ISDS

Emerging Health Threats Journal 2011, **4**:s136. **doi:** 10.3134/ehtj.10.136
© 2011 E Lanktree *et al.*; licensee Emerging Health Threats Journal.
www.eht-journal.org

ABSTRACT

Assessing address data quality for public health surveillance in Montreal

E Lanktree[1], C Jauvin[1,2], L Bédard[1], and D Buckeridge[1,2]

[1]*Bureau de surveillance épidémiologique, Direction de santé publique, Agence de santé et des services sociaux de Montréal, Montréal, Canada; and* [2]*Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, Canada*
E-mail: elanktre@santepub-mtl.qc.ca

Objective
To (1) validate an address verification algorithm (*Dracones qualité*, DQ, described by Zinszer *et al.*[1]) developed to improve data quality for public health mapping and (2) identify the origin of address errors.

Introduction
In Montreal, notifiable diseases are reported to the Public Health Department (PHD). Of 44, 250 disease notifications received in 2009, up to 25% had potential address errors. These can be introduced during transcription, handwriting interpretation and typing at various stages of the process, from patients, labs and/or physicians, and at the PHD. Reports received by the PHD are entered manually (initial entry) into a database. The archive personnel attempts to correct omissions by calling reporting laboratories or physicians. Investigators verify real addresses with patients or physicians for investigated episodes (40–60%).

The DQ address verification algorithm compares the number, street and postal code against the 2009 Canada Post database. If the reported address is not consistent with a valid address in the Canada Post database, DQ suggests a valid alternative address.

Methods
Individual disease episodes for Montreal residents reported to the PHD between 26 January and 9 May 2009 and followed by PH nurses were included ($n = 798$). Exclusion criteria were H1N1 episodes, those whose investigator address was completely different from the initially reported address, invalidated or non-nominal episodes, for a sample of 408.

Complete investigator addresses treatable by DQ ($n = 342$) were verified manually by a member of the research team (EL) using the Canada Post website.[2] These addresses were also scanned by DQ. DQ and Canada Post results were compared to determine whether DQ correctly identified valid and invalid addresses and proposed valid suggestions.

Calls were made to labs, hospitals or clinics, between 9 April and 9 July 2010, to confirm reported addresses. A list was compiled by selecting all cases that had a reported address deemed invalid by DQ, ($n = 66$) and a 20% random sample[3] of the 181 cases that had reports with hand-written addresses.

Results
DQ detected an error rate of 11.8% (47 of 398) upon initial data entry, 10.0% (40 of 401) at final entry, after archive verification, 18.0% from investigator addresses (60 of 333), and 36.5% from calls to reporting labs/physicians (23 of 63). All addresses corrected at data entry corresponded to DQ suggestions.

Of the 336 episodes with investigator addresses corresponding exactly to initial entry, 334 (99.4%) DQ suggestions correspond to Canada Post. For the two others, both numbered avenues (that is, ninth), DQ proposed one suggestion and Canada Post proposed two.

The investigator address was chosen as the gold standard, as it was closest to the case's actual address. Among initial entries, DQ detected three types of errors: 27 postal code, five street number and six street number or postal code errors. Most corresponding investigator addresses were identical. All other addresses (11 of 27, three of five, and one of six, respectively) that were corrected during the investigator calls corresponded to DQ suggestions.

Conclusions
The DQ algorithm is valid but short street names seem problematic. The algorithm will be refined using a street name length and edit distance solution.

Approximately two percent of errors detected at initial data entry were corrected through data entry and/or archive team efforts, but could be corrected by DQ if integrated into the electronic chart database, saving manpower and time. Labs and physicians should update patient addresses regularly to decrease error rates. Surveillance and intervention

OPEN ACCESS This is an Open Access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/2.5) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

would benefit as reliable data would improve disease cluster identification and communication with cases.

Acknowledgements

This paper was presented as a poster at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010.

References

1 Zinszer K, Jauvin C, Verma A, Bedard L, Allard R, Schwartzman K, *et al*. Residential address errors in public health surveillance data: a description and analysis of the impact on geocoding. *Spat and Spatio-temporal Epidemiolo* 2010;**1**:163–8.
2 Canada Post Corporation 2010. http://www.canadapost.ca/cpotools/apps/fpc/personal/findByCity?execution=e1s1.
3 StataCorp 2009. STATA 10.1.