Knowing what we know: Comparing and consolidating empirical findings

Solomon M. Hsiang
UC Berkeley

BITSS
Berkeley, June 3rd 2014

"Doing empirical research is like making sausage. Doing meta-analysis is like using sausage to make sausage"

"Doing empirical research is like making sausage. Doing meta-analysis is like using sausage to make sausage"

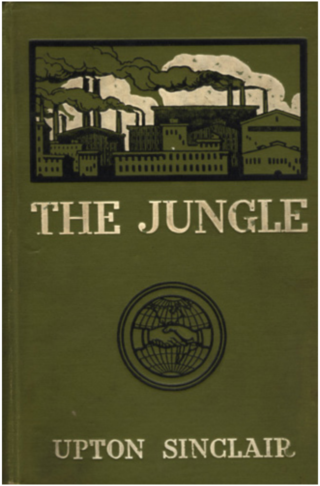
This is true. But it is not a reason to punt on meta-analysis.

"Doing empirical research is like making sausage. Doing meta-analysis is like using sausage to make sausage"

This is true. But it is not a reason to punt on meta-analysis.

(1) Complicated things always look like making sausage until you understand how to do it. But complexity is not a reason to not do something important.

E.g. Most people think all of statistics (or academic research generally) looks like making sausage.



(2) Lots of people eat lots of sausage. Somebody has to look out for them. If we don't make safe sausage, somebody else will make crappy sausage and feed it to all those hungry people.

Sausage, Italian, Pork, Cooked

italian sausage, meat, sausage, dinner, pork

C-
Grade

286
Calories

Nutrition Facts

Serving Size 1 link, 4/lb (83 g)

Per Serving	% Daily Value*
Calories 286	
Calories from Fat 204	
Total Fat 22.7g	**35%**
Saturated Fat 7.9g	**40%**
Polyunsaturated Fat 2.7g	
Monounsaturated Fat 9.9g	
Cholesterol 47mg	**16%**
Sodium 1002mg	**42%**
Potassium 252.32mg	**7%**
Carbohydrates 3.5g	**1%**
Dietary Fiber 0.1g	**0%**
Sugars 0.7g	
Protein 15.9g	

Vitamin A 1% · Vitamin C 0%

Calcium 2% · Iron 7%

(3) Sausage contains lots of good stuff! It's a waste to throw out tidbits of research just because they aren't the filet mignon. The public should at least get to use all of the research that it paid for.

Why do research?

The objective of research is to learn about the world.

Settling armchair debates requires only that somebody is right and somebody is wrong (i.e. hypothesis tests).

Designing welfare-improving public policy requires that we know what we know and that our quantitative values are right (or as good as we can get them).

Why do research?

The objective of research is to learn about the world.

Settling armchair debates requires only that somebody is right and somebody is wrong (i.e. hypothesis tests).

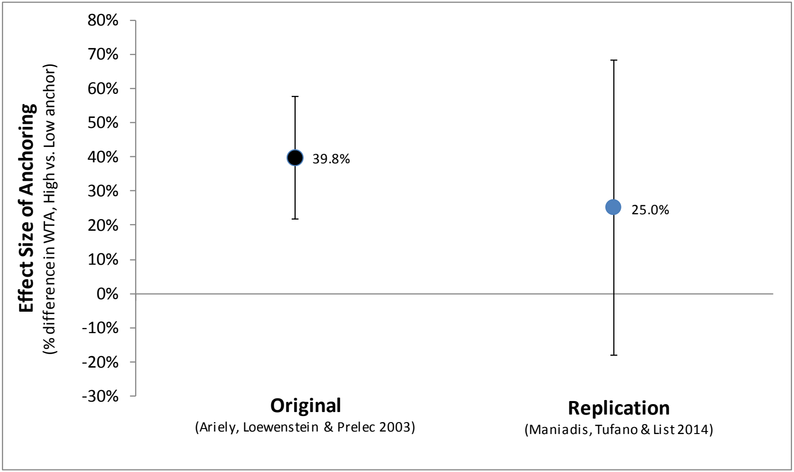
Designing welfare-improving public policy requires that we know what we know and that our quantitative values are right (or as good as we can get them).

Knowledge accumulates study by study.

Our collective knowledge is some composite of prior studies.

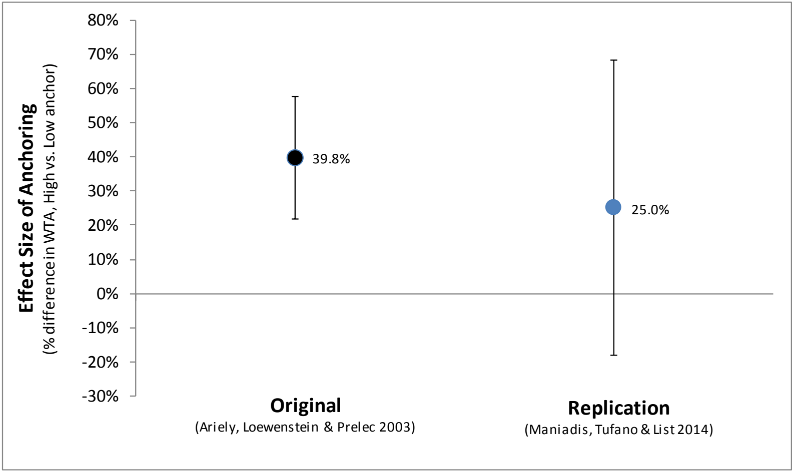
By formalizing how we combine information from studies, we can be clear and precise about what we mean by knowledge and our grasp of it.

Example: Does anchoring affect valuation?



Probably. List et al. should not have claimed to refute Ariely et al.

Example: Does anchoring affect valuation?

