Influential Reinhart-Rogoff economics paper suffers spreadsheet error

April showers bring … database errors?

The other day, we wrote about two retractions in the Journal of the American College of Cardiology, and another in the American Heart Journal, stemming from database errors.

Seems to be catching.

The Economist (among other outlets) is reporting about a similar mother database glitch — not, we'll admit, a retraction — involving a landmark 2010 paper by a pair of highly influential economists. The controversial article, "Growth in a Time of Debt," by Harvard scholars Carmen Reinhart and Kenneth Rogoff, argued that countries that took on debt in excess of 90% of their gross domestic product suffered sharp drags on economic growth. That evidence became grist for the austerity mill, including Paul Ryan.

Turns out, that conclusion was based in part on an Excel error. As the business press explained, a trio of researchers at the University of Massachusetts, Thomas Herndon, Michael Ash and Robert Pollin, found that the Reinhart/Rogoff analysis had excluded a handful of critical data points by basically lopping them off the spreadsheet. The result: Their claims about the deleterious effects of debt on growth are somewhere — indeed, substantially, in a way — overstated.

As the U. Mass economists note in their rebuttal paper, "Does High Public Debt Consistently Stifle Economic Growth? A Critique of Reinhart and Rogoff":

> *Herndon, Ash and Pollin replicate Reinhart and Rogoff and find that coding errors, selective exclusion of available data, and unconventional weighting of summary statistics lead to serious errors that inaccurately represent the relationship between public debt and GDP growth among 20 advanced economies in the post-war period. They find that when properly calculated, the average real GDP growth rate for countries carrying a public-debt-to-GDP ratio of over 90 percent is actually 2.2 percent, not -0.1 percent as published in Reinhart and Rogoff. That is, contrary to RR, average GDP growth at public debt/GDP ratios over 90 percent is not dramatically different than when debt/GDP ratios are lower.*
>
> *The authors also show how the relationship between public debt and GDP growth varies significantly by time period and country. Overall, the evidence we review contradicts Reinhart and Rogoff's claim to have identified an important stylized fact, that public debt loads greater than 90 percent of GDP consistently reduce GDP growth.*

In the interest of full disclosure, we'll also quote the two corrections to this paper:

> *(1) The notes to Table 3: "Spreadsheet refers to the spreadsheet error that excluded Australia, Austria, Belgium, Canada, and Denmark from the analysis." is corrected to read: "Spreadsheet refers to the spreadsheet error that excluded Australia, Austria, Belgium, Canada, and Denmark from the analysis."*
>
> *(2) Page 13: "Thus, in the highest, above-90 percent public debt/GDP, GDP growth of 4.1 percent per year in the 1950–2009 sample declines to only 2.5 percent per year in the 1980–2009 sample" is corrected to read "Thus, in the highest, above-90 percent public debt/GDP category, GDP growth of 4.1 percent per year in the 1950–2009 sample declines to only 2.5 percent per year in the 1980–2009 sample."*

Reinhart and Rogoff, for their part, have acknowledged the error:

> *We literally just received this draft comment, and will review it in due course. On a cursory look, it seems that Herndon Ash and Pollin also find lower growth when debt is over 90% (they find 0-30=4.2% growth; 30-60, 3.1%; 60-90, 3.2%; 90-120, 2.4%; and over 120, 1.6%). These results are, in fact, of a similar order of magnitude to the detailed country by country results we present in table 1 of the AER paper, and to the median results in Figure 2. And it is in much of a large and growing literature, including our Journal of Economic Perspectives paper (joint with Vincent Reinhart)… However, these strong similarities are not what these authors choose to emphasize.*
>
> *The 2012 JEP paper largely anticipates and addresses any concerns about aggregation (the main bone of contention here). The JEP paper not only provides individual country averages (as we already featured in table 1 of the 2010 AER paper) but it goes further and provide episode by episode averages. Not surprisingly, the results are broadly similar to our original 2010 AER table 1 and averages and to the median results that also figure prominently. It is hard to see how one can interpret these tables and individual country results as showing that public debt overhang over 90% is clearly benign.*
>
> *The JEP paper with Vincent Reinhart looks at all public debt overhang episodes for advanced countries in our database, dating back to 1800. The overall results of the paper is that public debt overhang episodes (over 90% or more are associated with 1.2 percent lower growth as compared to growth when debt is under 90%. (We also include in our tables the small number of shorter episodes.) Note that because the historical public debt overhang episodes last an average of over 20 years, the cumulative effects of small growth differences are large. It is utterly misleading to speak of a 1% growth differential that lasts 10-25 years as small.*
>
> *By the way, we are very careful in all our papers to speak of "association" and not "causality" since like THIS TIME IS DIFFERENT showed that debt explodes in the immediate aftermath of financial crises. This is why we restrict attention to longer debt overhang periods in the JEP paper, though as noted there are number of short ones. Moreover, we have generally emphasized the 1% differential median result in all our discussions and subsequent writing, precisely to be understated and cautious, and also in recognition of the results in our core Table 1 (AER paper).*
>
> *Lastly, our 2012 JEP paper cites the papers from the BIS, IMF and OECD (among others) which virtually all find very similar conclusions to original findings, albeit with many different nuances of differences in threshold, and many nuances of alternative interpretation. These later papers, by the way, use a variety of methodologies for dealing with non-linearity and also for trying to determine causation. Of course much further research is needed as the data we developed and is being used in these studies is new. Nevertheless, the weight of the evidence to date —including this latest comment — seems entirely consistent with our original interpretation of the data in our 2010 AER paper.*
>
> *Carmen Reinhart and Kenneth Rogoff*
> *April 16, 2013*

