Commentary on

Analysis of official deceased organ donation data casts doubt on credibility of China's organ transplant reform

by Matthew P. Robertson, Raymond L. Hinde and Jacob Lavee: January 26, 2019.

Professor Sir David Spiegelhalter FRS OBE

Submitted to China Tribunal
March 19 2019

1. I am a statistician, formally Professor for the Public Understanding of Risk at the University of Cambridge. I retired from this post in December 2018, and am currently Chair of the Winton Centre for Risk and Evidence Communication in the Centre for Mathematical Sciences in the University of Cambridge, which aims to improve the way that statistical evidence is used by health professionals, patients, lawyers and judges, media and policy-makers. I advise organisations and government agencies on risk communication and am a regular media commentator on statistical issues, with a particular focus on communicating uncertainty.

   I have over 200 refereed academic publications and am co-author of 6 textbooks, as well as *The Norm Chronicles* (with Michael Blastland), *Sex by Numbers,* and *The Art of Statistics*. I presented the BBC4 documentaries '*Tails you Win: the Science of Chance*' and the award-winning "*Climate Change by Numbers*". I was elected Fellow of the Royal Society in 2005, and knighted in 2014 for services to medical statistics. I was President of the Royal Statistical Society for 2017-2018.

2. I have reviewed the evidence presented in the report *Analysis of official deceased organ donation data casts doubt on credibility of China's organ transplant reform*
   by Matthew P. Robertson, Raymond L. Hinde and Jacob Lavee, dated January 26, 2019 (subsequently referred to as Robertson et al, 2019). My approach has been that of a sceptical statistical referee for a medical paper. I have looked only at the data analysis between pages 17 and 44 in the report, and I have not checked the provenance of any of the data or examined any background material. I have not examined, and cannot comment on, the subsequent claims made about individual surgeons and hospitals. I have repeated some of the authors' analysis and carried out further analyses myself.

COTRS data

3. Figure 1 in the paper shows a smooth curve for the growth in transplant numbers from 2010 to 2016 reported by COTRS (*China Organ Transplant Response System).* The authors fit a quadratic curve which shows a very close fit, as measured by the R^2 statistic. Comparison with trends in other countries

do not show such close alignment with a quadratic curve (Additional file 1). When adding in COTRS 2018 data, they show that an even simpler quadratic fits well, where the number of transplants is close to 107.98 * (year -2000)^2: this is shown in my Figure 1 below. They show in their Figure 3 that if the curve is allowed to have any power form, then the closest fit is a power of 2.01, very close to 2.

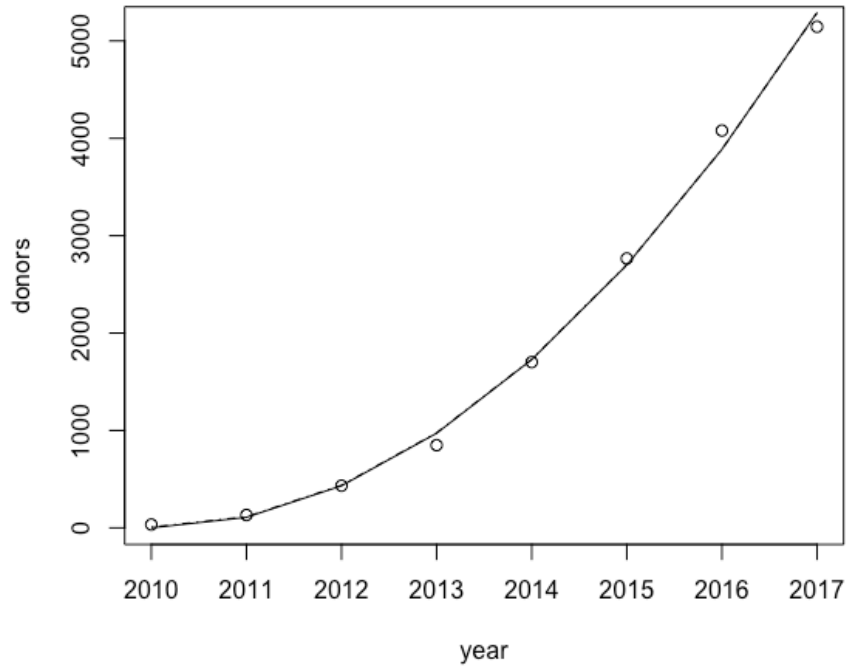

Figure 1. Reported donors each year in the COTRS between 2010 and 2017, showng the close fit of a quadratic curve. The simple quadratic 107.98 * (year -2000)^2 *is indistinguishable from a general quadratic curve.*

