Reviewer's report

Title: Cost and efficiency of public sector sexually transmitted infection clinics in Andhra Pradesh, India

Version: 1 Date: 8 August 2005

Reviewer: Reijo Sund

Reviewer's report:

As a statistician, I have certain comments concerning the methodology utilised in the current version of the manuscript: Cost and efficiency of public sector sexually transmitted infection clinics in Andhra Pradesh, India by Dandona et al.

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

Major Compulsory Revisions:

1. The use of multiple regression in assessing total economic cost - as reported in the current version of the manuscript - is not justified enough to be methodologically sound. First, authors explain the absolute economic costs having absolute fixed costs and numbers of initial and other visits as independent variables. Why not first subtract the fixed costs from the economic costs (imagine that the parameter for fixed cost term is fixed to one) and then try to "extract" the costs of visits from the "residual" costs? However, even in that case the absolute costs are explained using the STI clinic size dependent number of visits resulting in a very high coefficient of determination. A solution could be to divide the "residual" costs (c) as well as the number of initial visits (x) and other visits (y) by the number of total visits (n). Since x/n+y/n=1, the model is not identifiable if the constant term is included. However - by dropping the constant - the model has the form c/n=a*x/n+b*y/n, i.e. the mean visit cost (c/n) equals to the sum of [estimated cost of initial visit (a) times the proportion of initial visits (x/n)] and [the estimated cost of other visits (b) times the proportion of other visits (y/n)]. An appropriate weighting should be used in order to take into account the volume effect.

In other words, the model utilised in the current version of the manuscript must be revised or justified much more carefully. Are there any examples of similar model in the literature? What is the justification to use absolute values instead of the (volume) standardized variables in the model? What is the interpretation of fixed costs term with estimated parameter greater than one or of the constant term in the model? (Are those some kind of non-fixed costs which are not attributable to visits or to random variation [if so, give an example]?) Estimates seem to change (I extracted required data from the tables for estimation) from 56INR to bigger than 70INR (initial visits) and from 4INR to about 10INR (other visits), so the effect is not negligible.

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

Minor Essential Revisions:

2. Authors mention that the data collection included a history of the evolution of the STI clinic (page 5). However, it remains unclear whether these potential historical changes had any effect on the actual data production or analysis. Were there any changes in demand or supply of services during the year?

3. Description of the output data production (page 8) is unclear. Authors also refer to output data problems in page 9. Were there monthly cross-tabulations with suitable stratification available?
Were the individual based medical records reviewed? Was there diagnosis recorded for each visit? What is the coverage and accuracy of written reports? Please describe the output data production more carefully and also move the exclusion justification from page 9 to this section.

4. The variation in reported STIs is mentioned in page 10, but is this variation attributable to 1) differences in absolute sizes of catchment populations or 2) to differences in (volume adjusted) demand?

5. Authors report a significant direct relation between the number of STI case treated per doctor FTE and the cost-efficiency for each STI case treated (abstract and page 11). I would prefer the wording non-linear relation instead of direct relation. The use of power function should be theoretically justified (What is the interpretation of power function in this case? Why is some other function - such as exponential growth curve of type $Y=a \exp(b \cdot X)$ - not a better choice?). If there is no sound theoretical interpretation for the function you decide to use, it is more appropriate to talk about smoothing (p-value and R-squared need not to be reported). It could also be a good idea to make sensitivity analysis by excluding the potentially influential "outlier" observation (the one with highest costs) from data and fitting the curve again (How much the estimates change?). By estimating the parameters for power function (using data available in tables 3 and 4) I did not end up with the same estimates as reported in figure 2 but in the curve $Y=2841570 \cdot X^{1.23742}$ with $R^2=0.81926$ (i.e. higher R-squared but poor visual fit). Please describe your estimation method and utilised parameter restrictions or revise the model.

6. In the tables 3 and 5 the order of STI clinics is not the same as in the table 1.

Discretionary Revisions:

7. Do median and mean values of reported STIs (page 10) give any useful information? How about mean after proportioning STI clinic specific numbers to corresponding catchment populations?

8. In the page 12 the authors talk about inadequate demand. In my opinion, less "provocative" wordings (as in the abstract) could be used.

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

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: Yes

Declaration of competing interests:

I declare that I have no competing interests.