SCIENTIFIC FRAUD & THE REPRODUCIBILITY CRISIS:

WHAT ROLE FOR RESEARCH ETHICS?

R. Grant Steen, *PhD*

Professor, Orthopaedic Surgery, LSU Medical Center

Medical Affairs, Bioventus LLC

President, MediCC!, Medical Communications Consultants LLC

Grant.Steen@bioventusglobal.com

Recent high-profile cases of research fraud in biomedicine

- Hwang and human stem cell fakery—this seems quaint now
- Wakefield and the MMR vaccine that does not cause autism
- Reuben had 25 papers retracted, many of them RCTs
- Stapel has 58 retractions to date for making up data
- Boldt currently has 94 papers retracted for fabrication
- Fujii has183 papers retracted for faking data

Is there an epidemic of bad science?

To put this in a broader context….

- In July, 2012, GSK agreed to pay $3B in a fraud suit
  - Antidepressants Paxil and Wellbutrin were promoted off-label
  - Safety data relating to diabetes drug Avandia went unreported
- In May, 2012, Abbott Labs settled a suit on deceptive marketing of the antiepileptic Depakote for $1.6B
- In November, 2013, J&J was fined $2.2B for off-label promotion of the antipsychotic Risperdal
- In December, 2016, Pfizer was fined $104.2M for unfair drug pricing

Is there just an epidemic of dishonesty?

Is there an epidemic of dishonest science?

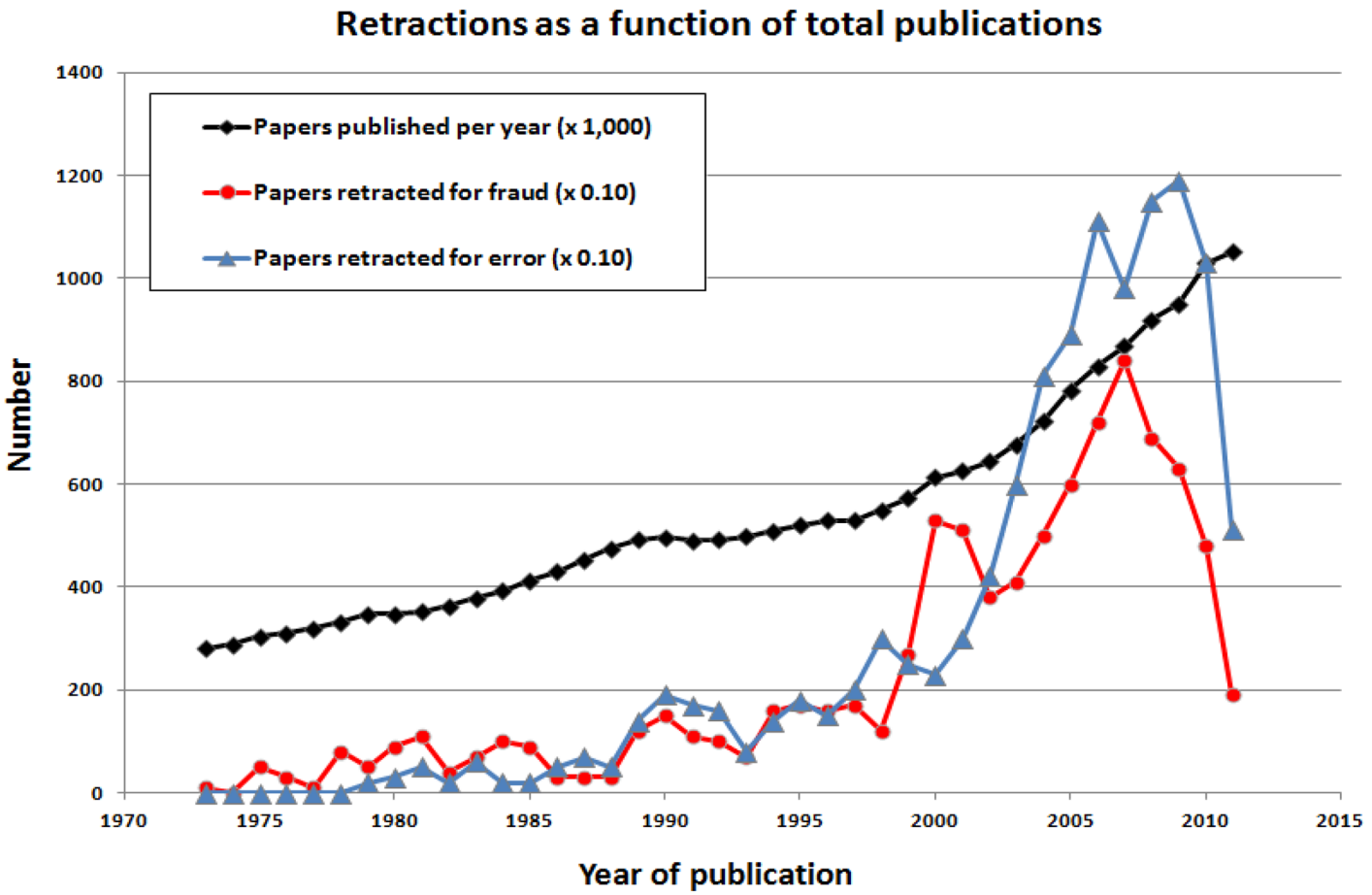
- Analysis of all articles listed as retracted in PubMed
  - Retrospective review done in May, 2012
  - Colleagues were Drs. Ferric Fang and Arturo Casadevall
- There were 2,047 retracted studies listed from 1973 to 2011
- There were ~21.2 million articles published during the same period
- **Retraction rate = 0.0097%**
- 1 in ~10,357 papers was retracted—this is no epidemic!

A key question remains:

How many retraction-worthy papers have not yet been retracted?

