How to conduct a classical model study

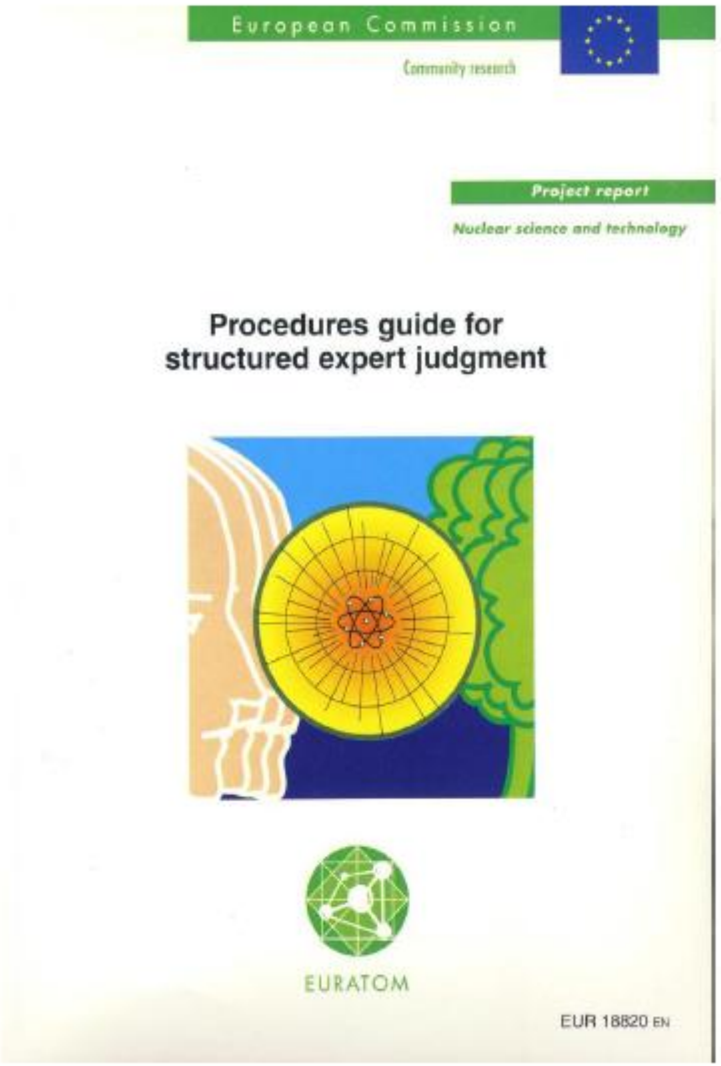
COST TRAINING SCHOOL

WARSAW

MARCH 16-18, 2016

ABIGAIL COLSON

UNIVERSITY OF STRATHCLYDE


European Commission
Community research
Project report
Nuclear science and technology
Procedures guide for
structured expert judgment
EURATOM
EUR 18820 EN

1

Content based on:

- Procedures guide
- Presentations by Roger Cooke
- Personal experience

Procedures Guide

The elicitation process

Pre-Elicitation

1. Define case structure.
2. Identify variables of interest.
3. Identify calibration variables.
4. Identify and select experts.
5. Write the elicitation protocol.
6. Pilot test the protocol.
7. Train the experts.

Elicitation

8. Conduct elicitation session(s).

Post-Elicitation

9. Combine expert assessments.
10. Conduct discrepancy and robustness analysis.
11. Provide feedback to experts.
12. Analyze the processed data.
13. Document the results.

Pre-Elicitation:
1. Define case structure.

- What values are uncertain?
- Is there historical or measurement data?
- What hypothetical measurements could be used?

Breastfeeding: achieving the new normal

Jeff Greenough/Blend Images/Corbis

See Comments pages 413 and 416

See Series pages 475 and 491

For more on the breastfeeding Series see http://www.thelancet.com/series/breastfeeding

For the Series on maternal and child nutrition see http://www.thelancet.com/series/maternal-and-child-nutrition

For more on breastfeeding and the Affordable Care Act see http://www.cdc.gov/breastfeeding/pdf/BF-Guide-508.PDF

Breastmilk makes the world healthier, smarter, and more equal: these are the conclusions of a new *Lancet* Series on breastfeeding. The deaths of 823 000 children and 20 000 mothers each year could be averted through universal breastfeeding, along with economic savings of US$300 billion. The Series confirms the benefits of breastfeeding in fewer infections, increased intelligence, probable protection against overweight and diabetes, and cancer prevention for mothers. The Series represents the most in-depth analysis done so far into the health and economic benefits that breastfeeding can produce.

However, although the Series is comprehensive, the message is not new. In 2013, a *Lancet* Series on maternal and child nutrition established that 800 000 child deaths could be prevented through breastfeeding, and called for further support. Despite consolidation of evidence for breastfeeding's benefits in recent years, in particular the economic gains to be reaped, global action has stalled. Why has so little progress been made?

Rates of breastfeeding vary wildly; it is one of the few health-positive behaviours more common in poor countries than rich ones. In low-income countries, most infants are still breastfed at 1 year, compared with less than 20% in many high-income countries and less than 1% in the UK. The reasons why women avoid or stop breastfeeding range from the medical, cultural, and psychological, to physical discomfort and inconvenience. These matters are not trivial, and many mothers without support turn to a bottle of formula. Multiplied across populations and involving multinational commercial interests, this situation has catastrophic consequences on breastfeeding rates and the health of subsequent generations.

There are glimmers of hope. Despite—or perhaps, because of—the execrable provision for paid maternity leave in the USA, the Affordable Care Act provides protected nursing breaks and insurance cover for breast pumps. Such allowances, the Series predicts, could increase breastfeeding by 25%. But, more importantly, genuine and urgent commitment is needed from governments and health authorities to establish a new normal: where every woman can expect to breastfeed, and to receive every support she needs to do so. ■ *The Lancet*

Lancet, The. 2016. "Breastfeeding: Achieving the New Normal." *The Lancet* 387 (10017): 404. doi:10.1016/S0140-6736(16)00210-5.

Social Science & Medicine 109 (2014) 55–65

Contents lists available at ScienceDirect

Social Science & Medicine

journal homepage: www.elsevier.com/locate/socscimed

Is breast truly best? Estimating the effects of breastfeeding on long-term child health and wellbeing in the United States using sibling comparisons

Cynthia G. Colen*, David M. Ramey

Department of Sociology, Ohio State University, United States

ARTICLE INFO

Article history:
Available online 29 January 2014

Keywords:
Breastfeeding
Child health
Race
Socioeconomic status
Life course epidemiology
Sibling comparisons

ABSTRACT

Breastfeeding rates in the U.S. are socially patterned. Previous research has documented startling racial and socioeconomic disparities in infant feeding practices. However, much of the empirical evidence regarding the effects of breastfeeding on long-term child health and wellbeing does not adequately address the high degree of selection into breastfeeding. To address this important shortcoming, we employ sibling comparisons in conjunction with 25 years of panel data from the National Longitudinal Survey of Youth (NLSY) to approximate a natural experiment and more accurately estimate what a particular child's outcome would be if he/she had been differently fed during infancy. Results from standard multiple regression models suggest that children aged 4 to 14 who were breast- as opposed to bottle-fed did significantly better on 10 of the 11 outcomes studied. Once we restrict analyses to siblings and incorporate within-family fixed effects, estimates of the association between breastfeeding and all but one indicator of child health and wellbeing dramatically decrease and fail to maintain statistical significance. Our results suggest that much of the beneficial long-term effects typically attributed to

Colen, Cynthia G., and David M. Ramey. 2014. “Is Breast Truly Best? Estimating the Effects of Breastfeeding on Long-Term Child Health and Wellbeing in the United States Using Sibling Comparisons.” *Social Science and Medicine* 109: 55–65. doi:10.1016/j.socscimed.2014.01.027.

