Quantities used to generate Figure 2

Let q denote the probability that a proposition is true, β the probability that a researcher is prone to fishing, and γ the probability that a non-fishing researcher adopts modified modules (for example due to learning). Let p_T^o and p_F^o denote the probability that a positive result is obtained under an "orthodox" plan (set prior to data analysis) given that the statement is true or false respectively and let p_T^f and p_F^f and p_T^m and p_F^m denote the corresponding probabilities that a fisher will find positive results (given that positive results were not obtained under the orthodox model), and the probability that a nonfisher would receive positive results (conditional on adoption of a modified model). The "compliers" are those that do not submit results that deviate from the orthodox model, this group contains both unsuccessful fishers and nonfishers who do not implement modified models. To focus on the sorting function of partial registration, assume that the incentives of researchers and journals are not affected by the form of compliance required. In this case, the share of positive results, positive results that are false, and negative results that are false are given in the Tables below.

Share of positives that are false	$\frac{(1-q)\Big((\beta+(1-\beta)(1-\gamma))p_F^o+\beta(1-p_F^o)p_F^f+(1-\beta)\gamma p_F^m\Big)}{(\beta+(1-\beta)(1-\gamma))(qp_T^o+(1-q)p_F^o)+\beta(q(1-p_T^o)p_T^f+(1-q)(1-p_F^o)p_F^f)+(1-\beta)\gamma(qp_T^m+(1-q)p_F^m)}$
Share of negatives that are false	$\frac{q\Big(\beta(1-p_T^o)(1-p_T^f)+(1-\beta)(1-\gamma)(1-p_T^o)+(1-\beta)\gamma(1-p_T^m)\Big)}{\beta[q(1-p_T^f)(1-p_T^o)+(1-q)(1-p_F^f)(1-p_F^o)]+(1-\beta)(1-\gamma)[q(1-p_T^o)+(1-q)(1-p_F^o)]+(1-\beta)\gamma[q(1-p_T^m)+(1-q)(1-p_F^m)]}$
Share of findings that are positive	$(\beta+(1-\beta)(1-\gamma))(qp_T^o+(1-q)p_F^o)+(1-\beta)\gamma(qp_T^m+(1-q)p_F^m)+\beta(q(1-p_T^o)p_T^f+(1-q)(1-p_F^o)p_F^f)$

Table 1: Outcomes from *No Registration*. These quantities are illustrated in the first column of Figure 2.

Share of positives that are false	$\frac{(1-q)p_F^o}{qp_T^o+(1-q)p_F^o}$
Share of negatives that are false	$\frac{q(1-p_T^o)}{q(1-p_T^o)+(1-q)(1-p_F^o)}$
Share of findings that are positive	$qp_T^o + (1-q)p_F^o$

Table 2: Outcomes from *Binding* registration. These quantities are illustrated in the second column of Figure 2.

Share of positives that are false:

Compliers	$\frac{(1-q)p_F^o}{qp_T^o+(1-q)p_F^o}$
Noncompliers	$\frac{(1-q)(\beta(1-p_F^o)p_F^f+(1-\beta)\gamma p_F^m)}{q(\beta(1-p_T^o)p_T^f+\gamma(1-\beta)p_T^m)+(1-q)(\beta(1-p_F^o)p_F^f+\gamma(1-\beta)p_F^m)}$

Share of negatives that are false:

Compliers	$\frac{q(1-p_T^o)((1-p_T^f)\beta+(1-\beta)(1-\gamma))}{q(1-p_T^o)((1-p_T^f)\beta+(1-\beta)(1-\gamma))+(1-q)(1-p_F^o)((1-p_F^f)\beta+(1-\beta)(1-\gamma))}$
Noncompliers	$\frac{q(1-p_T^m)}{q(1-p_T^m)+(1-q)(1-p_F^m)}$

Share of findings that are positive:

Compliers	$\frac{((1-q)p_F^o+qp_T^o)(\beta+(1-\beta)(1-\gamma))}{(1-\beta)(1-\gamma)+\beta[q(p_T^o+(1-p_T^o)(1-p_T^f))+(1-q)(p_F^o+(1-p_F^o)(1-p_F^f))]}$
Noncompliers	$\frac{\beta(q(1-p_T^o)p_T^f+(1-q)(1-p_F^o)p_F^f)+(1-\beta)\gamma(qp_T^m+(1-q)p_F^m)}{\beta(q(1-p_T^o)p_T^f+(1-q)(1-p_F^o)p_F^f)+(1-\beta)\gamma}$

Table 3: Outcomes from *Nonbinding* registration, by group. These quantities are illustrated in the final two columns of Figure 2.