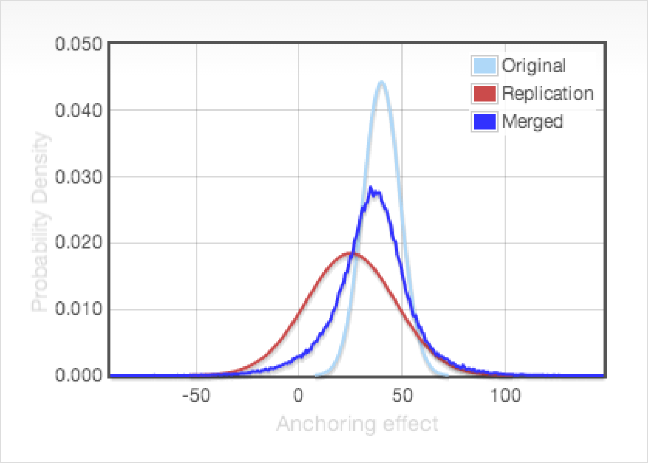


Probably. List et al. should not have claimed to refute Ariely et al.

But what is the best estimate, now that we have more information?

Example: Does anchoring affect valuation?

But what is the best estimate, now that we have more information?

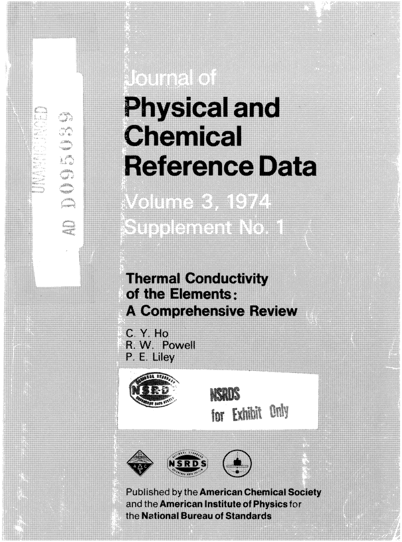


Setting the bar

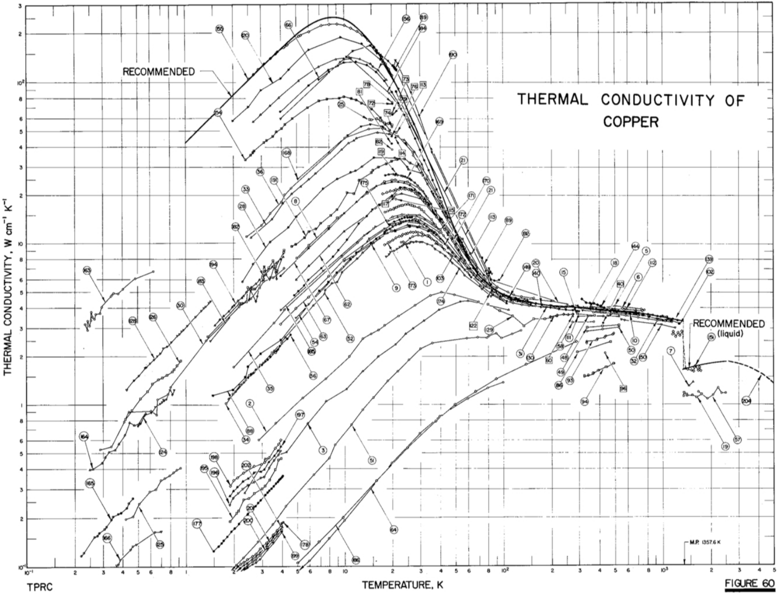
Did you use a cell phone, computer, or light bulb today?

Setting the bar

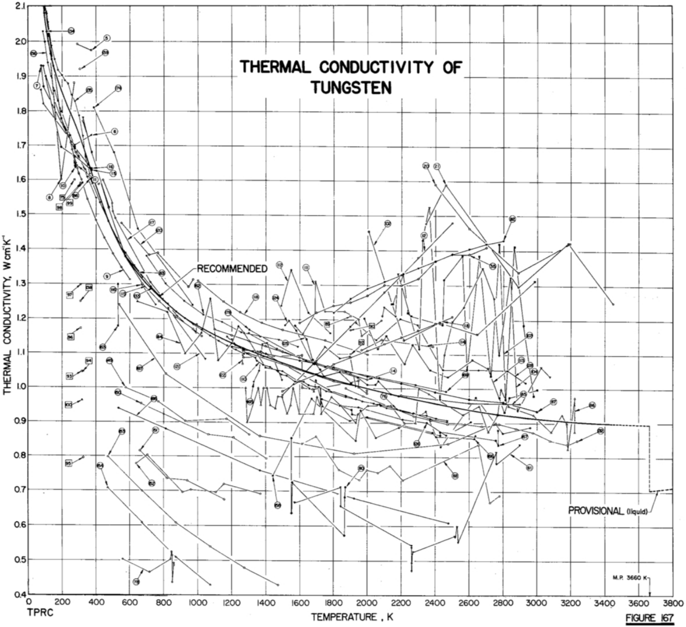
Did you use a cell phone, computer, or light bulb today?



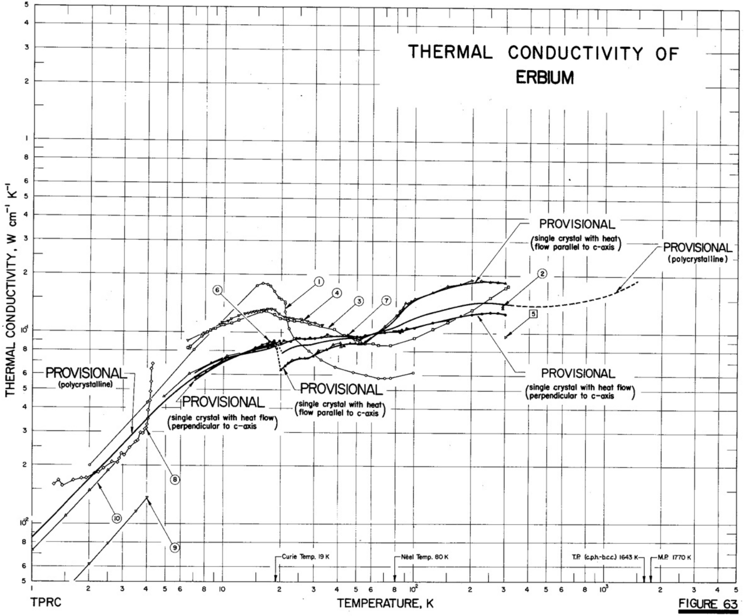
Setting the bar



Setting the bar



Setting the bar



Warming increases the risk of civil war in Africa
Burke, Miguel, et al. (PNAS, 2009)