Case study: Introduction

Breastfeeding definitely has high health benefits!

In some places…

Breastfeeding more common in high income families!

In some places…

However, based on the current evidence, WHO recommends exclusive breastfeeding for 6 months, with partial breastfeeding until 24 months.

Case study: Introduction

There’s a lot of data…from a few places (mostly US and UK).

Current studies struggle with confounding and self-selection bias.

Case study: Introduction

There's a lot of data…from a few places (mostly US and UK).

Current studies struggle with confounding and self-selection bias.

So we have some data, but it's not *exactly* the data we want.
Sounds like a case for expert judgment!

Case study
1. Define the structure.

- The study is focused on the impact (if any) of breastfeeding on cognitive development in three countries: USA, India, and China.
- We will use an IQ-type test as a proxy measurement for cognition.
- We are narrowly focused on this. We don't want to capture:
  - Benefits of breastfeeding instead of using low-quality formula.
  - Benefits from the mother-child interaction of the act of breastfeeding.

Pre-Elicitation:
2. Identify variables of interest.

You can't use SEJ for everything, so how do you choose?

- Is it uncertain?
- Is there data?
- Does uncertainty on this parameter impact the final endpoint?

Carefully specify these variables: you don't want questions that different experts interpret differently.

There's no rule of thumb for the best number of variables of interest.

Case study
2. Identify variables of interest.

Questions 12 through 23 concern a hypothetical ideal perfectly randomized experiment with a very large number of subjects from each of three countries. We select India and China because their populations are important from a global health perspective and yet estimates of effects of breastfeeding on cognitive performance from long-term longitudinal studies appear to be sparse for these countries. We include the U.S. because the published literature includes multiple studies of associations between breastfeeding and cognitive performance, using different data.

All infants are randomly assigned to one of four feeding cohorts.

Case study

2. Identify variables of interest.

Feeding	Feeding Patterns by Age			
	Cohorts			
	1	2	3	4
Breastfeeding, Exclusive	None	3 months	6 months	6 months
Breastfeeding, Any	None	3 to 9 months	None	6 to 24 months
Infant Formula, Exclusive	6 months	None	None	None
Infant Formula, Any	6 to 15 months	3 to 15 months	6 to 15 months	None
Complementary Foods	From 6 months	From 6 months	From 6 months	From 6 months

Case study
2. Identify variables of interest.

All formula is approved by the U.S. Food and Drug Administration and provided by the mother while holding the infant in a position where breastfeeding could have occurred.

All children are tested at age ten with the Wechsler Intelligence Scale for Children, Revised, (WISC) or its foreign equivalent, properly normed. The overall average WISC-R, (IQ) score (within each country and cohorts) is 100, st dev = 15.

You may consider the following data while developing your responses. The reported values are for the most recent data that are publicly available.

Pre-Elicitation:
3. Identify calibration variables.

	predictions	retrodictions
Domain	+++	++
Adjacent field	++	+

Avoid almanac-type questions or questions that are “google-able”.

Rule of thumb: have at least 10 seed questions.

Case study
3. Identify calibration variables.

In the NLSY79-C the average Peabody Picture Vocabulary Test (PPVT) mean score, among the children with scores, is 90.660. What is the average among first-born children with at least one PPVT score?

In the 2005-06 Demographic Health Survey for India, what is the 50th percentile for duration of breastfeeding (in months), among children who were breastfed and who were not still breastfeeding at the time of the survey?

In NLSY79-C the average age in weeks when breastfeeding ended is 9.12. What is the average age in weeks when breastfeeding ended among the 1583 only children who were breastfed?

Pre-Elicitation:
4. Identify and select experts.

- Identify potential experts through a round robin or snowflake process.

- Aim for 5-10 experts.
  - 4 *can* work
  - Returns drop off after 10+ experts

Too few experts

Expert scores: Nebraska

Results of scoring experts

Bayesian Updates: no   Weights: global   DM Optimisation: yes

Significance Level: 0.03297   Calibration Power: 1

Nr.	Id	Calibr.	Mean relative total	Mean relative realization	Numb real	UnNormalized weight	Normaliz.weigh without DM	Normaliz.weig with DM
1	NE01	7.341E-009	2.1	2.156	10	0	0	0
2	NE02	0.0003719	0.847	1.002	10	0	0	0
3	NE03	0.03297	1.282	1.447	10	0.04771	1	0.5
4	NE04	7.832E-005	1.122	1.735	10	0	0	0
5	PW	0.03297	1.282	1.447	10	0.04771		0.5
6	EW	0.3684	0.3958	0.6951	10	0.2561		0.8415

Microsoft Word Document

Too many experts

Microsoft Word Document

Expert scores: CDC ROI Final

Results of scoring experts

Bayesian Updates: no Weights: global DM Optimisation: yes

Significance Level: 0.7203 Calibration Power: 1

Nr.	Id	Calibr.	Mean relative total	Mean relative realization	Numb real	UnNormalized weight	Normaliz.weigl without DM	Normaliz.weig with DM
1	01	0.7203	2.597	2.305	10	1.66	1	0.5
2	02	1.602E-005	1.904	1.655	10	0	0	0
3	03	1.273E-006	2.344	3.49	10	0	0	0
4	04	5.559E-006	2.961	2.719	10	0	0	0
5	05	0.4988	2.341	1.39	10	0	0	0
6	06	0.01651	1.39	1.355	10	0	0	0
7	07	2.181E-007	2.09	3.345	10	0	0	0
8	08	0.4988	3.825	1.737	10	0	0	0
9	09	0.1321	4.623	1.719	10	0	0	0
10	10	1.273E-006	3.08	3.071	10	0	0	0
11	11	0.02366	4.113	1.82	10	0	0	0
12	12	0.00917	2.797	2.304	10	0	0	0
13	13	0.007147	2.758	2.063	10	0	0	0
14	14	0.0001328	3.815	3.279	10	0	0	0
15	15	0.1249	3.843	2.66	10	0	0	0
16	16	0.0003053	2.79	2.057	10	0	0	0
17	17	0.02919	2.471	1.745	10	0	0	0
18	18	1.428E-006	2.727	3.155	10	0	0	0
19	19	0.4988	2.032	1.91	10	0	0	0
20	20	0.04675	2.208	2.183	10	0	0	0
21	PW	0.7203	2.597	2.305	10	1.66		0.5
22	EW	0.2328	1.117	1.23	10	0.2864		0.0543

Pre-Elicitation:
4. Identify and select experts.

