Employment and Commuting Data for the Simulation of Pandemics

Moshe Feder, D.Sc.[†]

[†]***RTI International, Research Triangle Park, NC***

OBJECTIVE

The objective in this study was to explore data on employment and commuting from different sources, using statistical analytic techniques together with geographical experts to obtain information to be provided to modelers in order to help them improve the employment and commuting component of their models, determine potential issues related to these data, and identify problem areas where further investigation is needed.

BACKGROUND

Evidence suggests that transmission within the workplace contributes significantly to the magnitude of a pandemic flu epidemic. A significant number of large organizations have a pandemic plan in place which may help in controlling this manner of transmission. These plans typically include telecommuting and other measures to reduce the need to physically commute to the workplace. Good data are needed in order to obtain valid results from simulation models and to be able to assess the effect of reductions in commuting.

In general, the commuting population of workers lives in one tract and work in a second tract. There are over 65,000 Census tracts in the US. These tracts are very heterogeneous in their demographic, geographic and economic makeup. A realistic commuting model component should correctly account for this heterogeneity.

Currently the MIDAS models mixing subpopulations of workers without regard to occupation, income etc., and many sources of heterogeneity in this population of tracts that are not fully accounted for, which raises the question whether these population traits are important in predicting transmission.

There are alternative sources of data that have not been incorporated into the model component of commuting behavior.

With an explicit commuting model component in place a number of important questions could be addressed such as: What will be the impact on disease spread if business air travel is restricted? What will the impact be of more liberal telecommuting policies? What will be the impact of business practices such as replacing face-to-face meeting with teleconferencing?

METHODS

We analyzed commuter data from different sources. All data were aggregated to the Census tract level. First, univariate outliers were identified and a random sample of these was drawn for examination by geography experts. This was followed by principal components analysis (PCA) leading to a classification of the tracts into a number of strata. A random sample of multivariate outliers (based on PCA scores) was also examined manually by geography experts. Regression analysis was carried out on the remaining data in order to identify explanations for differences in employment data between data sets. Regression residual analysis revealed some additional outliers which were also examined manually.

RESULTS

There is significant heterogeneity in the tracts. The sometimes large discrepancy in the employment data between the data sets we analyzed, is associated with certain characteristics of these tracts. When these were manually examined, it was found out that many of these were tracts where airports were located, prisons, large campuses, lakes and parks, heavy industry, etc.

CONCLUSIONS

In order to make valid inference on the impact of commuting on the dynamics of a pandemic in the U.S, and assess the effects on changes in its patterns, valid data are needed. Using data from multiple sources, together with manual examination of statistical outliers can be a good step in this direction.

Further information:
Moshe Feder, mfeder@rti.org