, 7th Edition. Upper Saddle River (NJ): Prentice Hall.

Greenland, S. (2005). Multiple-bias modelling for analysis of observational data (with discussion). *Journal of the Royal Statistical Society Series A* 168(1): 267–306.

Hermann, M. Kjellman, H. Ljunggren, C. (1984). Antenatal prophylaxis of Rh immunization with 250 μg anti-D immunoglobulin. *Acta Obstetricia et Gynecologica Scandinavica* Supplem. 124: 1–15.

Ioannidis, J.P.A. (2011). Commentary: Adjusting for bias: a user's guide to performing plastic surgery on meta-analyses of observational studies. *International Journal of Epidemiology* 40(3): 777–779.

Kaizar, E.E. (2011). Estimating treatment effect via simple cross design synthesis. *Statistics in Medicine* 30: 2986–3009.

28

Kynn, M. (2008). The 'heuristics and biases' bias in expert elicitation. *Journal of the Royal Statistical Society Series A* 171(1): 239–264.

Lash, T.L. Fox, M.P. MacLehose, R.F. Maldonado, G. McCandless, L.C. Greenland, S. (2014). Good practices for quantitative bias analysis. *International Journal of Epidemiology* 43(6): 1969–1985.

McCandless, L.C. Gustafson, P. Levy, A.R. Richardson, S. (2012). Hierarchical priors for bias parameters in Bayesian sensitivity analysis for unmeasured confounding. *Statistics in Medicine* 31: 383–396.

Manisera, M. Zuccolotto, P. (2014). Modeling rating data with Nonlinear CUB models. *Computational Statistics & Data Analysis* 78: 100–118.

Nasser, M. van Weel, C. van Binsbergen, J.J. van de Laar, F.A. (2012). Generalizability of systematic reviews of the effectiveness of health care interventions to primary health care: concepts, methods and future research. *Family Practice* 29: 94–103.

Oakley, J.E. O'Hagan, A. (2007). Uncertainty in prior elicitations: a nonparametric approach. *Biometrika* 94: 427–441.

Schnell-Inderst, P. Iglesias, C. Arvandi, M., Ciani, O. Matteucci Gothe, R. Peters, J. Blom, A. W., Taylor, R. S., Siebert, U. (2017). A Bias-Adjusted Evidence Synthesis of RCT and Observational Data: The Case of Total Hip Replacement. *Health Economics* 26(1): 46–69.

Smith, T.C. Spiegelhalter, D.J. Thomas, A. (1995). Bayesian approaches to random-effects meta-analysis: a comparative study. *Statistics in Medicine* 14: 2685–2699.

Spiegelhalter, D.J. Best, N.G. (2003). Bayesian approaches to multiple sources of evidence and uncertainty in complex cost-effectiveness modelling. *Statistics in Medicine* 22: 3687–3709.

Thompson, S. Ekelund, U. Jebb, S. Lindroos, A.K. Mander, A. Sharp, S. Turner, R.M. Wilks, D. (2011). A proposed method of bias adjustment for meta-analyses of published observational studies. *International Journal of Epidemiology* 40: 765–777.

Turner, R.M. Lloyd-Jones, M. Anumba, D.O.C. Smith, G.C.S. Spiegelhalter, D.J. Squires, H. Stevens, J.W. Sweeting, M.J. Urbaniak, S.J. Webster, R. Thompson, S.G. (2012) Routine antenatal anti-D prophylaxis in women who are Rh(D) negative: meta-analyses adjusted for differences in study design and quality. *PLoS ONE* 7(2): 1–10.

Turner, R.M. Spiegelhalter, D.J. Smith, G.C.S. Thompson, S. G. (2009). Bias modelling in evidence synthesis. *Journal of the Royal Statistical Society Series A* 172(1): 21–47.

Verde, P.E. Ohmann, C. (2015). Combining randomized and non-randomized evidence in clinical research: a review of methods and applications. *Research Synthesis Methods* 6: 45–62.

Welton, N.J. Ades, A.E. Carlin, J.B. Altman, D.G. Sterne, J.A.C. (2009). Models for potentially biased evidence in meta-analysis using empirically based priors. *Journal of the Royal Statistical Society Series A* 172(1): 119–136.

Wilks, D.C. Mander, A.P. Jebb, S.A. Thompson, S.G. Sharp, S.J. Turner, R.M. Lindroos, A.K. (2011). Dietary energy density and adiposity: Employing bias adjustments in a meta-analysis of prospective studies. *BMC Public Health* 48(11): 1–9.

