

The Evidence-to-Policy Pipeline

Eva Vivalt

University of Toronto

July 31, 2025

Motivation

- Policymakers make very consequential policy decisions
- In order to have evidence-based policymaking, we need three things:
 1. Evidence
 2. Ability to interpret the evidence
 3. Willingness and capacity to act on that evidence
- There may be barriers at all three stages

Motivation

- Policymakers make very consequential policy decisions
- In order to have evidence-based policymaking, we need three things:
 1. Evidence
 2. Ability to interpret the evidence
 3. Willingness and capacity to act on that evidence
- There may be barriers at all three stages



Today

- Discuss the barriers in the evidence-to-policy pipeline drawing on three papers in this research agenda:
 - Vivalt (2020). "How Much Can We Generalize From Impact Evaluations?" *Journal of the European Economics Association*
 - Vivalt and Coville (2023). "How Do Policymakers Update Their Beliefs?" *Journal of Development Economics*
 - Vivalt, Coville, and KC (2025). "Local Knowledge, Formal Evidence, and Policy Decisions", *Journal of Development Economics*

```
graph LR; A[Generating Evidence] --> B[Interpreting Evidence]; B --> C[Acting on Evidence];
```

Generating
Evidence

Interpreting
Evidence

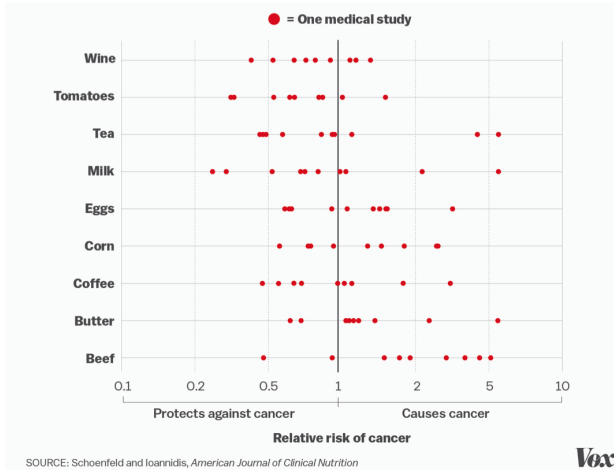
Acting on
Evidence

Generating Evidence

- Recently, there has been an explosion of rigorous studies in economics
- However, this evidence may not always be the most applicable to problems policymakers face
 - Academics may not focus on the highest-value topics
 - Even when work is done on the right topics, are the experiments designed for the problem at hand?

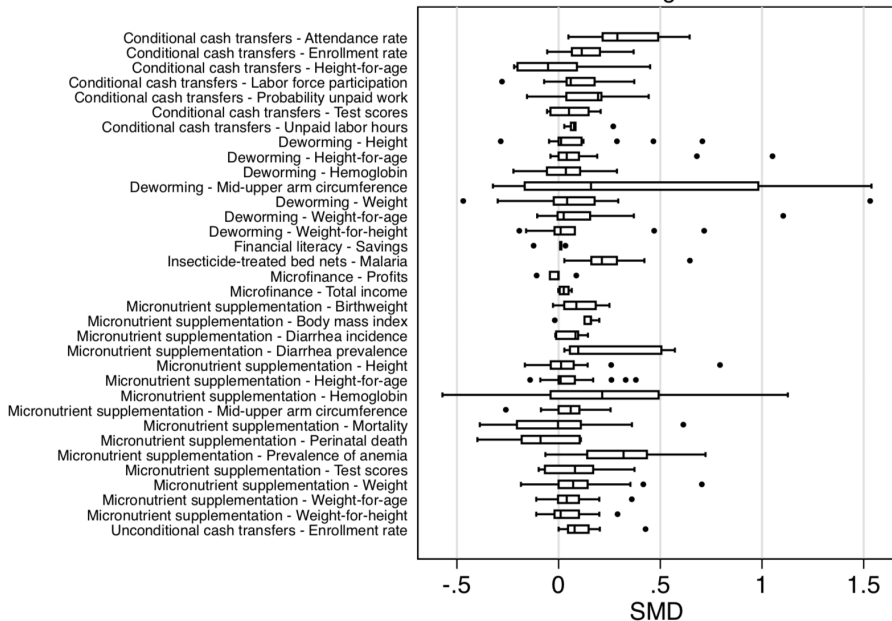
Results May Not Generalize

Figure: “Everything both causes and cures cancer”