After identifying experts, tell them:

- Purpose of study
- Format of elicitations
- Payment details
- Use of experts' names
  - Link between name and assessments (or qualitative information) *preserved but not published*
  - List of experts and affiliations published

Pre-Elicitation:
5. Write the elicitation protocol.

Include:

- The motivation for the study
- The questions (calibration questions can be labelled or not)
- May want to include a briefing book

Microsoft Word Document

Pre-Elicitation:
6. Pilot test the protocol.

With a substantive expert (who wasn't involved in writing the protocol), check:

- Are the questions clear?
- Does the structure make sense?
- Is additional information needed to make sure we're capturing what we want to capture?
- Is the timing appropriate?

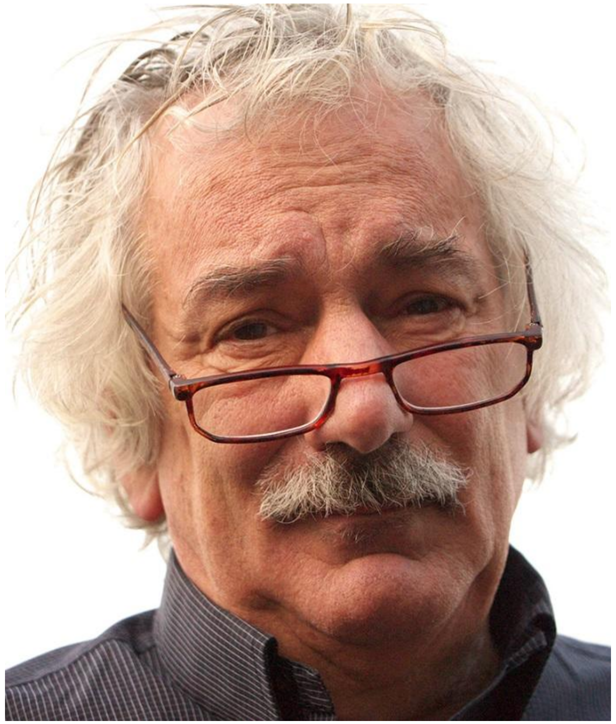
Pre-Elicitation:
7. Train the experts.

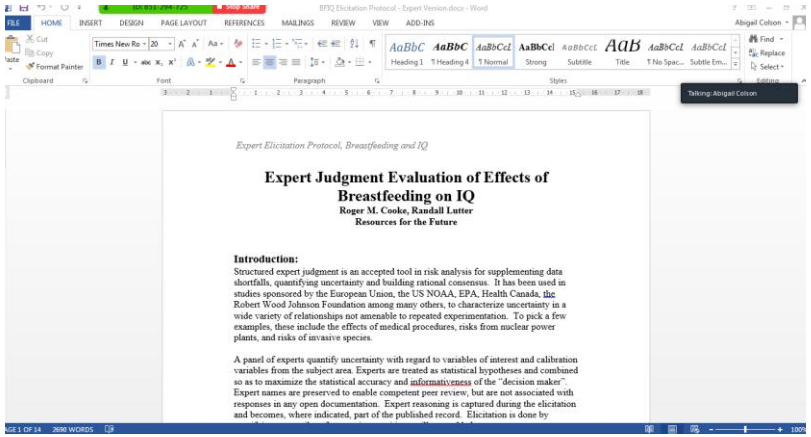
DEPENDS ON TIME, BUDGET, LOCATION OF THE EXPERTS, AND COMPLEXITY OF THE ELICITATION.

- 30 minute, 1:1 training session
- Webinar
- Half day group meeting
- Multi-day workshop

- Discuss case structure
- Explain method and scoring
- Discuss over-confidence

Elicitation:
8. Conduct elicitation session(s).





Expert Elicitation Protocol, Breastfeeding and IQ

Expert Judgment Evaluation of Effects of Breastfeeding on IQ

Roger M. Cooke, Randall Lutter

Resources for the Future

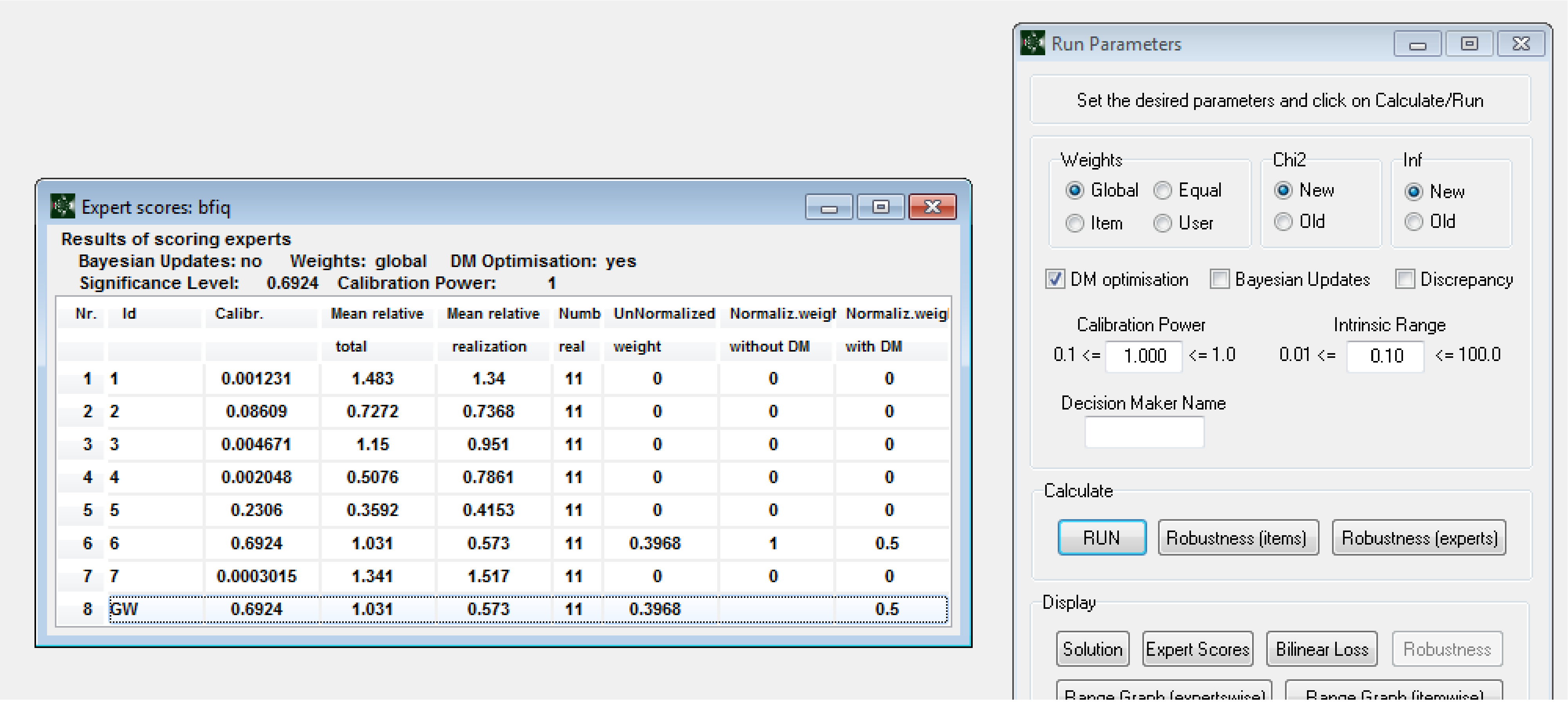
Introduction:

Structured expert judgment is an accepted tool in risk analysis for supplementing data shortfalls, quantifying uncertainty and building rational consensus. It has been used in studies sponsored by the European Union, the US NOAA, EPA, Health Canada, the Robert Wood Johnson Foundation among many others, to characterize uncertainty in a wide variety of relationships not amenable to repeated experimentation. To pick a few examples, these include the effects of medical procedures, risks from nuclear power plants, and risks of invasive species.