Proofs of claims for the model discussed in Section 5.3

Assumptions

Let the researcher's objective function be given by:

$$U(n|\overline{u},\underline{u},\overline{\beta}) = \underline{u} + (\overline{u}-\underline{u})\int_{\overline{\beta}-a}^{\overline{\beta}+a}\pi(\beta,n)f(\beta)d\beta - c(n) \tag{1}$$

where c is a strictly convex cost function, π is the power of the test, and f denotes prior beliefs about β which we assume to be a uniform distribution over $[\overline{\beta}-a,\overline{\beta}+a]$ for $0<\overline{\beta}<a$. We examine a simple case where treatment and control groups are of equal size and potential outcomes in both groups are distributed normally with unit variance; in this case the power function is given by:

$$\pi(\beta,n) = 1-\Phi\left(t-\beta\sqrt{\frac{n}{4}}\right) \tag{2}$$

where $\Phi()$ is the standard normal cdf and t is the t statistic employed for this test. Note that we assume that one sided tests are employed and that the 95% confidence level is adopted, implying that t takes a value of approximately 1.65.

Note that:

$$\frac{\partial\pi}{\partial n} = \phi\left(t-\beta\sqrt{\frac{n}{4}}\right)\frac{\beta}{4\sqrt{n}} \tag{3}$$

And so power is increasing in n if β is positive and decreasing if β is negative.

However a useful result is that for positive β:

$$\frac{\partial\pi(\beta,n)}{\partial n}+\frac{\partial\pi(-\beta,n)}{\partial n} = \left(\phi\left(t-\beta\sqrt{\frac{n}{4}}\right)-\phi\left(t+\beta\sqrt{\frac{n}{4}}\right)\right)\frac{\beta}{4\sqrt{n}} > 0 \tag{4}$$

The inequality follows from the fact that $t-\beta\sqrt{\frac{n}{4}}$ is closer to 0 than $t+\beta\sqrt{\frac{n}{4}}$. It follows from our assumptions on the distribution of β that the expected marginal effect of n on power is positive.

$$\int_{\overline{\beta}-a}^{\overline{\beta}+a}\frac{\partial\pi(\beta,n)}{\partial n}f(\beta)d\beta > 0 \tag{5}$$

Concavity of the objective function

We first establish that this objective function is strictly concave. Taking the second derivative of the objective function we have:

$$\frac{\partial^2 U}{\partial n^2} = (\overline{u} - \underline{u}) \int_{\overline{\beta}-a}^{\overline{\beta}+a} \frac{\partial^2 \pi(\beta, n)}{\partial n^2} f(\beta) d\beta - c''(n) \tag{6}$$

From convexity $c''(n) > 0$ and hence we focus on $\int_{\overline{\beta}-a}^{\overline{\beta}+a} \frac{\partial^2 \pi(\beta,n)}{\partial n^2} f(\beta) d\beta$.

Claim A: $\int_{\overline{\beta}-a}^{\overline{\beta}+a} \frac{\partial^2 \pi(\beta,n)}{\partial n^2} f(\beta) d\beta \leq 0$

Claim A is difficult to establish because under our assumptions on t the power function is concave for $\beta > 0$ and convex for $\beta < 0$. Our strategy is instead to show that for any $\beta > 0$ $\frac{\partial^2 \pi(\beta,n)}{\partial n^2} + \frac{\partial^2 \pi(-\beta,n)}{\partial n^2} \leq 0$.[1]

Proof. The second derivative of the power function is given by:

$$\pi'' = -\frac{\beta}{(4n)^{\frac{3}{2}}} \left[\beta^2 \frac{n}{4} - t\beta \frac{\sqrt{n}}{2} + 1\right] \phi\left(t - \beta\sqrt{\frac{n}{4}}\right) \tag{7}$$

