Bayesian Meta Analysis and Bias Modeling: A Case Study with Relative Clause Processing in Mandarin Chinese

Shravan Vasishth

Department Linguistik, Universität Potsdam
Centre de Recherche en Mathématiques de la Décision (CEREMADE), Université Paris-Dauphine, PSL Research University
vasishth@uni-potsdam.de
http://www.ling.uni-potsdam.de/∼vasishth

February 5, 2016

Meta-analysis: Why synthesize evidence?

Meta-analysis (evidence synthesis) is an important tool for theory development and evaluation, but it remains essentially unutilized in cognitive science.

A nice example is the Chinese relative clause problem. I will skip the details today, but see:

Shravan Vasishth, Zhong Chen, Qiang Li, and Gueilan Guo. Processing Chinese Relative Clauses: Evidence for the Subject-Relative Advantage. PLoS ONE, 8(10):1-14, 10 2013.

The research question

Chinese relative clauses

Suppose we are interested in determining whether a particular effect (say, reading time in milliseconds) has a positive or negative sign.

The data (15 studies)

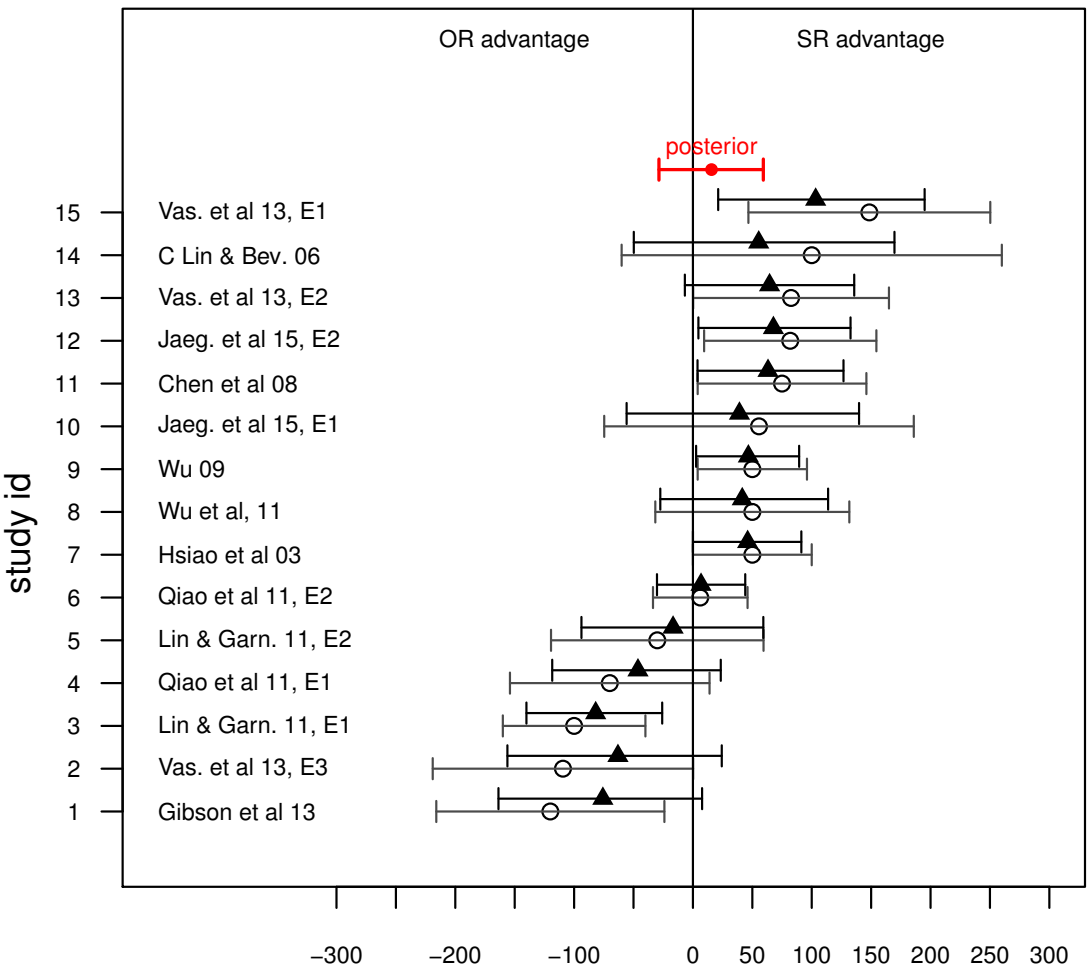
	study	y (ms)	se	nsubj	nitem	qacc	method
1	Gibson et al 12	-120	48	37	15	91	SPR
2	Vas. et al 13, E3	-109.40	54.80	40	15	87	SPR
3	Lin & Garn. 11, E1	-100.00	**30.00**	48	80	88	SPR
4	Qiao et al 11, E1	-70.00	42.00	32	24		GMaze
5	Lin & Garn. 11, E2	-30.00	44.63	40	80		SPR
6	Qiao et al 11, E2	6.19	19.90	24	30		LMaze
7	Hsiao et al 03	50.00	25.00	35	20	70	SPR
8	Wu et al, 11	50.00	**40.74**	48			SPR
9	Wu 09	50.00	23.00	40			SPR
10	Jaeg. et al 15, E1	55.62	65.14	49	16	85	SPR
11	Chen et al 08	75.00	35.50	39	23	86	SPR
12	Jaeg. et al 15, E2	81.92	36.25	49	32	80	ET
13	Vas. et al 13, E2	82.60	41.20	61	24	82	SPR
14	C Lin & Bev. 06	100.00	80.00	48	24		SPR
15	Vas. et al 13, E1	148.50	50.90	60	20	82	SPR

