ENHANCED IDENTIFICATION OF OUT-OF-STATE CASES BY UTILIZING 'TEXT – PLACE OF DIAGNOSIS'

Xiuling Zhang, Francis P. Boscoe

New York State Cancer Registry, Bureau of Cancer Epidemiology, New York State Department of Health, Albany, NY

NEW YORK STATE | Department of Health

Introduction and Aims

Patients reported to a central cancer registry are not always residents of the states, provinces, or territories to which they are reported. Patients routinely cross state or national borders to receive better or more economical care, and may provide the address of a relative or hotel rather than their permanent address. Counting such patients inflates cancer rates and risks their being counted in multiple registries. These patients elude being included in interstate data exchanges because they are not obviously out-of-state patients. One way to identify such patients is to make use of the field 'TEXT-PLACE OF DIAGNOSIS' (hereafter shortened to 'place text'), which often contains useful clues about residential status. We presented some preliminary findings on this topic at the 2012 NAACCR meeting, and this poster includes extended detail.

Methods

We identified all tumors in the New York State Cancer Registry (NYSCR) among purported NYS residents (that is, where ADDR AT DX-STATE = "NY") that also contained place text. Using the INDEX function in SAS, we searched the place text for any occurrence of city, state, province, territory, or country names suggestive of being outside New York and flagged these cases for review. Specifically, the search consisted of these steps:

1. We searched for the names of all countries and names and abbreviations of all provinces, states, and territories in the current national data standard (see Appendix B of the 2014 SEER Program Coding and Staging Manual 2014). Various combinations of spacing and punctuation marks were included (e.g., " INDIA ", " INDIA.", " INDIA," and ",INDIA"). For countries consisting of multiple words (e.g., Trinidad and Tobago), each of the key words was included in the search. Common words ("NORTH") and abbreviations that are also common words or abbreviations ("TX", "MD") were excluded, except where they followed a comma (as in "PLANO, TX"). Commonly used character strings that include country names ("BETH ISRAEL MEDICAL CENTER", "JAMAICA, QUEENS") were excluded.

2. We searched for the names of the largest non-US cities in the world (tinyurl.com/cq3w5g) and the largest non-New York cities in the US (tinyurl.com/me973ad), again including various combinations of spacing and punctuation. Names duplicating place names in wide use in New York State (such as Rome, Delhi, and Madison) were excluded.

Cases flagged as potential non-NY residents were first sent to LexisNexis (LN), an online demographic information service, in an attempt to obtain an out-of-state address. The LN search yielded three possible positive results:

(1) A New York State address was returned, in which case the patient was assumed to be a NYS resident and no further action was taken.

(2) An out-of-state address from the state indicated in the place text was returned, in which case the patient was identified as a probable out-of-state resident and given to clerical staff for manual review. Manual review was required as the LN results were not considered definitive on their own.

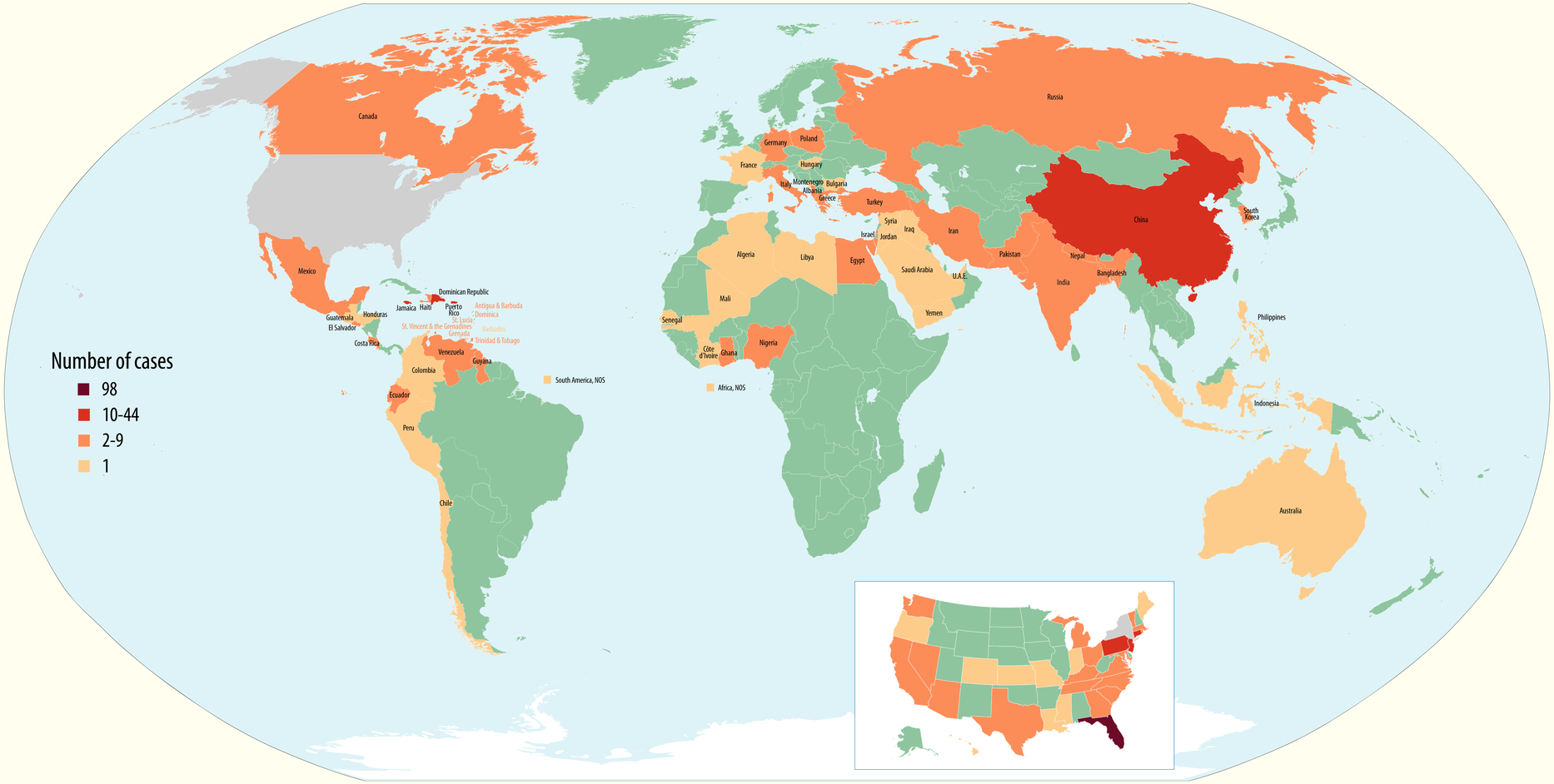
(3) An out-of-state address from a different state than indicated in the place text was returned, in which case the LN result was considered a possible false match and the patient was left as a NYS resident (this outcome was rare).