1945-2009				
	RR (2010)		HAP (2013)	
Debt/GDP	Mean	Median	Mean	Median
0 to 30	4.1	4.2	4.2	NA
30 to 60	2.8	3.9	3.1	NA
60 to 90	2.8	2.9	3.2	NA
Above 90	-0.1	1.6	2.2	NA
1800-2011				
1800-2011				
0 to 30	3.7	3.9	NA	NA
30 to 60	3.0	3.1	NA	NA
60 to 90	3.4	2.8	NA	NA
Above 90	1.7	1.9	NA	NA
RRR (2012), 1800-2011				
	Mean			
Below 90	3.5			
Above 90	2.4			

Of course, this isn't a retraction, at least not yet. And as we've noted, retractions are rare in economics. Still, a number of people have sent us tips about this paper, so we thought it was worth a post.

Update, 10:15 a.m. Eastern, 4/19/13: Word in headline corrected from "database" to "spreadsheet," as per comment below from Neil Saunders.

SHARE THIS:

Email Bluesky LinkedIn Facebook X

RELATED

Error-laden database kills paper on extinction patterns
December 27, 2016
In "australia retractions"

Pfizer database errors cause two voluminous retractions for JACC statin-biomarker papers
April 16, 2013
In "australia retractions"

Errors in govt database prompt authors to retract and replace paper in JAMA journal
October 3, 2017
In "doing the right thing"

April 18, 2013 · Adam Marcus · economics, investigator error

39 thoughts on "Influential Reinhart-Rogoff economics paper suffers spreadsheet error"

nsaunders
April 18, 2013 at 8:08 pm

I'm not seeing the database here. I hope you are not referring to Excel as a database. It most emphatically is not.

REPLY

Otto
April 19, 2013 at 12:17 am

This is basically the whole problem: Excel *invites* the mixing of data and methods and then obscures the methods on top of it.

REPLY

littlegreyrabbit
April 19, 2013 at 1:13 am

Of course Excel can be used to hold and manage a simple database and it supplies a perfectly adequate set of tools to answer many questions. I suppose the problem with Excel is any ordinary Joe can use it.

The problem with this paper was not Excel but whether the authors were knowingly manipulating the data for political ends. There has been some discussion elsewhere how only 1 year of data from NZ was used – that of 1951 – because it showed -7.8% growth. But actually 1951 was the start of the great Wool Boom in the New Zealand economy where sky-high wool prices flowing out of the Korean War had the New Zealand economy anchored in #4 position in the OECD GDP/capita rankings for the next decade.
The only reason that it showed negative growth that year was because for 150 days the ports were paralysed by the Waterfront strike. The Wages Arbitration Board had ordered a general 15% wage rise for New Zealand workers. However the warfies were not covered by that decision and the port owners offered them a 9% rise, the warfies responded by refusing to work overtime and the port authorities declared that this was a strike and locked the workers out. Since New Zealand was an export orientated country 150 days of little maritime traffic in 1951 was going to deliver a strong blow to the accounts.
So the large retraction in GDP that year had nothing to do with debt to GDP ratio and everything to do with a strike largely arising out of how to distribute out the rising prosperity – the rest of the decade the economy boomed.