And so we seek to establish for $\beta > 0$:

$$\begin{aligned}
\pi''(\beta) + \pi''(-\beta) &\leq 0 \\
&\leftrightarrow \\
-\frac{\beta}{(4n)^{\frac{3}{2}}} \left[\beta^2 \frac{n}{4} - t\beta \frac{\sqrt{n}}{2} + 1\right] \phi\left(t - \beta\sqrt{\frac{n}{4}}\right) + \frac{\beta}{(4n)^{\frac{3}{2}}} \left[\beta^2 \frac{n}{4} + t\beta \frac{\sqrt{n}}{2} + 1\right] \phi\left(t + \beta\sqrt{\frac{n}{4}}\right) &\leq 0 \\
&\leftrightarrow \\
-\left[\beta^2 \frac{n}{4} - t\beta \frac{\sqrt{n}}{2} + 1\right] e^{\beta t\sqrt{n}} + \left[\beta^2 \frac{n}{4} + t\beta \frac{\sqrt{n}}{2} + 1\right] &\leq 0
\end{aligned}$$

Defining $\kappa = \beta\sqrt{n}$, we then seek to show:

$$G \equiv -e^{t\kappa}\left(1 + \frac{\kappa^2}{4} - \frac{\kappa t}{2}\right) + \left(1 + \frac{\kappa^2}{4} + \frac{\kappa t}{2}\right) \leq 0 \tag{8}$$

We now show that for $t \leq \sqrt{3}$ (which includes the range of interest for this analysis) G is decreasing in κ and takes on its maximum value at $\kappa = 0$; at this maximum value the condition holds with equality.

[1] Note that we prove this for our assumption on t but note that the result does not necessarily obtain for values outside of this range. As an illustration set $\beta = .5, n = 4, t = 2$. Then $\pi''(\beta) = -0.00025$ and $\pi''(-\beta) = 0.00031$.

We have:

$$\begin{aligned}\frac{\partial G}{\partial \kappa} &= \frac{\kappa + t}{2} - e^{\kappa t}\left(\frac{\kappa - t}{2}\right) - te^{\kappa t}\left(1 + \frac{\kappa^2}{4} - \frac{\kappa t}{2}\right) \\ &= \frac{1}{2}\left(\kappa + t - e^{\kappa t}\left(\frac{t\kappa^2}{2} - \kappa t^2 + \kappa + t\right)\right)\end{aligned}$$

Note that $\frac{\partial G}{\partial \kappa} = 0$ for $\kappa = 0$.

We now show that $\frac{\partial G}{\partial \kappa} \leq 0$ for all κ by showing that $\frac{\partial^2 G}{\partial \kappa^2} \leq 0$

We have:

$$\begin{aligned}\frac{\partial^2 G}{\partial \kappa^2} &= \frac{1}{2}\left(1 - e^{\kappa t}\left(1 + \frac{(\kappa t)^2}{2} + \kappa t(2 - t^2)\right)\right) \\ &= \frac{1}{2}\left(1 - e^{\kappa t}\left(1 + \psi \kappa t^* + \frac{(\kappa t)^2}{2}\right)\right)\end{aligned}$$

where $\psi = 2 - t^2$ takes a value in $(-1, 2)$ for $t \in [0, \sqrt{3}]$.

Now from the expansion of e^x we have:

$$e^{\kappa t} \geq 1 + \kappa t + \frac{(\kappa t)^2}{2} \tag{9}$$

Note that $\left(1 + \psi \kappa t^* + \frac{(\kappa t)^2}{2}\right)$ is positive (to check this note that the smallest value this expression can take in this range is for $t = \sqrt{3}$ and at this value the minimum is achieved when $\kappa = \frac{1}{\sqrt{3}}$ which yields a value of $\frac{1}{2}$ for the expression). We then have:

$$e^{\kappa t}(1 + \psi \kappa t + \frac{(\kappa t)^2}{2}) \geq (1 + \kappa t + \frac{(\kappa t)^2}{2})(1 + \psi \kappa t + \frac{(\kappa t)^2}{2}) = 1 + (1 + \psi)(\kappa t^* + (\kappa t)^2 + \frac{(\kappa t)^3}{2}) + \frac{(\kappa t)^4}{4} \geq 1 \tag{10}$$

And so:

$$\frac{\partial^2 G}{\partial \kappa^2} \leq 0 \tag{11}$$

Thus the highest value G can take is for $\kappa = 0$ in which case $G = 0$ and we are done. ■

Comparative statics

The first order condition defines the optimum implicitly:

$$\frac{\partial U(n^*|\overline{u},\underline{u},\overline{\beta})}{\partial n} = (\overline{u}-\underline{u})\int_{\overline{\beta}-a}^{\overline{\beta}+a}\frac{\partial \pi(\beta,n^*)}{\partial n}f(\beta)d\beta - c'(n^*) = 0 \tag{12}$$