A panel of experts quantify uncertainty with regard to variables of interest and calibration variables from the subject area. Experts are treated as statistical hypotheses and combined so as to maximize the statistical accuracy and informativeness of the "decision maker". Expert names are preserved to enable competent peer review, but are not associated with responses in any open documentation. Expert reasoning is captured during the elicitation and becomes, where indicated, part of the published record. Elicitation is done by

Capture qualitative reasoning alongside the quantitative judgments.

Post-Elicitation:
9. Combine expert assessments.

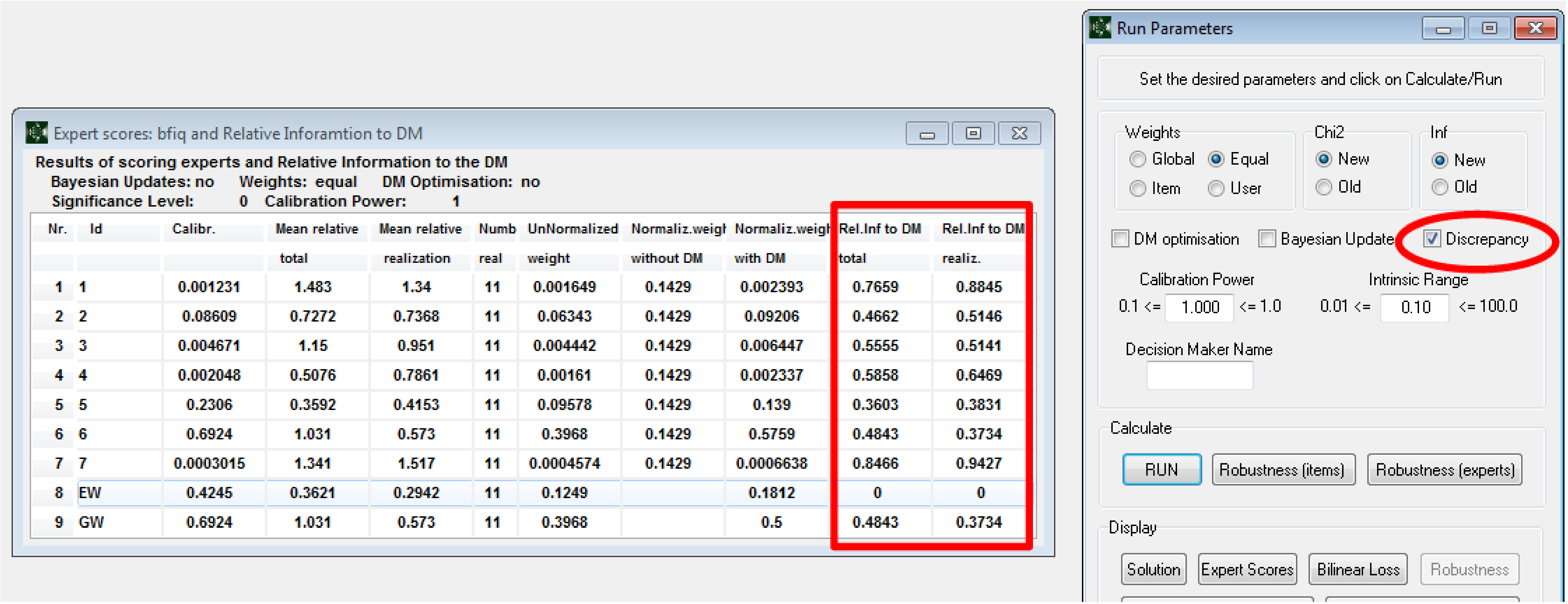


Expert scores: bfiq

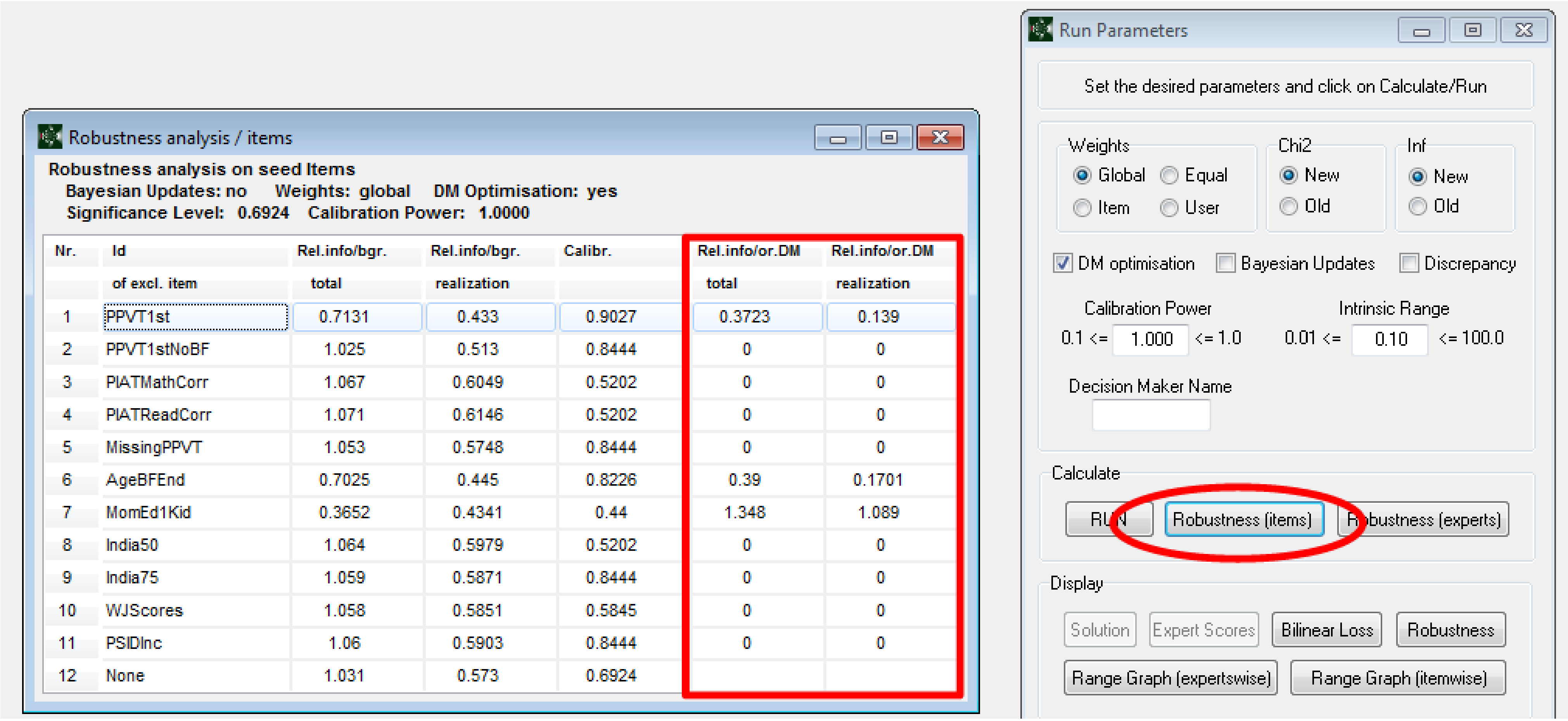
Results of scoring experts
Bayesian Updates: no Weights: global DM Optimisation: yes
Significance Level: 0.6924 Calibration Power: 1