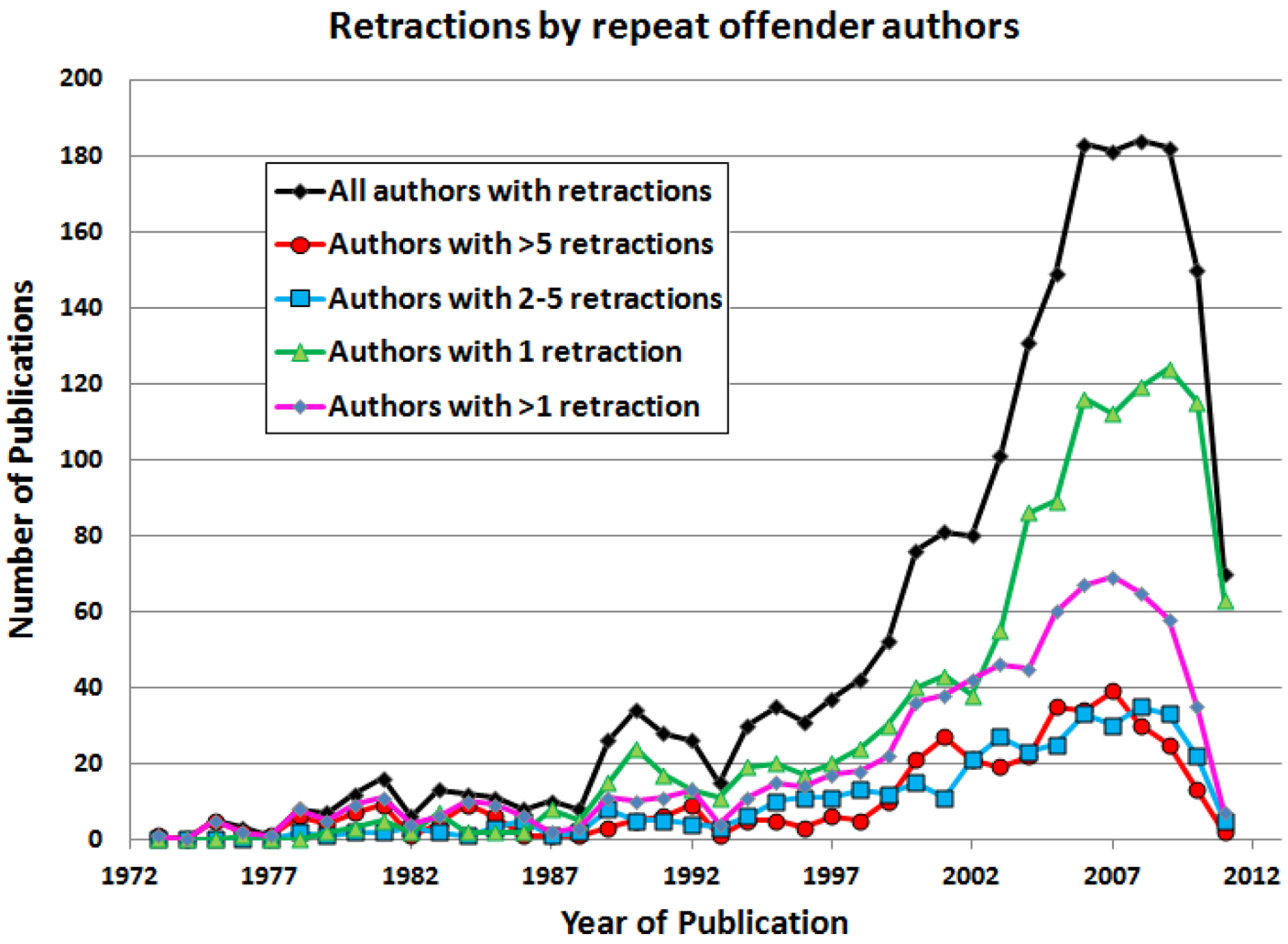
Steen *et al.* '13. *PLoS ONE 8*(7)

Rate of retraction has increased



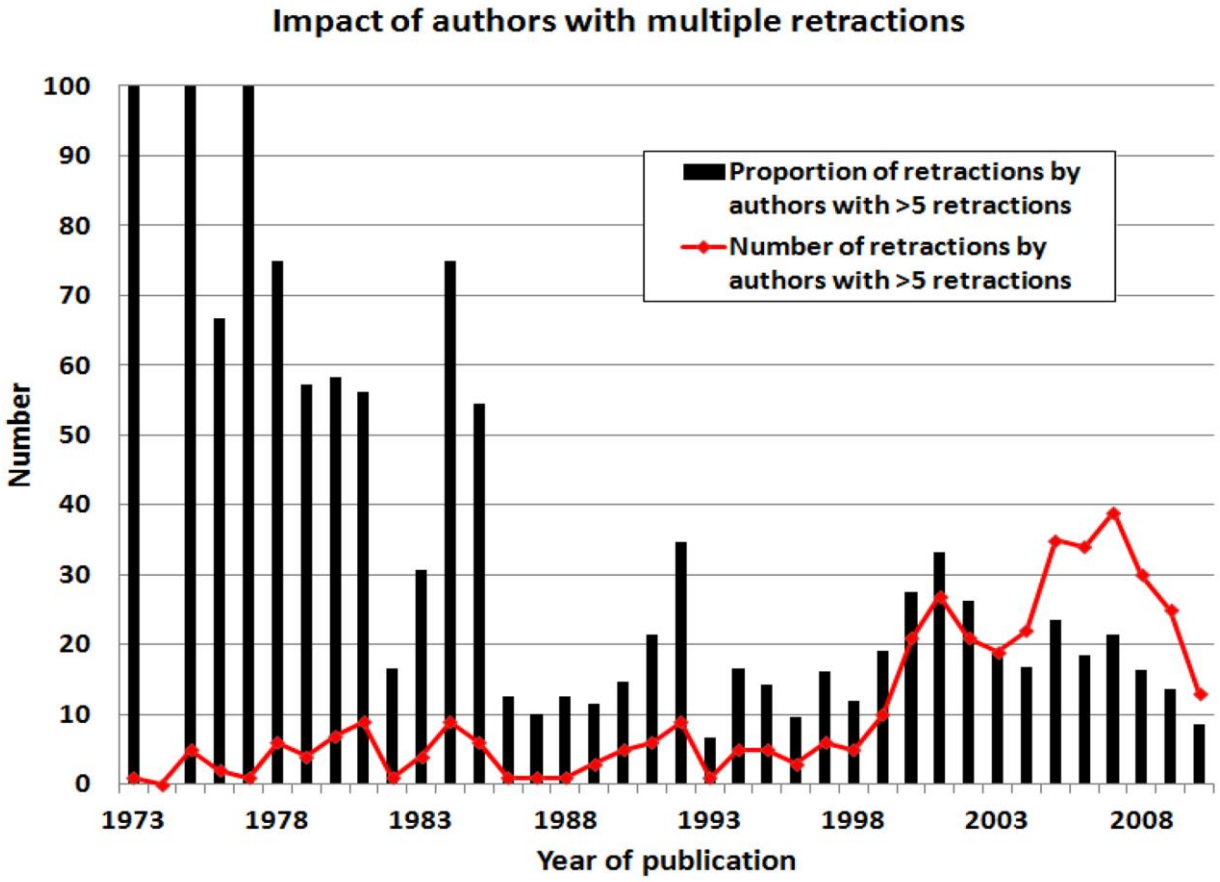
Steen *et al.* '13. *PLoS ONE 8*(7)

