

# Plagiarism in the electronic age

**Harold C. Sox, M.D. MACP**

**Editor Emeritus**

**Annals of Internal Medicine**

**Professor of Medicine and of The  
Dartmouth Institute (emeritus, active)**

**Dartmouth Medical School**

**A REMINDER**

The scientific record should be  
sacrosanct

Fixing errors and removing fakery is  
an *obligation* of the scientific  
community

# Investigations

Alleged Misconduct

```
graph TD; A[Alleged Misconduct] --> B[Guilty verdict]; B --> C[Investigate other work];
```

Guilty verdict

Investigate *other work*

# The scope of inquiry

- The rule of thumb: **every article is suspect until proven otherwise**
- **This rule can mean thousands of hours of work reviewing articles and interviewing co-authors.**

# CASE 1

# Stanford Investigates Plagiarism Charge

*It is looking into the unattributed use of copyrighted material in a textbook chapter written by the chairman of the department of medicine.*

Stanford University is investigating allegations of academic misconduct that have been levelled against the chairman of the Department of Medicine, Kenneth Melmon. The allegations stem from the fact that Melmon incorporated large chunks of copyrighted material from a book he helped edit into a chapter he wrote for another textbook. The material was used without attribution and apparently without permission.

Melmon says he incorporated the material reluctantly, at the insistence of the textbook's academic editor, only after he had been assured that permission had been granted. He also says that his manuscript contained prominent attribution for the reprinted text and he was "stunned" when he found it was omitted from the published version.

The two books in question, Goodman and Gilman's *The Pharmacological Basis of Therapeutics* and Williams's *Textbook of Endocrinology*, are considered the standard works in their fields. Melmon was an associate editor of the sixth edition of the Goodman and Gilman book, which was published by Macmillan in 1980, and he wrote a chapter in the sixth edition of Williams, which was published a year later by W. B. Saunders. Some 15 pages of Goodman and Gilman, taken from eight chapters by four different authors, were incorporated into Melmon's 73-page chapter.

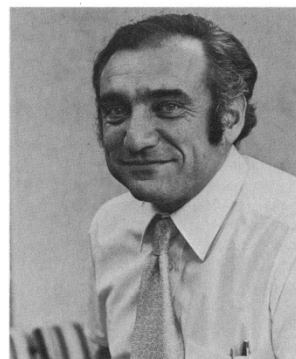
Although the Williams book has been on the market for more than 2 years, the unattributed use of the material did not come to light until early February, when William W. Douglas, a pharmacologist at Yale University School of Medicine, noticed some of his own text from Goodman and Gilman in Melmon's chapter. Douglas discovered it when he was looking through the literature to update his chapters for the next edition of Goodman and Gilman. "I was just the first of a group of outraged authors who spotted this," he says.

Douglas called Alfred Goodman Gilman to ask why Gilman had never notified him or sought his permission for the material to be used. Gilman, a pharmacologist at the University of Texas, who edited the book with his father, Alfred Gilman, and his father's longtime collaborator, Louis Goodman, says it was the first he knew of the matter. He checked with Macmillan and then notified Domi-

nick Purpura, dean of the medical school at Stanford, and asked for an investigation. Gilman also called Melmon.

Melmon says he had not examined the published version of the Williams book until Gilman called. When he found no attribution for the reprinted material and learned that no permission had been given, he was "just dumbfounded. It washed me away."

Purpura referred the matter to the medical school's Committee on Ethical Scientific Performance, which he chairs,



**Kenneth Melmon**

*Says he was stunned when allegations arose.*

and the committee began an internal investigation. In mid-March, just a few weeks before the investigation would be completed, photocopied pages from the two books were sent anonymously to *Science*. An article about the investigation appeared in the 22 March edition of the San Jose *Mercury News*, and the university then issued a press statement.

According to Melmon's reconstruction of the events, the whole sorry business stemmed from a combination of a breakdown in communication, Melmon's failure to read the proofs of his chapter himself, and editorial chaos following the sudden death of Robert Williams, the editor of the *Textbook of Endocrinology*. Williams died in November 1979 when the book was still in progress. Melmon was also overcommitted and in the process of moving to Stanford from the University of California at San Francisco. "He is not trying to put the blame

on anybody else. He recognizes there are problems. He recognizes he should not have relied on other people," says Jack Friedenthal, a Stanford law professor who is acting on a voluntary basis as Melmon's legal counsel.