Nr.	Id	Calibr.	Mean relative total	Mean relative realization	Numb real	UnNormalized weight	Normaliz.weigt without DM	Normaliz.weig with DM
1	1	0.001231	1.483	1.34	11	0	0	0
2	2	0.08609	0.7272	0.7368	11	0	0	0
3	3	0.004671	1.15	0.951	11	0	0	0
4	4	0.002048	0.5076	0.7861	11	0	0	0
5	5	0.2306	0.3592	0.4153	11	0	0	0
6	6	0.6924	1.031	0.573	11	0.3968	1	0.5
7	7	0.0003015	1.341	1.517	11	0	0	0
8	GW	0.6924	1.031	0.573	11	0.3968		0.5

Post-Elicitation:
10. Conduct discrepancy and robustness analysis.

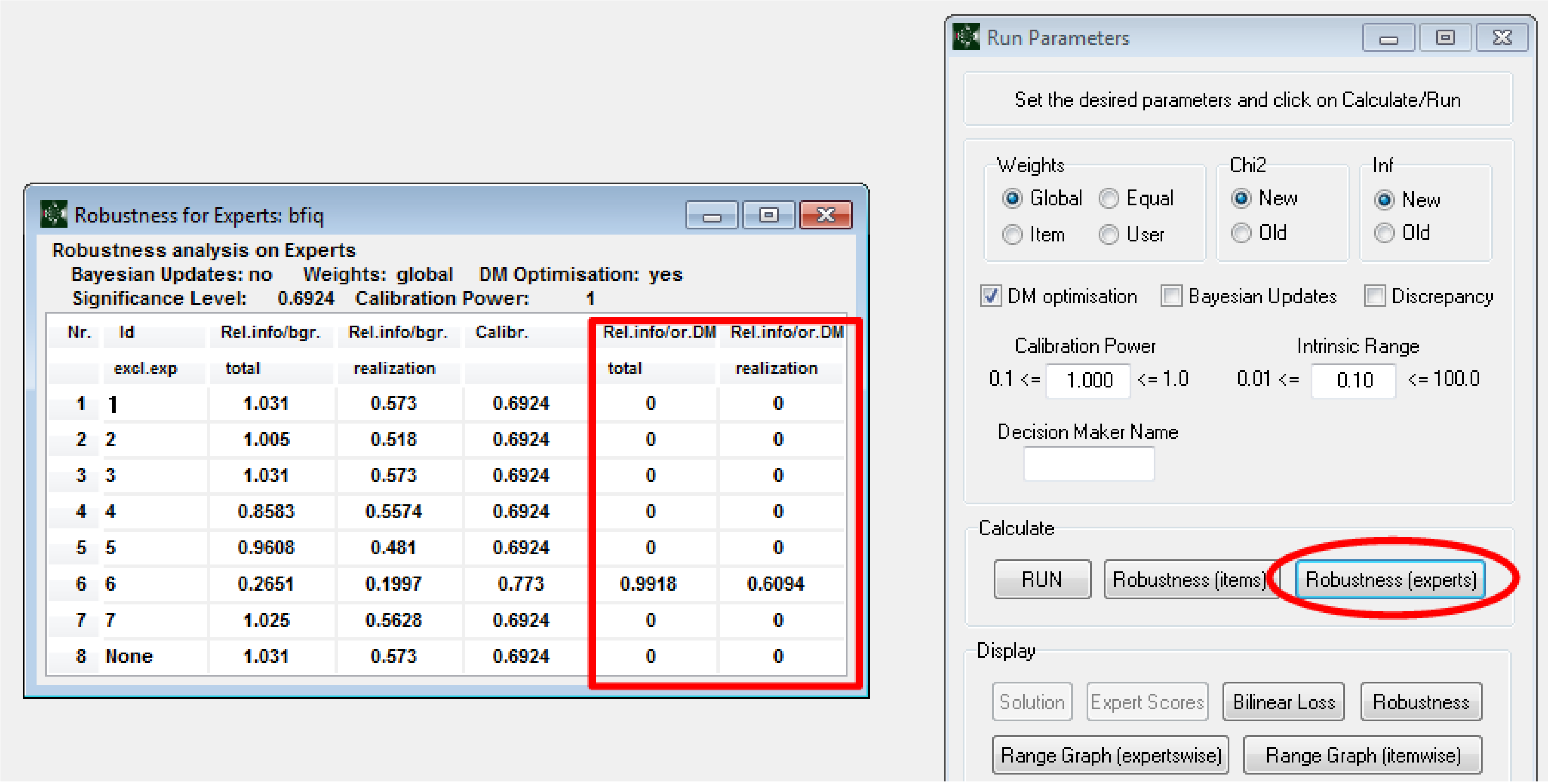


Post-Elicitation:
10. Conduct discrepancy and robustness analysis.



Nr.	Id of excl. item	Rel.info/bgr. total	Rel.info/bgr. realization	Calibr.	Rel.info/or.DM total	Rel.info/or.DM realization
1	PPVT1st	0.7131	0.433	0.9027	0.3723	0.139
2	PPVT1stNoBF	1.025	0.513	0.8444	0	0
3	PIATMathCorr	1.067	0.6049	0.5202	0	0
4	PIATReadCorr	1.071	0.6146	0.5202	0	0
5	MissingPPVT	1.053	0.5748	0.8444	0	0
6	AgeBFEnd	0.7025	0.445	0.8226	0.39	0.1701
7	MomEd1Kid	0.3652	0.4341	0.44	1.348	1.089
8	India50	1.064	0.5979	0.5202	0	0
9	India75	1.059	0.5871	0.8444	0	0
10	WJScores	1.058	0.5851	0.5845	0	0
11	PSIDInc	1.06	0.5903	0.8444	0	0
12	None	1.031	0.573	0.6924		

Post-Elicitation:
10. Conduct discrepancy and robustness analysis.


