



Department of Epidemiology
Johns Hopkins Bloomberg School of Public Health
615 N. Wolfe Street
Baltimore, Maryland 21205

1 December 2016

Memorandum

To: Trialists
Fr: Curtis Meinert
Re: Draining the swamp?

When you are up to your ass in alligators it's hard to remember that you started out to drain the swamp.

That is the situation I and some of my colleagues are in trying to write a paper entitled "The State and Nature of NIH-funded Trials", based on information in ClinicalTrials.gov. The swamp is a dataset where the most important variable for what we are trying to do is missing! There is no variable in which registrants are asked to indicate funding source. Funder type in CT.gov is derived from what respondents report as "sponsor" and "collaborator"; defined, respectively, in CT.gov as:

The organization or person who oversees the clinical study and is responsible for analyzing the study data.

An organization other than the sponsor that provides support for a clinical study; may include funding, design, implementation, data analysis, or reporting.

If either field includes the NIH, the registration is coded as having NIH as a funder type otherwise it is not, even if the trial is, in fact, NIH-funded.

How many trials are we missing? Is the "miss" rate different over time and different by NIH agency? Are the characteristics of missed trials the same as those identified as NIH-funded?

One can advance ad hominem arguments that there are no reasons to believe registration behavior has changed over time, but that argument is at odds with what a colleague at the NIH tells me:

Either the FDA regulation changed, or NIH belatedly realized, that the SPONSOR of trials registered in ClinicalTrials.gov had liability for complying with the FDA registration requirements - that's when we were told to take ourselves out of the loop completely with regards to what was submitted to ClinicalTrials.gov and, if asked, to steer clear of being labeled sponsor.

But, pigs is pigs and data is data and ad hominem arguments are neither so one can try estimating the "miss" rate. One approach is to compare how publication of randomized controlled trials, indexed as having an NCT registration number and also indexed as having been NIH-funded are coded in CT.gov in regard to funding, but that comparison involves as many assumptions as made by pollsters in the recent election – and we know how that turned out.

But, nonetheless, for what they are worth, counts by year of publication of the number indexed as RCTs, having an NCT registration number, and indexed as involving NIH funding compared to how they are registered are in Table 1 below.

Indications are that less than two-thirds of published trials so indexed are registered as having been funded by NIH but publications are an obvious biased sample of all trials since only a fraction are published and only about 20% of those are published with registration numbers.

A better estimate of the “miss” rate can be obtained from counts of trials in NIH RePORTER but not registered in CT.gov as NIH-funded. NIH RePORTER is a dataset described as follows on its website:

The information found in RePORTER is drawn from several extant databases—eRA databases, Medline, PubMed Central, the NIH Intramural Database, and iEdison—using newly-formed linkages among these disparate data sources. RePORTER includes information on research projects funded by the NIH as well as the Centers for Disease Control and Prevention (CDC), Agency for Healthcare Research and Quality (AHRQ), Health Resources and Services Administration (HRSA), Administration for Children and Family (ACF), and U.S. Department of Veterans Affairs (VA). RePORTER also includes links to publications and patents citing support from these projects. The comprehensiveness of these databases varies, as does the quality of the linkages formed among them. We expect that the quality of RePORTER data will improve over time as a result of changes in both data collection (e.g., implementation of the NIH Public Access policy) and the increased ability to identify missing information that comes from making these data accessible to more people.

Table 2 gives counts of active trials listed in NIH RePORTER for fiscal years 2005, 2010, and 2015 and the percentage of those trials not also listed in CT.gov as NIH-funded.

Keep in mind that the counts are not independent. For example, a trial active in FY 2005 could be active in FY 2010 and FY 2015 and, hence, counted in all three years.

The NIH institutes and agencies represented in Table 2 are institutes and agencies identified via CT.gov as having ≥ 100 trials registered over the time interval 2005 - 2014.

The “miss” percentages in Table 2 range from 1.3% to 60.2%. The last three lines in the Table give summaries for the years represented. The “miss” percentage for 2015 (9.8%) is twice that for 2005 (4.6%).

The median “miss” percentage are 7.3%, 13.5%, and 16.6% for years 2005, 2010, and 2015, respectively.

The obvious fix to reduce the “miss” rates is to add a variable to allow registrants to indicate funding sources rather than database operators deriving funding from the concatenation of two noisy variables subject to varying interpretation, depending on persons doing the registering.

Table 1: Fraction of publications indexed in PubMed to the publication type "randomized controlled trial" [RCT], indexed as having an NCT no. as a secondary Id, and indexed as involving NIH funding by calendar year

	A No. pub indexed [RCT] & NCT SI	B No. minus multiple pub	C No. in CT.gov w NIH as funder type	D $100 * C / B$
2005	12	12	7	58.3
2006	101	90	71	78.9
2007	188	172	105	61.0
2008	337	301	196	65.1
2009	478	417	258	61.9
2010	658	582	391	67.2
2011	832	729	439	60.2
2012	994	837	494	59.0
2013	1,114	949	562	59.2
2014	1,222	1,026	595	58.0
2015	1,245	1,078	584	54.2
Total	7,181	6,193	3,702	59.8

Table 2: Percentage of active trials (interventional studies) by fiscal year listed in NIH RePORTER that do not have NIH as a funder type in CT.gov*

	A	B	C	
	No. active listed in NIH RePORTER	No. in CT.gov with NIH as funder type	No. in CT.gov not with NIH as funder type	“miss” % 100*C/A
NCI				
2005	3,132	3,085	47	1.5
2010	4,141	4,086	55	1.3
2015	3,101	3,054	47	1.5
NIMH				
2005	340	314	26	7.6
2010	433	371	62	14.3
2015	314	220	94	29.9
NHLBI				
2005	334	311	23	6.9
2010	459	404	54	11.8
2015	304	261	43	14.1
NIDA				
2005	329	298	31	9.4
2010	454	385	69	15.2
2015	322	244	78	24.2
NIAID				
2005	100	93	7	7.0
2010	142	138	4	2.8
2015	110	102	8	7.3
NIDDK				
2005	205	180	25	12.2
2010	333	294	39	11.7
2015	255	214	41	16.1
NICHD				
2005	200	184	16	8.0
2010	284	243	41	14.4
2015	252	209	43	17.1
NIA				
2005	128	115	13	10.2
2010	201	172	29	14.4
2015	131	101	30	22.9
NCCIH				
2005	120	113	7	5.8
2010	131	106	25	19.1
2015	80	60	20	25.0
NINDS				
2005	70	65	5	7.1
2010	103	97	6	5.8
2015	74	64	10	13.5

Table 2 (con't)

	A	B	C	
	No. active listed in NIH RePORTER	No. in CT.gov with NIH as funder type	No. in CT.gov not with NIH as funder type	"mis" % 100*C/A
NIAAA				
2005	90	79	11	12.2
2010	107	89	18	16.8
2015	92	64	28	30.4
NINR				
2005	64	49	15	23.4
2010	94	75	19	20.2
2015	62	47	15	24.2
NCRR				
2005	125	112	13	10.4
2010	140	118	22	15.7
2015	93	37	56	60.2
NEI				
2005	55	52	3	5.5
2010	61	57	4	6.6
2015	62	60	2	3.2
NIAMS				
2005	66	64	2	3.0
2010	71	62	9	12.7
2015	36	33	3	8.3
Totals				
2005	5,358	5,114	244	4.6
2010	7,154	6,697	456	6.4
2015	5,288	4,770	518	9.8

* Different projects counted to the same NCT counted only once in totals