Melmon wrote chapters in both the fourth and fifth editions of Williams's *Textbook of Endocrinology* and agreed to contribute to the sixth edition. He says he was late with his manuscript, and Williams would constantly call to ask how it was coming along and make suggestions about what should be included. When the chapter was completed, but before it was sent off, Melmon says Williams "started peppering me with requests" to put in more pharmacology.

Melmon formerly worked with Williams at the University of Washington—where Williams remained until his death—and regarded him as something of a mentor. He thus found it difficult to resist Williams's request. Nevertheless, Melmon says he told Williams that the pharmacology was well covered in Goodman and Gilman and that a reference in his chapter to that work would be sufficient. As an alternative, Melmon suggested that Williams could take his manuscript and give it to somebody else to add in the pharmacology.

According to Melmon, Williams went off and read the fifth edition of Goodman and Gilman and called back a few days later even more insistent that more pharmacological details should be added. He then suggested, Melmon recalls, that Melmon should include some of the material he was working on as associate editor of the sixth edition of Goodman and Gilman.

At that point, Melmon says he told Williams that it would be "an awful lot to ask" for permission to put this material into his own chapter. He says he told Williams he would include the material only if Williams himself obtained the permission. "I was hoping they would say no," Melmon says. "but it didn't work." Williams called back later to say permission had been granted.

Gilman, Douglas, and Norman Weiner, a pharmacologist at the University of Colorado whose text was also reproduced in Melmon's chapter, all told *Science* that their permission was neither sought nor given. An attorney at Macmillan says the company has found no

**Stanford University is investigating allegations of academic misconduct that have been leveled against the chairman of the Department of Medicine, Kenneth Melmon. The allegations stem from the fact that Melmon incorporated large chunks of copyrighted material from a book he helped edit into a chapter he wrote for another textbook. The material was used without attribution and apparently without permission....**

Source: Science. 1984; 224:36-37.



**“The two books in question, Goodman and Gilman's The Pharmacological Basis of Therapeutics and Williams's Textbook of Endocrinology, are considered the standard works in their fields....**

**Some 15 pages of Goodman and Gilman, taken from eight chapters by four different authors, were incorporated into Melmon's 73-page chapter.”**

Source: Science. 1984; 224:36-37.

**“Melmon says that when he cut-and pasted the material into his manuscript, he added handwritten notations detailing where the text came from. These notations were supposed to have been printed in the body of his chapter.....”**

**“Nevertheless, Melmon says he delegated responsibility for checking the galleys to assistants in San Francisco.” [HS note: he also did not look at the chapter after it had been published]**

Source: Science. 1984; 224:36-37.

**“The medical school ethics committee is expected to send a report of its investigation to Stanford president Donald Kennedy in the next few weeks. It will then be up to Kennedy to determine what action, if any, should be taken.”**

Source: Science. 1984; 224:36-37.

## **POSTSCRIPT:**

**Melmon lost his departmental chairmanship but retained his tenured professorship. He served the School of Medicine in various second-tier administrative roles for 15 years and remained active in writing for publication. He retired in his mid-60's and died in April 2002.**

# MORALS OF THE STORY

- Plagiarism, even when attribution was intended, can bring down the mighty.
- If you want to use someone else's work , don't delegate responsibility for asking permission.
- Writing for publication is serious business. Check your own proofs.

# Etymology of “Plagiarism”

- derives from Latin *plagiārius*, "kidnapper", equivalent to *plagium*, "kidnapping"

# Plagiarism defined

According to the Merriam-Webster Online Dictionary, to "plagiarize" means

- to steal and pass off (the ideas or words of another) as one's own
- to use (another's production) without crediting the source
- to commit literary theft
- to present as new and original an idea or product derived from an existing source.

**In other words, plagiarism is an act of fraud. It involves both stealing someone else's work and lying about it afterward.**

# Plagiarism defined

According to the Merriam-Webster Online Dictionary, to "plagiarize" means

- to steal and pass off (the ideas or words of another) as one's own
- to use (another's production) without crediting the source
- to commit literary theft
- to present as new and original an idea or product derived from an existing source.

In other words, **plagiarism is an act of fraud**. It involves both stealing someone else's work and lying about it afterward.