Schoenfeld and Ioannidis 2013

Variation in Programs' Effects



Problems

- If trying to predict the treatment effect of a similar study using only the mean treatment effect in an intervention-outcome combination:
 - An inference about another study will have the correct sign about 60% of the time
 - The median ratio of the root-MSE to that mean is 2.5 across intervention-outcome combinations
- You can often find *a* study to support your views which may not be what the totality of the evidence shows
- There is often a real trade-off between internal and external validity via who is included in the sample
 - More diverse groups included → increased variance in the outcome variable, possibly increased research costs for some hard-to-reach groups → decreased statistical power, but better ability to speak to impacts for those groups

Research Question

- How do policymakers value advice from local experts versus formal evidence from impact evaluations when making policy decisions? (Vivalt, Coville and KC 2025)

Setting

- Can't bring policymakers to a lab
- Bring the lab to policymakers - collaboration leveraging World Bank and Inter-American Development Bank workshops
- Sample: 156 policy professionals from 5 workshops
 - Policymakers are government officials, either providing technical support (e.g. monitoring and evaluation specialist) or with decision-making power
 - Policy practitioners are mostly operational bank staff, some NGO counterparts

Method

Discrete choice experiment:

Now imagine that you need to provide a recommendation to a counterpart agency in your country on which of two programs to implement. A study was done on each program, with the results below. Please select which program you would recommend.

	<i>Study on Program A</i>	<i>Study on Program B</i>
Method	Observational	Experimental
Location	Same country	A different country in the same region
Impact on enrollment rates, with margin of error (95% confidence interval)	+10 percentage points, +/-10 percentage points	+5 percentage points, +/-1 percentage point

A local expert tells you that they believe Program A would perform better in your context.

Which program do you recommend?

Attributes and Levels

Attributes	Levels
Method	Experimental, Quasi-experimental, Observational
Location	Same country, Different country in the same region, Different region
Impact	0, +5, +10 percentage points
Confidence Interval	+/-1, +/-10 percentage points
Recommended	Yes, No

Willingness-to-Pay in Terms of Estimated Impact

	Pooled	Policymaker	Policy Practitioner
	(1)	(2)	(3)
Quasi-Experimental	-1.55 (1.61)	0.66 (3.10)	-2.89 (1.84)
Experimental	-4.83*** (1.84)	-5.26 (3.42)	-4.70** (2.18)
Different country, same region	-2.07 (1.57)	-1.03 (2.95)	-2.88* (1.74)
Same country	-6.70*** (1.73)	-5.77* (2.97)	-7.55*** (2.09)
Recommended by local expert	-3.11** (1.23)	-4.99* (2.84)	-2.01* (1.12)
Small C.I.	-3.67*** (1.24)	-1.21 (2.21)	-5.13*** (1.49)
Observations	818	407	411

```
graph LR; A[Generating Evidence] --> B[Interpreting Evidence]; B --> C[Acting on Evidence];
```

Generating
Evidence

Interpreting
Evidence

Acting on
Evidence

Interpreting Evidence

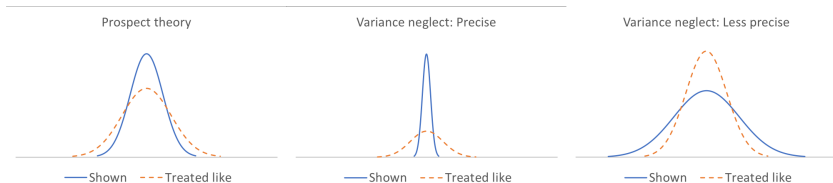
- We know people have behavioural biases
- Policymakers may be special:
 - They may have undergone particular selection pressures
 - They may have adopted norms
 - They may face certain types of information
- Even if not, the decisions they make can be very important
- In work with Aidan Coville (2023), look at two biases of interest: asymmetric optimism and variance neglect

