
Classical Studies

— Analyzing Misinformation — in Research Studies

By: Raghu Srinivasan, Jonathan
Alexander, Henry Wang

Problem

- Analyze misinformation in scientific studies
 - **Misinformation in research:** Published studies that are the product of misconduct including fabrication of results, invalid scientific method, manipulation of data, invalid conclusion
- Problem
 - Researchers unknowingly cite retracted articles
 - Media outlets make claims based on flawed research
- Identify patterns between valid and invalid research papers
- Difficulties
 - Creating a workflow to gather data
 - Determining which features are relevant and worth exploring

Current Practice

- Linguistic analysis for obfuscation^[1]
- Analysis of journal instead of individual article^{[2][5]}
- Statistical analysis for detecting fabrication of data^[3]
- Style-based analysis of fake news in general^[4]
- Limitations:
 - Trusting the journal itself
 - Trusting the reason for retraction in the retraction notice
 - Mainly relying on human confirmation

Novelty

- Comparing linguistic/style-based analysis of fraudulent/non-fraudulent studies over multiple datasets
- Include other metrics:
 - Number of citations
 - Number of review board members (or lack of review board)
 - Lack of information on review staff

Stakeholders

- Editors of scientific journals
- Scientists conducting research
- People who make decisions based off studies
- Media outlets writing stories based on scientific studies

Impact

- More transparency in scientific studies
- More accurate information being supplied to the public
- Better decision making through access to more reliable information
- Reduction in fraudulent studies and predatory journals
- More meaningful scientific advancement
- Success can be measured by identifying differences between retracted articles and non-retracted articles

Challenges and Payoffs

Challenges:

- Difficulty of multi-feature analysis
- Non-standardization of research papers

Payoffs:

- Be able to identify the key features and distinctions that separate fraudulent from non-fraudulent papers
- Be able to train a classifier based off of these distinctions

Deliverables and Cost

Midterm (4/1):

- Acquisition of datasets (PubMed and Retraction Watch)
- Refined set of metrics

Final (5/2):

- Measurements of desired metrics
- Analysis and interpretation of data

Plan of Activities

- Data preparation
 - Acquire data using PubMed database
 - Format and clean data
- Explore data
 - Analyze available attributes
- Determine features to use for analysis
- Data analysis
 - Statistical analysis
- Draw conclusions
- Final report

Preliminary Data Analysis

TSHZ3 and SOX9 Regulate the Timing of Smooth Muscle Cell Differentiation in the Ureter by Reducing Myocardin Activity (BLS) Anatomy/Physiology; (BLS) Biochemistry; (BLS) Biology - Cellular; <i>PLoS One</i> --- <i>PLoS</i>	+Concerns/Issues About Data +Concerns/Issues About Image
Aix-Marseille Université, CNRS, IBDM UMR7288, Marseille, France Institut für Molekularbiologie, Medizinische Hochschule Hannover, Hannover, Germany	
Bone Mesenchymal Stem Cell-Conditioned Medium Regulates the Differentiation of Neural Stem Cells Via Notch Pathway Activation (BLS) Anatomy/Physiology; (BLS) Biochemistry; (BLS) Biology - Cellular; <i>Cellular Reprogramming</i> --- <i>Mary Ann Liebert</i>	+Concerns/Issues About Data +Concerns/Issues About Results
Department of Orthopedics, The First Affiliated Hospital, Anhui Medical University, Hefei, China	
Organocatalytic stereoselective synthesis of passifloricin A (PHY) Chemistry; (PHY) Crystallography/Spectroscopy; <i>Organic & Biomolecular Chemistry</i> --- <i>Royal Society of Chemistry (RSC)</i>	+Concerns/Issues About Data +Concerns/Issues About Image
Division of Organic Chemistry, National Chemical Laboratory (CSIR), Pune, India Chemistry Department, University of Pune, Pune, India http://retractionwatch.com/2017/09/15/chem-journal-cautions-readers-data-three-papers/	+Investigation by Company/Institution
Organocatalytic stereoselective synthesis of passifloricin A (PHY) Chemistry; (PHY) Crystallography/Spectroscopy; <i>Organic & Biomolecular Chemistry</i> --- <i>Royal Society of Chemistry (RSC)</i>	+Concerns/Issues About Data +Concerns/Issues About Image
Division of Organic Chemistry, National Chemical Laboratory (CSIR), Pune, India Chemistry Department, University of Pune, Pune, India http://retractionwatch.com/2017/09/15/chem-journal-cautions-readers-data-three-papers/	+Investigation by Company/Institution
The mitochondrial genome of Pomacea maculata (Gastropoda: Ampullariidae) (BLS) Genetics; (BLS) Zoology; <i>Mitochondrial DNA Part A</i> --- <i>Taylor and Francis</i>	+Concerns/Issues About Data +Concerns/Issues About Results
Zhejiang Provincial Key Laboratory of Biometrology and Inspection & Quarantine, College of Life Science, China Jiliang University, Hangzhou, China Department of Entomology, China Agricultural University, Beijing, China Markey Cancer Center, University of Kentucky, Lexington, KY, USA	

```

<pub-date pub-type="collection">
  <year>2000</year>
</pub-date>
<pub-date pub-type="epub">
  <day>19</day>
  <month>7</month>
  <year>2000</year>
</pub-date>
<volume>1</volume>
<fpage>1</fpage>
<lpage>1</lpage>
<ext-link ext-link-type="uri" xlink:href="http://www.biomedcentral.com/1471-2091/1/1"/>
</history>
<date date-type="received">
  <day>8</day>
  <month>6</month>
  <year>2000</year>
</date>
<date date-type="accepted">
  <day>19</day>
  <month>7</month>
  <year>2000</year>
</date>
</history>
<permissions>
  <copyright-statement>
    Copyright © 2000 Gaillard and Strauss; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this notice is preserved along with the article's original URL.
  </copyright-statement>
  <copyright-year>2000</copyright-year>
  <copyright-holder>
    Gaillard and Strauss; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this notice is preserved along with the article's original URL.
  </copyright-holder>
</permissions>
<abstract>
  <sec>
    <title>Background</title>
    <p>
      Alternative DNA conformations are of particular interest as potential signals to mark important sites on the genome. These are repetitive poly(CA) · poly(TG) DNA sequences spread in all eukaryotic genomes as tracts of DNA. The structure of poly(CA) · poly(TG) can vary markedly from the classical right handed DNA double helix and adopt formation and the structure of an alternative DNA structure, named Form X, which was observed previously by poly(CA) microsatellite poly(CA) · poly(TG) but had not yet been characterized.
    </p>
  </sec>
</abstract>

```

Retractiondatabase.org (left), Pubmed study data (right)

References

- [1] <http://journals.sagepub.com/doi/abs/10.1177/0261927X15614605>
- [2] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5493175/>
- [3] <https://www.sciencedirect.com/science/article/pii/S1743919106000471>
- [4] <https://arxiv.org/pdf/1812.00315.pdf>
- [5] https://link.springer.com/chapter/10.1007%2F978-3-030-03402-3_35