Source: [http://www.plagiarism.org/plag\\_article\\_what\\_is\\_plagiarism.html](http://www.plagiarism.org/plag_article_what_is_plagiarism.html)



# What acts are plagiarism?

- turning in someone else's work as your own
- copying words or ideas from someone else without giving credit
- failing to put a quotation in quotation marks
- giving incorrect information about the source of a quotation
- changing words but copying the sentence structure of a source without giving credit
- copying so many words or ideas from a source that it makes up the majority of your work, whether you give credit or not (see our section on "fair use" rules)

Source: [http://www.plagiarism.org/plag\\_article\\_what\\_is\\_plagiarism.html](http://www.plagiarism.org/plag_article_what_is_plagiarism.html)

# Plagiarism in the electronic age

- Easier to commit: electronic cut and paste
- Easier to detect: electronic text-matching
- Annals: one instance of plagiarism in the period from July 2001 to July 2009

# CASE 2

## Plasma Level of a Triggering Receptor Expressed on Myeloid Cells-1: Its Diagnostic Accuracy in Patients with Suspected Sepsis

Sébastien Gibot, MD; Marie-Nathalie Kolopp-Sarda, PharmD, PhD; Marie C. Béné, PharmSci, PhD; Aurélie Cravoisy, MD; Bruno Levy, MD, PhD; Gilbert C. Faure, MD, PhD; and Pierre-Edouard Bollaert, MD, PhD

**Background:** Previous experimental studies have suggested that the triggering receptor expressed on myeloid cells-1 (TREM-1) is specifically upregulated in the presence of microbial products.

**Objective:** To evaluate the diagnostic value of plasma levels of the soluble form of TREM-1 in patients admitted with clinical suspicion of infection.

**Design:** Prospective, noninterventional study conducted between July and September 2003.

**Setting:** Medical adult intensive care unit at a university hospital in France.

**Participants:** 76 consecutive newly admitted patients who presented with clinically suspected infection and fulfilled at least 2 criteria of the systemic inflammatory response syndrome.

**Measurements:** Sensitivity and specificity of plasma soluble TREM-1 levels at admission for the diagnosis of infection. Two independent intensivists blinded to the results of soluble TREM-1 assays retrospectively classified patients as having the systemic

inflammatory response syndrome, sepsis, severe sepsis, or septic shock.

**Results:** The systemic inflammatory response syndrome was diagnosed in 29 patients (38%), and sepsis, severe sepsis, or septic shock was diagnosed in the remaining 47 (62%). A plasma soluble TREM-1 level higher than 60 ng/mL was more accurate than any other clinical or laboratory finding for indicating infection (sensitivity, 96% [95% CI, 92% to 100%]; specificity, 89% [CI, 82% to 95%]; positive likelihood ratio, 8.6 [CI, 3.8 to 21.5]; negative likelihood ratio, 0.04 [CI, 0.01 to 0.2]).

**Limitations:** The study did not enroll patients with mild infections not requiring intensive care unit hospitalization, patients older than 80 years of age, or patients who were immunocompromised.

**Conclusion:** In newly admitted critically ill patients, measurement of plasma levels of soluble TREM-1 could help to rapidly identify those with infection.

*Ann Intern Med.* 2004;141:9-15.

www.annals.org

For author affiliations, see end of text.

Sepsis is a common cause of morbidity and death in intensive care units (1). Clinical and laboratory signs of systemic inflammation, including changes in body temperature, tachycardia, or leukocytosis, are neither sensitive nor specific enough for the diagnosis of sepsis. These signs can also be misleading because critically ill patients often present with the systemic inflammatory response syndrome but no infection (2–4). This issue is of paramount importance, since therapy and outcome differ greatly between patients with and those without sepsis. Moreover, the widespread use of antibiotics for all such patients is likely to increase antibiotic resistance, toxicity, and costs (5).