Robustness for Experts: bfiq

Robustness analysis on Experts
Bayesian Updates: no   Weights: global   DM Optimisation: yes
Significance Level: 0.6924   Calibration Power: 1

Nr.	Id excl.exp	Rel.info/bgr. total	Rel.info/bgr. realization	Calibr.	Rel.info/or.DM total	Rel.info/or.DM realization
1	1	1.031	0.573	0.6924	0	0
2	2	1.005	0.518	0.6924	0	0
3	3	1.031	0.573	0.6924	0	0
4	4	0.8583	0.5574	0.6924	0	0
5	5	0.9608	0.481	0.6924	0	0
6	6	0.2651	0.1997	0.773	0.9918	0.6094
7	7	1.025	0.5628	0.6924	0	0
8	None	1.031	0.573	0.6924	0	0

Post-Elicitation:
11. Provide feedback to experts.

Have the experts review:

- What you captured of their reasoning
- The combined decision maker assessments
- Their scores (not needed, but experts often ask)

Post-Elicitation:
12. Analyze the processed data.

Post-Elicitation:
13. Document the results.

Koch, Benjamin J., Catherine M. Febria, Roger M. Cooke, Jacob D. Hosen, Matthew E. Baker, Abigail R. Colson, Solange Filoso, et al. 2015. “Suburban Watershed Nitrogen Retention: Estimating the Effectiveness of Stormwater Management Structures.” *Elementa: Science of the Anthropocene* 3 (July): 000063. doi:10.12952/journal.elementa.000063.

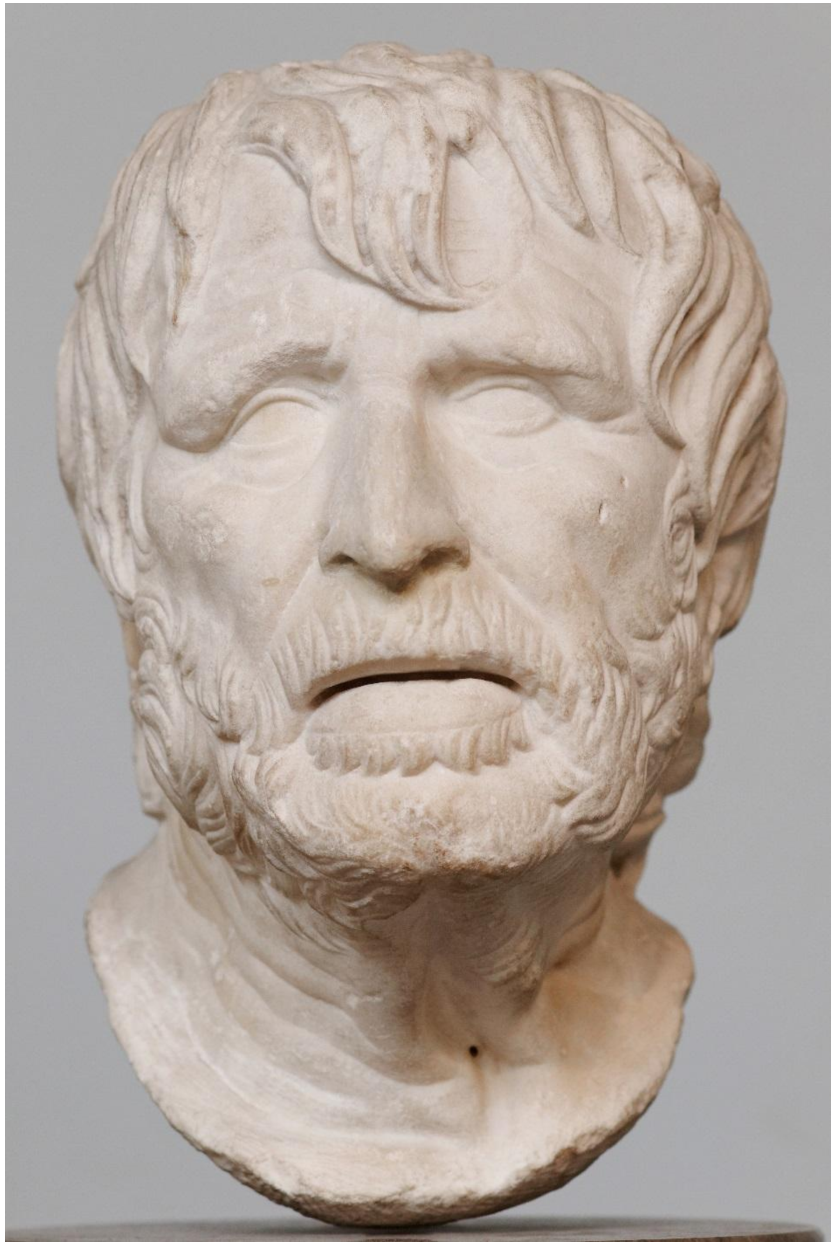
Wittmann, Marion E., Roger M. Cooke, John D. Rothlisberger, and David M. Lodge. 2014. “Using Structured Expert Judgment to Assess Invasive Species Prevention: Asian Carp and the Mississippi—Great Lakes Hydrologic Connection.” *Environmental Science & Technology* 48 (4): 2150–56. doi:10.1021/es4043098.

Wittmann, Marion E., Roger M. Cooke, John D. Rothlisberger, Edward S. Rutherford, Hongyan Zhang, Doran M. Mason, and David M. Lodge. 2015. “Use of Structured Expert Judgment to Forecast Invasions by Bighead and Silver Carp in Lake Erie.” *Conservation Biology* 29 (1): 187–97. doi:10.1111/cobi.12369.

