

# The impact of missing or incorrect addresses

Peter Pudney<sup>1</sup>      Gentry White<sup>2</sup>

(Received 20 November 2015; revised 31 July 2016)

## Abstract

Addresses are critical for finding locations. They are used by emergency services, by utilities, for delivery of mail and parcels, and for our daily navigation. The inability to find a location because an address is missing from a database or points to the wrong location incurs an economic cost, and these costs can be significant. There are two key challenges to estimating the potential cost of missing and incorrect address information. The first challenge is to estimate the number of addresses that are missing or incorrect in the various databases used by different agencies. The second challenge is to estimate the cost of missing or incorrect address; these are not reported by the various agencies that incur these costs. Over the course of the week-long MISG-2015 workshop these challenges were explored. We had insufficient data to make accurate estimates of the extent or cost of addressing errors. However, seven confirmed cases where ambulance delay was a factor in the cause of death indicate that the cost of addressing errors is likely

---

[DOI:10.21914/anziamj.v57i0.10263](https://doi.org/10.21914/anziamj.v57i0.10263) gives this article, © Austral. Mathematical Soc. 2016. Published August 17, 2016, as part of the Proceedings of the 2015 Mathematics and Statistics in Industry Study Group. ISSN 1445-8810. (Print two pages per sheet of paper.) Copies of this article must not be made otherwise available on the internet; instead link directly to the DOI for this article.

to exceed \$1 million per year in Queensland. More data is required to estimate the costs of addressing errors to other service providers and to estimate how effective corrections are in reducing the costs of addressing errors. Our findings should help focus efforts on developing processes to detect and correct errors.

# Contents

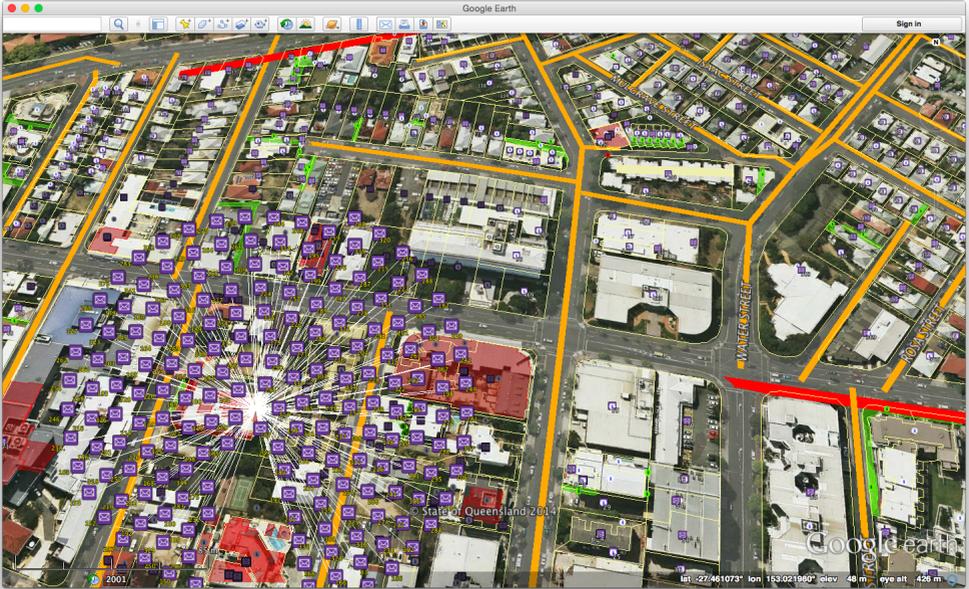
<b>1 Introduction</b>	<b>M116</b>
<b>2 Data</b>	<b>M118</b>
<b>3 Questions and approaches</b>	<b>M121</b>
3.1 Categories of address errors . . . . .	M122
3.2 Economic impacts . . . . .	M123
<b>4 Recommendations, discussion and conclusions</b>	<b>M124</b>
<b>References</b>	<b>M126</b>

## 1 Introduction

Addresses are used to identify locations. Every location where some type of activity may occur should have an address, and may even have multiple addresses. Each address should point to exactly one location. Figure 1 shows an example of locations in a city. Each location has an address, and one location—a commercial building—has multiple tenants and hence multiple addresses.