Thus, there is an as-yet-unsatisfied need for clinical or laboratory tools allowing clinicians to distinguish between the systemic inflammatory response syndrome and sepsis. Among the potentially useful markers of sepsis, procalcitonin has been suggested as the most promising (6–8). However, several investigators have questioned the diagnostic and prognostic accuracy of routine procalcitonin measurements, reporting inconsistent and variable results depending on the severity of illness and infection in the patient sample studied (9–11). The triggering receptor expressed on myeloid cells-1 (TREM-1) is a member of the immunoglobulin superfamily, and its expression is upregulated on phagocytic cells in the presence of bacteria or fungi (12). Several experiments by Bouchon and colleagues (13) showed that TREM-1 mediates the acute inflammatory response to microbial products. Human tissues infected

with bacteria are infiltrated with neutrophils and macrophages that express high levels of TREM-1. Conversely, TREM-1 is only weakly expressed in samples from patients with noninfectious inflammatory disorders (13). In addition, TREM-1 is shed from the membrane of activated phagocytes and can be found in a soluble form in body fluids. The presence of a soluble form of TREM-1 in samples of bronchoalveolar lavage fluid from mechanically ventilated patients has been shown to be a good indicator of infectious pneumonia (14). In this study, we prospectively investigated the diagnostic value of an assay measuring the plasma level of soluble TREM-1 in distinguishing sepsis from severe systemic noninfectious inflammation among newly admitted critically ill patients with suspected infection.

### METHODS

#### Study Sample

All consecutive patients who were newly hospitalized in the medical intensive care unit of a teaching hospital in France between July and September 2003 were prospectively enrolled in the study if they had clinically suspected infection and fulfilled at least 2 criteria of the systemic inflammatory response syndrome (Appendix Table 1) (15). Clinically suspected infection was defined as an explicit statement by the attending physician indicating suspicion of an ongoing infection. In all enrolled patients, diagnostic work-up was performed to identify or rule out

# Process in the Gibot case

- Annals published Gibot article in July 2004
- September 2004: Swiss author alleges plagiarism of 2 paragraphs in the Discussion.
  - Synonyms substituted for some words.
  - Two paragraphs became three.
- Annals editors confirm strong similarity of the text in the two articles.
- Autumn 2004: HS writes to dean at Universite' de Nancy:
  - “reply in 6 months, or Annals will publish a notice that explains its process in following up on the Swiss authors' complaint.”

## Process in the Gibot case

- Dean conducts investigation. Dr. Gibot acknowledges cut-and-paste plagiarism.
- Dean communicates findings to Swiss author.
- Apology from French senior author to Swiss author.
- Apology accepted by the Swiss
- Spring 2005: Dean write to HS
- Spring 2005: Annals publishes [Correction](#).

# Correction: Plasma Level of a Triggering Receptor Expressed on Myeloid Cells-1

The third, fourth, and fifth paragraphs in the Discussion of an article published in *Annals of Internal Medicine* [\(1\)](#) contain unattributed material similar to 2 paragraphs in the Discussion of another article [\(2\)](#). The authors of the *Annals* article have acknowledged their error, and the authors of the earlier article have accepted their apology. None of the text in question contains any factual errors.

# Follow-up

- Dr. Gibot's Annals article has been cited 153 times.
- He continues to publish actively in the critical care literature



# MORALS OF THE STORY

- In this instance of cut-and-paste plagiarism, the perpetrator was a fellow with weak English language skills (but he was experienced in writing for publication).
- The moral(s) of the story:
  - **Alert, interested readers have a role. Should they be reviewers?**
  - **Not all plagiarism ends badly for the perpetrators. Collegiality has its place, sometimes.**
  - **Senior authors, carefully check the work of your fellows. And maintain cordial relations with rivals.**

# CASE 3

# Serial plagiarism; the Kurjak case

- Late 1980's: systematic reviewers discovers plagiarized article by Kurjak.
  - Half of content, some data identical to earlier article
  - Kurjak's co-author knew nothing of the article.
- Ian notifies journal, school dean, WHO.
  - WHO suspends K, director of a WHO center.
  - **Journal editor did not retract**
  - **Dean begged for discrete handling**
- 2001: K publishes book chapter with cribbed material.
  - Book republished without chapter.
  - **No action by University of Zagreb.**

Chalmers I. **Role of systematic reviews in detecting plagiarism: case of Asim Kurjak.**  
**BMJ.** 2006;333:594-596

## MORALS OF CASE 3

- Journal editors must retract fraudulent articles.....but often don't.
- Academic leaders must realize that failure to investigate alleged misdoing brings greater institutional shame than acting and accepting the consequences.
- Fraud is often serial, another reason to “check every article.”