4. I have rerun the regression analysis for the China data and my results exactly matched that of the authors. I also tried fitting an exponential model, which had a markedly poorer fit than the quadratic.

5. It should be noted that R^2 would not be the standard measure of agreement with a statistical model for such data, and the standard method would be a Poisson regression whose fit would be assessed by a deviance measure. Were this approach to be taken, I found that many other countries obey a quadratic model closer than China, in terms of the statistical significance when testing a null quadratic model: in particular, the trend in the United States, the country with the largest activity, shows good agreement with a quadratic model when tested formally.

   However, Robertson *et al* argue that it is not appropriate to fit a formal probability model and test its concordance with a quadratic form, as we have no reason to believe that a plausible true model would have an exact quadratic

form. I would agree with the authors' claim that it is not appropriate to fit formal probability models, as the interest rather lies on whether a simple mathematical function fits the reported numbers. Therefore I believe it is reasonable to use R^2 as a measure of fit.

6. R^2 may be expected to be larger for countries with larger numbers of donors and transplants. R^2 is given by *1 – (RMS error/RMS Total)*^{2} [i] and Figure B (reproduced as Figure 2 below) in the authors' Additional File 1 shows these individual quantities plotted (on a logarithmic scale) for a quadratic curve fitted to the numbers of donors, kidneys and livers transplanted each year from 2010 to 2016 in 53 countries. The points marked C represent the results from China, which is clearly an outlier.