> *Temperature variables are strongly related to conflict incidence over our historical panel, with a 1 C increase in temperature in our preferred specification leading to a 4.5% increase in civil war in the same year and a 0.9% increase in conflict incidence in the next year.*

Warming increases the risk of civil war in Africa
Burke, Miguel, et al. (PNAS, 2009)

> *Temperature variables are strongly related to conflict incidence over our historical panel, with a 1 C increase in temperature in our preferred specification leading to a 4.5% increase in civil war in the same year and a 0.9% increase in conflict incidence in the next year.*

Climate not to blame for African civil conflict
Buhaug (PNAS, 2010)

> *Scientific claims about a robust correlational link between climate variability and civil war do not hold up to closer inspection.... The challenges imposed by future global warming are too daunting to let the debate on social effects and required countermeasures be sidetracked by atypical, nonrobust scientific findings and actors with vested interests.*

Warming increases the risk of civil war in Africa
Burke, Miguel, et al. (PNAS, 2009)

> *Temperature variables are strongly related to conflict incidence over our historical panel, with a 1 C increase in temperature in our preferred specification leading to a 4.5% increase in civil war in the same year and a 0.9% increase in conflict incidence in the next year.*

Climate not to blame for African civil conflict
Buhaug (PNAS, 2010)

> *Scientific claims about a robust correlational link between climate variability and civil war do not hold up to closer inspection.*

Reconciling disagreement over climate-conflict results in Africa
Hsiang & Meng (PNAS, 2014)

> *We reexamine this apparent disagreement by comparing the statistical models from the two papers using formal tests. When we implement the correct statistical procedure, we find that the evidence presented in the second paper is actually consistent with that of the first.*

"Non-robust sign and magnitude" using different outcome variables

Table 2. Alternative measures of civil war

	Model 5: incidence 1,000+	Model 6: outbreak 1,000+	Model 7: incidence 25+	Model 8: outbreak 25+	Model 9: outbreak 100+
Temperature	–0.006	–0.005	0.015	–0.009	0.016
	(0.021)	(0.013)	(0.040)	(0.026)	(0.024)
Temperature_{t-1}	–0.025	–0.009	–0.031	–0.004	–0.018
	(0.028)	(0.015)	(0.032)	(0.026)	(0.017)
Precipitation	0.062	–0.012	0.129*	0.055	–0.014
	(0.061)	(0.052)	(0.072)	(0.068)	(0.074)
$\text{Precipitation}_{t-1}$	0.056	0.003	0.024	0.018	–0.010
	(0.062)	(0.035)	(0.069)	(0.071)	(0.060)
Intercept	0.358	0.448	–0.112	0.214	0.138
	(1.231)	(0.531)	(1.521)	(0.891)	(0.911)
Country fixed effects	Yes	Yes	Yes	Yes	Yes
Country time trends	Yes	Yes	Yes	Yes	Yes
R^2	0.76	0.09	0.65	0.13	0.10
Civil war observations	169	11	226	46	23
Observations	889	889	889	889	769

Data are OLS regression estimates with country fixed effects and country-specific linear time trends; SEs are in parentheses. Models 5–8 apply different operationalizations of civil war from the same conflict database (11); model 9 uses civil war data from an alternative source (12).
$**P < 0.05$, $*P < 0.1$.

Buhaug (PNAS, 2010)

Units must be standardized, differences must be tested

Table 2. Testing for disagreement between results when alternative conflict variables are used

	Burke et al. (1) war years 1000+ (standardized)	Buhaug (8) model 5 incidence 1000+ (standardized)	Buhaug model 6 outbreak 1000+ (standardized)	Buhaug model 7 incidence 25+ (standardized)	Buhaug model 8 outbreak 25+ (standardized)	Buhaug model 9 outbreak 100+ (standardized)
Probability of occurrence	0.110	0.190	0.012	0.254	0.052	0.030
Temperature_t	0.390	−0.030	−0.408	0.060	−0.165	0.532
	(0.197)	(0.110)	(1.046)	(0.156)	(0.504)	(0.790)
Temperature_{t-1}	0.120	−0.130	−0.755	−0.121	−0.083	−0.598
	(0.211)	(0.147)	(1.233)	(0.128)	(0.505)	(0.581)
Precipitation_t	−0.209	0.326	−1.001	0.508	1.065	−0.455
	(0.471)	(0.318)	(4.212)	(0.281)	(1.316)	(2.465)
$\text{Precipitation}_{t-1}$	0.227	0.296	0.205	0.093	0.352	−0.321
	(0.443)	(0.324)	(2.847)	(0.271)	(1.370)	(2.017)
Observations	889	889	889	889	889	769
R-squared	0.657	0.765	0.090	0.652	0.130	0.099
		Testing Whether Coefficients Differ from Burke et al. using SUR (P value)				
Temperature_t		0.0558	0.4388	0.1299	0.2638	0.8558
Temp_t, temp_{t-1}		0.1392	0.4563	0.1598	0.4276	0.3700
All four variables		0.1290	0.2843	0.1453	0.4429	0.4333

This table replicates Buhaug table 2. All regressions contain country fixed effects and country-specific trends with standard errors clustered by country, shown in parentheses. The unconditional probability of occurrence is shown and is used to standardized each conflict outcome. For regression coefficients shown, a 0.1 effect implies a 10% change relative to average risk levels. We estimate Buhaug models 5–9 simultaneously with the Burke et al. model using seemingly unrelated regression (SUR) to test a null hypothesis that coefficients from the two models are the same in bottom panel.