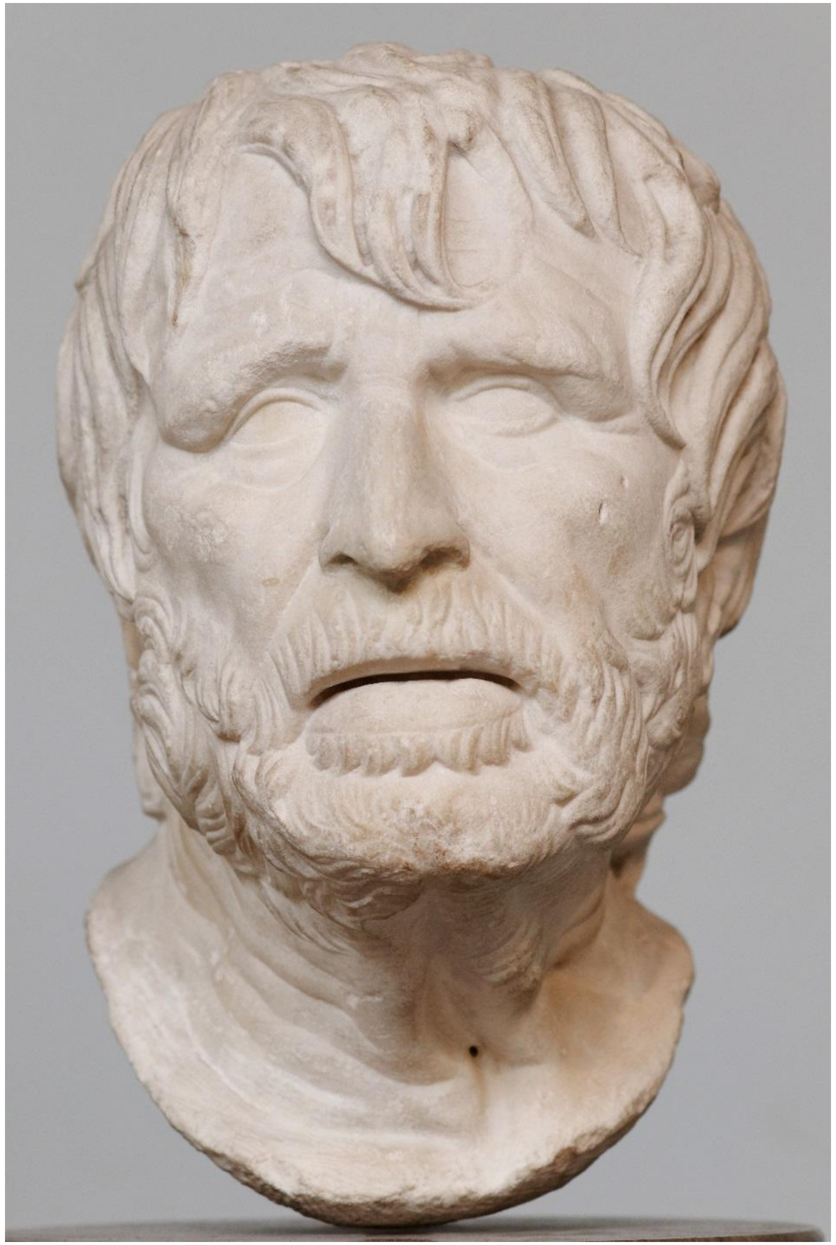
Documentation Bibliography

Frequently Heard Comments & Questions

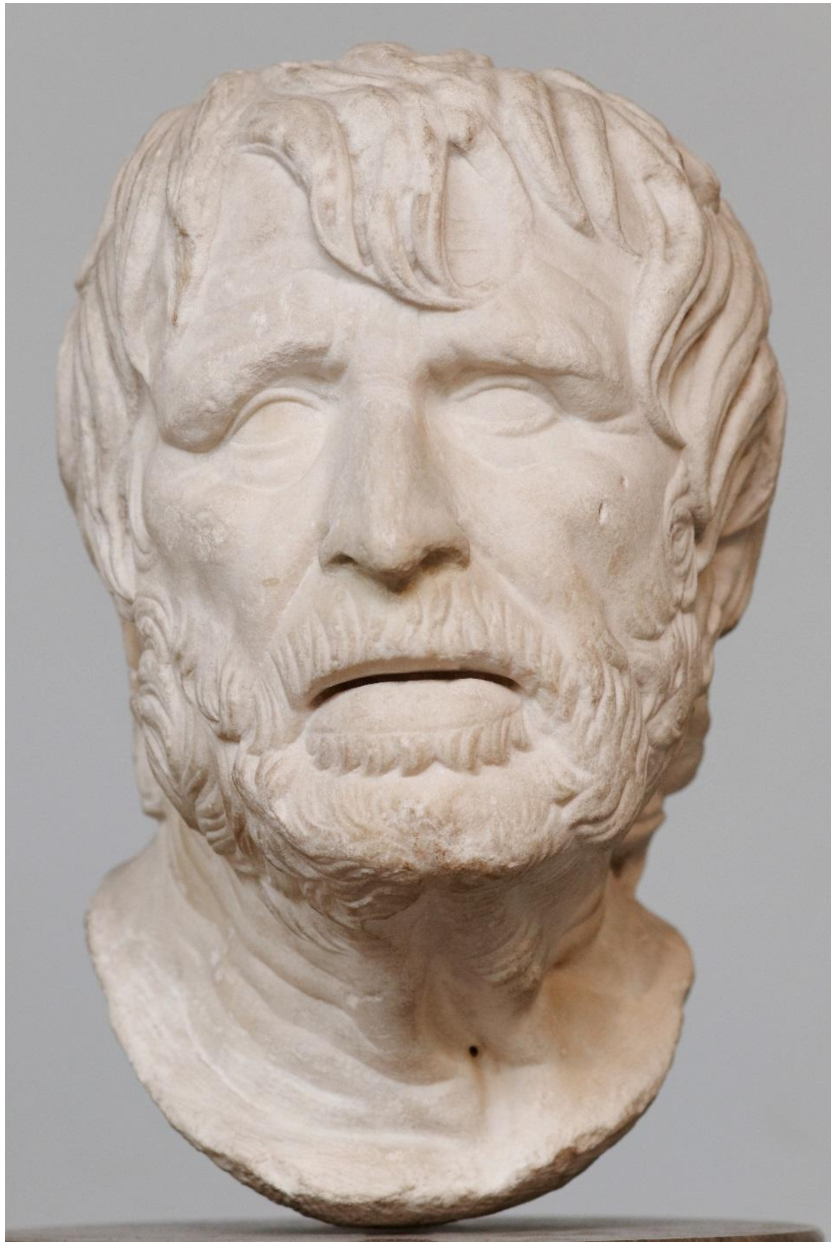
FROM EXPERTS AND PROBLEM OWNERS



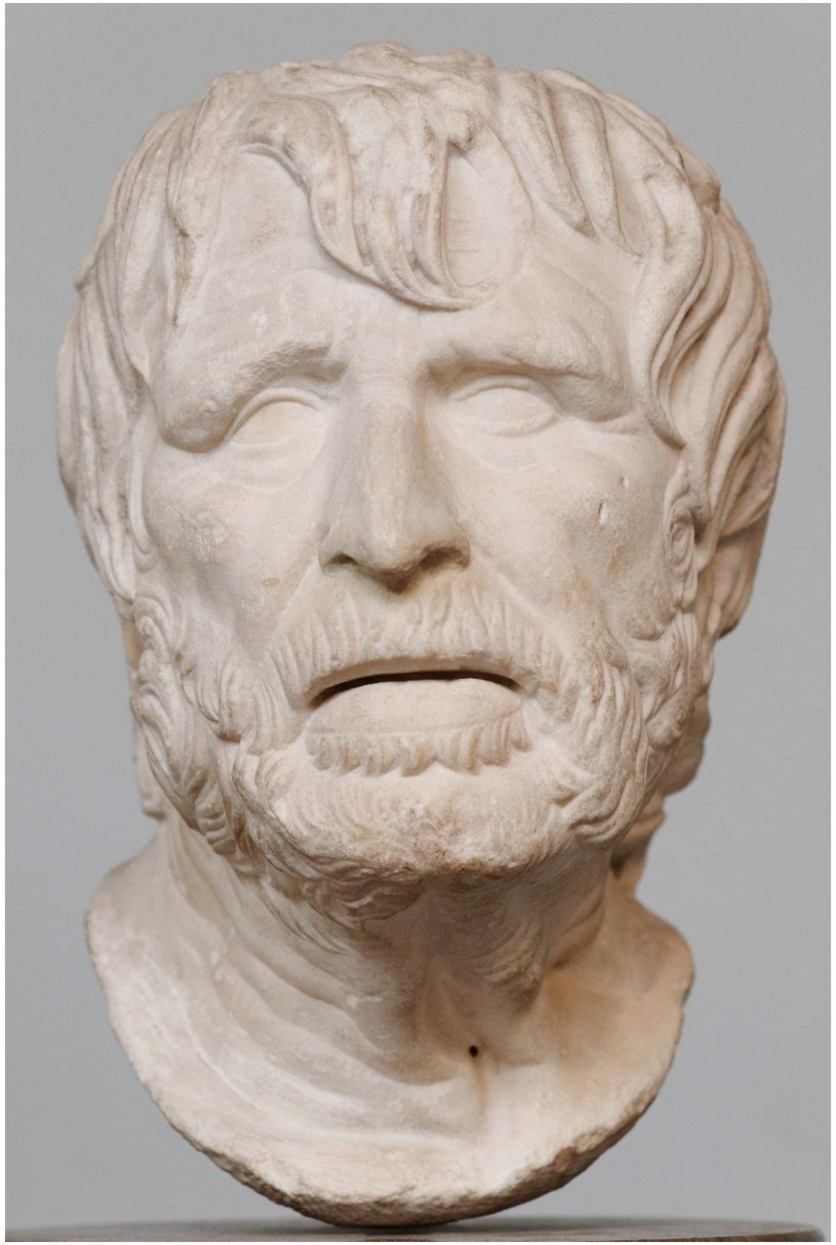
I don’t know that!?!?