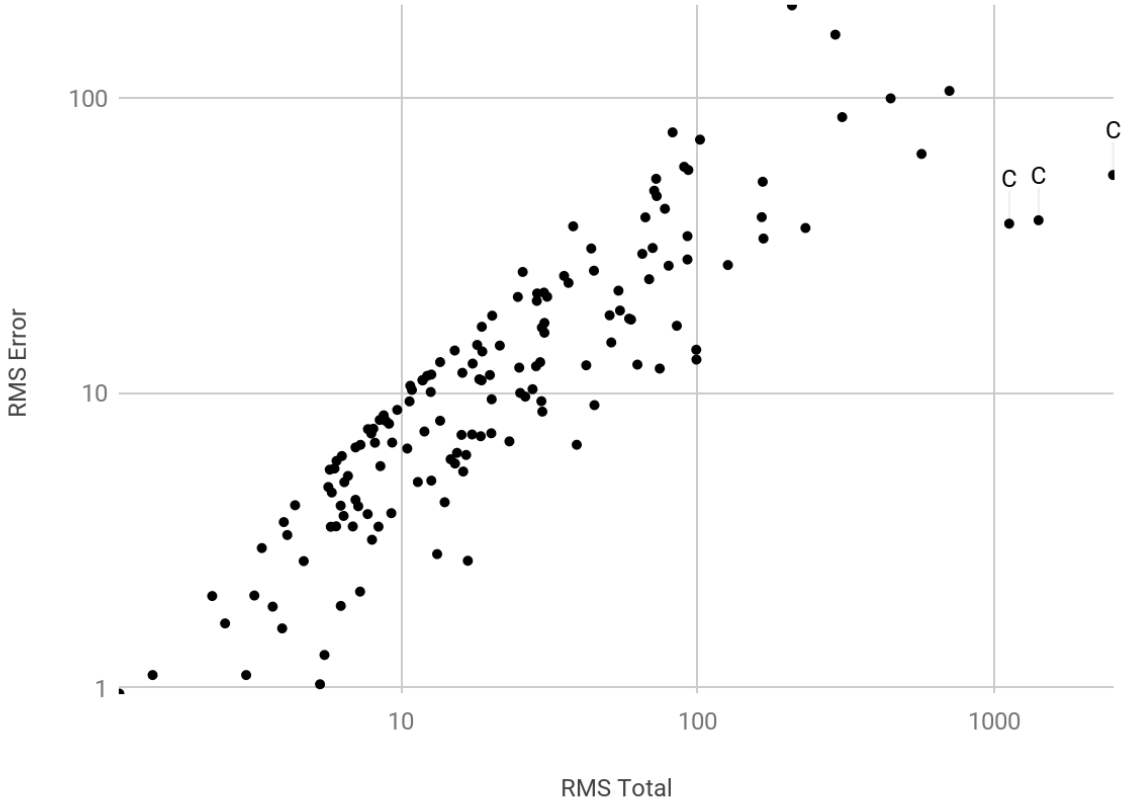

Figure 2. Figure B from Additional File 1, showing outlying points from China.

[i] *RMS Error* is the square root of the mean squared deviation of the points from the fitted line. *RMS Total* is the square root of the mean squared deviation of the points from their average.

7. Most countries show a fairly stable activity between 2010 and 2016, whereas China reported a 40-fold increase in annual donors. It is therefore reasonable to compare with other countries which also reported a substantial increase in activity. The four countries with the largest increase in donors between 2010 and 2016 are shown below: their progress does not follow a smooth curve.

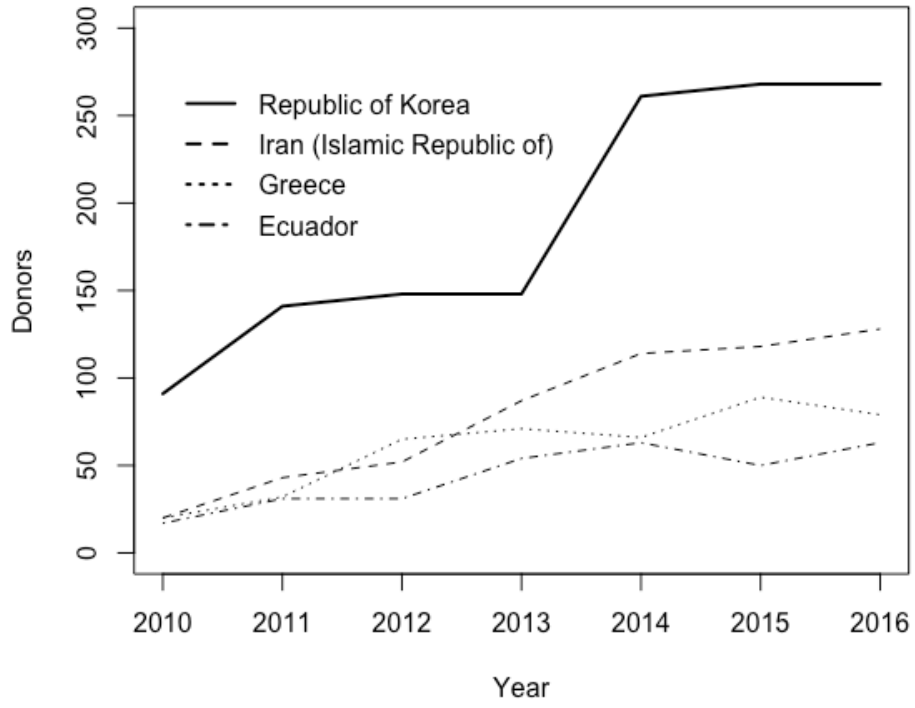

Figure 3. The trends in donors between 2010 and 2016 for the four countries with the biggest increase in activity (except China). Data from the authors' Additional file 1.

Central Red Cross Data

8. The authors constructed cumulative data from Central Red Cross Data website. They found sporadic unusual observations, labelled A to D. In chronological order:

9. D – 18th December 2015. After a period with higher than average transplants per donor, over a 25-day interval there were 659 donors and 1415 transplants, a lower rate of 2.15 per donor, and the highest recorded daily rate of new donors. This restored the cumulative rate to its long-term average of 2.75 per

donor. This could in principle be explained by a backlog of donors being entered onto the system.

