Reviewer’s report

Title: Health behaviour modelling for prenatal diagnosis in Australia: A geodemographic framework for health service utilisation and policy development

Version: 2 Date: 14 March 2006

Reviewer: Jane McElroy

Reviewer’s report:

General
In this manuscript “Health behaviour modeling for prenatal diagnosis in Australia” the authors stated goal are 1) to investigate a range of sociodemographic characteristics of women in Victoria Australia who have prenatal diagnosis and 2) evaluate the association between testing and down’s syndrome incidence. A third goal is applying consumer behavior modeling techniques to health data. Both of the primary goals may be of sufficient breadth to warrant separate manuscripts that would allow adequate space to address each goal. For example, I was dissatisfied with the level of detail provided in the methods section probably due to space concerns which led me to suggest two separate papers. Further, the described analyses did not match or could not answer the research questions and will need to be redone.

The geographic unit needs to be much more thoroughly described (and justified) and some attention needs to be given to non-Australian readers. A map of the Victoria with the postcode and LGA’s boundaries might be illuminating, if this stays as the geographic unit of analysis.

-------------------------------------------------------------------------------

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

It is insufficient to describe the aggregation process for those postcode boundaries that bisected the LGA, to merely say the postcode was assigned proportionally based on census data. The assignment of these is critical to the analysis and the technique used needs to be transparent to the reader. Further, the magnitude of possible misclassification needs to be described. What percent of postcodes in the final dataset bisected LGA boundaries?

When looking up the geographic units on the web, a question arose as to why not use Postal Area (previously Collection District Derived Postcodes). According to the internet site these “Postal Areas are approximations of Australia Post Postcodes and are created by allocating Collection Districts (CDs) to existing postcodes on a best fit basis. A Postal Area boundary is then created to encompass all of the allocated CDs and the data for the CDs are aggregated to form Postal Area data.” Also from the internet the “The Local Government Area (LGA) is a geographic area under the responsibility of an incorporated local government council. The number of LGAs and their boundaries can change over time. The LGAs applicable to the 1996 Census output are those which existed at 1 July 1996.” Is there any concern that the LCG boundaries have changed from 1997 to 2002 and if so, how were these changes addressed in the assignment of study participants? This needs to be clarified. Is there any issue with non-incorporated areas in Victoria being excluded from the analysis? It seems these LGA are more or less cities and shires. Why not run the analysis using these cities and shires instead of aggregating up from postcodes if postal areas can't be used? Is city or shire residence not included in the data? But really, it was not clear why these data needed to be aggregated up at all. Why not run these analyses using postcodes?

The geodemographics of the population needs to be much more thoroughly described. Listing the different groupings (table1) without providing any descriptive statistics is pretty much a black box. Furthermore, it does a disservice to the wealth of information that undoubtedly the authors are privy
to but have not shared with the readers. Although the software is proprietary, some detail on how these groups were created needs to be described. For example, adding birth data, percent of female population in child bearing years, percent of population below poverty or some income level, median household income, etc would be helpful.

The statistic (analysis section) is opaque. I can’t figure out what analysis were performed on what data—what the numerator and denominator was for each analysis. From my best guess of what was done, I am concerned with the analysis. It seems that SIRs need to be calculated. Not the proportion analysis. From my best read, this was not done.

I don’t know what the screening analysis tell the reader since terminations were excluded from the analysis; and there is no indication of how many women who had a ds child chose to NOT have an amnio since they had no intention of terminating the pregnancy.

Why was the data truncated to include only women who were expected to deliver in 1998 and 2002 (for those who had the amnio or cvs)? There doesn’t seem to be any rationale for that.

I couldn’t figure out what data was used for which analysis. There was the 1998 and 2002 screening data; and there was the 1997-2002 birth data. Clearly state what years of data were used in the different analysis. The sentence “Denominator data were obtained from all recorded births (Down syndrome prevalence) or confinement (prenatal diagnosis) at the PDCU in 1998 and 2002.” left me puzzled. Denominator data for what analysis? I have no idea what all recorded births = down syndrome prevalence means; and confinement = prenatal diagnosis (of what)?

It makes little sense to me to toss out data. Less than 1% means <600 births per geographic area. That is a lot of births. There are statistical techniques to deal with small numbers in geographic regions (if indeed this is the case for this study—600 is not small).

The two sample test of proportions is improper statistical analysis. The way the first proportion was calculated to the best of my understanding of the paragraph is incorrect.

Referencing a model to describe how the expected rates were calculated does not provide sufficient information to the reader. This model needs to be fully described in the manuscript.

Modelled observed rates of prenatal diagnosis (of what? Of different geographic areas or of the different geodemographic groups) were compared to State average rates…

Please provide information regarding those women who are in the higher risk category (>37 yo) who opted not to have an amnio or cvs. This would change the screening rate, if amnio refusal varied by geography.

In the results, I don’t know what “uptake of prenatal diagnosis” means. Were the authors referring to screening?

Table 2 and 3 do not add to the paper. Table 2: why show these data by age groups that are not used in the analysis? Table 3: these are not part of the analysis (unless these covariates are used in the model); Too general since the data are analyzed by geodemographic region.

Fig 1, 2, 3 and 4 are not illustrating continuous data so should not be drawn with lines. In all of the line graphs, CI should be included. Reporting the names of some of the regions and not others is inconsistent.
Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)
It seems confinement means a pregnancy which results in at least one live birth (from internet definition). If this is the correct definition of this term, then the definition needs to be included in the manuscript.

Discretionary Revisions (which the author can choose to ignore)

What next?: Reject because scientifically unsound

Level of interest: An article of limited interest

Quality of written English: Needs some language corrections before being published

Statistical review: Yes

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