Repeat offenders are not the biggest problem



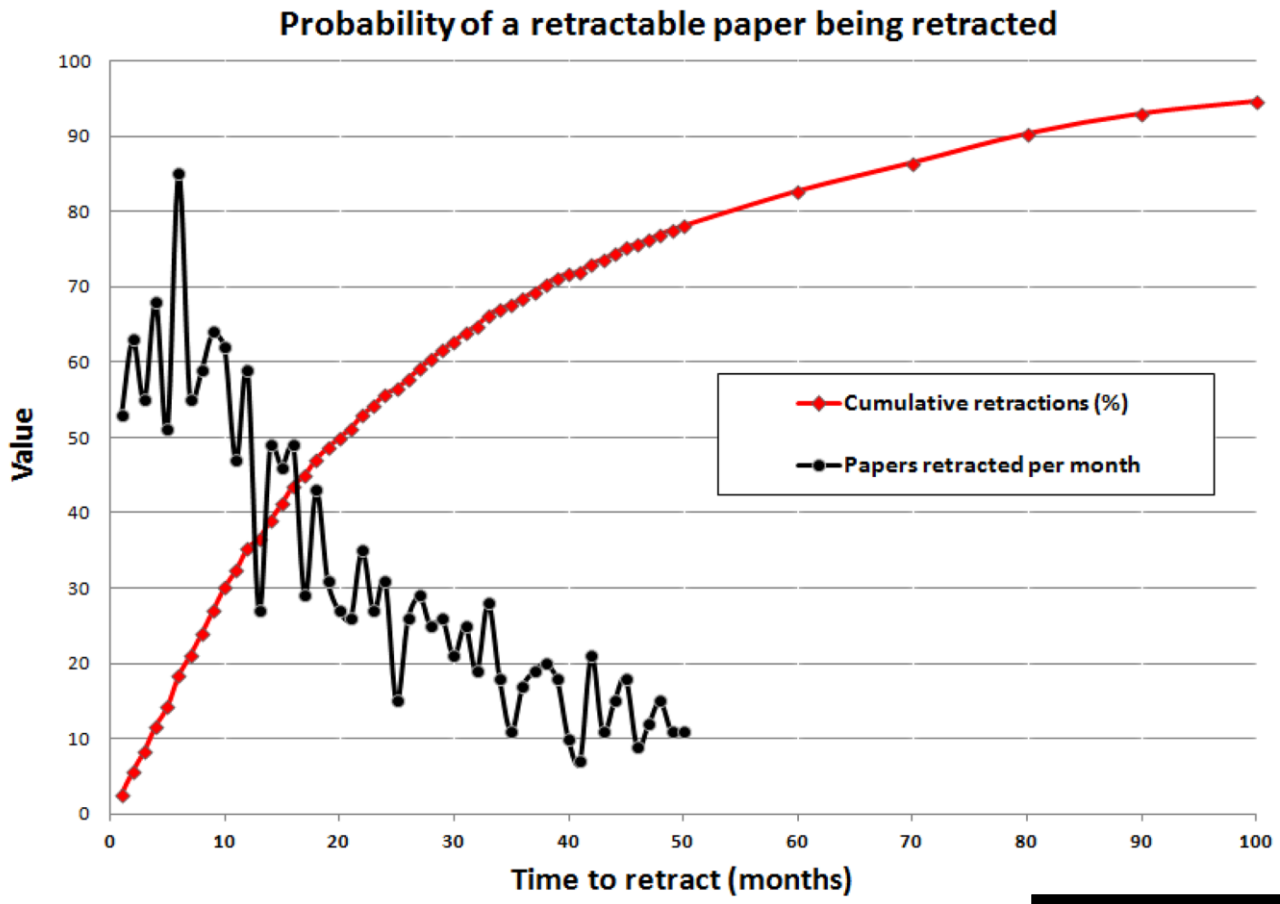
Steen *et al.* '13. *PLoS ONE 8*(7)

Proportional impact of single offenders has increased



Steen *et al.* '13. *PLoS ONE 8*(7)

Will the number of retractions increase in future?



This analysis assumes that all retractable papers are retracted after ~200 months.

Steen *et al.* '13. *PLoS ONE 8*(7)

Can retractions be predicted?