Hsiang & Meng (PNAS, 2014)

Table 3. Alternative climate parameters and controls

	Model 10: outbreak 25+	Model 11: outbreak 25+	Model 12: outbreak 25+	Model 13: outbreak 25+
Temperature deviation	−3.917	−12.631	−18.977	−130.35
	(10.146)	(12.144)	(12.899)	(113.69)
Temperature deviation$_{t-1}$	3.112	−6.180		
	(12.635)	(11.517)		
Precipitation deviation	−0.238	0.509		
	(0.519)	(0.578)		
Precipitation deviation$_{t-1}$	−0.792	−0.169		
	(1.674)	(0.915)		
Political exclusion$_{t-1}$	0.760*	0.820**	0.774*	0.823**
	(0.409)	(0.396)	(0.399)	(0.399)
Temperature deviation × political exclusion$_{t-1}$			11.519	
			(12.382)	
Ln GDP capita$_{t-1}$	−0.482**	−0.547**	−0.532**	−0.557**
	(0.236)	(0.263)	(0.243)	(0.265)
Temperature deviation × ln GDP capita$_{t-1}$				−15.932
				(14.559)
Post-Cold War	0.893**	1.017**	1.013**	1.066**
	(0.381)	(0.423)	(0.407)	(0.418)
Intrastate conflict$_{t-1}$	−0.726	−0.690	−0.718	−0.690
	(0.552)	(0.549)	(0.555)	(0.528)
Intercept	−0.122	0.295	0.188	0.327
	(1.768)	(1.978)	(1.794)	(1.923)
Pseudo R^2	0.05	0.05	0.05	0.05
Civil war observations	45	45	45	45
Observations	866	867	867	867

Data are logit regression estimates; robust SEs clustered on countries in parentheses. The climate parameters measure deviation from previous year's estimate (model 10) and deviation from the long-tem normal annual level (models 11–13). Ln indicates natural logarithm of values.

$^{**}P < 0.05$, $^{*}P < 0.1$.

Buhaug (PNAS, 2010)

Models must be apples to apples

Convert logic and linear probability models to a common metric: relative risk ratios

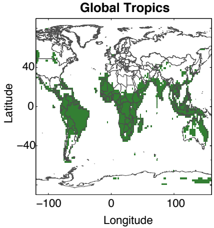
Table 3. Relative risk ratio from +1 °C

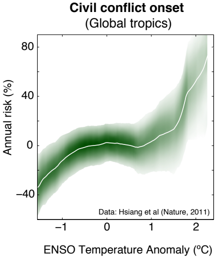
	Burke et al. (1) implied war years 1000+	Buhaug (8) model 10 outbreak 25+	Buhaug model 11 outbreak 25+	Buhaug model 12 outbreak 25+	Buhaug model 13 outbreak 25+
Upper bound effect (95% CI)		8.62×10^{6}	7.10×10^{4}	546.9	1.45×10^{40}
Average effect of temperature	1.39	0.0199	3.27×10^{-6}	5.73×10^{-9}	2.46×10^{-57}
Lower bound effect (95% CI)		4.60×10^{-11}	1.51×10^{-16}	6.01×10^{-20}	4.20×10^{-154}

This table replicates Buhaug table 3. Estimates are relative risk ratios from +1 °C. Models described in Buhaug. CI, confidence interval.

(Hsiang & Meng, PNAS, 2014)

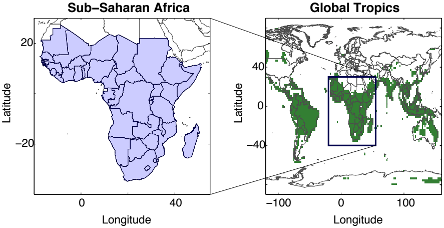
Modern climate and violence in Africa

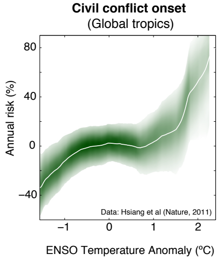




Hsiang, Burke, Miguel (Science, 2013)

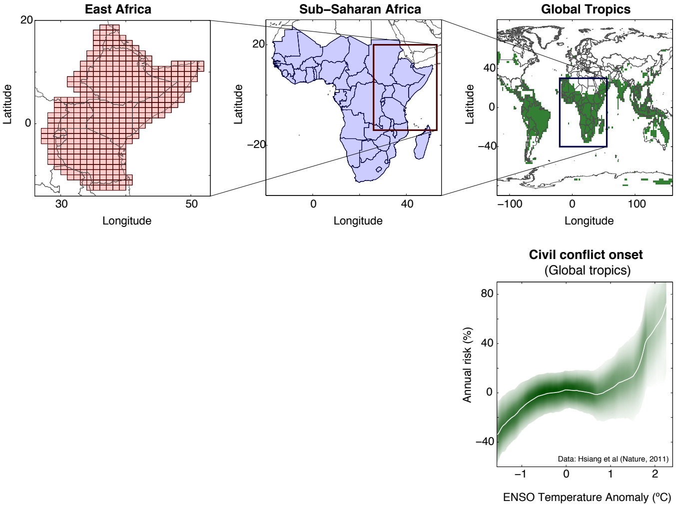
Modern climate and violence in Africa





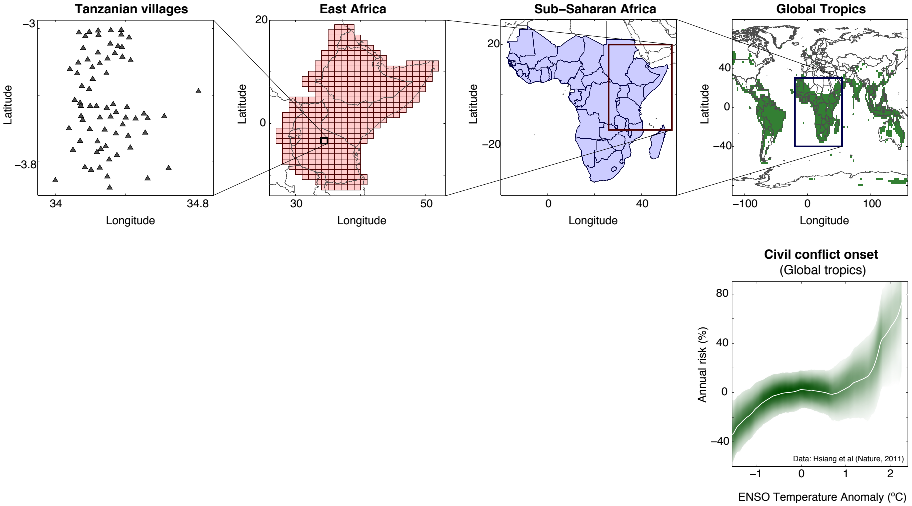
Hsiang, Burke, Miguel (Science, 2013)