Random-effects meta-analysis

$$
\begin{aligned}
y_i \mid \theta_i, {\sigma_i}^2 &\sim N(\theta_i, {\sigma_i}^2) \quad i = 1, \ldots, n \\
\theta_i \mid \theta, \tau^2 &\sim N(\theta, \tau^2), \\
\theta &\sim N(0, 100^2), \\
1/\tau^2 &\sim Gamma(0.001, 0.001) \\
\text{OR} : \tau &\sim Uniform(0, 200) \\
\tau &\sim Normal(0, 200^2) I(0,)
\end{aligned} \tag{1}
$$

1. y_i is the effect size in milliseconds in the i-th study.
2. θ is the true (unknown) effect, to be estimated by the model.
3. ${\sigma_i}^2$ is the true variance of the sampling distribution; each σ_i is estimated from the standard error in study i.
4. The variance parameter τ^2 represents between-study variance.

Random effects meta-analysis of the 15 studies



Discussion of Random Effects Meta-Analysis

1. The posterior probability of the effect being positive is approximately 0.78.
2. Note that the studies may be biased.
   The term bias here refers to systematic (as opposed to random) error or deviation from the true value, which either leads to an overestimate or an underestimate.
3. We will now take this bias into account quantitatively in the meta-analysis.

Our approach is based on
Turner, Rebecca M., et al. "Bias modelling in evidence synthesis." Journal of the Royal Statistical Society: Series A (Statistics in Society) 172.1 (2009): 21-47.

Potential sources of bias in a study

See separate sheet.

Steps for modeling bias

Turner et al 2008

1. **Define the target question** and the target experimental manipulation, including the population being studied, and the outcome of interest.
2. Define an **idealized version** of each source study and write down a mini-protocol that lists each component of the idealized study.
3. Compare the details of the completed source study against the mini-protocol defined in the previous step.
4. **These steps help in identifying internal and external bias** by comparing each idealized study with the target study.

Adjusting means and variances by incorporating biases

If there were no *internal* biases, the generating distribution would be

$$y_i \sim Normal(\theta_i, {s_i}^2) \tag{2}$$

- i indexes the study
- θ_i is the true study-level effect such that $\theta_i \sim Normal(\theta, \tau^2)$
- ${s_i}^2$ is the variance for the sampling distribution of the mean of the i-th study.

We assume throughout that both internal and external biases are independent of the magnitude of the effect (additive biases).

Incorporating potential sources of bias in a study

Assume that ${\mu_i}^I$ and ${\mu_i}^E$ are total internal and external bias means with variances $({\sigma_i}^I)^2$ and $({\sigma_i}^E)^2$, then

$$y_i \sim N(\theta + {\mu_i}^I + {\mu_i}^E, {s_i}^2 + ({\sigma_i}^I)^2 + \tau^2 + ({\sigma_i}^E)^2) \qquad (3)$$

τ^2 is unexplained between-study heterogeneity.

