PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	**Consent to organ donation should not be presumed before exploring regional variations**
AUTHORS	McGlade, Donal; Rae, Gordon; McClenahan, Carol; Pierscionek, Barbara

VERSION 1 - REVIEW

REVIEWER	***Professor Dave Collett*** Associate Director Statistics and Clinical Audit NHS Blood and Transplant Bristol
REVIEW RETURNED	10-Jan-2011

THE STUDY	The analysis of the data on donation rates is based on 570 pairwise comparisons of proportions (Table 2) without any allowance for multiple comparisons. If this is done, using the Bonferonni correction, an overall 5% sig level would require individual tests to be significant at the 0.00009 level, for which a z score greater than 5 is needed. If this is used, none of the associations for solid organs are significant, but there remain differences between cornea rates between the nations (apart from N vs S). This means that much of the discussion in the paper is not correct. Of course, an analysis based on nearly 600 p-values is not the right thing to do, but comparisons of trends over time would be more informative. Also note that no interval estimates for donation rates etc have been given. There is little consideration of factors that may account for any differences seen (eg no programme for donation after cardiac death in Ireland, variation in ethnicity across the UK). It also appears as though the changes in population size over the study period have not been accounted for. Much of the article relies on a reference to Pocock (2010), which is not listed in the references, but is email correspondence between a statistician at NHS Blood and Transplant and one of the authors. I suspect that Dr Pocock was unaware of the author's intentions to publish his remarks.
RESULTS & CONCLUSIONS	See above for comments on the statistical shortcomings of the paper.

REVIEWER	***Greg Atkinson*** Professor RISES Liverpool John Moores University
REVIEW RETURNED	28-Feb-2011

THE STUDY	In this manuscript, many pairwise comparisons of two population proportions have been undertaken. There are a number of concerns with the analysis approach in this research.

	1. It is unclear exactly which type of test of proportions has actually been employed. Was it Fisher's exact test or a normal approximation of the chi-squared test? 2. There does not seem to be any information on sample size for the various samples from which the proportion data have been collected? 3. There is a serious risk of type I error inflation in this study. From the information presented in Table 2, it appears as though 19 years x 5 organs x 6 comparisons = 570 z-tests have been performed. It would be expected that approximately 5% (n=28) of these tests would be statistically significant when in fact the null hypothesis should not be rejected. 4. The above analysis approach is surprising since there are procedures for analysing multifactorial differences between sample proportions. For example, Zar (199) details a procedure for comparing more than two proportions. He also details another approach based on the Scheffe control of type I error. I think the authors need to adopt these procedures for controlling for type I error.
RESULTS & CONCLUSIONS	Without a sound statistical analysis, especially in terms of an appropriate control of type I error, the reliability of inferences is compromised in this study.
REPORTING & ETHICS	CONSORT not relevant since not a randomized controlled trial
GENERAL COMMENTS	The research seems interesting and clinically-relavant but the data should be analysed with more approriate methods

VERSION 1 – AUTHOR RESPONSE

Barbara K Pierscionek (on behalf of all authors)

Reviewer(s)' Comments to Author:
FIRST REVIEW
Reviewer: Professor Dave Collett
Associate Director Statistics and Clinical Audit
NHS Blood and Transplant
Bristol

REVIEW ON APPEAL
Thank you for the opportunity to comment on the author's responses to my review of the statistical component of their paper.

I first note that the referees for the BMJ did not include a statistician.

Turning to the author's points in turn, let me first comment on the appropriateness of the Bonferonni correction. I agree that this is controversial, that the correction can be conservative, and that there is scope for arguing about how many comparisons should be considered in applying the correction. I also agree that other approaches such as that proposed by the author or the false discovery rate (Benjamini and Hochberg) may be helpful. However, it is simply bad statistical practice to use 570 individual statistical tests to analyse a data set consisting of 380 rates or proportions. The 570 tests cannot be assumed independent, because a given donor may contribute different organs, and the fact that there are only 3 possible independent comparisons between the four countries for each organ x year combination. The authors make no allowance for this, simply interpreting every z score that exceeds 1.96 on face value.

Response: In response to Professor Collett's comment about multiple comparisons and the potential for a proliferation of Type I errors we have adopted Professor Atkinson's suggestion and used a test proposed by Zar (see below). Regarding the issue of independence it should be noted that the comparisons that we have carried out are those within each combination of organ and year which are indeed independent. No comparisons are being made between, say, the proportion of Scots donating corneal tissue and those donating lungs which could violate independence.

