



What's in a name? Improving Geocoding Quality using the NAACCR Geocoder Municipal Alias Tables in New Jersey

NAACCR 2019 Annual Conference June 9-13, 2019

Pamela K. Agovino MPH¹, Gerald Harris PhD², Daniel Goldberg PhD⁴, Payton Baldrige BS⁴, Recinda Sherman PhD⁵, Lisa Paddock, PhD^{2,3}, Antoinette Stroup, PhD^{2,3}
¹New Jersey Department of Health, Trenton, NJ
²Rutgers Cancer Institute of New Jersey, New Brunswick, NJ, ³Rutgers School of Public Health, Piscataway, NJ
⁴Texas A&M University, College Station, TX, ⁵North American Association of Central Cancer Registries, Springfield, IL



Background

Geocoding New Jersey State Cancer Registry (NJSCR) case addresses has presented challenges particular to NJ. Often addresses cannot be accurately geocoded because 1) NJ zip codes often do not follow municipal boundaries, 2) the common use of historical or alternative place names in lieu of the city names used by current street address databases, and 3) common misspellings of city names. In collaboration with the NAACCR Geocoding Work Group, the authors proposed the creation of a NJ-specific Municipal Alias (NJ-MA) list to address these challenges.

Collaboration

Through the NAACCR Geocoding Workgroup the NJSCR partnered with Texas A&M University and NAACCR leadership to develop a plan to create, implement and evaluate a NJ Municipal Alias List to improve geocoding accuracy. We would like to acknowledge Information Management Services (IMS) for their valuable support in creating updates to the DMS system that made implementation and evaluation possible.

Objectives

Evaluate changes in geocoding quality after implementation of the NJ-MA list.

Methods

The NJ-MA list was assembled using a collection of historical local names that linked to at least 30 address municipal values in the NJSCR from 1979-2016. This process also identified common misspellings of local and municipal names that occurred in NJSCR which were used to supplement the NJ-MA list. The geocoding scientists from Texas A&M University (TAMU) modified the NAACCR geocoder to include an option for the address search to account for the NJ-MA table. Using a subset of cases diagnosed between 2000-2017 (N=955,868), we compared the results of geocoding with vs. without the NJ-MA table to determine if there was an improvement in the NAACCR Development Geocoder (NDG).

Acknowledgments

Data were collected by the New Jersey State Cancer Registry under contract HHSN 261201300021 and control No. N01PC-2013-00021 from the Surveillance, Epidemiology, and End Results Program of the National Cancer Institute, and under cooperative agreement 5U58DP003931-02 from the National Program of Cancer Registries, Centers for Disease Control and Prevention, in addition to funding from the State of New Jersey, and the Rutgers Cancer Institute.

Methods con't

Creation of the NJ-MA Listing

- Linked a pre-existing historical listing of NJ local municipal names associated with a listing of Diagnosis City names for all NJSCR cases diagnosed from 1979-2017.
- Retain all records that linked an existing local name to Diagnosis City a minimum of 30 times. Higher Frequencies will yield more reliable results and can be useful for testing purposes.
- The final NJ-MA list is manually assembled using these results ensuring that NJ-MA listings are unique to each county.
- This process will capture common municipal misspellings and abbreviations, which will be useful as NJ-MA listings.
- Examples of common municipal name variations

Buena Vista Township will also appear as

Buena Vista Town
 Buena Vista Towns
 Buena Vista Twp
 Collings Lake
 Milmay
 East Vineland

- The NJ-MA listing was tested in NJSCR DMS Test Site for cases diagnosed from 2000-2017.
- A comparison of the DMS Development Geocoder results without NJ-MA verses with NJ-MA was performed.

Results

Table 1

- The addition of the NJ-MA listing increased the number of matched addresses by 19,369 (2.0%); the number of addresses needing review and the number of addresses not matching decreased by 18,803 and 566, respectively.

Table 1			
Match Status Change with Alias NJSCR Cases 2000-2017, N=955,839			
Address Match Status	Town Name Aliasing		
	Off	On	Difference
Match	757,285	776,654	19,369
Non-Match	60,849	60,283	-566
Review	137,705	118,902	-18,803
Total	955,839	955,839	0

References

Historical listing of NJ local municipal names associated with an incorporated municipality released by the NJ Department of Transportation in 2004
<https://www.state.nj.us/transportation/publicat/localnames2004.pdf>

Results con't

Table 2

- The addition of the NJ-MA listing, resulted in 19,591 NJSCR cases that converted from Review to Match and 570 NJSCR cases that converted from Non-Match to Review.
- A review of the small number of NJSCR cases that declined in match status from Match to Review N=221, showed that these cases were true Matches. This information lead to adjustments to the DMS Development Geocoder to avoid this decline in the future.
- A review of the small number of NJSCR cases that declined in match status from Match/Review to Non-Match N=4, showed that the cases did not change Match status with the DMS Development Geocoder adjustments mentioned above.

Table 2				
Match Status Crosstabs NJSCR Cases 2000-2017, N=955,839				
Aliasing Off	Aliasing On			Total
	Match	Review	Non-Match	
Match	757,063	221	1	757,285
Review	19,591	118,111	3	137,705
Non-Match	0	570	60,279	60,849
Total	776.654	118.902	60.283	955.839

Match Status Change with Alias
 No Change (Blue) Improve (Green) Decline (Red)

Table 3

- The introduction of the NJ-MA listing has improved geocoding accuracy by as much as 2.7% of the total cases processed by the NJSCR annually.

Table 3		
Match Status Improvements with Alias By Dx Year Grouping		
Dx Year	Average Annual Count	% of Total Average Annual
2013-2016	1126	2.1%
2008-2012	1480	2.7%
2003-2007	908	1.7%
2000-2002	683	1.3%

Note- Only complete Dx Years used.

Conclusions

- Inclusion of an MA table has been shown to improve geocoding accuracy and efficiency.
- Inclusion of a Municipal Alias list is like to improve geocoding outcome and reduce staff resources among state cancer registries.