Wilks, D.C. Sharp, S.J. Ekelund, U. Thompson, S.G. Mander, A.P. Turner, R.M. Jebb, S.A. Lindroos, A.K. (2012). Objectively measured physical activity and fat mass in children: A bias-adjusted meta-analysis of prospective studies. *PLoS ONE* 6(2): 1–8.

Wolpert, R.L. Mengersen, K.L. (2004). Adjusted likelihoods for synthesizing empirical evidence from studies that differ in quality and design: effects of environmental tobacco smoke. *Statistical Science* 19: 450–471.

Zapata-Vásquez, R.E. O'Hagan, A. Soares Bastos, L. (2014). Eliciting expert judgements about a set of proportions. *Journal of Applied Statistics* 41(9): 1919–1933.

Žliobaite, I. Custers, B. (2016). Using sensitive personal data may be necessary for avoiding discrimination in data-driven decision models. *Artificial Intelligence and Law* 24: 183–201.

Appendix: Bias for μ using an elicitation scale for log relative risk

To aid exposition, assume that $\bar{b}$ is known to be equal to 1. Then the elicitations of bias for assessor i become, from Eq. (24):

$$\mu_i^e = \frac{1}{2}\log(a_i^e). \tag{30}$$

for $i = 1, \ldots, M$. We assume that the elicitations are random with expectation equal to $\bar{a}$, which equals the true value of the lower range of the elicitation scale. Again, to aid exposition, we assume that the density function for a^e has a clear upper and lower-bound (a^l and a^u). The mapping from a^e to μ^e via the nonlinear function h is shown in Figure 5.

The true bias adjustment is $h(E[a^e]) = h(\bar{a})$. Under the transformation given by h, the expectation of the μ_i^es will be biased downward, that is, $E[h(a^e)] < h(E[a^e]) = h(\bar{a})$, as shown. Other things equal, the further to the left lies the centre of mass of the distribution of a^e, that is, the greater the true bias $\bar{a}$, the greater the divergence between $E[h(a^e)]$ and $h(E[a^e])$, meaning the greater is the
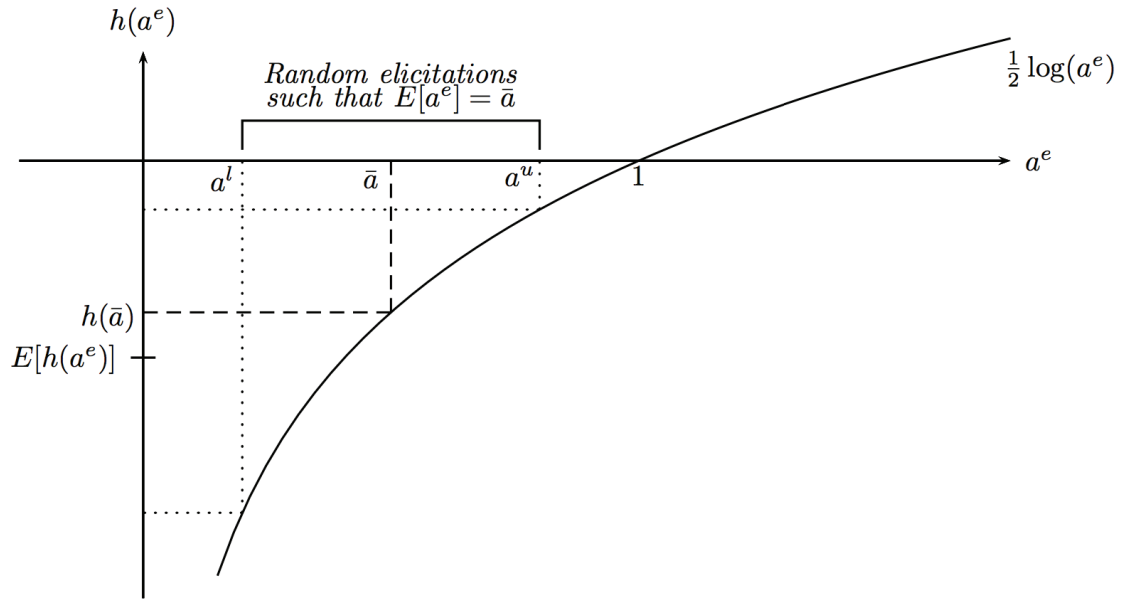
Figure 5: Difference between $\mu = E[h(a^e)]$ and $h(E[a^e]) = h(\bar{a})$ for Eq. (28) and the function $h = (1/2)\log(a^e)$, where we assume $\bar{b}$ is known to equal 1

bias in the bias elicitation. This is due to the increasing, strictly concave nature of h and explains the shape of the level curves in Figure 3.