Revolutions

Learn more about using open source R for big data analysis, predictive modeling, data science and more from the staff of Revolution Analytics.

September 12, 2011

Why you should care about reproducible research

This week's Economist has an in-depth article on the consequences of failures reproducible research, adding more detail to the report in the New York Times in July. Errors in data analysis by researchers at Duke University led to patients in clinical trials being assigned the wrong drug:

> Dr Potti and his colleagues had mislabelled the cell lines they used to derive their chemotherapy prediction model, describing those that were sensitive as resistant, and *vice versa*. This meant that even if the predictive method the team at Duke were describing did work, which Dr Baggerly and Dr Coombes now seriously doubted, patients whose doctors relied on this paper would end up being given a drug they were less likely to benefit from instead of more likely.

One startling aspect of this saga is that the story only came to prominent attention after revelations that Dr Potti falsely claimed to be a Rhodes Scholar on his resume, despite repeated attempts by Baggerly (who presented the problems at the R BioConductor conference more than 2 years ago) to raise awareness of this serious problem.

> [Baggerly] noted that in addition to a lack of unfettered access to the computer code and consistent raw data on which the work was based, journals that had readily published Dr Potti's papers were reluctant to publish his letters critical of the work. *Nature Medicine* published one letter, with a rebuttal from the team at Duke, but rejected further comments when problems continued. Other journals that had carried subsequent high-profile papers from Dr Potti behaved in similar ways. ... Eventually, the two researchers resorted to publishing their criticisms in a statistical journal, which would be unlikely to reach the same audience as a medical journal.

Said Dr Baggerly: "I find it ironic that we have been yelling for three years about the science, which has the potential to be very damaging to patients, but that was not what has started things rolling."

The Economist: An array of errors

Posted by David Smith at 16:13 in life sciences, R, statistics | Permalink

TrackBack

TrackBack URL for this entry:
http://www.typepad.com/services/trackback/6a010534b1db25970b014e8b7fc35f970d
Listed below are links to weblogs that reference Why you should care about reproducible research:

Comments

You can follow this conversation by subscribing to the comment feed for this post.

Wow, that really is very scary indeed. But how can things like this be prevented without micromanagement of every lab researcher?

Posted by: Heather Devlin | September 13, 2011 at 05:05

Of course there is the opposite problem as well. Gelmen did a post on Chinese papers reproducing results that were too good.

http://andrewgelman.com/2011/05/suspicious_patt/

Posted by: kirk mettler | September 13, 2011 at 06:40