10. C – 20th May 2016. There is a rate of 21.3 transplants per donor during a 10-day interval, due to excessive recording of 640 transplants from 30 donors. This is clearly impossible. The interval also has comparatively low number of transplants (3 a day). Previous intervals had showed a lower number of transplants per donor. The authors show that this impossible data puts transplants-per-donor rate back to 2.75. These data could in theory be explained as a backlog of transplant cases that had not been entered into the database, whereas donors had been.

11. B - 31st December 2016. The cumulative transplant total decreases from 27647 to 27613 (I have checked the linked screengrabs). This could be a transcription error – if 27613 is changed to 27913 (which it has been in Additional File 2), the anomaly disappears. The authors note that this error did, however, serve to again bring the transplants-per-donor rate back to 2.75.

12. A – 28th May 2017. This is an apparent transcription error: the anomaly disappears if 31849 is corrected to 32849 (as in Additional File 2)

13. Implausibly large sudden jumps are identified in the number of registered volunteers are identified. These could in principle be explained by recording a batch of volunteers from a new source.

Comparison between central Red Cross and COTRS data

14. The authors note that, although the two data sources largely agree on the number of donors, in 2016 the Central Red Cross data recorded 11,620 transplants, while COTRS recorded 10,481 kidney/liver transplants. This gap of 1,139 does not appear explainable by lung and heart transplants. While discrepancies between data-bases would not be unusual, it is surprising that over the same period the number of donors matches exactly in the two databases.

Comparison between Red Cross data and hospital activity in different regions

15. The Red Cross data in five regions — Henan, Guangxi, Shaanxi, Sichuan, and Zhejiang — was compared with recorded hospital activity. It is important to note that the authors selected these regions because of apparent anomalies in their data, and so they do not represent 'typical' examples.

16. Henan province data showed rapidly increasing numbers in donors and transplants, almost exactly doubling over three successive years. This was not matched by reports from the main hospitals. Data from Guangxi, Shaanxi, Sichuan province both show some implausible transplant rates, and a substantial number of unexplained donors. The authors claim that in each province the majority of transplants are performed in a few large centres, and it is not plausible that lesser hospitals could make up the substantial missing

numbers. In contrast, hospitals in Zhejiang reported substantially more transplants than recorded in the Red Cross data.

17. I have not examined the claims made in the section **Additional sources and considerations** about voluntary/non-voluntary misclassfications and non-reporting of non-voluntary transplants.

Conclusions

18. As far as I can ascertain, the analyses conducted by Robertson *et al* have been carried out appropriately. From a methodological perspective, the main issue that could be contested is their use of R^2 as a measure of model-fit, when their data would be generally considered as Poisson variables. But in this context, the numbers are not being considered as random variables, but as numbers being checked for veracity. I therefore feel their analyses are appropriate.

19. Routine collected data is rarely perfect: even in the data provided in Additional file 1, the Dominican Republic shows 102 donors in 2010 with only 7 kidneys transplanted. There are almost inevitably recording errors and delays in data provided by any bureaucracy. Nevertheless, the anomalies in the data examined by Robertson *et al* follow a systematic and surprising pattern.

20. The close agreement of the numbers of donors and transplants with a quadratic function is remarkable, and is in sharp contrast to other countries who have increased their activity over this period. This could, of course, just be coincidence, although it is difficult to quantify how surprising such a pattern is without a model for the 'natural' development of a transplant programme. But I cannot think of any good reason for such a quadratic trend arising naturally.

21. The Central Red Cross data shows anomalies that could, in principle, be explainable by inadequacies in the reporting system, with batch uploads of delayed notifications. However the authors identify that anomalies B, C D, all tended to bring the series back to a norm of 2.75 transplants per donor.

22. Some individual provinces show sharp disagreement between the data available from the major hospitals and the larger numbers reported in the Red Cross data.

23. I am unable to ascribe any specific reasons for the anomalies described above, but feel they are certainly worthy of further investigation.