

# Groundbreaking text mining project highlights 'gender gap' in scientific research

March 3 2016

---

A project at The University of Manchester to analyse 15,000 mouse studies - the largest of its kind ever undertaken - has revealed that about half of these studies failed to report the sex and age of the mice involved, despite these being recognised as key variables that can affect the outcome of scientific studies. The project utilised text mining software developed at the University, which can analyse large collections of documents to unearth information which would otherwise have been virtually impossible to discover. The software relies on a number of rules, which automatically scan the method section of papers to identify mentions of gender and age.

The results of the project, published this week in the eLife journal, highlight the issue of reproducibility of [scientific research](#) - around £20 billion is spent every year on research which is not reproducible, and over 80% of potential therapeutics fail in humans after being tested in [mice](#). Previously published studies have suggested that research done on female animals may not be applicable for men, and in many of the studies analysed in this project, the animals used were overwhelmingly female. This may be due to female mice being less aggressive, which makes them easier to use in the studies. This is important, because the sexes can have markedly different responses to the same investigations - for example, in infection research. This may significantly reduce the reliability of studies, and lead to drugs that won't work for half of the population.

The reproducibility of studies often focuses on the interpretation of

statistics, but this project has highlighted that the methods used may not be reported rigorously enough to assess whether they were done correctly. By looking at the methods used, it is possible to infer whether or not the statistics produced are sound, and reproducible in the future. Without knowing these methods, this cannot be inferred at all, which hampers cross-disciplinary research and longevity of data.

The project has produced a vital tool to measure the reproducibility of scientific studies, but there is a long way to go - failure to consider gender in research is still very much the norm, and according to one analysis of scientific studies published in 2009, only 45% of animal studies involving depression or anxiety and only 38% involving strokes used females, even though these conditions are more common in women.

"The opportunity to use text mining to cover such a broad portfolio of research was brilliant, and vital to see the bigger picture," said Sheena Cruickshank, Senior Lecturer in Immunology at The University of Manchester. "We are an interdisciplinary team, and it was this which enabled us to spot this issue and then explore it. The paper builds on several pieces of work we have done together, and highlights the importance of the scientific community to come together and define what is important in the current reproducibility crisis."

"This study has demonstrated how state-of-the-art computer science technology is instrumental for a large-scale and systematic analysis of literature," said Dr Goran Nenadic from The University of Manchester's School of Computer Science. "It avoids small sample bias, and allows us to explore the research landscape on a large scale to identify key issues in reporting details of scientific methodologies, which are necessary for reproducibility, transparency and fidelity of research."

Provided by University of Manchester

Citation: Groundbreaking text mining project highlights 'gender gap' in scientific research (2016, March 3) retrieved 24 April 2024 from <https://phys.org/news/2016-03-groundbreaking-text-highlights-gender-gap.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.