1. The challenge is to quantify the external and internal biases in each study.
2. Experts are then recruited to deliver the priors for these biases by using a **prior elicitation framework** such as SHELF (Sheffield Elicitation Framework): http://www.tonyohagan.co.uk/shelf/

Example elicitation from two experts

From the SHELF help page

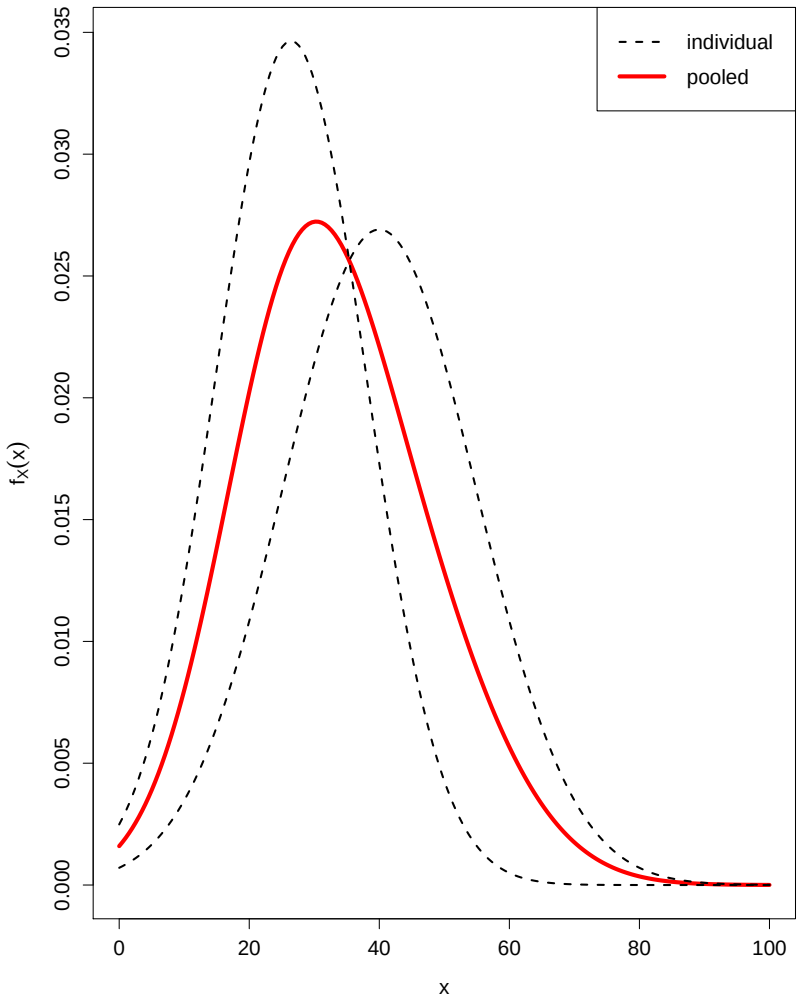
Elicit judgements regarding each bias from two experts individually:

- Expert 1 states
  $P(X < 30) = 0.25, P(X < 40) = 0.5, P(X < 50) = 0.75$
- Expert 2 states
  $P(X < 20) = 0.25, P(X < 25) = 0.5, P(X < 35) = 0.75$

Both experts state $0 < X < 100$.

O'Hagan, Anthony, Caitlin E. Buck, Alireza Daneshkhah, J. Richard Eiser, Paul H. Garthwaite, David J. Jenkinson, Jeremy E. Oakley, and Tim Rakow. Uncertain judgements: eliciting experts' probabilities. John Wiley & Sons, 2006.