Premise:

The best predictor of the future is the immediate past.

$$\chi = \rho / \pi$$

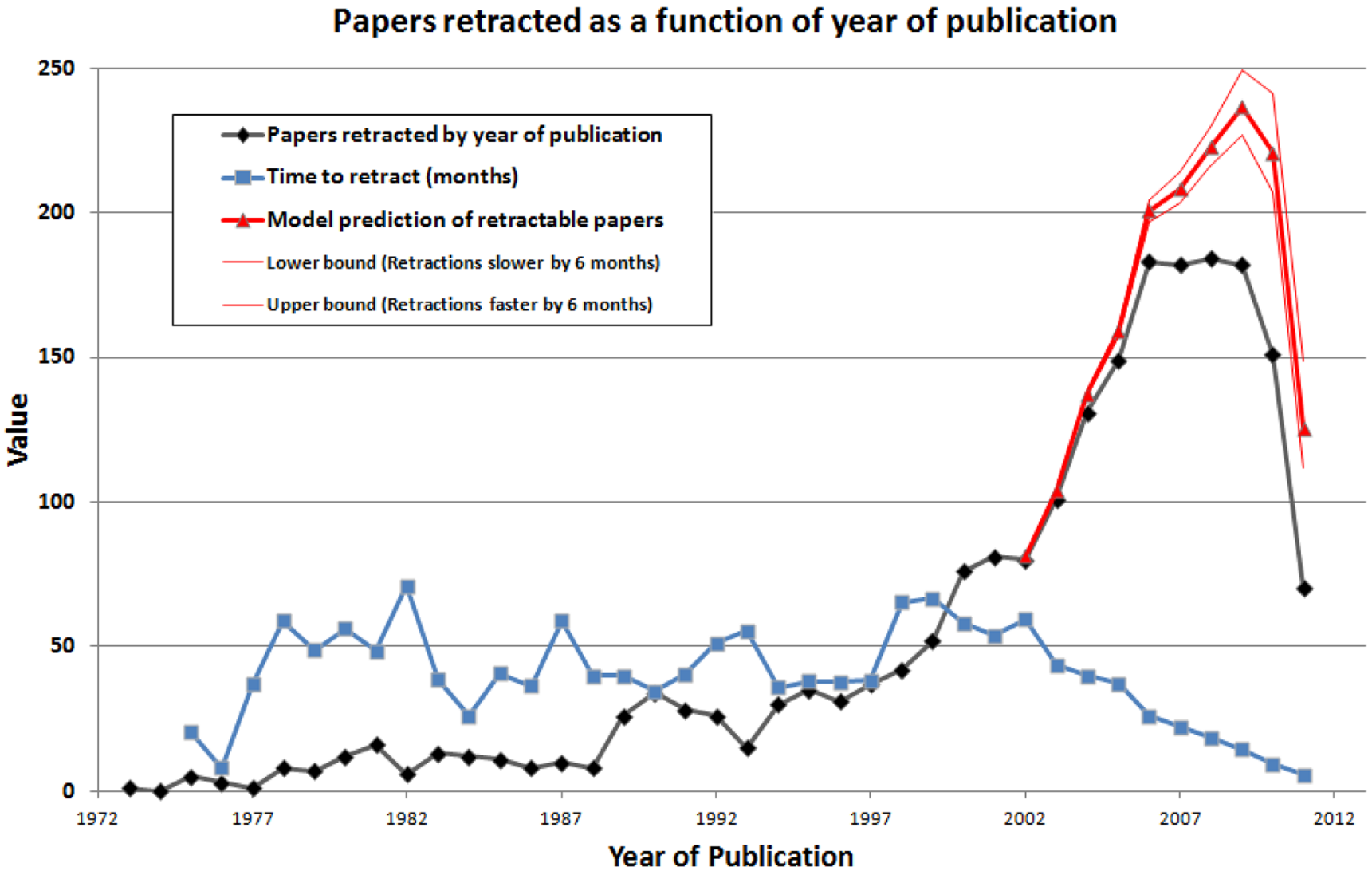
χ = Corrected number of retractions in any given year

ρ = Number of articles retracted by year *Y* since publication

π = Cumulative probability of retraction by year *Y*

NB: Assumes 100% probability of retraction within 200 months

Will the number of retractions increase in future?



Steen *et al.* '13. *PLoS ONE 8*(7)

Some scientists *are* actively trying to deceive

- Among 2,047 retracted papers, retractions were for:
  - **Misconduct in 67.4% of cases**
    - Includes fraud, suspected fraud, duplicate publication, plagiarism
  - Scientific error in 21.3% of cases
  - Some reasons for retraction are unknown
    - Some retraction notices are totally cryptic
- Most retracted papers came from the USA
  - We cannot blame anyone else for retractions

Do papers retracted for fraud differ from papers retracted for error?

Fang *et al.* '12. *PNAS* 109:17028

Fraud appears to be deliberate

Parameter	Fraud	Error	P value
Number of retractions (n)	881	982	---
Target journal IF	8.75	6.77	< 0.0001
Time to retract (months)	43.2	25.7	< 0.0001
"Repeat offender" author	56.8 %	21.3 %	< 0.0001

→ Numbers recently updated from Steen '11. *J Med Ethics* 37: 113

Error is not like fraud;
Retraction notices for error often suggest embarrassment

Consequences of fraudulent clinical research

- Patients may receive risky experimental therapy
- Time and money may be wasted by clinicians and patients in both the primary and any secondary RCTs
- False information may pervade the literature
- Risky therapy may be accepted quickly and used widely

Fraudulent RCTs put patients at risk

- Example: COOPERATE, an RCT published in *Lancet* in 2003
  - Treated hypertension in patients with non-diabetic renal disease
  - Patients received ACE, ARB, or combination (ACE+ARB) therapy
  - Retracted for data falsification and ethical violations in 2009
- Primary study treated 263 patients
  - Study cited 581 times, including 173 review articles and meta-analyses
- Secondary clinical studies enrolled 35,929 patients
  - 31,239 patients received risky ACE+ARB (combination therapy)
  - These studies were legitimate efforts to replicate the earlier study, so they really did put patients at risk

Many patients may be put at risk in retracted studies

Summary of the impact of 180 retracted clinical papers

	Number	Average per retraction
Subjects enrolled	28,783	160.8
Patients at risk	17,783	99.3
Patients treated	9,189	51.3

→ NB: Some of these patients may have been fabricated

Steen '11. *J Med Ethics* 37: 688

Many patients are also put at risk in secondary studies that are based on retracted papers

A total of 851 secondary papers were inspired by the 180 clinical papers retracted from 2000 to 2010

→ NB: Apparently none of these patients were fabricated

	Number	Average per retraction
Subjects enrolled	445,064	2,472.6
Patients at risk	165,588	919.9
Patients treated	70,501	391.7

Steen '11. *J Med Ethics* 37: 688

Retractions for fraud may put more patients at risk than retractions for error

	Fraud retractions	Error retractions	P value
Subjects per study	147.0	163.2	0.78
Patients at risk	125.9	84.4	0.20
Treated patients	96.2	24.2	0.01

Why are more patients treated in fraudulent RCTs?