All cases flagged as potential out-of-country residents, as well as those that did not match to LN, were given to clerical staff for manual review, as the LN database does not contain international address information.

The clerical review was performed by an experienced geocoding clerk, and consisted of reviewing other text fields in the cancer registry record, searching the New York State Department of Motor Vehicles database, using online resources such as peoplefinders.com, and/or querying the Social Security Death Index. This proved to be highly labor-intensive and so we only report results for those cases diagnosed in 2012.

Results

Of 1.3 million tumors containing place text in the NYSCR, 14,957 were identified as potentially associated with non-NY residents. Of these, 1,214 were diagnosed in 2012 and included in the present analysis. About half (610) were flagged for clerical review, and 466 (76%) of these were determined to have a residence at diagnosis in another state, territory, or country (Table 1). International patients most often hailed from the Dominican Republic, Jamaica, other Caribbean islands, China, or India. Domestic patients were most often from Florida, Puerto Rico, New Jersey, Connecticut, or Pennsylvania. As seen in the maps at right, many other parts of the country and world were also represented.



Distribution of patients who were determined to reside outside of New York State

Results - continued

Patients determined to be residents of other countries were disproportionately coded as having been from New York City: 79%, compared with 34% of state residents overall (Table 2). Patients determined to live in other states were much more likely to have interstate data exchange sources: 15.8% versus 3.1% for cases generally (Table 3).

Conclusions

Many out-of-state cases can be identified utilizing the field 'TEXT--PLACE OF DIAGNOSIS'. About 75% of cases reviewed were determined to reside outside of the state, enough to reduce overall crude rates for 2012 by about 0.4%. The labor-intensive nature of the clerical review process means that it might not be practical for registries to review all of their historical data in this manner, but it might be possible for them to keep pace going forward. States lacking international borders and international airports may not see as many patients from other countries in their registries, but out-of-state patients are likely to be an issue just about everywhere. Closer attention to the place text field can also improve the identification of cases that should be part of interstate data exchange.

Table 1. Distribution of cases that were reviewed and determined to reside outside of New York State

Category	Cases with non-NY residence suggested in text (total)	Cases with non-NY residence suggested in text (2012 Dx)	Cases selected for clerical review	Cases verified to be non-NY residents	% of cases verified to be non-NY residents
Other countries	3,426	273	273	220	80.6%
Cities of other Countries	129	8	8	5	62.5%
Other states	9,607	796	283	213	75.3%
Cities of other states	1,795	137	46	28	60.9%
Total	14,957	1,214	610	466	76.4%

Acknowledgements

We acknowledge Tina Maddalla of the NYSCR for her assistance in reviewing cases. This work was supported in part by the Centers for Disease Control and Prevention's Cooperative Agreement U58/DP003879, awarded to the New York State Department of Health through the National Program of Cancer Registries. It would not have been possible without the diligence of the dedicated Certified Tumor Registrars of the hospitals of New York or the coding of the New York State Cancer Registry.

Table 2. Influence of New York City residence

Type of Cases	Total	Count of cases from NYC	% of cases from NYC
All cases diagnosed in 2012	120,186	41,231	34.3%
Cases determined to be out-of-state residents	241	90	37.3%
Cases determined to be out-of-country residents	225	178	79.1%

Table 3. Influence of interstate data exchange sources

Type of cases	Total	Cases that have at least one interstate data exchange source		Cases that have only interstate data exchange sources	
		Count	%	Count	%
All cases diagnosed in 2012	120,186	3,674	3.1%	1,509	1.3%
Cases determined to be out-of-state residents	241	38	15.8%	23	9.5%
Cases determined to be out-of-country residents	225	3	1.3%	2	0.9%