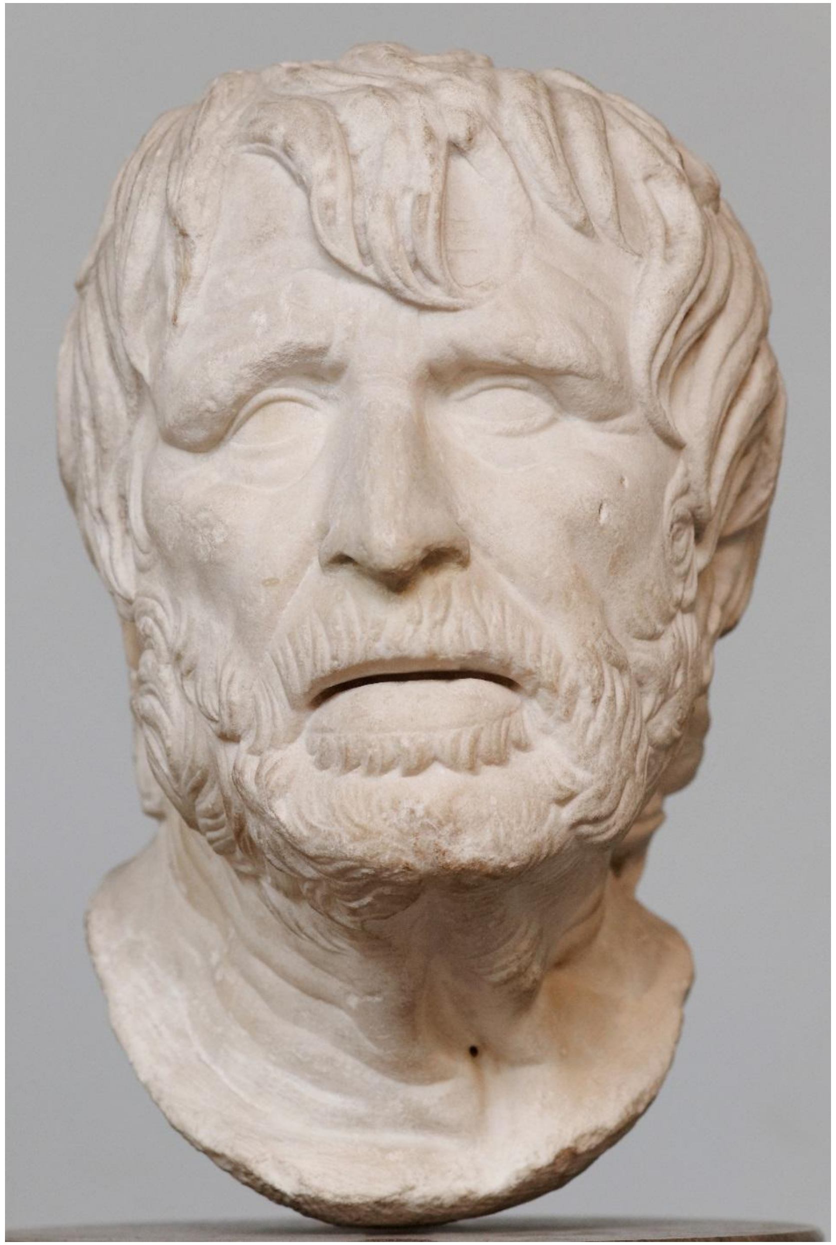
I need more information to assess this.



Does this answer look ok?



I can’t do this.



I can’t do this.*

*Not frequently heard.

So you test them like school children?!

So you test them to see who's *really* an expert?

Why am I paying for this expert and then giving her zero weight?

Why am I paying for so many experts and only giving weight to one?!?!

That assessment is crazy! Who said that?

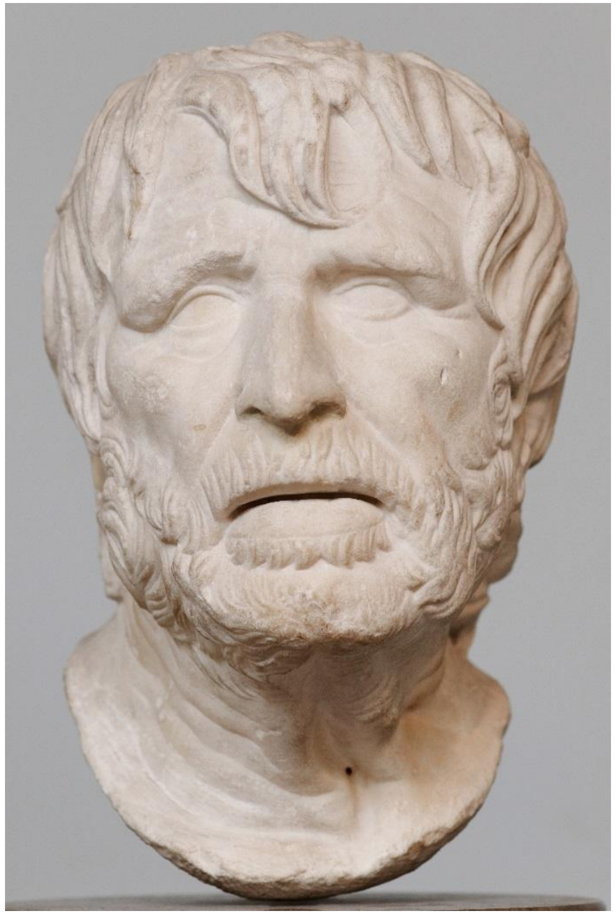
That assessment is crazy! Who said that?*

*Not frequently heard.

Ok…but I just want to use equal weights after all.

Ok…but I just want to use equal weights after all.*

*Not frequently heard.



Questions?