# Plagiarism at other journals

- NEJM: 1 episode of publishing plagiarized material in ~70,000 submissions on Jeff Drazen's watch.
- Author (Barbaro) published faked data.
  - He had turned a published image upside down, backward, and converted it to black and white.
  - Author of original article spotted the fraud and had the original slide.
  - Author refused to retract, so Jeff retracted.
- Follow-up to check other Barbaro articles: unknown to me

## Incidence of published, proven plagiarism is probably low

- Annals: one case in ~24,000 submissions
- NEJM: one case in ~70,000 submissions
- Common feature: the author whose work was plagiarized spotted the fraud.
- **What is the incidence of unpublished or undetected plagiarism?**

# Measuring the prevalence of plagiarism in science

- Do a survey of scientists
- Use software that compares a manuscript to articles in the published literature to see if they contain similar strings of words.
  - Decide which strings of words are plagiarism (true positives) and which are not (false-positives).

# Prevalence of plagiarists among scientists

- Anonymous survey of 3247 American biomedical researchers
- Findings
  - Plagiarism: 1.4%
  - Duplicate publication: 4.7%

Matinson BC et al. Nature. 2005;435:737-8 (described in Rifai N, Clinical Chemistry 2008;54:777-78) .



# Systematic review of surveys about scientific misconduct

- Misconduct = fabricate, falsify, or modify data  $\geq 1x$
- 21 surveys in SR; 18 in meta-analysis.
- Result: 1.97% (N=7) admitted misconduct
- More common in medical/pharm researchers
- **Excluded questions about plagiarism.**
- Likely an underestimate

# Prevalence of duplicate publication

- Text-matching software
- 280,000 references in arXiv, an open access data base of publications in **math, physics, biology, statistics, and computer science**
- Results:
  - 0.2% suspected of plagiarism
  - 10.5% suspected of duplicate publication (i.e., by the same authors).
- Caveat: no investigation of suspect articles.

# Duplicate publication 1975-2008

	Articles published	Suspected duplicates	True duplicates
NEJM	11,779	14	0
Clin Chem	8,867	27	1 (?2others)*
Lancet	68,948 (incl. letters)	24	4 (?4 others)**

Authors used text similarity software and on-line databases of publications. They manually verified suspected duplicates.

- One in foreign language; in the other, the original article was not retrievable
- \*\* original in a foreign language.

Rifai et al. Clinical Chemistry. 2008;54:777-78.

# The Problem

- Scientific misconduct is quite common
- Purging the literature of published fraudulent work is a huge task.
- Prevention is far better than cleaning up the mess....but how?
  - Inculcate scientists with a sense of high moral purpose.
  - Detect misconduct before it gets into the literature and refuse to publish it.
  - **What is the yield of pre-publication screening?**