The Queensland Department of Natural Resources and Mines (DNRM) is responsible for collecting and verifying addresses in Queensland for use by Queensland government services, including the Queensland Police Service

Figure 1: Addresses in a city, indicated by purple dots.



(QPS) and Queensland Fire and Emergency Services (QFES). Misspecified or missing address data hampers the ability of these agencies to serve the people of Queensland, resulting in increased service call times and cost to services, increased risk to public safety, and in some cases preventable serious injury or death.

The problem and implications of misspecified addresses has been studied [10, 24, 21, 2]. Most of the research focused on the impact of address misspecification on epidemiology and spatial disease mapping [17, 8, 23, 11] and on public health [3, 22]. Some research does exist concerning the cost of address misspecification on commerce [9, 4, 5] and on the impacts on public services [1, 20] but the effectiveness of tools for measuring these costs is limited [18]. Despite this there is a plethora of anecdotes [14, 12] indicating

that the misspecification has serious implications and potentially high costs.

Over the course of the Mathematics in Industry Study Group workshop, held at Queensland University of Technology in 2015, a group of researchers worked with representatives of the Queensland Department of Natural Resources and Mines to identify causes of address errors, to estimate the number of address errors in a set of three databases, and to estimate one significant cost component—the cost of fatalities due to incorrect addresses.

## 2 Data

The Queensland Department of Natural Resources and Mines (DNRM) collects address data from Queensland councils and creates a database called the Queensland Address Management Framework (QAMF). DNRM then sends its data to the Geodetic National Address File (G-NAF), where it is combined with address data from the Australian Electoral Commission (AEC) and Australia Post. The G-NAF database then collates all the data received and creates a master database of addresses. The DNRM then uses the G-NAF database to identify and investigate addresses where agencies disagree, and where necessary, update the DNRM database. The data flows, and processing durations, are shown in Figure 2.

Table 1 shows the number of Queensland addresses contributed by agency to the national database from DNRM’s Queensland Address Management Framework (QAMF), the Australian Electoral Commission (AEC), and from Australia Post (AP), as of November 2014.

The first section of Table 1 indicates the number of addresses contributed by one agency only, the second section indicates the number of addresses contributed by two agencies, and the final row before the total shows the number of addresses contributed by all three agencies.

There are a total of over 2.7 million addresses in the G-NAF. Of these over

Figure 2: Address data flows, with processing durations.