- Is it just easy for fraudulent authors to fabricate patients?
- Don't fraudulent authors care that patients are put at risk?

Steen '11. *J Med Ethics* 37: 688

Misinformation from retracted papers corrupts the literature

Summary of the impact of 180 clinical papers retracted from 2000 to 2010

	Number	Average per retraction
Research-related citations	5,143	28.6
Post-retraction citations	1,973	11.0
Retraction-related citations	360	2.0
Review papers + Metas	1,372	7.6
Studies that enrolled patients	851	4.7

Steen '11. *J Med Ethics* 37: 688

Reality check

According to PubMed (search in July 2016):

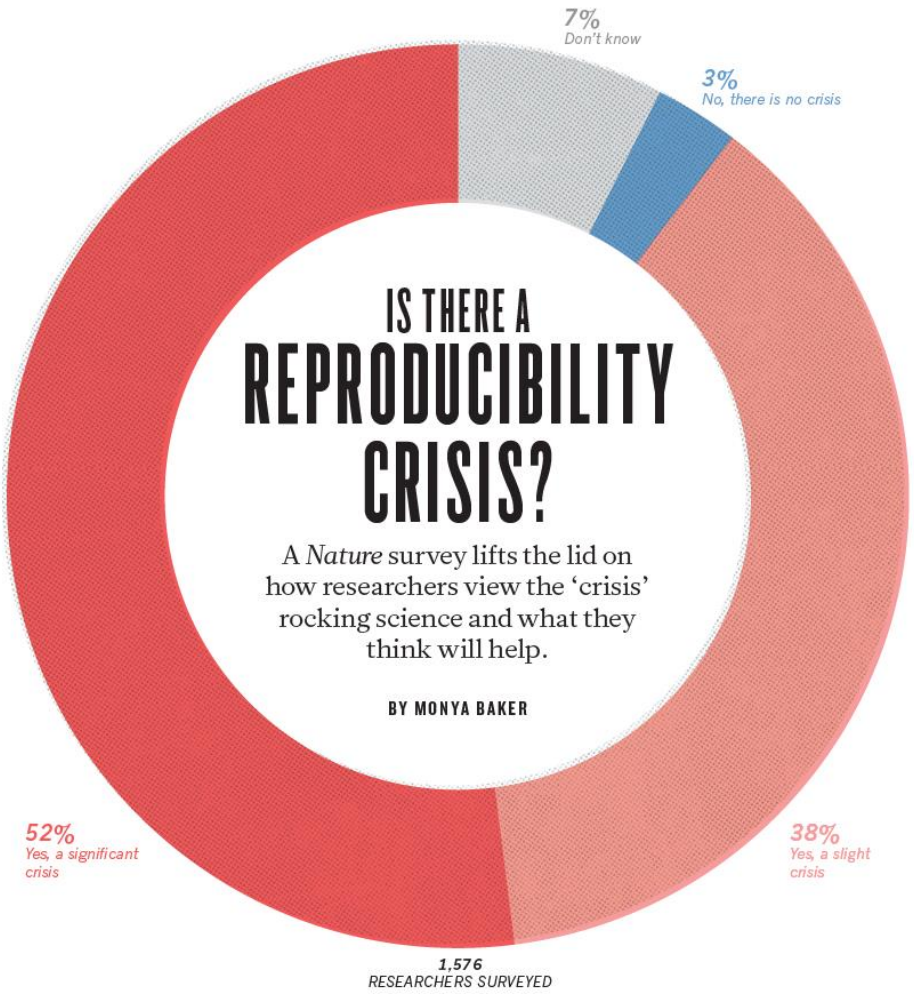
- **20.2 Million journal articles have been published in English**
- **4,380 of these articles were retracted**
- **Thus, 1 in every 4,622 articles is retracted**
- **This represents 0.0216% of the literature**

→ Is this analogous to 12.7 yrs of perfect weather forecasts?

Is this a crisis? I think not….

However, people have lost faith in their institutions

The reproducibility crisis in science



How much of the "Reproducibility Crisis" might be explained by fraud?

To address this question, we evaluated the types of mistakes that scientists make

Nature, May 2016

What mistakes do authors make?