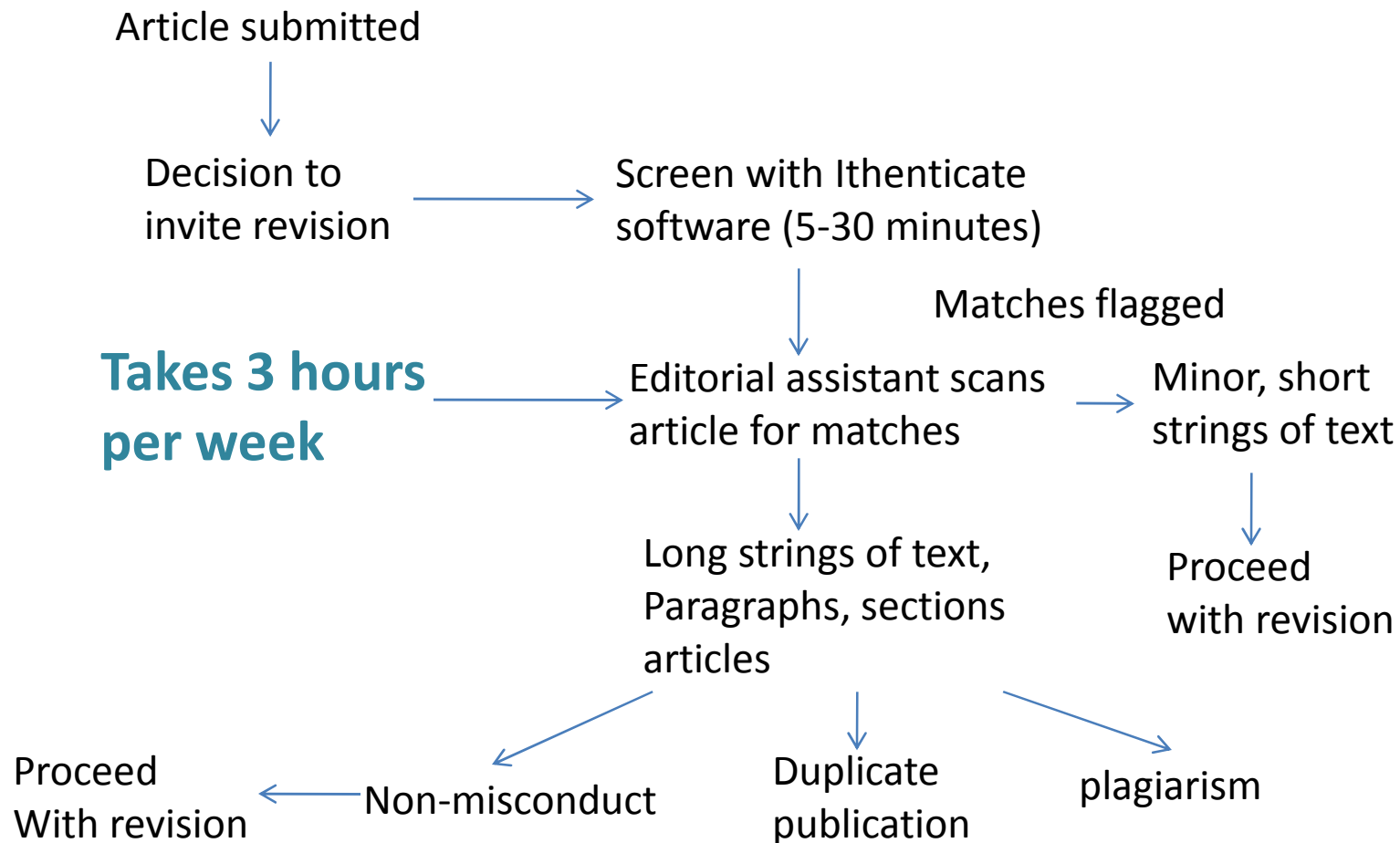
# Pre-publication detection of plagiarism

- Thanks to Jeff Drazen and Steve Morrissey for providing an account of the NEJM experience.
- Steve and his NEJM colleague, Sarah Fishkin, participated in an evaluation of Ithenticate™
- They liked it and signed up to include it in their manuscript management software package.