Modern climate and violence in Africa



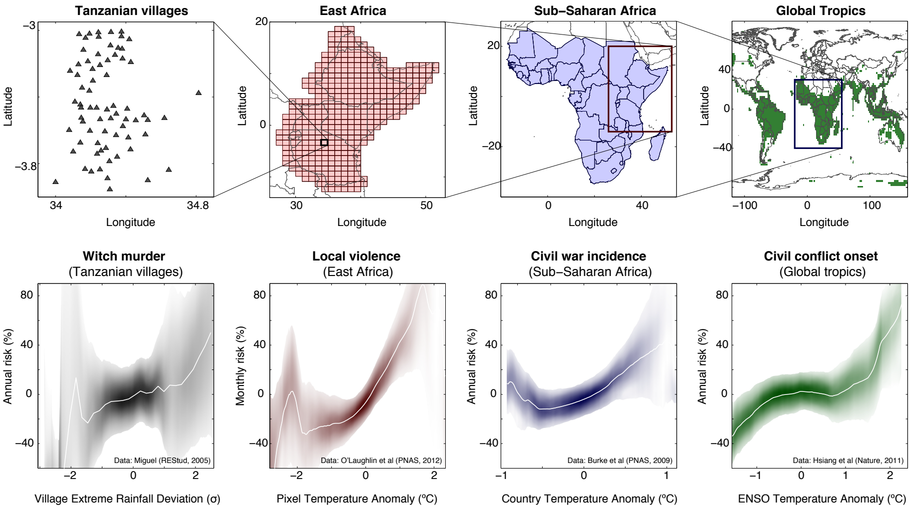
Hsiang, Burke, Miguel (Science, 2013)

Modern climate and violence in Africa



Hsiang, Burke, Miguel (Science, 2013)

Modern climate and violence in Africa



Hsiang, Burke, Miguel (Science, 2013)

When can we compare or consolidate results?

- Must have (reasonably) comparable units.
- Units of measure must be comparable (e.g. standardized to %).
- Models must be structurally similar enough for comparison (e.g. local linearization).
- Methods should not have systematic bias relative to one another.
- Must have limited publication bias.

When can we compare or consolidate results?

- Must have (reasonably) comparable units.
- Units of measure must be comparable (e.g. standardized to %).
- Models must be structurally similar enough for comparison (e.g. local linearization).
- Methods should not have systematic bias relative to one another.
- Must have limited publication bias.

Many times comparisons are a bad idea.

But sometimes they are essential (e.g. policy design) and should be done carefully and thoughtfully.

Replications with same effect and same error structure

Obs. i in multiple experiments indexed by j, with outcome variable y:

$$y_{ij} \sim N(\beta, \sigma^2)$$

where estimates are

$$\hat{\beta}_j = \frac{1}{n} \sum_i y_{ij}, \qquad \hat{\sigma}_j^2 = \frac{\sigma^2}{n_j}$$

Replications with same effect and same error structure

Obs. i in multiple experiments indexed by j, with outcome variable y:

$$y_{ij} \sim N(\beta, \sigma^2)$$

where estimates are

$$\hat{\beta}_j = \frac{1}{n} \sum_i y_{ij}, \qquad \hat{\sigma}_j^2 = \frac{\sigma^2}{n_j}$$

If experiments only differ by sample size n_j (i.e. σ^2 and β are the same for all j), then we should pool observations into one mega-experiment:

$$\tilde{\beta} = \frac{\sum_j \frac{1}{\hat{\sigma}_j^2} \hat{\beta}_j}{\sum_j \frac{1}{\hat{\sigma}_j^2}}$$

Replications with same effect and same error structure

Obs. i in multiple experiments indexed by j, with outcome variable y:

$$y_{ij} \sim N(\beta, \sigma^2)$$

where estimates are

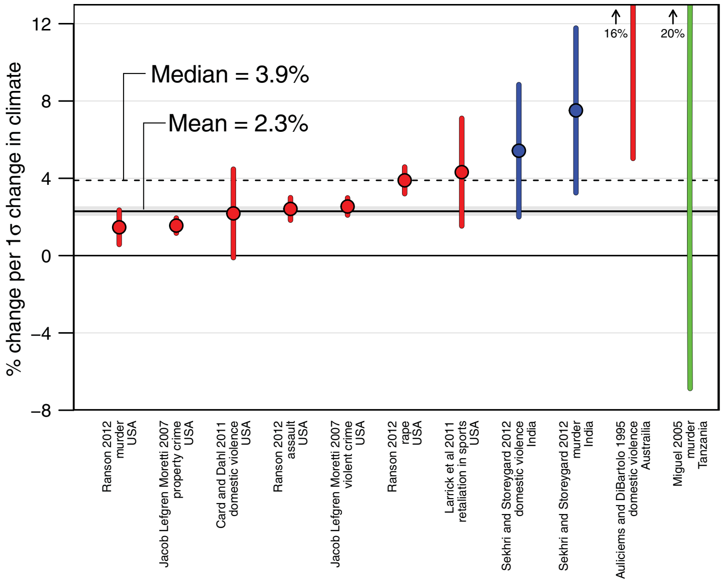
$$\hat{\beta}_j = \frac{1}{n} \sum_i y_{ij}, \qquad \hat{\sigma}_j^2 = \frac{\sigma^2}{n_j}$$

If experiments only differ by sample size n_j (i.e. σ^2 and β are the same for all j), then we should pool observations into one mega-experiment:

$$\tilde{\beta} = \frac{\sum_j \frac{1}{\hat{\sigma}_j^2} \hat{\beta}_j}{\sum_j \frac{1}{\hat{\sigma}_j^2}} = \frac{\sum_j \frac{n_j}{\sigma^2} \hat{\beta}_j}{\sum_j \frac{n_j}{\sigma^2}} = \frac{\sum_j n_j \hat{\beta}_j}{\sum_j n_j}$$

$\frac{1}{\sigma_j^2}$ is called **the precision** of $\hat{\beta}_j$.