REPLY

StatObserver
April 19, 2013 at 7:54 pm

I really disagree. For calculating something quick and dirty, OK. But for science? No way. Excel sucks badly. It is not auditable. This is the key – how can you inspect the code which is used to perform a task?

Excel should NEVER be used for science, and products like GraphPad Prism should be removed from the scientific workbench.

REPLY

StatObserver
April 19, 2013 at 7:56 pm

I will say, as a statistician, that I work with Excel spreadsheets every day. For better or worse, they are used to move data around for small projects, and I work with a lot of those. I tell people how to send me the data, and can read these in. I will NOT perform calculations using Excel tools within the spreadsheet. This is both unethical and dangerous.

No scientist should ever perform calculations within Excel. You cannot audit or check these.

REPLY

Stewart
April 20, 2013 at 3:03 am

StatObserver,

Whilst I agree we ought to be rather careful with statistical analysis there are several programs open to most if not all University Researchers such as GraphPad or SPSS perfectly suited, but just as easily misused as Excel.

In all my days using Excel I never mixed up data-sets or moved the wrong column here or there. Those who do 'errors' in Excel will no doubt do 'errors' in GraphPad and SPSS.

I don't see using excel or any other software dangerous or unethical. PIs should always have access to the raw data and check the appropriate statistical analysis was applied.

REPLY

StatObserver
April 20, 2013 at 9:49 am

GraphPad is not an acceptable tool for a lot of statistical analysis. For single observation studies, possibly. With repeated measures data, it is 30+ years out of date. Totally unacceptable, and it is not in compliance with reproducible research standards.

Stewart
April 20, 2013 at 11:41 am

'GraphPad is not an acceptable tool for a lot of statistical analysis. For single observation studies, possibly. With repeated measures data, it is 30+ years out of date. Totally unacceptable, and it is not in compliance with reproducible research standards' – Statsobserver

Can you explain why it is unacceptable?

StatObserver
April 22, 2013 at 2:24 pm

GraphPad uses a model for repeated measures analysis which was shown, in 1980, to be unacceptable. You can easily see this by their discussion of compound symmetry. No one today who uses correct models discusses compound symmetry. This is only a requirement if you have no other approach.

GraphPad is not acceptable for analysis of repeated measures data when more than 2 measures are used. I would accept the results for 2 measures.

amarcus41
April 19, 2013 at 10:05 am

Neil – Thanks for your comment. I suppose the term "spreadsheet" would have been more appropriate, so we edited the headline and added a note at the bottom.
Best,
Adam

REPLY

R. Grant Steen
April 18, 2013 at 10:51 pm

This post was a bit hard to follow, though I don't fault RW for that. The response by Reinhardt and Rogoff (RR) seems rather blinkered, in the sense that they highlight the similarities between themselves and HAP, and elide the only difference that makes a difference. Looking at the means in the table showing growth as a function of debt/GDP ratio, the only place where's much difference between RR and HAP is when the debt ratio exceeds 90%. This is precisely what got the original notice for the RR paper because it would seem to argue for austerity. And yet their response doesn't really talk about this difference; instead, RR focus on the broad similarities. Yet no one cares about these similarities….

If there is no tipping point in growth, there is no story.

REPLY

StatObserver
April 19, 2013 at 7:57 pm

It's really stunning that anyone pays any attention to anything generated using Excel spreadsheet calculations.

REPLY

Marco
April 20, 2013 at 3:43 am

Why? There's plenty of stuff for which Excel is perfectly appropriate.

REPLY

StatObserver
April 20, 2013 at 9:50 am

Yes, I'm sure that's true, but none of those should result in publication by a journal or quoting as science.

REPLY

Marco
April 20, 2013 at 12:09 pm

I wanted to use General McAuliffe's reply here, but I'm afraid I would be censored.

REPLY

R. Grant Steen
April 20, 2013 at 5:26 pm

This is silly…. You can analyze your results with an abacus and have good science at the end, if you're careful enough.

I think we've missed the point here. It doesn't matter how the results were analyzed. It matters that the results were wrong and policy decisions were based on wrong answers.

Austerity is a sham, if the new results are reliable.

REPLY

StatObserver
April 22, 2013 at 2:26 pm

You cannot do reproducible research with an abacus. I would never accept such a research tool.

The point is that reproducible methods, in which the code to perform the analysis is tracable and auditable, cannot be done using Excel.