Asymmetric Optimism

- Intuition: Update more when see more positive values relative to priors
- Or: Updating is fine, just misinterprets signal (Rabin and Schrag, 1999)

Variance Neglect

- Similar to extension neglect, “law of small numbers”
- Does not appropriately update on information about the uncertainty of the results - updates “too much” on results with large confidence intervals and “too little” on results with small confidence intervals



Sample

- 180 from World Bank/Inter-American Development Bank workshops, 68 from post-workshop videoconferences, 152 from headquarters surveys (400 total)
- Workshops:
 - ~ 1/3 policymakers
 - ~ 1/3 policy practitioners
 - ~ 1/3 researchers

Strategy

- Ask policymakers to provide priors as to the effects of various programs
- Present them with results from academic studies
- Elicit posteriors

Experiment

Consider a conditional cash transfer (CCT) program in which a household is provided with the equivalent of \$20 USD per month as long as all their children between age 6 and 16 stay in school. The program targets rural areas. Just before the CCT program is implemented, 90% of these children were enrolled in school.

Please provide your best estimate of how much the CCT increased enrolment (in percentage points). Remember that an increase by X percentage points is not the same thing as an increase by X percent!

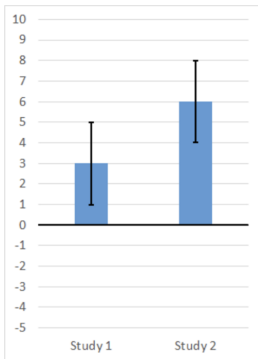
Experiment

Please use the sliders below to let us know how likely you think the program was to have had a certain impact. The number of points you assign to each row will directly correspond to how likely you think the true impact was to have fallen within that range. Place more points on the ranges that you think are very likely and fewer points on the ranges you think are unlikely. You can also enter or revise your estimates by entering numbers in the right-hand column.

	0	10	20	30	40	50	60	70	80	90	100	
9 to 9.99											0	
8 to 8.99											0	
7 to 7.99											0	
6 to 6.99											0	
5 to 5.99											0	
4 to 4.99											0	
3 to 3.99											0	
2 to 2.99											0	
1 to 1.99											0	

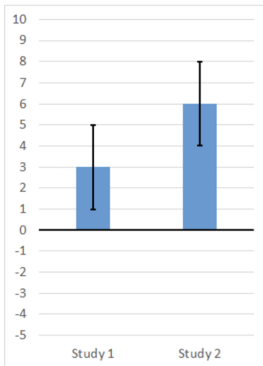
Experiment

Suppose that 2 independent studies of this program were conducted. Each study followed the exact same design, but you do not know in which order they were done. One study found that the program increased enrolment by 3.0 percentage points, plus or minus 2.0 (this means that the 95% confidence interval was between 1.0 and 5.0 percentage points). The other study found that the program increased enrolment by 6.0 percentage points, plus or minus 2.0. A graphical depiction of the results of the 2 studies is provided below:



If a third study were done on this program following the exact same design, what effect do you think it would find? Please provide your best estimate.

Experiment



Now please use the sliders below to indicate how likely the study would be to find an effect within a given range. As before, use the sliders to place more points on the ranges that you think are very likely and fewer points on the ranges you think are unlikely.