Inter-personal conflict and climate



Replications with same effect but different error structure

Obs. i in multiple experiments indexed by j, with outcome variable y:

$$y_{ij} \sim N(\beta, \sigma_j^2)$$

We look for a weighted average of prior estimates:

$$\tilde{\beta} = \sum_j \omega_j \hat{\beta}_j$$

where ω_j is the weight for study j.

$$Var(\tilde{\beta}) = \sum_k \sum_j \left[\omega_k \omega_j Cov(\hat{\beta}_k, \hat{\beta}_j) \right]$$

If the studies are independent, then $Cov(\hat{\beta}_k, \hat{\beta}_j) = 0$ for all $k \neq j$ and

$$Var(\tilde{\beta}) = \sum_j \omega_j^2 {\hat{\sigma}_j}^2$$

Replications with same effect but different error structure

Obs. i in multiple experiments indexed by j, with outcome variable y:

$$y_{ij} \sim N(\beta, \sigma_j^2)$$

We look for a weighted average of prior estimates:

$$\tilde{\beta} = \sum_j \omega_j \hat{\beta}_j$$

where ω_j is the weight for study j.

$$Var(\tilde{\beta}) = \sum_k \sum_j \left[\omega_k \omega_j Cov(\hat{\beta}_k, \hat{\beta}_j)\right]$$

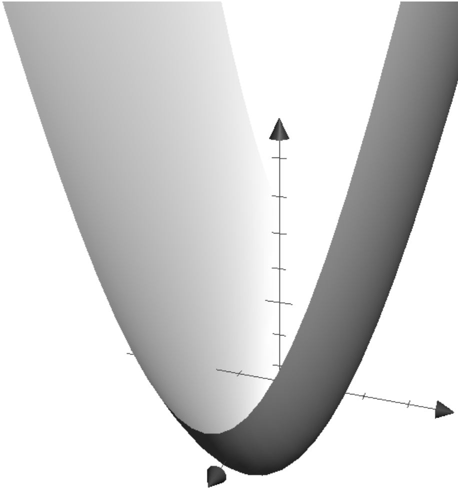
If the studies are independent, then $Cov(\hat{\beta}_k, \hat{\beta}_j) = 0$ for all $k \neq j$ and

$$Var(\tilde{\beta}) = \sum_j \omega_j^2 \hat{\sigma}_j{}^2$$

A reasonable goal: Minimize $Var(\tilde{\beta})$ subject to the constraint $\sum \omega_j = 1$.

Problem: $\sigma_1^2 = 1, \sigma_2^2 = 3$

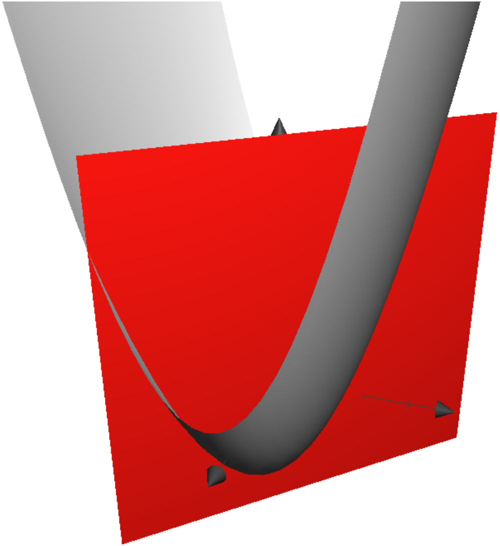
Minimize: $Var(\tilde{\beta}) = \omega_1^2 \hat{\sigma_1}^2 + \omega_2^2 \hat{\sigma_2}^2 = \omega_1^2 + 3\omega_2^2$
Subject to: $\omega_1 + \omega_2 = 1.$



Problem: $\sigma_1^2 = 1, \sigma_2^2 = 3$

Minimize: $Var(\tilde{\beta}) = \omega_1^2 \hat{\sigma_1}^2 + \omega_2^2 \hat{\sigma_2}^2 = \omega_1^2 + 3\omega_2^2$

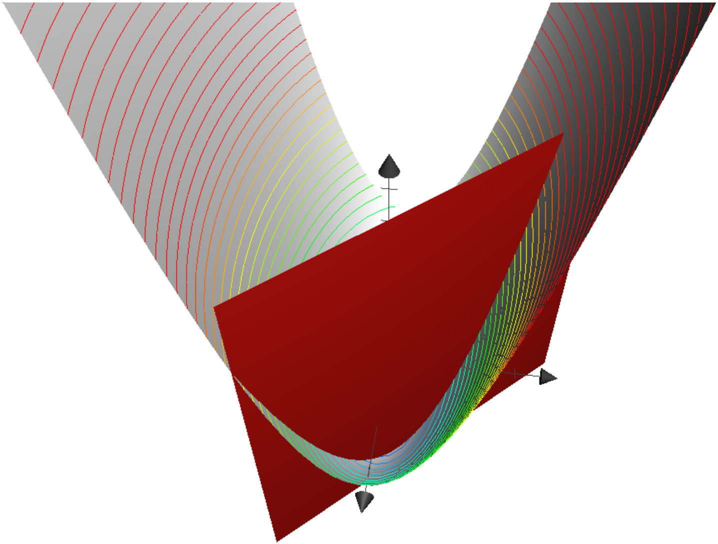
Subject to: $\omega_1 + \omega_2 = 1.$



Problem: $\sigma_1^2 = 1, \sigma_2^2 = 3$

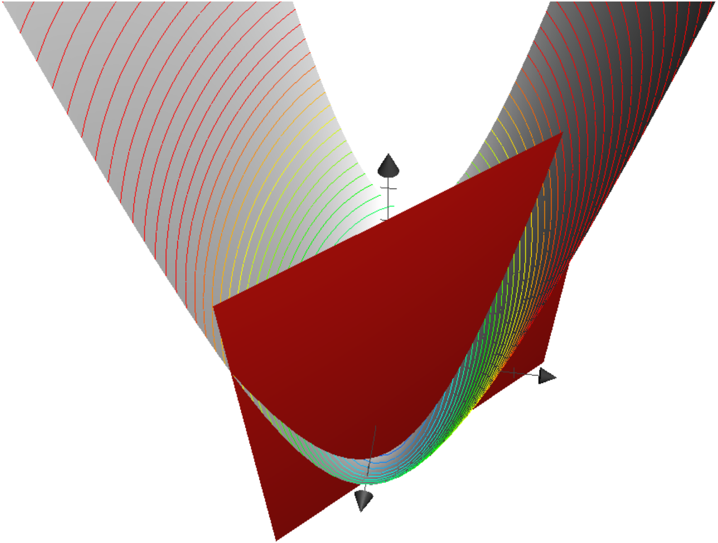
Minimize: $Var(\tilde{\beta}) = \omega_1^2 \hat{\sigma_1}^2 + \omega_2^2 \hat{\sigma_2}^2 = \omega_1^2 + 3\omega_2^2$