The second point concerned interval estimation. I disagree that interval estimates add nothing. They convey much more information about the extent of a difference in two proportions than is provided by a significance test. However, I would not suggest that the 570 p-values be replaced by 570 confidence intervals! An analysis that is firmly based on the data structure would be more sensible. This could include some simple graphs showing trends in donor rates over time by country and organ. But note that standard analysis of variance (with an error term for the F statistic), as proposed by the authors, is inappropriate as the raw data are rates. A Poisson regression model, with a linear (or indeed polynomial) term in time and interactions with organ and country seems appropriate and may well be informative. The authors might like to obtain guidance on this.

Response: We have now adopted a completely new and far more conservative analysis of the data following the recommendation of the second reviewer, Professor Atkinson (see below).

I still think it discourteous of the authors to refer to Dr Pocock's contribution on six separate occasions without informing him of their intentions regarding publication. After sending in my report, I had an opportunity to talk to Dr Pocock, in confidence, about the extent of his contribution, and the analyses that had been carried out. He was understandably concerned to be seen to be associated with a paper that had such poor statistical analyses. He would like any reference to his email correspondence deleted from the paper, and replaced by an acknowledgement to NHSBT for the provision of data, but the authors should really discuss this with him directly.

Response: The corresponding author has discussed the paper with Dr Pocock and the reference to NHSBT data is now acknowledged with Dr Pocock's agreement.

The authors have not responded to my other points (little consideration of other factors that may affect the rates, and no account taken of changes in population size).

Response: The change in population size has been taken into account in our analyses and this is now addressed in the third paragraph of the section titled 'Variations in donation according to organ type'. Other factors mentioned in this reviewer's previous recommendations include the fact that Ireland has no cardiac death programme and that there is no mention of ethnic differences. As the paper deals with data from the UK, donation programmes for Ireland are not considered. Whilst ethnic differences may exist in approach to donation, ethnicity as a factor was not part of this study, which aimed to consider only whether regional variations exist and whether there may be variation depending on the organ donated.

I stand by my earlier criticisms of the paper, that inappropriate statistical techniques have been used. If Table 2 and the conclusions based on it were replaced by a more appropriate statistical summary, the paper may be acceptable from the statistical viewpoint.

Response: As noted earlier, a different and more robust statistical analysis has now been applied and the original Table 2 has been replaced.

Reviewer: Greg Atkinson
Professor

RISES
Liverpool John Moores University

In this manuscript, many pairwise comparisons of two population proportions have been undertaken. There are a number of concerns with the analysis approach in this research.

1. It is unclear exactly which type of test of proportions has actually been employed. Was it Fisher's exact test or a normal approximation of the chi-squared test?

Response: The original test was based on the assumption that the sampling distribution of the difference between proportions is Normal. The test is described in Altman, D.G. Practical Statistics for Medical Research (1991) Chapman Hall, London, page 234. However, we have now applied a different set of statistical tests as recommended by this reviewer (see response to point 4).

2. There does not seem to be any information on sample size for the various samples from which the proportion data have been collected?

Response: This information is now presented in the section titled: 'Variations in donation according to organ type'.

3. There is a serious risk of type I error inflation in this study. From the information presented in Table 2, it appears as though 19 years x 5 organs x 6 comparisons = 570 z-tests have been performed. It would be expected that approximately 5% (n=28) of these tests would be statistically significant when in fact the null hypothesis should not be rejected.

Response: This has now been addressed using Zar's approach suggested below. It might also be worth noting that if Type I errors were a problem then one would expect them to be randomly distributed in Table 2. Clearly this is not the case.

4. The above analysis approach is surprising since there are procedures for analysing multifactorial differences between sample proportions. For example, Zar (199) details a procedure for comparing more than two proportions. He also details another approach based on the Scheffe control of type I error. I think the authors need to adopt these procedures for controlling for type I error.
Without a sound statistical analysis, especially in terms of an appropriate control of type I error, the reliability of inferences is compromised in this study.

Response: The authors thank the reviewer for drawing our attention to this procedure by Zar which we have now applied. The post hoc tests recommended by Zar in his book Biostatistical Analysis (1999) adopt the Tukey HSD method of paired comparisons and this is what we have used.

The research seems interesting and clinically-relavant but the data should be analysed with more approriate methods.

Response: this has now been done.