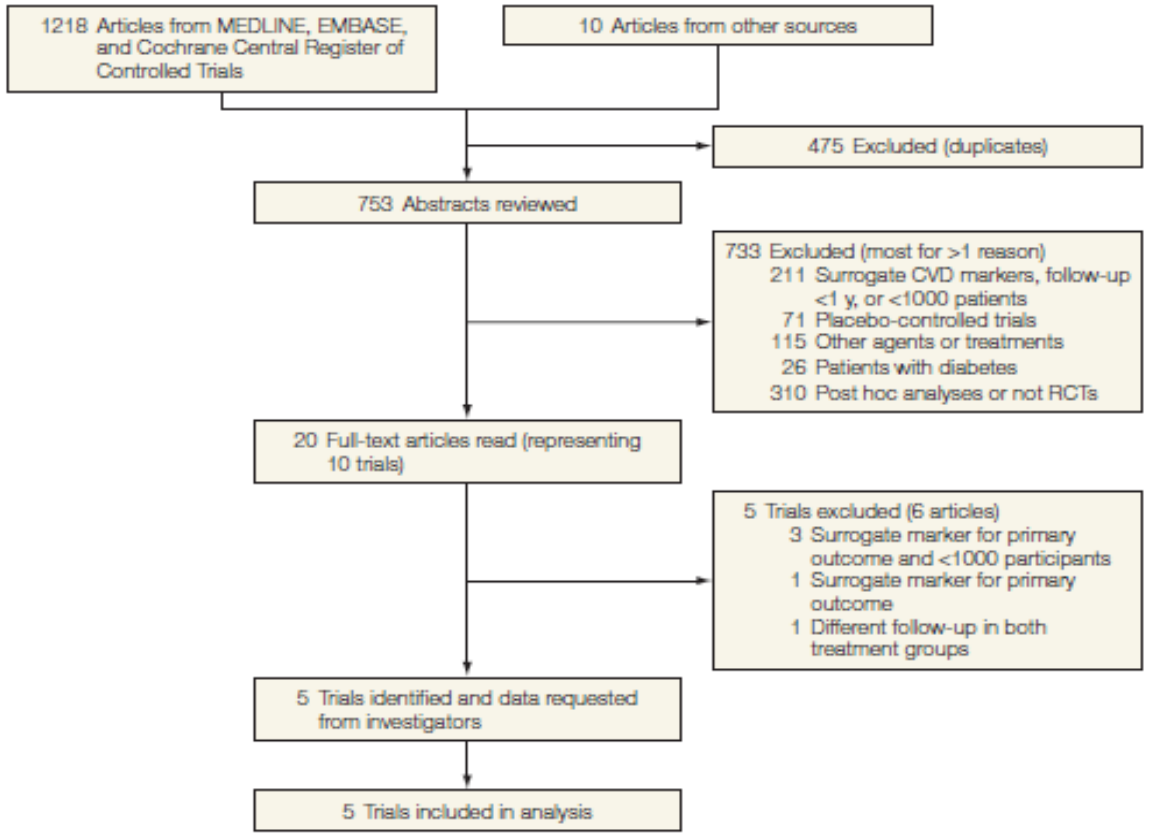
A QUORUM figure summarizes the process of data collation used in writing a meta-analysis:

All papers evaluated are classified into groups:

- Evaluated for inclusion
- Given detailed review
- Excluded because of non-relevance
- Excluded for any of several errors
- Included in meta-analysis

Example:
QUORUM figure from Priess *et al* '11 *JAMA*

Figure 1. Flow Diagram of the Literature Search



CVD indicates cardiovascular disease; RCTs, randomized controlled trials.

What kind of mistakes did Preiss *et al* find?

753 clinical studies were evaluated for inclusion in a meta-analysis of risk of diabetes following moderate or intensive statin therapy

Mistake	Excluded	Percent
Not an RCT or evaluated *post-hoc* hypotheses	310	41 %
FU < 1 yr or <1,000 pts or surrogate marker used	211	29 %

Preiss *et al* '11. JAMA 305: 2556

We used QUORUM figures to evaluate errors in a broad sample of the literature

We basically did a meta-analysis of meta-analyses:

- We evaluated all metas in the 20 most highly-cited journals
- Every meta published from 2009 to 2010 was included
- Sample size = **316 meta-analyses**
- Over 785,000 papers were evaluated in these metas
- A total of 56,911 papers were given detailed review
- 11,346 papers were included in the 316 metas

→ 81.1% of papers given detailed review were excluded

Question: Why are papers excluded from meta-analysis?

Journal of origin of the 316 meta-analyses

Journal	Impact Factor	# Metas Published
BMJ	13.5	72
Lancet	33.6	36
J Am Coll Cardiol	14.3	30
Lancet Infect Dis	16.1	29
JAMA	30.0	21
15 other journals	23.1	128
Average=	22.7	15.8

Work done in collaboration with Stephen Dager, *MD*, U Washington

How many clinical studies are high quality?

Meta-meta-analysis in Steen & Dager '13 *FASEB J.* 27:3430

A total of 316 recent meta-analyses are summarized

Parameter	Number	Percent
Total papers evaluated by meta-authors	785,478	---
Total papers given detailed review	56,620	100 %
Papers excluded for non-relevance	34,005	60 %
Papers excluded for a specific error	11,412	20 %
Papers included in meta-analyses	11,203	20 %

Note that only 1.4% of papers evaluated were included in a meta!

What are the most common errors in clinical studies?

Type of Error	Number	Percent
Insufficient detail given to replicate	3,317	29.1 %
Duplicative analysis or publication	1,775	15.6 %
Improperly or inadequately randomized	1,676	14.7 %
Improper controls used	1,648	14.4 %
Inadequately blinded	871	7.6 %

The remaining 19.6% of flawed papers had small *N*, inadequate follow-up, or were too poorly written.

Steen & Dager '13 *FASEB J.* 27:3430

Sources of error in the retracted scientific literature

Retraction notices for 423 articles were evaluated:

- 236 retractions (55.8%) were for laboratory error
  - 30.3% were unique to the article
  - 17.5% were due to contamination
  - 7.1% were DNA-related
- 80 retractions (18.9%) for analytical error
- 68 retractions (16.1%) for lack of reproducibility
- 39 retractions (9.2%) were for “Other” reasons

Casadevall et al. 2014 *FASEB J* 28:3847

Many unretracted papers are known to be wrong

Steinschneider. 1972. *Pediatrics* 50:646

A description of multiple cases of SIDS in a single family

→ The mother was subsequently convicted of murder

Wolf-Simon *et al*. 2011. *Science* 332:1163

Bacterium uses arsenate not phosphate in nucleic acid

→ Bacterium does not contain arsenate

Chow *et al*. 1993. *Nature* 361:650

Multidrug resistance incompatible with HIV replication

→ Multiresistant HIV can nonetheless replicate

Casadevall et al. 2014 *FASEB J* 28:3847

Most errors in the literature are never retracted

For the record: I do not believe that all errors should be retracted….

- How often is "accepted wisdom" really true?
  - How often are highly-cited papers contradicted?
- Of 49 papers published in high-IF journals & cited >1,000 times
  - 45 of 49 studies touted a successful ("positive") intervention
  - How often were positive findings replicated?

→ **Data from Ioannidis ('05 *JAMA* 294: 218)**

How often are high-impact studies replicated?