Subject to: $\omega_1 + \omega_2 = 1.$



Solution: Precision weights!

$$\omega_j = \frac{\frac{1}{\hat{\sigma}_j^2}}{\sum_j \frac{1}{\hat{\sigma}_j^2}}$$



Precision weights are a simple and general solution

The combined estimate

$$\tilde{\beta} = \sum_j \omega_j \hat{\beta}_j, \qquad \omega_j = \frac{\frac{1}{\hat{\sigma}_j^2}}{\sum_j \frac{1}{\hat{\sigma}_j^2}}$$

is optimal if effects are the same across studies, regardless of whether or not error structure is the same across studies.

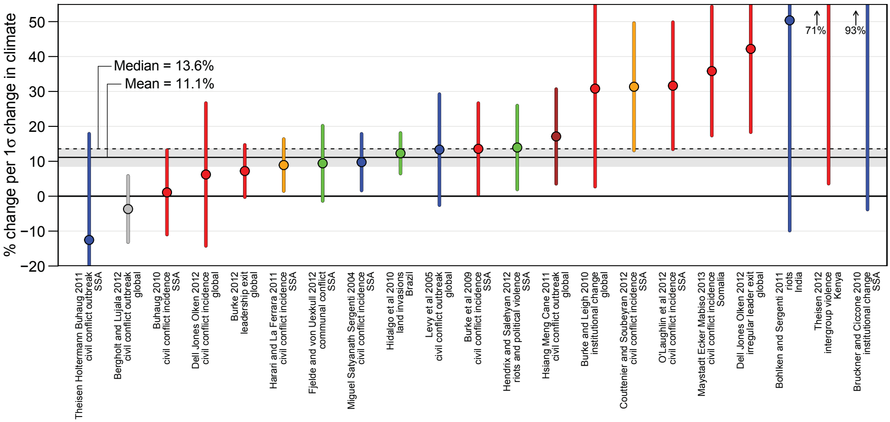
When do error structures change across studies?

- More orthogonal controls reduce residual variance

- Populations are subject to different disturbances

- Observational units are aggregated differently across samples

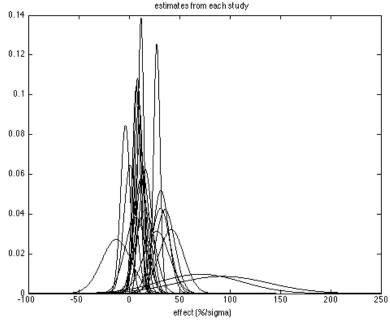
Inter-group conflict and climate

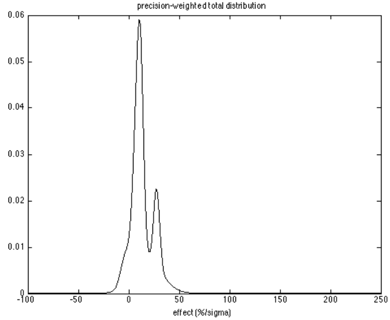


Going beyond the mean

A natural extension is to combine the full probability distribution for effects (rather than just the mean):

$$\tilde{B}_\beta = \sum_j \omega_j N_\beta(\hat{\beta}_j, \hat{\sigma}_j)$$





Summarizing results for climate and conflict

Table: **Summary statistics for the distribution of effects across studies**

	Median	$\tilde{\beta}$	$\sigma(\tilde{\beta})$	Percentiles of $\tilde{B}_\beta$				
				5%	25%	50%	75%	95%
Intergroup	13.56	11.12	1.34	-4.60	5.80	10.20	14.30	32.00
Interpersonal	3.89	2.29	0.12	1.20	1.50	2.20	2.60	4.00

Hsiang, Burke, Miguel (Science, 2013)

Sometimes cross-study differences seem inconsistent with previously estimated within-study sampling variability

School	Estimated treatment effect, y_j	Standard error of effect estimate, σ_j
A	28	15
B	8	10
C	−3	16
D	7	11
E	−1	9
F	1	11
G	18	10
H	12	18

Table 5.2 *Observed effects of special preparation on SAT-V scores in eight randomized experiments. Estimates are based on separate analyses for the eight experiments. From Rubin (1981).*

Precision-weighted $\tilde{\beta} = 7.9$ (± 4.2)

Hierarchical (random effects) model of research findings

Observations i in experiment j

$$y_{ij} \sim N(\beta_j, \sigma^2)$$

Which let's us estimate $\hat{\beta}_j$ for each study. The true β_j's for the studies **are not the same**, but have a distribution:

$$\beta_j \sim N(\mu, \tau^2)$$

μ and τ are called **hyperparameters**, they have an unknown (possibly non-normal) distribution.

Interpretation

Studies really do differ in substantive ways unrelated to sampling variability in y_{ij}, however some component of their results is common across studies (μ).

τ describes the extent to which studies describe fundamentally different results.