VERSION 2 - REVIEW

REVIEWER	***Dave Collett***
REVIEW RETURNED	20-May-2011

THE STUDY	This is now the third time that I have been sent this paper, and still the approach used for the analysis is based on 570 comparisons of

	just 380 proportions. The authors have now used a method that controls the overall Type I error rate, and this has of course led to far fewer significant differences. The authors pick on a few differences amongst solid organs as significant, but there is no pattern in these, and no explanation as to why proportions may be different on one year but not in another (as in lines 26-52 of page 9). On the basis of this (inappropriate) analysis, the conclusion is that there are no systematic differences in the solid organ donor rates, and the only real differences appear to be amongst cornea donation rates. In spite of criticism of the statistical approach used in my two previous reports, the authors have not carried out an analysis based on statistical models for the relation between a donation rate and other factors. I therefore continue to doubt the value of this work.
RESULTS & CONCLUSIONS	See above comments

REVIEWER	***Greg Atkinson***
REVIEW RETURNED	23-May-2011

THE STUDY	The authors have attempted to answer my previous questions about the statistical approach that was adopted. The amendemts have gone some way in addressing the concerns, but there are still many statistical tests applied to the dataset. I agree with the other referee who indicates that the dataset should really be analysed with a single statistical model. I suppose one key disclaimer would be one of an exploratory nature in this research, possibly in order to generate hypotheses for future research but I still see several contradictions in this respect. First there is a strong emphasis on statistical rather than clinical significance and in fact no data for the prmary findings are presented in the abstract at all. Second, I find that the statement-type title of the study does not represent the research that has been undertaken and it really doesn't make sense to me how this statement has been arrived at. What is meant by saying that these data should be explored before patient consent is given? How does all this influence whether a patient should consent or not? It just seems out of bounds of the data that have been analysed?
RESULTS & CONCLUSIONS	See above concerns about how the data represent the statement that is made in the title of the study.

VERSION 2 – AUTHOR RESPONSE

The responses are given below each respective additional comment by the reviewers that was sent to the authors subsequent to the comments given above.

Reviewer Professor Collett
The author's analysis is very inefficient as it is testing 570 hypotheses about how donor rates compare. I would have adopted a modelling approach, using a Poisson regression model.

Suppose that the true donor rate pmp is r_ijk for the i'th country (i=1,2,3,4), the j'th year (j=1 to 18) and the k'th organ (k=1 to 5). The data available are values y_ijk (the number of deceased donors) and n_ijk (population size divided by 10^6) and the observed rates given in Table 1 are y_ijk/n_ijk, which are estimates of the r_ijk. We might then suppose that the counts of the numbers of deceased donors y_ijk have a Poisson distribution with mean m_ijk = E(y_ijk). A log linear model for the underlying true rates r_ijk is natural, and so we would take log r_ijk = country_i + year_j + organ_k

Since m_ijk=r_ijk x n_ijk, a log linear model for the Poisson mean m_ijk is given by

log m_ijk = log n_ijk + country_i + year_j + organ_k

This is a Poisson regression model for the number of deceased donors, that includes what is known as an 'offset' (log n_ijk), which is an x-type variable with a known coefficient of 1. This model is easily fitted using standard software for a generalised linear model. By adding country to a model that includes year and organ effects (as well as the offset), we can determine the evidence for differences between countries, adjusted for year and organ effects. See Breslow and Day (1987), my own book on modelling binary data (Collett, 2003), other books on modelling, or google 'Poisson regression model for rates' on the internet for more details.

Here, the model might be extended by adding interactions between organ type and country, to investigate whether donor rates for a given organ are consistent over countries, and if not where the differences lie.

There are still problems with this analysis, as no account is taken of correlation between the numbers of deceased donors in a given year and country - someone donating a kidney may also donate a liver etc. This criticism can be applied to the authors analysis as well. This could be allowed for, informally, by undertaking a separate analysis for each organ.

Hope this helps the authors in producing a much more appropriate analysis of the data.

Response: The statistical analysis described by Professor Collett has now been applied and this has been done for each organ separately to avoid problems caused by multiple organ donation. The authors are grateful to Professor Collett for his recommendations and assistance in explaining what was required.

Reviewer Professor Atkinson
The authors statement regarding 'the statistical test of Zar' is a perplexing illustration of perhaps how far away they are in applying an appropriate analysis to their dataset. In my first review, I merely asked then to consult Zar's textbook (in which he covers a multitude of tests) to help them apply more appropriate analyses. For example, Zar describes and references an approach for a three-factor chi-square analysis which may help to actually model (in one or at most a few steps) their dataset. None of these approaches are 'statistical tests of Zar' - Zar is the author of a pretty well-respected textbook in which various approaches are described.

