Reviewer’s report

Title: Managing Re-identification Risk for Small Geographic Areas

Version: 2 Date: 9 December 2009

Reviewer: Tom Burr

Reviewer’s report:

Authors:
The revision looks much improved and your responses are thorough.

Minor:
Font: page 9: There is at least one "X" used for multiplication.
Page 10: define the bracket symbol (greatest integer in?) and I don't quite follow the transition from Dirichlet drawer principle to Sweeney’s approach description.
Throughout: when you mean "because," use "because," not "since"
page 13: logit( ) ~ POP + MaxCombs + (POP x MaxCombs) is not typical way to write a fitted model (there are constants/parameters to fit so usually write for example:
logit() ~ a x POP + b x MaxCombs + c x (POP x MaxCombs)
your style is used in some computing languages such as R and Splus.
BTW, what did you use to fit the LR model?
page 14: "scaled the independent" variables should be "scaled the predictors by dividing by 10,000# ----> use "predictor" rather than "independent" as the 3 predictors are obviously not independent (you included interaction predictor) and "predictor variable " is the modern version of "independent variable"
page 14: roughly how many outliers/influential observations were found
page 16: define confusion matrix (it’s standard, but define it)
I still don't really know why you don't subsample from the 20% sample to help validate (build LR model using 1%, 3%, 5%, etc. from the 20% sample), but I think your cross validation approach is OK.
Why not give confidence intervals around predicted probs at each of 5% or 20%?
Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

I declare that I have no competing interests