# Electronic detection of duplicated text

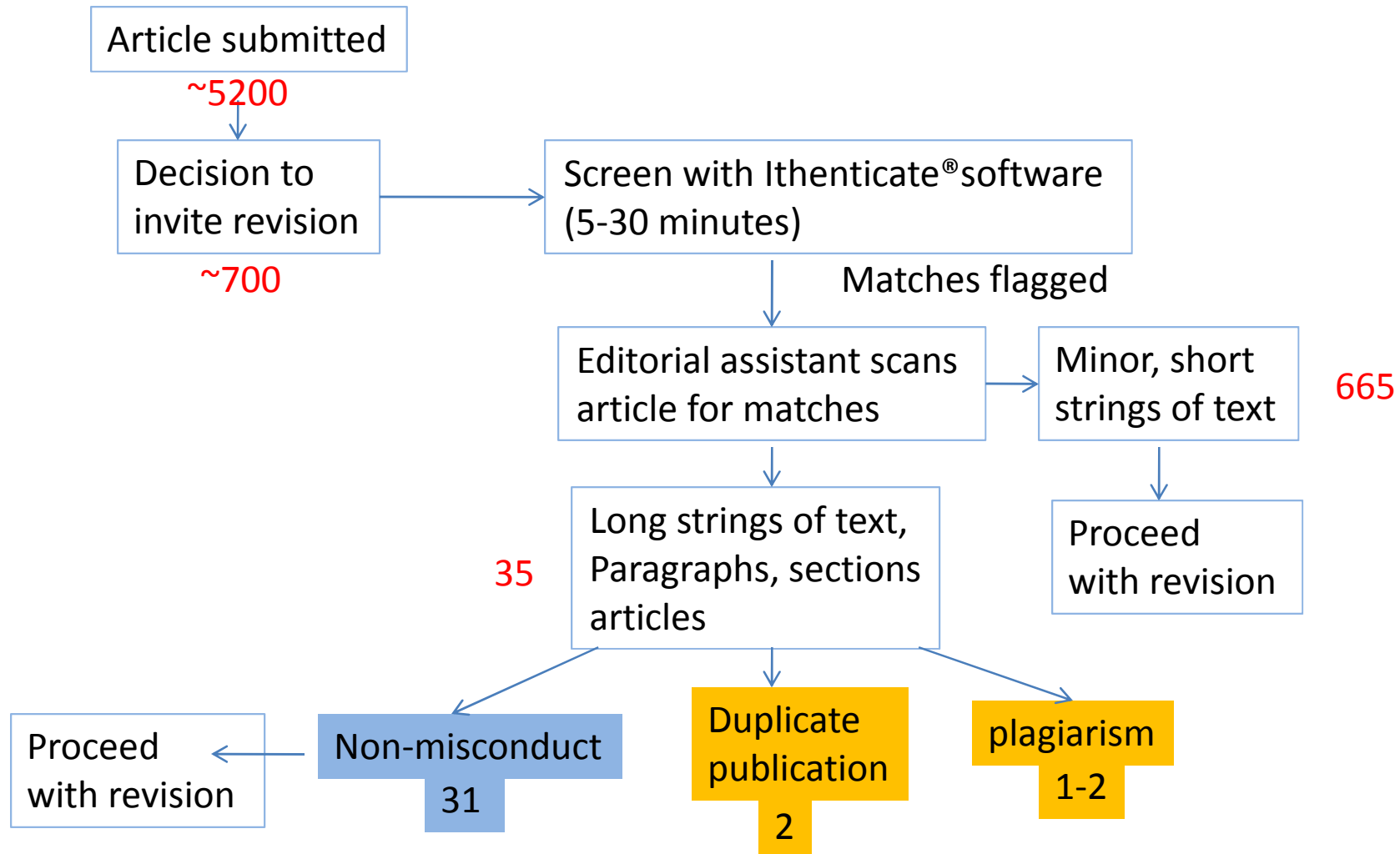
- Terminology:
  - Test manuscript
  - Source articles.
- Software detects source article text that is similar to test manuscript text.
- Marks it in the test manuscript.
- Other speakers will provide details.

# Screening for plagiarism at the NEJM



Source: Stephen Morrissey and Sarah Fishkin of NEJM, personal communication

# Screening for plagiarism at the NEJM



Source: Stephen Morrissey and Sarah Fishkin of NEJM, personal communication



# Plagiarism is common at the NEJM!

- First series (to my knowledge) of unpublished manuscripts examined for plagiarism.
- **Plagiarism:** 1 or 2 cases per 700 examined MS.
- **Duplicate publication:** 2 cases/700
- Caveat: Sample limited to articles the journal was interested in publishing.
- 3 hours per week to screen ~20 manuscripts

# Conclusions

- Scientific misconduct is quite common , and it should trigger an investigation to clean up the scientific literature. Better to prevent than to clean up.
- Publication of faked data and other people's ideas is preventable...by detecting it and not publishing it.
- **Should screening for plagiarism be routine journal practice?**

# Conclusions

- Software that detects similar word sequences in supposedly unrelated articles should be sensitive but is not specific.
  - Its main function is to mark places in the text for a trained editor to evaluate.
  - Theoretically, an unaided human is not sensitive but is highly specific.
  - Theoretically, the combination of these two tests should be very effective at detecting kidnapped prose.

# Conclusions

- Can the software detect ideas that are similar but expressed differently?
  - Can it detect the most insidious form of plagiarism: the piracy of ideas?
- Should we care? Is this a “no harm, no foul” situation?
  - The date of publication usually settles issues of priority.
  - People who publish later fulfill the same key function as scientists, who, by replicating new findings, validate them.

# Questions for the future

- Is plagiarizing data/images a greater offense than plagiarizing text?
- Should journals require authors to attest that they have searched for duplicated text before submission?
- Screen all submissions or just ones you want to publish?
- What should be the threshold for asking local authorities to investigate alleged plagiarism?
- What is the threshold first offense to trigger checking all of an author's publications?