REPLY

R. Grant Steen
April 22, 2013 at 3:23 pm

The idea that scientific research can only be done with computer software blessed by statisticians is totally ridiculous. Good science was done long before there were either statisticians or software packages.

REPLY

StatObserver
April 23, 2013 at 10:01 am

Read about Anil Potti. Read many of the contributions and issues here. The problems are often 1) inability by non-specialists to know what the heck they were doing and 2) manipulating images. These are both exactly the same fundamental problem – the ability of trashy interactive software to do things with data (numbers, pictures) that no one can keep track of. It's all reproducibility problems. You can say whatever you want, but if you are using an interactive program (abacus, GraphPad, Paint, graphical software) to produce something, you are not doing science.

R. Grant Steen
April 23, 2013 at 10:18 am

Cheaters cheat. Software can't prevent that and an audit trail provides no guarantees that the underlying data are legitimate. I believe that Jon Sudbø used SAS for his analyses in Lancet, NEJM, and JCO, but he made up data for about 250 patients.

Good scientists can do good science with an abacus; fraudulent scientists can commit fraud with SAS.

StatObserver
April 24, 2013 at 7:32 pm

These days, the SHEER SIZE of data makes even SAS incapable of handling it. And this site is filled with cases that tell us that you are wrong.

StatObserver
April 24, 2013 at 7:34 pm

It's really amazing. With commercial enterprises, what stops them from cheating? Easy – audits.

We will have audits in science. It's started already. And for such studies, Excel will put you in science jail.

Claudio
April 29, 2013 at 2:45 pm

I strongly desagree with StatObserver. You can replicate anything that someone else did at Excel as long as you have the database. That's exactly what happens with RR. Someone tried to replicate their results and found that they were wrong. Of course, it is easier to read a code, but you can replicate anything someone else did at the abacus (or replicate the results using another software, what's the problem? You should have the same results).

CarolynS
April 20, 2013 at 1:44 am

In the early 2000s, there was a big kerfluffle involving the estimates of deaths due to obesity in a paper put out by the Centers for Disease Control. That paper also turned out to have errors that were due to using Excel for the analysis and making mistakes in data entry (covered up as a "software glitch" by the agency) and to have vastly overestimated the numbers.

REPLY

Otto
April 22, 2013 at 6:06 am

"That paper also turned out to have errors that were due to using Excel for the analysis and making mistakes in data entry (covered up as a "software glitch" by the agency)"

Perhaps you could illustrate where either the Excel or "covered up" parts of this statement are demonstrated. They're certainly not claimed in the Mokdad et al. erratum (PMID 15657315). The original (PMID 15010446) plainly states that "[w]e used SAS (version 6.2, SAS Institute Inc, Cary, NC) and SUDAAN (version 8.0, Research Triangle Institute, Research Triangle Park, NC) statistical software."

REPLY

Otto
April 22, 2013 at 6:26 am

I should also note, with respect to the "vastly overstated" part, the Mokdad et al. correction was from ~400,000 to 365,000. The later estimate of 112,000 deaths in 2000 (which for some reason seems to get "revised" to ~24,000 by the likes of the Washington Times) was in Flegal et al. (PMID 15840860) and used a completely different data set that, at first glance, allowed for better control of confounders.

REPLY

CarolynS
April 22, 2013 at 4:02 pm

I have seen the original Excel spreadsheet used and I know what the data entry problem was. Excel was used for the calculation of deaths due to obesity. http://www.foxnews.com/story/0,2933,144802,00.html describes CDC referring to this as a computer software error. The correction is on the order of 80,000 deaths. If you go back and look at the Mokdad paper, they averaged their value with a previous value to arrive at their final number of 365,000, but the averaging halves the size of their error. (Suppose I have two numbers 8 and 4 and average them to get 6. Now suppose 8 is wrong and it should be 7 so my error is 1. Now I average 7 and 4 together and get 6.5. It looks like my error is 0.5,but it was really 1.0).

REPLY

Otto
April 23, 2013 at 6:29 am

Are you relying on an AP report to support the notion that the error, which was described in a published correction, was "covered up as a 'software glitch' by the agency"? How much detail do you really expect Gerberding (who was the senior author) to provide in a brief quote to a reporter? I don't know what halving "the size of the error" means. These are *uncertainties*; they propagate in quadrature. If you have 8 ± 1 and 4 ± 1, the mean for a normal distribution is 6 ± 0.7.