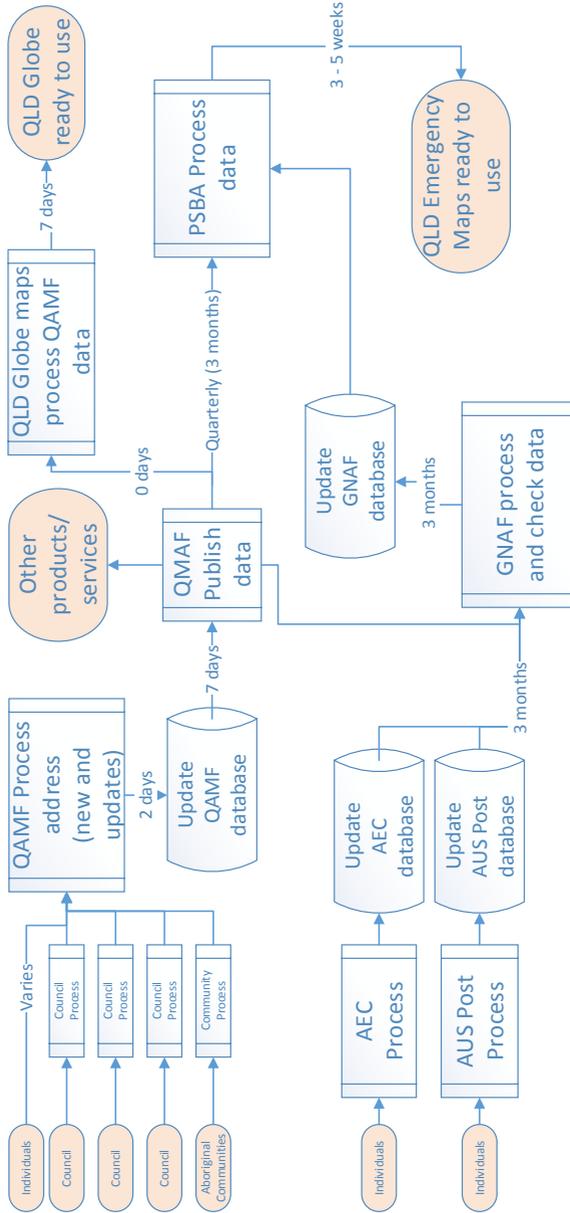


Table 1: Number of Addresses Reported to G-NAF by Agency

Reporting Agency	Records
QAMF	363,404
AEC	100,881
AP	282,388
QAMF and AEC	77,337
QAMF and AP	228,026
AEC and AP	93,843
QAMF and AEC and AP	1,653,969
Total Records in G-NAF	2,799,848

1.6 million addresses appear in all three of the AEC, Australia Post, and QAMF databases. The over 1 million remaining cases (particularly those in the AEC and Australia Post databases but not the QAMF database) are subject to investigation and potential inclusion in the QAMF database.

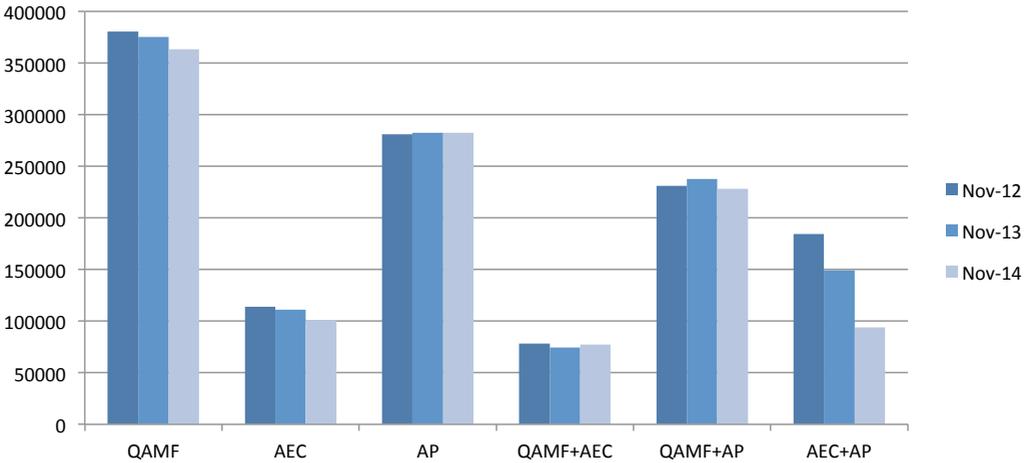
The AEC contributed over 100,000 addresses that Australia Post or QAMF did not have. Many of these occur because AEC generated addresses that should not exist by filling in gaps in address ranges along roads, regardless of the existence of voters (or even properties) in those gaps. Australia Post also has multiple addresses for non-residential properties such as office blocks, industrial sites and shopping centres. Australia Post has addresses for commercial and industrial properties, which should not have voters.

Both the AEC and Australia Post have addresses for individual dwellings on multi-dwelling sites for which QAMF has only a single site address. These should be incorporated into QAMF.

There should be relatively few valid voter addresses that are not in QAMF or the Australia Post data. If we ignore the addresses that only AEC has, then there are about 2.7 million addresses in the database.

DNRM is gradually incorporating the more precise addresses into their QAMF

Figure 3: Number of records in the G-NAF, by time and category.



database. How the number of records changed with time is an important factor, for both G-NAF and QAMF. Figure 3 shows the number of records in the G-NAF that are not common to all three agencies, and how these numbers have changed over recent reporting periods.

