

Estimating the accuracy of geographical imputation

Henry KA, Boscoe FP.

BACKGROUND: To reduce the number of non-geocoded cases researchers and organizations sometimes include cases geocoded to postal code centroids along with cases geocoded with the greater precision of a full street address. Some analysts then use the postal code to assign information to the cases from finer-level geographies such as a census tract. Assignment is commonly completed using either a postal centroid or by a geographical imputation method which assigns a location by using both the demographic characteristics of the case and the population characteristics of the postal delivery area. To date no systematic evaluation of geographical imputation methods ("geo-imputation") has been completed. The objective of this study was to determine the accuracy of census tract assignment using geo-imputation.

METHODS: Using a large dataset of breast, prostate and colorectal cancer cases reported to the New Jersey Cancer Registry, we determined how often cases were assigned to the correct census tract using alternate strategies of demographic based geo-imputation, and using assignments obtained from postal code centroids. Assignment accuracy was measured by comparing the tract assigned with the tract originally identified from the full street address.

RESULTS: Assigning cases to census tracts using the race/ethnicity population distribution within a postal code resulted in more correctly assigned cases than when using postal code centroids. The addition of age characteristics increased the match rates even further. Match rates were highly dependent on both the geographic distribution of race/ethnicity groups and population density.

CONCLUSION: Geo-imputation appears to offer some advantages and no serious drawbacks as compared with the alternative of assigning cases to census tracts based on postal code centroids. For a specific analysis, researchers will still need to consider the potential impact of geocoding quality on their results and evaluate the possibility that it might introduce geographical bias.

Int J Health Geogr. 2008 Jan 23;7:3.