0 10 20 30 40 50 60 70 80 90 100

9 to 9.99 0

8 to 8.99 0

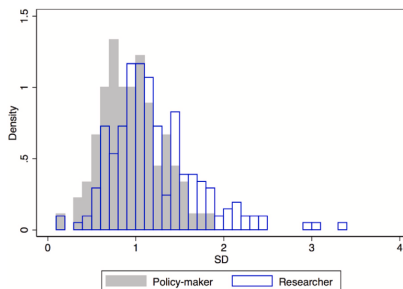
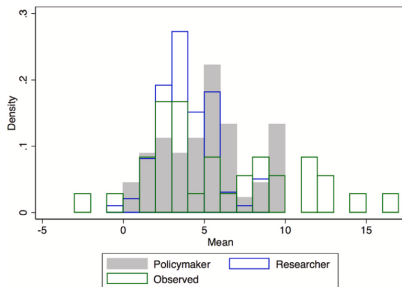
Experiment

Please indicate how you would prefer our funds to be allocated. Allocations must sum to 100%:

Cash transfers	0
School meals	0
Further research	0
Total	0

Prior Beliefs

- Policymakers expect larger effects and are more certain in their beliefs



Findings

- Both biases exist. In our setting, they would lead policymakers to believe the program had a 10% larger impact than it did
- Participants were also overconfident: if they were to construct a 95% credible interval in which they thought the true effect would fall, that interval would be 31% narrower than that of a Bayesian
- Policymakers, policy practitioners, and researchers seem to have the same biases, despite different priors

Information Changes Allocations

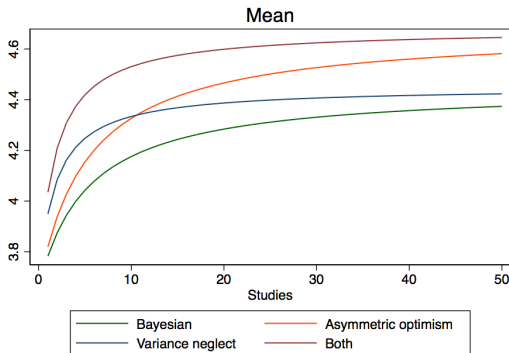
Table: Regressions of Allocations on Evidence Shown

	CCT Funding			Research Funding		
	(1)	(2)	(3)	(4)	(5)	(6)
Large Point Estimate	7.137*** (1.95)		7.450*** (2.68)	1.910 (2.48)		-0.347 (3.30)
Large Confidence Interval		-8.379*** (1.98)	-8.291*** (2.58)		5.741** (2.48)	3.132 (3.60)
Large Point Estimate * Large C.I.			-0.354 (3.78)			4.435 (4.94)
Observations	308	308	308	308	308	308

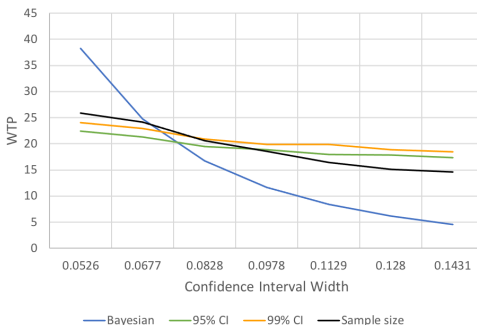
Implications

- Under these biases, one may not necessarily even converge to the evidence base over time

Figure: Simulations of biases affecting interpretation



Sample Sizes May Be Easier to Interpret



- In a follow-up incentive-compatible willingness-to-pay exercise, participants were willing to pay different amounts for results with 95% or 99% confidence intervals or sample sizes that represented the same amount of precision

```
graph LR; A[Generating Evidence] --> B[Interpreting Evidence]; B --> C[Acting on Evidence];
```

Generating
Evidence

Interpreting
Evidence

Acting on
Evidence

Acting on Evidence

- Even if you have evidence and interpret it perfectly, you may not act on it
- Many political economy questions
- Organizational inertia
- Misunderstanding goals
 - E.g., do policy professionals correctly predict others' preferences?
- Structural issues
 - E.g., if data are “too big” or “too complex” for easy analysis, can LLM tools help users engage and learn?

Parting Thoughts

- Can we develop broader tools that help people make better decisions?
- Social Science Prediction Platform: collects forecasts from experts about the effects of different programs
 - Connects forecasts with impact evaluation results
 - Can we improve forecasts and get more out of them?

www.socialscienceprediction.org

Parting Thoughts

- There is still a lot of room to do good research in this area!
- Stay in touch:

`eva.vivalt@utoronto.ca`