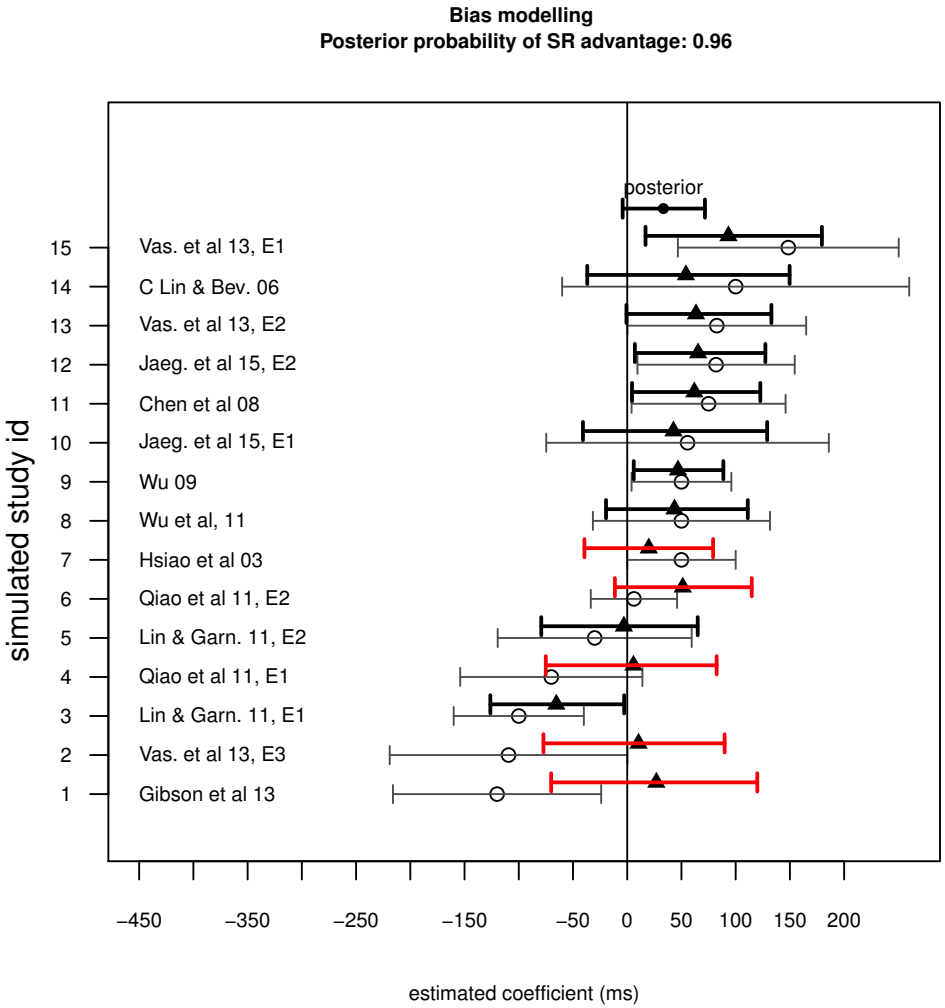
Example elicitation from two experts



Proof of concept: Bias modeling of five studies using one expert (SV)

Study	Paper	Type	Bias	Mean	SD
1	GW13	Internal	Selection	-107	64
1	GW13	Internal	Attrition	-25.5	15.8
2	Vas13E3	Internal	Selection	-90	25
4	QiaoE1	Internal	Other	-50	31
4	QiaoE1	External	Outcome	-25	17
6	QiaoE2	Internal	Other	-51	31
6	QiaoE2	External	Outcome	-55.6	33.6
7	HG03	Internal	Other	37.4	26.5

Bias modeling results



Some limitations of the present work

- Only one expert was used; in future work, we intend to elicit priors from two experts (four would be ideal, but impractical).
- Not all studies were independent; this has not yet been taken into account.

Concluding remarks

- Bias modeling seems like a very important and useful tool for evidence synthesis.
- One downside is the effort involved in identifying biases.
- It forces us to think more carefully about biases, and to quantify our uncertainty about biases; this may also help us to run better studies in the future.

Concluding remarks

For details, see:
Shravan Vasishth, Zhong Chen, Qiang Li, and Gueilan Guo. Processing Chinese Relative Clauses: Evidence for the Subject-Relative Advantage. PLoS ONE, 8(10):1-14, 10 2013.
Shravan Vasishth, A meta-analysis of relative clause processing in Mandarin Chinese using Bias Modelling, MSc Dissertation, Uni Sheffield, UK.

For code and data, please email me:
vasishth@uni-potsdam.de.