Bayesian solution

The conditional posterior

$$\beta_j | \mu, \tau, y \sim N(\breve{\beta}_j, V_j)$$

where

$$\breve{\beta}_j = \frac{\frac{1}{\hat{\sigma}_j^2}\hat{\beta}_j + \frac{1}{\tau^2}\mu}{\frac{1}{\hat{\sigma}_j^2} + \frac{1}{\tau^2}}, \qquad V_j = \frac{1}{\frac{1}{\hat{\sigma}_j^2} + \frac{1}{\tau^2}}$$

Bayesian solution

The conditional posterior

$$\beta_j | \mu, \tau, y \sim N(\breve{\beta}_j, V_j)$$

where

$$\breve{\beta}_j = \frac{\frac{1}{\hat{\sigma}_j^2}\hat{\beta}_j + \frac{1}{\tau^2}\mu}{\frac{1}{\hat{\sigma}_j^2} + \frac{1}{\tau^2}}, \qquad V_j = \frac{1}{\frac{1}{\hat{\sigma}_j^2} + \frac{1}{\tau^2}}$$

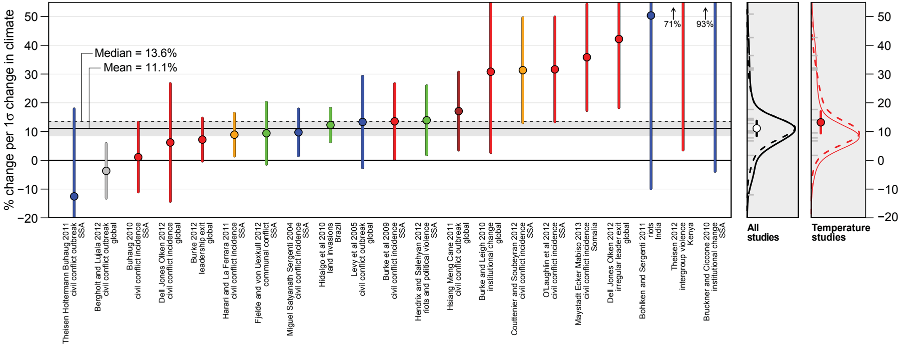
Common component of studies is μ

$$\mu | \tau, y \sim N(\hat{\mu}, V_\mu)$$

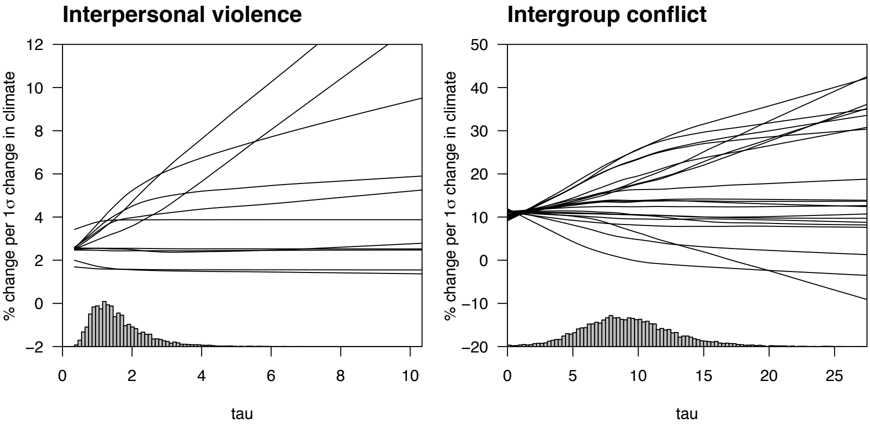
where

$$\hat{\mu} = \frac{\sum_j \frac{1}{\frac{1}{\hat{\sigma}_j^2} + \frac{1}{\tau^2}} \hat{\beta}_j}{\sum_j \frac{1}{\frac{1}{\hat{\sigma}_j^2} + \frac{1}{\tau^2}}}, \qquad V_\mu^{-1} = \sum_j \frac{1}{\frac{1}{\hat{\sigma}_j^2} + \frac{1}{\tau^2}}$$

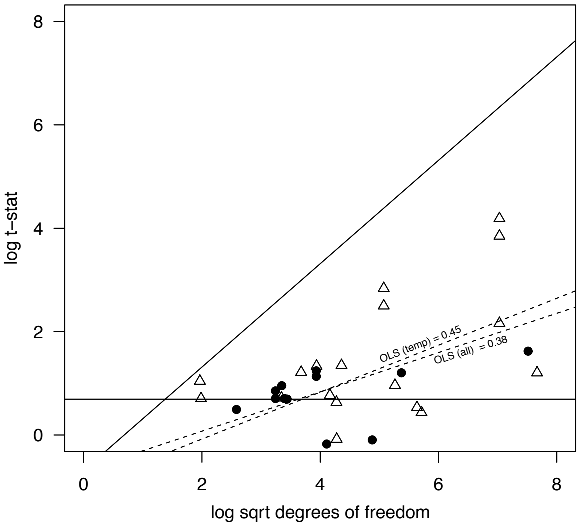
Meta-analysis: inter-group conflict



Predicting true study-specific effects β_j conditional on hyperparameter τ

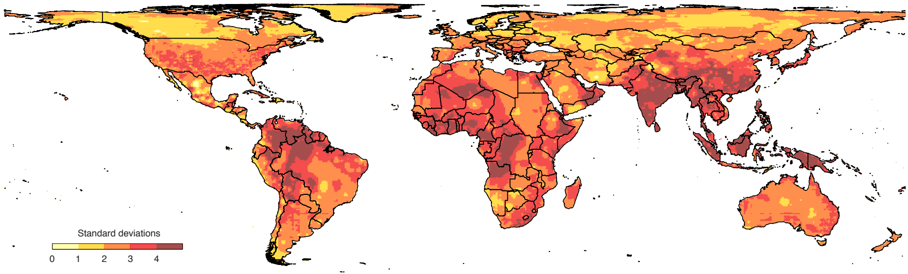


Publication bias is always a major issue – check with tests like p-curves



Pinning down numbers informs policy!

Standardized temperature change by 2050:
Most inhabited areas warm 2-4σ



Median temperature effects:

$+3.9\%/\sigma$ for interpersonal conflict

$+13.6\%/\sigma$ for intergroup conflict

Can we consolidate and unify all quantitative human knowledge in real time?

Solution: Crowd-sourcing empirical results from the researchers that produce them (think Wikipedia for empirical findings).

"Distributed Meta-Analysis System"
Rising & Hsiang (2014)

`dmas.berkeley.edu`