

A Nationwide Geo-Referenced Synthesized Agent Database for Infectious Disease Models

W.D. Wheaton, M.A., Bernadette Chasteen, M.A., James C. Cajka, M.A., Justine Allpress, M.A., Philip C. Cooley, MS, Laxminarayana Ganapathi, PhD, Joseph G. Pratt, M.P.M.

RTI International, Research Triangle Park, NC

OBJECTIVE

Our objective in this research is to develop a national, geospatially-explicit set of human agents for use in agent-based models. [The term 'agents', in agent-based modeling, refers to computerized entities that represent individuals who interact with each other and their environment.]

BACKGROUND

Agent-based models assign behaviors to individuals based on demographic, socio-economic, and location-based characteristics of the overall population. However, data at a sufficiently detailed individual-level are seldom available. In support of the Models of Infectious Disease Agent Study (MIDAS) funded by the National Institutes of General Medical Sciences (NIGMS), RTI is employing existing tools (developed for the TranSims transportation model developed at Los Alamos National Lab) and new methods to generate a nationwide synthesized agent database that represents the entire U.S. 2000 population and includes the assignments of agents to schools and workplaces as appropriate.

METHODS

Census Summary Tape File 3 (SF3), Public Use Microdata (PUMs), census block group boundaries, and Public Use Microdata Areas (PUMAs) are used as input to an Iterative Proportional Fitting (IPF)¹ technique that expands the PUMS 5% sample to encompass 100% of the household population. Using this method, the total households and household structure information contained in the PUMS data for the 5% sample is allocated across each geographic area and given x, y coordinates. This results in a synthesized population that is reflective of the actual population and household structures in different areas of the country. Since disease transmission has a higher probability of occurring in places where agents are in close contact, agents of school age are assigned to schools and agents across the

threshold of "working age" are assigned to places of work. School-aged agents are assigned to actual schools and grades such that school capacities by grade are filled. Working-aged agents are assigned to workplaces based on commuting patterns available in Census's Special Tabulations File 64 (STP64) which indicates, by census tract, a matrix of their place-of-work census tract and place-of-residence census tract.

RESULTS

For the synthesized human population, the process results in spatially-referenced synthesized agent populations where housing density matches census data at the block group level and household structure, demographic, and occupation data match SF3 counts at the county level. Schools include public, private, and parochial schools. A small percentage of students are assumed to be home-schooled and are therefore not assigned to any school.

CONCLUSIONS

These analyses result in generic, synthesized agents that can be used in agent-based models. The agent database is, in effect, 'infrastructure' for agent-based modeling. The agent database provides a solid foundation on which modelers can add additional behaviors and characteristics. For the human population, the resulting synthesized population is consistent with block group level census data. These data can be made available to researchers requesting data for defined geographical areas at www.midasmodels.org.

REFERENCES

^[1] Beckman, Richard J., Baggerly, Keith A., McKay, Michael D., "Creating Synthetic Baseline Populations", *Transportation Research*, Vol. 30, No. 6, pp. 415-429, 1996.

Further information:
Bill Wheaton, wdw@rti.org
www.midasmodels.org