I cannot analyse the author's data. The authors should be attempting to analyse their data in the most appropriate way. Binary type or ordinal type data can be modelled in one step (e.g. logistic regression, ordinal regression) with multifactorial type analyses so that all the multiple hypothesis testing (which was at it's worst in the first submission but it still present to a certain extent) is avoided. The authors should consult with a statistician about this.
I assume the authors understand my other points about clinical significance of differnces and the confusing statement-type title?

Response: The application of the previous statistical analysis, using the tests given in Zar’s textbook, has now been replaced with Poisson regression with a Bonferroni correction to significance levels where appropriate. The title of the paper has been changed and the abstract amended to reflect the change in emphasis.
The authors would like to clarify that this paper was never intended to be a clinical paper. The aim

was to explore whether regional differences in organ donation existed across the UK as these may be pertinent to the success of any legislative change from informed to presumed consent and the type of presumed consent (soft or hard) that may be introduced.

VERSION 3 - REVIEW

REVIEWER	***Dave Collett***
REVIEW RETURNED	15-Jul-2011

GENERAL COMMENTS	I am glad to see that the authors have finally used a sensible approach to the analysis, and the conclusions drawn seem appropriate. Now that the statistics in the manuscript have been sorted out, I have some comments on the material on donation and transplantation. On page 4, please remove reference [2] from the two places where it occurs and from the list of references. In the opening sentence of the Introduction, delete the statement on the percentage that 28% carry a card (out of date and unreliable information) and replace it by the percentage registered on the ODR. This is 30%, but up to date figures are on the organdonation (sic) website, which can be referenced. The second occurrence of [2] does not need a reference. In the middle of the Introduction, the authors hang presumed consent on to their analysis. They do not explain why differences between nations have to be considered before considering a change to the system is considered. This is certainly not stopping Wales considering introducing a system of presumed consent. On page 5, delete 'and to carry an organ donor card' as this is not part of the requirement. Individuals need to register on the ODR, and this Register is then consulted to ascertain if any potential donor is registered. Card carrying is too unreliable an indication. On page 6, replace '…because it was deemed highly probable that donors often give…' by '…because donors may donate more than …'. Table 2 is not really needed, as the essential points are given in the results section. In the discussion on page 10, some statements are made without any supporting references and sound more like anecdote. Specifically, the phrase about 'information regarding the effect of collecting eye tissue on facial tissue is not sufficiently well explained'. Whose view is this? Where is the evidence? Similarly, at the bottom of page 10, what is the evidence that the presence of an eye bank is the most likely means of increasing corneal donation? Finally, in the last sentence of the Results section (page 11), the authors seem unaware that there are procedures already in place for facilitating donation from EDs in certain circumstances.

REVIEWER	***Greg Atkinson***

REVIEW RETURNED	17-Jul-2011

GENERAL COMMENTS	Thank you for employing more appropriate statistical analyses in your research. 1. There is still no reall attempt to communicate clinical significance in this research. The desired 20 donations per million population is mentioned but the differences in proportions are still not really considered with clinical significance in mind, e.g. what does a regional difference of say 3 donations per million represent in terms of clinical significance. Data should be actually presented in the abstract so that a reader can judge clinical significance. 2. How much crossover is there between regions in terms of donations? For example, can a donation from someone in England or Wales actually find it's way to Northern Ireland. What is the impact of this factor? 3. How much impact does health status of each particular organ have? I note that corneal donations are actually the highest (despite stated cultural limitations that were discussed). Is this representative of health status of organ being a major factor, i.e. cornea from most donators are relatively healthy whereas hearts and livers from many donators might not be healthy enough to be considered using? 4. Can the authors be more clear in saying exactly how the extent of regional differences in donation that are reported lead to impact on how successful future legislation might be in increasing donation rates?

VERSION 3 – AUTHOR RESPONSE

Reviewer: Professor Dave Collett
Associate Director, Statistics and Clinical Audit NHS Blood and Transplant
UK
No competing interests

I am glad to see that the authors have finally used a sensible approach to the analysis, and the conclusions drawn seem appropriate. Now that the statistics in the manuscript have been sorted out, I have some comments on the material on donation and transplantation.
Response: We thank Professor Collett for his guidance with the statistical analysis.

On page 4, please remove reference [2] from the two places where it occurs and from the list of references. In the opening sentence of the Introduction, delete the statement on the percentage that 28% carry a card (out of date and unreliable information) and replace it by the percentage registered on the ODR. This is 30%, but up to date figures are on the organ donation (sic) website, which can be referenced. The second occurrence of [2] does not need a reference.
Response: These changes have all been made.