For comparative statics, we use the implicit function theorem around n^*. Define $F(n,\overline{u},\underline{u},\overline{\beta}) = (\overline{u}-\underline{u})\int_{\overline{\beta}-a}^{\overline{\beta}+a}\frac{\partial \pi(\beta,n)}{\partial n}f(\beta)d\beta - c'(n)$. We have:

$$\frac{\partial n^*}{\partial \underline{u}} = -\frac{\frac{\partial F(n^*,\overline{u},\underline{u},\overline{\beta})}{\partial \underline{u}}}{\frac{\partial F(n^*,\overline{u},\underline{u},\overline{\beta})}{\partial n}} \tag{13}$$

Both the denominator and numerator are negative. For the numerator recall that we have from Equation (3) that $\int_{\overline{\beta}-a}^{\overline{\beta}+a}\frac{\partial \pi(\beta,n)}{\partial n}f(\beta)d\beta > 0$. The denominator is negative from the strict concavity of the objective function, as established above. This yields: $\frac{\partial n^*}{\partial \underline{u}} < 0$. Similar reasoning follows for $\frac{\partial n^*}{\partial \overline{\beta}} < 0$.

Comparative statics of the mean prior threshold

Assume now that data gathering costs are such that in some cases the costs of implementing the optimal design are so great that a researcher would do better forgoing a study entirely. In our model the expected value of any study is an increasing function of $\overline{\beta}$ and so it is helpful to focus on problems for which prior distributions with expectation greater than some value $\overline{\beta}^*$ yield positive expected value while distributions with expectation below $\overline{\beta}^*$ yield negative expected value.

In particular, let $V(\overline{u},\underline{u},a,\overline{\beta})$ denote the value of U when the optimal value of n is chosen and consider ranges in which $\overline{\beta}^*(\overline{u},\underline{u},a)$ is defined implicitly by the condition:

$$V(\overline{u},\underline{u},a,\overline{\beta}^*) = 0 \tag{14}$$

A result from the Envelope theorem is:

$$\frac{\partial V(\overline{u},\underline{u},\overline{\beta})}{\partial \overline{\beta}} = \frac{\partial U(n^*,\overline{u},\underline{u},\overline{\beta})}{\partial \overline{\beta}} + \frac{\partial U(n^*,\overline{u},\underline{u},\overline{\beta})}{\partial n}\times\frac{\partial n^*}{\partial \overline{\beta}} = \frac{\partial U(n^*,\overline{u},\underline{u},\overline{\beta})}{\partial \overline{\beta}} \tag{15}$$

Applying differentiation under the integral sign, note that:

$$\frac{\partial U(n^*,\overline{u},\underline{u},\overline{\beta})}{\partial \overline{\beta}} = (\overline{u}-\underline{u})(\pi(\overline{\beta}+a,n)-\pi(\overline{\beta}-a,n)) + \int_{\overline{\beta}-a}^{\overline{\beta}+a}\frac{\partial \pi(\beta,n)}{\partial \overline{\beta}}f(\beta)d\beta \tag{16}$$

The first quantity is positive since power is increasing in effect sizes. The second quantity is 0 since π does not depend on $\overline{\beta}$ except through the limits of integration.

We then have:

$$\frac{\partial \overline{\beta}^*}{\partial \underline{u}} = -\frac{\frac{\partial V}{\partial \underline{u}}}{\frac{\partial V}{\partial \overline{\beta}}} \tag{17}$$

From the envelope theorem the numerator is clearly positive and the denominator has already been shown to be positive. This establishes that: $\frac{\partial \overline{\beta}^*}{\partial \underline{u}} < 0$.

Humphreys, Sanchez de la Sierra, and Van der Windt, 2013.

Fishing. Online Appendix C.

Table C1: Overview of the fourteen WHO Primary Registries and ClinicalTrials.