Can you post the "original Excel spreadsheet used?"

REPLY

CarolynS
April 23, 2013 at 7:41 pm

The initial error was ~80,000 deaths, but the method of averaging with another number (which did not change, because it was from another paper) make it look like the error was ~40,000 deaths. There is no statistical uncertainty – it's just a calculation error that resulted from incorrect data entry. The only thing I have ever seen or heard from CDC was that it was a "computer software error." The economists have been more forthright. However I don't think further discussion of this issue is likely to be productive.

Otto
April 23, 2013 at 6:38 am

OK, now I get it: the objection is that the correction of the 2000 input estimate was from 494,921 to 414,423. But this *wasn't the output*; the same averaging was used in both cases to arrive at the final result, which is what was ultimately being corrected.

Eka-silicon
April 21, 2013 at 12:12 am

Any 'analysis' via Excel is no longer a valid piece of research, end stop. No audit trail, no script, it's nonsense. I've seen multiple cases of so-called Excel 'databases' completely hosed after someone sorted random rows. One of these required re-entering, by hand, 6+ years of clinical data. This massive error was only discovered after comparing different versions using real software (SAS in this case).

REPLY

littlegreyrabbit
April 21, 2013 at 1:23 am

Yes well I know of a case using R where they turned our a postgraduate year after year because some statistician forgot that she/he needed to remove the data frame headers before joining two datasets and got a rather famous -1 error

There are risks and errors inherent in any platform – they simply need to be managed. With the Excel example you simply need to archive a password-protected master copy and just distribute copies of that. Obviously you can't force anyone to do that – just as you can't force that as yet anonymous person at Duke University to remember whether they need to remove their column headers or not.

REPLY

Charles Hoogstraten
April 22, 2013 at 4:13 pm

As several commenters have said, it is of course true that errors or deliberate misuse are possible with any software package (or any procedure more generally). The real questions are twofold: how easy or difficult does a package make it to avoid those errors if one takes reasonable care, and to what extent does the package provide an audit trail of what was done to the data so that any errors or problems can be reconstructed? For real data analysis, Excel fails miserably on both counts. The fundamental problem is that it conflates tables designed for display with actual data structures; when you are manipulating how things are displayed, you are also manipulating the underlying data. Which means it is all too easy, when you "sort" rows or columns or "delete" what you think are gaps or do any number of other things, to scramble the underlying data completely, and once the operation is out of the "undo" queue there is absolutely no tracking it.

I have seen several examples of exactly such errors creeping in when Excel is used to calculate students' grades; in this case, the problem is immediately caught in a flurry of outraged emails from students doing 4.0 work who received 1.5 marks, and results only in a lot of work to correct things. If it is instead scientific data that is distorted, data points don't email you to tell you that you goofed them up, and things can propagate onward in too many cases uncaught. At the very least, an appropriate scientific data handling program will have data structures that are independent of tabular or graphical displays, so that one may change what one is looking at and how one is looking at it without risk.

REPLY

Daniel Lemire
April 24, 2013 at 8:47 am

I don't think a retraction is likely for another reason. The work was not published in a journal: it is mostly cited as a "working paper" (http://www.nber.org/papers/w15639).

REPLY

Brian W.
April 27, 2013 at 3:59 pm

And the fact that it was never published in a peer reviewed outlet should have given everyone (especially academic economists, but also policymakers) pause when reading the paper. It is, unfortunately, a prime example of why economics research is often regarded as "the dismal science."

REPLY

CAM
May 4, 2013 at 5:24 am

That is not correct. It has been published in 2010 in the American Economic Review, an influential academic journal; indeed, one of the most influential ones in the world.

REPLY

StatObserver
April 30, 2013 at 5:23 pm

I'm doubling down on my comments about Excel. You've just found out something alarming about Excel:

Value*blank cell = 0

If that does not alarm a person, nothing will. Excel calculations are not acceptable as science.

REPLY

Stewart
April 30, 2013 at 5:31 pm

May I suggest not including blank cells in any calculations?

REPLY

StatObserver
May 1, 2013 at 9:49 am

Basically Excel does not do a standard computation accurately. This is not OK with me.

REPLY

Leave a Reply

This site uses Akismet to reduce spam. Learn how your comment data is processed.

PREVIOUS
Calibration error sends moisture paper down the drain

NEXT
University of Virginia doctoral candidate plagiarizes in business ethics journal, but remains in program