In the middle of the Introduction, the authors hang presumed consent on to their analysis. They do not explain why differences between nations have to be considered before considering a change to the system is considered. This is certainly not stopping Wales considering introducing a system of presumed consent.
Response: The reason for investigating regional donation rates and whether differences existed was because these may have an effect on how well a system of presumed consent, which is most likely to

be a ‘soft’ one is welcomed and implemented. Such a system would require permission of relatives. If donor rates for different organs vary greatly across the regions, it may reflect attitudes to organ donation amongst other factors and would warrant further study. Professor Collett is right that this does not prevent any region considering such a change in legislation but it may affect how it is accepted and how successful it is. We have tried to make this clearer by removing the section to which Professor Collett refers and replacing it with:. “ Any variations may be influenced by factors that could subsequently affect whether or not a presumed consent system, especially one that included the consent of relatives, were to be successful or whether it may create regional inequalities.” This is the second last sentence in the Introduction. To further clarify, the following words have also been added to the last sentence:”.. in order to see whether such factors may be discernable.”

On page 5, delete ‘and to carry an organ donor card’ as this is not part of the requirement. Individuals need to register on the ODR, and this Register is then consulted to ascertain if any potential donor is registered. Card carrying is too unreliable an indication.
Response: This has been deleted.

On page 6, replace ‘…because it was deemed highly probable that donors often give…’ by ‘…because donors may donate more than …’.
Response: This amendment has been made.

Table 2 is not really needed, as the essential points are given in the results section.
Response: This Table has been removed.

In the discussion on page 10, some statements are made without any supporting references and sound more like anecdote. Specifically, the phrase about ‘information regarding the effect of collecting eye tissue on facial tissue is not sufficiently well explained’. Whose view is this? Where is the evidence?
Response: This statement has now been clarified and qualified with a reference. It now says :”…the effect of collecting eye tissue on facial appearance will require comprehensive explanation as studies show that procuring corneal tissue is erroneously considered to be a procedure that leads to disfigurement.”

Similarly, at the bottom of page 10, what is the evidence that the presence of an eye bank is the most likely means of increasing corneal donation?
Response: This has now been altered to “The presence of an eye bank would make it possible to procure corneal tissue locally, however, it has been found that appropriately trained and dedicated staff are the most likely means of increasing corneal donations.” References are provided.

Finally, in the last sentence of the Results section (page 11), the authors seem unaware that there are procedures already in place for facilitating donation from EDs in certain circumstances.
Response: This sentence and the reference have been removed.

Reviewer: Prof Greg Atkinson
Liverpool John Moores University, Biological Rhythm Research
No competing interests

Thank you for employing more appropriate statistical analyses in your research.
Response: We in turn appreciate the guidance given to us with this.

1. There is still no real attempt to communicate clinical significance in this research. The desired 20

donations per million population is mentioned but the differences in proportions are still not really considered with clinical significance in mind, e.g. what does a regional difference of say 3 donations per million represent in terms of clinical significance. Data should be actually presented in the abstract so that a reader can judge clinical significance.
Response: It was never the intention of this work to consider clinical significance but to investigate whether regional differences in organ donation exist and, if so, to what extent so that any decisions made to introduce presumed consent, which would most likely be a ‘soft’ form requiring relative’s consent, would take into account potential differences that may impede the success of such legislation. Clinical significance is important in many medical contexts such as transplantation and quality of life for transplant recipients and we are not underestimating this, but the paper was not concerned with this aspect.

2. How much crossover is there between regions in terms of donations? For example, can a donation from someone in England or Wales actually find it's way to Northern Ireland. What is the impact of this factor?
Response: Whilst there is a small chance that organs (such as the liver) can be recovered from donors in one region and transported to another for transplantation this would not be common due to rapid deterioration in organs with increased time (NHSBT 2010. Transplant activity in the UK 2009-2010. NHSBT). We have searched the literature and cannot find the impact of this factor. It would be more relevant for transplantation than for donation as it does not affect donation rates in the regions.

3. How much impact does health status of each particular organ have? I note that corneal donations are actually the highest (despite stated cultural limitations that were discussed). Is this representative of health status of organ being a major factor, i.e. cornea from most donators are relatively healthy whereas hearts and livers from many donators might not be healthy enough to be considered using?
Response: Indeed, not all organs recovered are suitable for transplantation and some discussion on this point has been added at the end of the section “Variations in donation according to organ type”. We thank Professor Atkinson for raising this point.

4. Can the authors be more clear in saying exactly how the extent of regional differences in donation that are reported lead to impact on how successful future legislation might be in increasing donation rates?
Response: This point is similar to the first point raised by Professor Collett and the response is given under that respective point.