Among 45 studies that touted a successful intervention:

- 44% were replicated in later studies (n=20)
- 16% were strengthened (n=7)
- 16% were contradicted (n=7)
- 24% were unreplicated (n=11)

Being unreplicated means being unconfirmed

- This is not good news!

Ioannidis '05. *JAMA* 294: 218

What characteristics define replicated RCTs?

	Conclusions challenged	Conclusions unchallenged	P value
Number of studies	9	30	----
Recent publication ('90-'95)	8	15	0.06
Median citations / year	149	214	0.07
Written about a mature field	4	13	1.00
Median study *N*	624	2,165	0.009

Sample size is crucial!

Ioannidis '05 *JAMA* 294: 218

Basic scientists can also fall victim to small *N*

Cell

Betatrophin: A Hormone that Controls Pancreatic β Cell Proliferation

Peng Yi,[1] Ji-Sun Park,[1] and Douglas A. Melton[1,*]
[1]Department of Stem Cell and Regenerative Biology, Harvard Stem Cell Institute, Howard Hughes Medical Institute, Harvard University, 7 Divinity Avenue, Cambridge, MA 02138, USA
*Correspondence: dmelton@harvard.edu
http://dx.doi.org/10.1016/j.cell.2013.04.008

RETRACTED

SUMMARY

Replenishing insulin-producing pancreatic β cell mass will benefit both type I and type II diabetics. In adults, pancreatic β cells are generated primarily by self-duplication. We report on a mouse model of insulin resistance that induces dramatic pancreatic β cell proliferation and β cell mass expansion. Using this model, we identify a hormone, betatrophin, that is primarily expressed in liver and fat. Expression of betatrophin correlates with β cell proliferation in other mouse models of insulin resistance and during gestation. Transient expression of betatrophin in mouse liver significantly and specifically promotes pancreatic β cell proliferation, expansion of β cell mass, and improves glucose tolerance. Thus, betatrophin treatment could augment or replace insulin therapy by increasing the number of endogenous insulin-producing cells in diabetics.

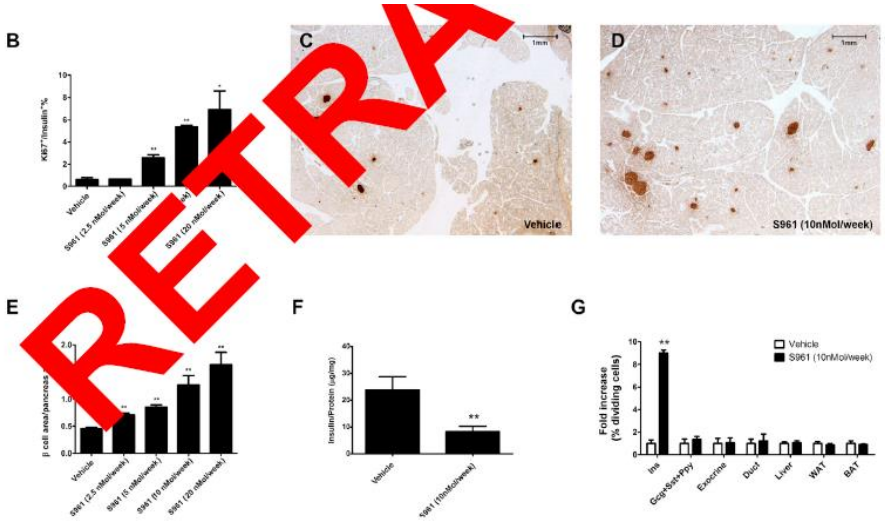
INTRODUCTION

Diabetes results from dysfunctional carbohydrate metabolism that is caused by relative deficiency of insulin. It has become a major threat to human health, the prevalence of which is projected to rise worldwide (171 million affected in 2000 and is predicted to rise to 4.4% (366 million) by 2030 (Wild et al., 2004). Around 10% of diabetics in the United States have T1DM, a disease caused by an autoimmune attack on pancreatic β cells and a consequent β cell deficiency. The majority of diabetics are type II, characterized by interrelated metabolic disorders that include decreased β cell function, peripheral insulin resistance, and, eventually, β cell failure and loss or dedifferentiation (Scheen and Lefèbvre, 1996; Talchai et al., 2012). Though the disease can be treated with antidiabetic drugs or subcutaneous insulin injection, these treatments do not provide the same degree of glycemic control as functional pancreatic β cells and do not prevent the debilitating consequences of the disease. Treatments that replenish β cell mass in diabetic patients could allow for the long-term restoration of normal glycemic control and thus represent a potentially curative therapy. Despite the fact that the primary causes for type I and type II diabetes differ, all diabetics will benefit from treatments that replenish their β cell mass.

Though there is evidence that new β cells can be derived from rare adult progenitors under extreme circumstances (Xu et al., 2008), the vast majority of new β cells are generated by simple self-duplication (Dor et al., 2004; Meier et al., 2008; Teta et al., 2007). After the rapid expansion in embryonic and neonatal stages, β cells replicate at an extremely low rate (<0.5% divide per day) in adult rodents (Teta et al., 2005) and in humans (Meier et al., 2008). However, pancreatic β cells retain the capacity to accelerate their replication rate in response to physiological challenges, including gestation (Parsons et al., 1992; Rieck et al., 2009), high blood sugar (Alonso et al., 2007), pancreatectomy (Cano et al., 2008; Nir et al., 2007), and peripheral insulin resistance (Brüning et al., 1997; Kulkarni et al., 2004; Michael et al., 2000; Pick et al., 1998).

The genetic mechanisms controlling β cell proliferation are incompletely understood. The cell-cycle regulators cyclin D1/D2 and CDK4 promote β cell proliferation (Georgia and Bhushan, 2004; Kushner et al., 2005; Rane et al., 1999), and cell-cycle-related transcription factors such as E2F1/2 are essential for pancreatic β cell proliferation (Fajas et al., 2004; Iglesias et al., 2004). On the contrary, cell-cycle inhibitors, including $p15^{Ink4b}$, $p18^{Ink4c}$, and $p27^{Kip1}$, repress β cell replication (Latres et al., 2000; Pei et al., 2004; Uchida et al., 2005). Other genes reported to regulate β cell proliferation include NFAT, Menin, p53, Rb, and Irs2 (Crabtree et al., 2003; Harvey et al., 1995; Heit et al., 2006; Kubota et al., 2000; Williams et al., 1994).