					Are the following specifically asked for by the registry?								Search for these registries in publications:							
	Registry	Website	Trials	Year	Hypotheses	Outcome	Dichotomize	Intervention	Covariates	Model	Het. Treats	Ethics	Search term	JSTOR	APSR	AJPS	PA	BMJ	The Lancet	JAMA
1	Australian New Zealand Clinical Trials Registry (ANZCTR)	anzctr.org.au	6,198	2005 (2007)	Yes	Yes	No	Yes	No	No	No	Yes	"ANZCTR" ("Australian New Zealand Clinical Trial Registry")	0/0 (0/0)	0 (0)	0 (0)	0 (0)	10 (1)	0 (1)	0 (10)
2	ClinicalTrials (NTC)	clinicaltrials.gov	122,567	2000 (2007)	Yes	Yes	No	Yes	No	No	No	Yes	"clinicaltrials"	0/153	0	0	0	243	446	448
3	Controlled-trials (ISRCTN)	isrctn.org	10,465	1998 (2007)	Yes	Yes	No	Yes	No	No	No	Yes	"ISRCTN"	0/30	0	0	0	189	48	42
4	Japan Primary Registries Network (JPRN)	umin.ac.jp/ctr	8,328	2007 (2008)	No	Yes	No	Yes	No	No	No	Yes	"JPRN"	0/18	0	0	0	0	0	0
5	The Netherlands National Trial Register (NTR)	trialregister.nl	3,187	2004 (2008)	Yes	Yes	No	Yes	No	No	No	No	"Netherlands National Trial Register"	0/0	0	0	0	0	0	0
6	Brazilian Clinical Trials Registry (ReBec)	ensaiosclinicos.gov.br/	131	2010 (2011)	No	Yes	No	Yes	No	No	No	No	"Brazilian Clinical Trials Registry"	0/0	0	0	0	0	0	0
7	Chinese Clinical Trial Registry (ChiCTR)	chictr.org/	1,983	2006 (2007)	No	Yes	No	Yes	No	No	No	Yes	"ChiCTR"	0/0	0	0	0	1	1	0
8	Clinical Research Information Service (CRiS)	cris.cdc.go.kr	378	1998 (2010)	Yes	Yes	No	Yes	No	No	No	Yes	"Clinical Research Information Service "	0/0	0	0	0	0	0	0
9	Clinical Trials Registry - India (CTRI)	ctri.nic.in	2,496	2007 (2007)	No	Yes	No	Yes	No	No	No	Yes	"Clinical Trials Registry - India"	0/1	0	0	0	1	1	0
10	Cuban Public Registry of Clinical Trials (RPCEC)	registroclinico.sld.cu/	392	2007 (2011)	No	Yes	No	Yes	No	No	No	Yes	"RPCEC"	0/0	0	0	0	0	0	0
11	EU Clinical Trials Register (EU-CTR)	clinicaltrialsregister.eu/	17,134	2004 (2011)	Yes	Yes	No	Yes	No	No	No	Yes	"EU-CTR" (EudraCT)	0/0 (0/6)	0 (0)	0 (0)	0 (0)	1 (19)	0 (13)	0 (1)
12	German Clinical Trials Register (DRKS)	germanctr.de/	829	2007 (2008)	No	Yes	No	Yes	No	No	No	Yes	"German Clinical Trials Register"	0/0	0	0	0	0	1	0
13	Iranian Registry of Clinical Trials (IRCT)	irct.ir/	2,451	2008 (2008)	No	Yes	No	Yes	No	No	No	Yes	"Iranian Registry of Clinical Trials"	0/0	0	0	0	0	0	0
14	Pan African Clinical Trial Registry (PACTR)	pactr.org/	97	2009 (2009)	No	Yes	No	Yes	No	No	No	No	"Pan African Clinical Trial Registry"	0/0	0	0	0	0	1	0
15	Sri Lanka Clinical Trials Registry (SLCTR)	slctr.lk/	71	2006 (2008)	No	Yes	No	Yes	No	No	No	Yes	"Sri Lanka Clinical Trials Registry"	0/0	0	0	0	0	0	0
													"trial registration"	1/225	0	0	0	386	50	381
													"ICTRP"	0/9	0	0	0	16	7	3

Requirements of major registries. "Trials" refers to the number of trials registered on 18 March 2012. "Year" refers to the start of the registry and when it was accepted as a WHO primary registry (in brackets). "Hypotheses", "Outcome", "Dichotomize", "Intervention", "Covariates", "Model", "Het. Treats" and "Ethics" refer to whether the registry specifically requests information about: hypotheses to be tested; the dependent variable of interest; how dependent variables are dichotomized (if relevant); covariates to be used in the analysis; statistical models to be used; sub-group analyses; whether the study received IRB approval. The JSTOR column shows the share of all JSTOR article search results mentioning these registries that are from political science journals. Results in the remaining columns are the raw number of hits for these terms. Data was searched on 18 March 2012. ICTRP is the WHO's International Clinical Trials Registry Platform.