In 30 months, DNRM added approximately 4800 new roads and 221,000 new addresses to the QAMF, and modified about 3200 roads and 224,000 addresses.

### 3 Questions and approaches

This project began with the simple statement of the problem that a large proportion of the G-NAF database contain errors. The extent and cost of these errors is difficult to assess. Some locations, particularly those beyond the built environment, do not have addresses. Locations without addresses were outside the scope of our investigation.

Analysis of the G-NAF identified several classes of address error, but did

not allow us to estimate the number of errors in the database. There is no clear application of statistical or mathematical modelling techniques to determine the cost of addressing errors or the benefits of correcting these errors. Instead, the challenge for the research group over the course of the MISG-2015 workshop was to refine and formulate potential research questions to guide DNRM's future research and planning.

### 3.1 Categories of address errors

Initial review of the data and discussions revealed that address errors can be classified into categories depending on the nature and cause of the errors. Understanding and identifying these categories is important in formulating specific solutions.

**Missing Addresses** The most fundamental error in the G-NAF is the occurrence of missing addresses, where a property exists but the G-NAF does not contain an address for that location. The cause for this is often simply a time-lag in registering of new addresses and the updating of the G-NAF database (see Figure 2).

**Imprecise Addresses** Imprecise addresses occur when there is a valid address for a location listed in the G-NAF database, but it is vague or indefinite. This often occurs for gated communities, where the street address for the entry to the community is listed as the location address for all locations within that community. This type of error causes serious problems for parcel and service delivery, particularly in the case of emergency services which can be extremely time sensitive.

**Obsolete Addresses** Obsolete addresses refer to properties that no longer exist or properties that never existed. In these cases the erroneous addresses could be used for fraudulent purposes.

**Non-standard Addresses** Addresses with non-standard numbering or other anomalies can cause confusion. These situations arise when residents

register their own “vanity” addresses with Australia Post, or developers create their own naming and numbering schemes for gated communities.

Each of these errors have differing causes and potential impacts. As a result they require differing approaches for resolution. The focus of our discussion and work during the MISG-2015 workshop was on clarifying and quantifying the economic impact of addressing errors.

## 3.2 Economic impacts

Assessing the impact of incorrect or misspecified addresses requires knowledge of both the types of errors in goods or service delivery that occur due to incorrect or misspecified addresses, and the cost of these errors. Some events such as failed or delayed postal deliveries are relatively common and yield costs associated with both restitution and customer complaint handling [15]. Other errors, including delays in access to emergency services, range from minor inconveniences to avoidable fatal tragedies. While there is much public discussion of these incidences [6, 19] and the potential for errors [7, 13], substantive data is not readily available.

Despite this, DNRM is aware of at least seven cases during the period 2006–2014 that listed ambulance delay due to address error as a factor in the cause of death. Unfortunately, determining culpability in cases of ambulance delay is difficult and typically requires litigation. As a result, with no clear mechanism for deciding, judgements by the coroner as to the contribution to cause of death are subjective and ad hoc. Thorough research is needed in this area. After a brief search we found seven verifiable cases where ambulance delay was a factor in the cause of death; it is likely that there are many more cases where ambulance delay contributed to further pain and suffering, potential treatment complications, and possible reductions in quality of life. Regardless of speculation, after a review of the coroner reports and public records we estimated the contribution of addressing delays to each death, and calculated that address errors were responsible for the equivalent of 3.1 deaths

in the previous nine years, or 0.34 deaths per year. This estimate is based on the verifiable data available during the course of the MISG-2015 workshop and is probably incomplete.

Valuing lives is difficult, as accepted actuarial values vary widely. A median value of \$2.9 million is suggested for Australia [16]. The seven confirmed cases where ambulance delay was a factor in the cause of death translates to approximately \$1 million per year, averaged over the last nine years, in costs to Queensland due to incorrect or misspecified addresses. These figures are most likely an underestimation of the true cost but do provide a concrete starting point to make further recommendations.

