Reviewer’s report

Title: Intra-cluster correlation coefficients in adults with diabetes in primary care practices: The Vermont Diabetes Information System Field Survey

Version: 2 Date: 16 March 2006

Reviewer: Martin Gulliford

Reviewer’s report:

Intracluster correlation coefficients in adults with diabetes in primary care

The reviewers’ comments were generally fairly consistent and the authors have submitted a greatly improved paper that addresses most of the reviewers’ concerns. However, the paper could still be improved by attention to detail; in particular the paper could be more carefully worded in many places. The authors could produce a higher quality end-product through attention to this aspect of the paper.

Abstract

ICCs varied around a median value of 0.0185 presumably.

It may not be true that an ICC of 0.001 is negligible, such an ICC may be quantitatively important if the number sampled per cluster is large – as is pointed out in the Discussion.

‘The values reported here’. This might read ‘The ICC values reported here’. However, sample size calculations will require values for standard deviations as well as ICCs, and this is why it may be more appropriate to include the SD rather than the SE in the Tables (see later).

Background

‘In a traditional, non-clustered study ‘ this would be better worded as ‘In a study based on simple random sampling or allocation’

‘The inflationary factor, sometimes called the design effect’ this would be better worded as ‘The design effect, sometimes referred to as the variance inflation factor’ because, in the context of cluster randomisation, the design effect is commonly referred to as the variance inflation factor.

‘How much correlation there is within each cluster’ - better as ‘the extent of correlation within clusters’

Statistical analyses (p6)

‘ICC represents the proportion of the total variation in the variable that is associated with the cluster’ better as ‘In the random effects model, the ICC is the proportion of the total variance that is between clusters’.

An algebraic form of the equation would be preferred in which the between- and within-cluster variance components are identified.

Specify the command used in Stata – ‘loneway’?. The method used to estimate confidence intervals
for the ICC should be identified. Further, as the confidence intervals are provided, the standard error for the ICC may not be necessary.

Rho is commonly used to denote the ICC, so it would be best to avoid this as the symbol for Spearman’s correlation coefficient in this presentation.

The standard error of the mean values for the variables of interest, is said to be ‘adjusted for clustering within practices’. This is inconsistent with the comment to the reviewers that the SD can be easily obtained from the standard error and the number in the sample.

The ICCs have been truncated at 0.001. It would be worth stating what proportion of ICCs were estimated to have negative values. Similarly, confidence intervals for the ICCs appear to be truncated at zero.

Results

The results are very briefly described. Table 2 does not appear to be cited.

Discussion (p7)

‘the average number of subjects per cluster’ – better as ‘the average number of subjects sampled per cluster’

Conclusions (p8)

Again, it does not seem safe to conclude that an ICC of <0.001 will always be negligible.

Table 1.

I still think it would be more appropriate to give the standard deviation for continuous variables. This is because it is the SD and not the SE that is required for sample size calculations.

I also think that the number sampled per cluster would be more appropriate, or at least a useful addition, to the total number sampled. In the one-way ANOVA, a weighted form of the average number of subjects per cluster is used (commonly represented by n-nought) and not the arithmetic mean which is referred to in the resubmission letter.

The SE for age appears to be in error.

The SE for number of cigarettes per day seems rather high. It might be more appropriate to only use current smokers to estimate number of cigarettes per day.

For dementia, leukaemia, lymphoma, metastatic cancer and HIV disease the total number of cases appears to be less than 10. It is not clear that the ICC could be estimated consistently for these variables.

What next?: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Needs some language corrections before being published
Statistical review: No

Declaration of competing interests:
'I declare that I have no competing interests'