In addition to the factors listed above, which are expressed in β cells themselves and act in a cell-autonomous fashion, there are several reports showing that systematic or circulating factors can regulate β cell replication and mass. Glucose itself is a β cell mitogen; infusion of glucose in rodents causes a mild increase in β cell replication (Alonso et al., 2007; Bernard et al., 1998; Bonner-Weir et al., 1989). And glucokinase defects significantly decrease the compensatory proliferation of pancreatic β cells in some contexts (Terauchi et al., 2007). In addition, genetic deletion of glucokinase in β cells can reduce replication rates, whereas pharmacological activation of this enzyme increases replication by 2-fold (Porat et al., 2011). Several hormones, including insulin, placental lactogen, and prolactin, also play a role in regulating β cell mass (Bernard et al., 1998; Paris et al., 2003; Parsons et al., 1992; Sachdeva and Stoffers, 2009). The incretin hormones glucagon-like peptide 1 (GLP-1) and glucose-dependent



RETRA

Figure 2. Administration of the Insulin Receptor Antagonist S961 Induces Pancreatic β Cell Proliferation and β Cell Mass Expansion
(A) S961 infused into adult mice at 10 nMol/week for 1 week induces pancreatic β cell proliferation (shown by costaining of Ki67 and insulin).
(B–E) (B) Proliferation rates of pancreatic β cells measured as percentage of dividing β cells 7 days after S961 treatment at different doses. S961 treatment significantly increases β cell area shown by insulin immunohistochemistry (brown) (representative sections shown in C and D; β cell area as a percentage of total pancreas area in E). n = 4 in each dosage group.

Basic research is often flawed by a tiny *N* used to make big claims.

"Truth" has a shelf-life

- All papers on cirrhosis or hepatitis were reviewed
- From 1944 to 1999, 474 papers were published
- Main conclusion of each paper was evaluated for "truth"
- In 2000, 60% of conclusions were still "true"
  - 19% of conclusions were obsolete or superseded
  - 21% of conclusions were considered to be "false"
    - None of these papers were retracted

Poynard '02. *Ann Int Med* 36: 888

"True" conclusions are likely to be recent

- **Main conclusion "true" in 43% of older papers**
- **Main conclusion "true" in 76% of recent papers**
- **Half-life of truth is ~45 years**
- **Medical science may be getting better**
  - Yet there has been less time to disprove recent studies

Poynard '02. *Ann Int Med* 36: 888

"True" conclusions arise from meta-analyses

- "True" conclusions in meta-analyses = 82%
- "True" conclusions in RCTs = 62%
- "True" conclusions in non-randomized studies = 50%

However, all meta-analyses were done after 1980, meaning that there has been less time to correct them.

Poynard '02. *Ann Int Med* 36: 888

Negative conclusions more likely to be "true"

50-year survival rate of "truth" was:

- 51% for 110 papers with negative conclusions
- 23% for 364 papers with positive conclusions

Negative conclusions are more frequent in:

- Therapeutic (29%) than in diagnostic studies (14%)
- RCTs (31%) than in non-randomized studies (16%)

Poynard '02. *Ann Int Med* 36: 888

Fraud *vs* error

What is the balance between fraud and error?

- **Ioannidis estimated that 50% of the literature is wrong**
  - **This was a theoretical argument…**
- **I estimate that 20% of the literature contains error**
  - **This was a pragmatic assessment….**
- **Probably less than 1% of the literature should be retracted**
  - **This is purely a guess….**

→ Error is at least 20-fold more common than fraud

How can we increase study reproducibility?

- Large sample size (N=4 is NEVER acceptable!)
- Clear statement of a testable hypothesis
- Reasonable inclusion / exclusion criteria
- Blinding of all study participants
- Long term follow-up of study participants
- Appropriate use of statistics
  - Rigorous test of a primary hypothesis, not a test of everything
- Publication in a prominent journal

How can misinformation in RCTs be minimized?

Key definitions:

Misinformation = False or incorrect information

- Large sample *N*
- Effective blinding
- Greater detail in the clinical trials registry
- Clear statistical criteria at every stage of a clinical trial
- Responsibility for data integrity that accrues to all authors
- Greater transparency as to how costs of research were paid

Registration in the Clinical Trials Registry

Registration of RCTs prevents some of the worst abuses:

- Non-reporting of a failed trial
- Non-reporting of a failed primary objective
- Reporting a secondary objective as if it were primary
- Changing inclusion / exclusion criteria after the fact
- Data fabrication might become harder

However, registration is not yet all that effective

Research ethics are a hedge against human fallibility

- We often actively deceive ourselves
  - The first person fooled is the easiest to fool
- Most misinformation is never retracted
  - A lot of what we think we know—we don't!
- Our cognitive castles are built on shifting sand
  - We crave certainty in a fundamentally uncertain world
- **Ethics are how we protect ourselves**
  - Often we most need protection from ourselves

Conclusions

- The research enterprise is fairly healthy overall
  - Yet improvement is always possible
- Some scientists are actively trying to deceive you
- Far more scientists make damaging errors
  - Does wishful thinking sometimes cause errors?
- Misinformation is rife in the literature
  - Ioannidis estimated that half the clinical literature is wrong
  - **I estimate that ~20% of the literature is flawed**

Can research ethics enhance reproducibility?

Open questions in research misconduct

- Should the term “research misconduct” be used for all infractions?
  - Should we conflate plagiarism with fabrication and falsification?
- Is it legitimate to separate scientific misconduct from publication misconduct?
  - Scientific misconduct is violation of the process of doing science
  - Publication misconduct is violation of the process of publication
  - However, publication can be considered part of the scientific process
- Are there circumstances in which misconduct is more likely?
  - How can those circumstances be mitigated?