## **4 Recommendations, discussion and conclusions**

The cursory review and exploration of the costs of missing or incorrect addresses in the DRNM's QAMF database from the MISG-2015 workshop reveals evidence that the costs merit the application of additional resources to improve the processes ensuring the accuracy of the QAMF database. Errors should be corrected if the expected cost of errors exceeds the cost of correction.

The expected cost of errors depends on the proportion of locations that have missing or incorrect addresses. Unfortunately, the number of incorrect or misspecified addresses in the QAMF database is not known, and is changing. Assuming a conservative estimate of an expected yearly cost of \$1 million, then it seems reasonable to us that DNRM allocate funding to correct QAMF errors and determine the effectiveness of the corrections.

During 2014, DNRM spent \$650 000 adding and correcting 180 000 addresses. The money spent making corrections was less than the likely cost of deaths due to addressing errors, but we do not know how much these corrections contributed to a reduction in the expected costs due to addressing errors. In

order to determine this, more accurate information about the number and cost of errors is needed.

The following recommendations to DNRM should help effect an improved estimate of both error rates and the cost of errors.

- Expand and relationships with service providers and other sources to collect better information regarding both errors in addresses and the cost of these errors. These providers include Australia Post, NBNC0 (who are discovering errors as they deploy the National Broadband Network), courier companies, and QPS and QFES.
- Contact the Australian Securities and Investments Commission (ASIC) and the Australian Business Register for information regarding the costs of criminal activities associated with fraudulent addresses.
- Categorise any incorrect addresses that are found in the G-NAF and identify the cause, cost of correction, and cost of error for each category. Prioritise corrections based on this cost information.

Apart from economic considerations, address errors, especially those that result in fatalities, have immeasurable impacts. An accurate database of addresses in the QAMF is the ultimate desired outcome for DNRM, but should be viewed as an aspirational goal, as some inconsistencies will always exist. This should not be viewed as discouraging efforts to correct the QAMF database, instead it should help focus efforts on developing processes to detect and correct errors as efficiently as possible.

**Acknowledgments** We thank the industry representatives Rob Duffy and Matt Higgins from the Queensland Department of Natural Resources and Mines, for bringing this problem to MISG and for their enthusiastic participation during the week, and contributors Kaye Marion, Mahsa Keshtkaran, Monika Buljan, Harish Sankaranarayanan, Martin O'Hely and Joe Maisano for their work on this problem.

## References

- [1] P. Beynon-Davies. Information systems ‘failure’: the case of the London Ambulance Service’s Computer Aided Despatch project. *European Journal of Information Systems*, 4:171–184, 1995. **M117**
- [2] Michael R. Cayo and Thomas O. Talbot. Positional error in automated geocoding of residential addresses. *International Journal of Health Geographics*, 2(10), 2003. **M117**
- [3] SM Dearwent, RR Jacobs, and JB Halbert. Locational uncertainty in georeferencing public health datasets. *Journal of exposure analysis and environmental epidemiology*, 11(4):329–334, 2001. **M117**
- [4] Julia Edwards, Alan McKinnon, Tom Cherrett, Fraser McLeod, and Liying Song. The impact of Failed Home Deliveries on Carbon Emmisions: Are Collection/Delivery Points Environmentally-Friendly Alternatives. In *14th Annual Logistics Research Network Conference*, 2009. **M117**
- [5] Julia Edwards, Alan McKinnon, Tom Cherrett, Fraser McLeod, and Liying Song. Carbon Dioxide Benefits of Using Collection-Delivery Points for Failed Home Deliveries in the United Kingdom. *Transportation Research Record: Journal of the Transportation Research Board*, 2191:136–143, 2010. **M117**
- [6] Christine Flatley. Widow sues government over faulty ambulance equipment. *Brisbane Times*, 11 March 2010. **M123**
- [7] Peter Foley. Paramedics plead for house numbers on rural properties. *Queensland Times*, 20 February 2015. **M123**
- [8] David I Gregorio, Ellen Cromley, Richard Mrozinski, and Stephen J Walsh. Subject loss in spatial analysis of breast cancer. *Health & Place*, 5(2):173–177, 1999. **M117**

