

Statistical Comparisons of Positional Accuracies of Geocoded Databases for Use in Medical Research

Tonny J. Oyana, Ph.D.
Department of Geography,
Environmental Resource and Policy Program,
Southern Illinois University,
Carbondale, IL 62901-4514, USA.
tjoyana@siu.edu

Abstract

The aim of this study is to statistically compare inaccuracies in previously geocoded features used in medical research. The study employs two widely used street-level reference data: ESRI USA StreetMap 2003 and Dynamap/2000 Street Network. We hypothesize that street-level reference data propagate positional errors in health outcomes, of which decision makers should be aware. A 10% random sample (sample size of asthma [$N = 619$], obstructive sleep apnea [$N = 425$], and gastroenteritis [$N = 489$]) is drawn from three geocoded disease files. The geostatistical model shows that streets in downtown areas were better matched than streets located in the suburbs or farther away. Both reference files meet the fitness-for-use criteria of data quality, but Dynamap/2000 performs better than StreetMap 2003. The findings have implications for health outcomes and the way study areas are delineated for spatial analysis and spatial modeling. Positional errors should clearly be communicated along with study findings, and strong emphasis should be placed on maximizing data-quality statements.