- [9] S.D. Jones, S. Eagleson, F.J. Escobar, and G.J. Hunter. Lost in the Mail: the Inherent Errors of Mapping Australia Post Postcodes to ABS Derived Postal Areas. *Australian Geographical Studies*, 41(2):171–179, 2003. **M117**
- [10] Nataliya Kravets and Wilbur C. Hadden. The accuracy of address coding and the effects of coding errors. *Health & Place*, 13(1):293–298, 2007. Part Special Issue: Environmental Justice, Population Health, Critical Theory and GIS. **M117**
- [11] Andrew B. Lawson. Editorial. *Spatial and Spatio-temporal Epidemiology*, 1(2–3), 2010. **M117**
- [12] Stephanie Linning. Dirty ambulance satnav and staff taking down the wrong address . . . two in a catalogue of 999 blunders that led to patient deaths, according to investigation. *Daily Mail Australia*, December 7 2014. **M117**
- [13] Nev Madsen. Seconds count: Ambos call for visible rural numbering. *The Chronicle*, 5 March 2014. **M123**
- [14] Chris Marshall. Thousands of ambulances sent to wrong addresses. *The Scotsman*, January 5 2015. **M117**
- [15] John McMilan. Australia Post: Determining Levels of Compensation for Loss or Damage of Postal Items. Technical report, Postal Industry Ombudsman, 2010. **M123**
- [16] Office of the Australian Safety and Compensation Council. The Health of Nations: The Value of a Statistical Life. Technical report, Australian Safety and Compensation Council, February 2008. **M124**
- [17] M. Norman Oliver, Kevin A. Matthews, Mir Siadaty, Fern R. Hauck, and Linda W Pickle. Geographic bias related to geocoding in epidemiologic studies. *International Journal of Health Geographics*, 4(29), 2005. **M117**

- [18] Peter C. Smith and Andrew Street. Measuring the efficiency of public services: the limits of analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168(2):401–417, 2005. [M117](#)
- [19] Tuck Thompson. Widow accuses Queensland Ambulance Service of covering up botched responses. *Courier-Mail*, 2 April 2013. [M123](#)
- [20] David Wastell and Michael Newman. Information system design, stress and organisational change in the ambulance services: A tale of two cities. *Accounting, Management and Information Technologies*, 6(4):283–300, 1996. [M117](#)
- [21] Eric A. Whitsel, P. Miguel Quibrera, Richard L. Smith, Diane J. Catellier, Duanping Liao, Amanda C. Henley, and Gerardo Heiss. Accuracy of commercial geocoding: assessment and implications. *Epidemiologic Perspectives & Innovations 2006*, 3:8, 3, 2006. [M117](#)
- [22] Paul A. Zandbergen and Joseph W. Green. Error and Bias in Determining Exposure Potential of Children at School Locations Using Proximity-Based GIS Techniques. *Environmental Health Perspectives*, 115(9):1363–1370, 2007. [M117](#)
- [23] Dale L. Zimmerman, Xiangming Fang, and Soumya Mazumdar. Spatial clustering of the failure to geocode and its implications for the detection of disease clustering. *Statistics in Medicine*, 27(21):4254–4266, 2008. [M117](#)
- [24] Dale L. Zimmerman, Xiangming Fang, Soumya Mazumdar, and Gerard Rushton. Modeling the probability distribution of positional errors incurred by residential address geocoding. *International Journal of Health Geographics*, 6(1), 2007. [M117](#)

## Author addresses

1. **Peter Pudney**, University of South Australia, School of Information Technology and Mathematical Sciences, Adelaide, South Australia, 5000  
<mailto:Peter.Pudney@unisa.edu.au>
2. **Gentry White**, Queensland University of Technology, Mathematical Sciences School, Brisbane, Queensland, 4001  
<mailto:gentry.white@qut.edu.au>