Error Consideration for Geocoding Police Reported Collision Data in California

Word Count: 5,063 words + 3 tables/figures (250 words each) = 5,802 words
Submission Date: July 25, 2014

John Bigham, MPH
University of California, Berkeley
Safe Transportation Research and Education Center
2614 Dwight Way #7374
Berkeley, CA 94720-7374
Phone: 510-643-1777
Fax: 510-643-9922
jbigham@berkeley.edu

Sang Hyouk Oum, MS *
University of California, Berkeley
Safe Transportation Research and Education Center
2614 Dwight Way #7374
Berkeley, CA 94720-7374
Phone: 510-642-5553
Fax: 510-643-9922
shoum@berkeley.edu

* Corresponding author

ABSTRACT

Geographic Information Systems (GIS) are frequently used to analyze collision data. In order to utilize GIS, the data must be geocoded, or assigned a latitude and longitude coordinate by translating a descriptive location onto street network data. However, the ability for accurate spatial analysis can be limited by geocoding errors that may occur due to limitations in data collection technologies, incorrect data entry due to human error, or inaccurate street reference data. In the state of California there is an increased opportunity for data entry errors, given the long sequence of events and resulting paper trail that is required prior to finalizing each collision record. Data entry errors can occur during the initial traffic collision report completion, statewide database entry, state highway reference location input, or during a separate process to geocode fatal collisions. These data entry errors are incorporated into any geocoding process and frequently cause geocoding errors; but even in the absence of data entry errors, discrepancies in street network data can also result in geocoding inaccuracies. The objective of this paper is to summarize the sources of geocoding errors that occur before and after collision data is compiled into the California state database and the federal database of fatal collisions. Consideration is also given to the potential errors that can arise from the use of Global Positioning System coordinates as an alternative to geocoding. Finally, the impact of geocoding errors on traffic safety analysis is discussed in the context of specific applications currently available in California.

INTRODUCTION

Motor vehicle collisions are a significant public health problem, and resulted in 32,561 fatalities and 2,362,000 injuries in the United States in 2012 *(1)*. Various efforts are being made at the local, state, and national level to reduce the number and severity of collisions occurring on the nation's roadways. New technologies are continually being developed to aid these efforts, and the emergence of Geographic Information Systems (GIS) is an example of a powerful tool that can be applied to traffic safety. GIS is implemented in a number of ways—from simple pin mapping on a website to elaborate spatial analyses. GIS provides an automated means to analyze data with a spatial component, and in the traffic safety field the most valuable data source is collision data. However, most collision data must first be geocoded before it can be directly used in a GIS.

Geocoding refers to the process of assigning latitude and longitude coordinates to a descriptive street location via street network data. For traffic collisions, police officers responding to the scene must complete a report that typically includes the primary road on which the collision occurred, and the distance and direction from the nearest intersecting road. This information can then be translated via the geocoding process into coordinates that pinpoint the collision's location in a GIS. There are numerous applications of geocoded collision data. Examples include micro level statistical analyses of individual collisions *(2-4)*, analysis of trends and relationships at aggregate levels *(5-8)*, data input for web based query and analysis applications *(9-11)*, and for evaluations of traffic safety programs such as Safe Routes to School *(12-14)*. While some of these applications can take advantage of previously geocoded collision data sources, others require separate geocoding processes.

The success of geocoding collision data varies widely by the source of collision and street network data, geographic scale of the data, ability to incorporate manual data reviews, and the desired precision. When a collision record fails to geocode properly, however, it is frequently the result of data entry error that misrepresents the collision's location. This scenario can be due either to limitations of data collection technologies or simply human error when entering the information. In the state of California, the opportunity for data entry errors is further amplified by the long sequence of events and paper trail that is required before each collision record is finalized and made available through the Statewide Integrated Traffic Records System (SWITRS). When a police officer begins creating a report at the scene of the collision, it is merely the first step in a multi-stage process in which various factors can contribute to eventual geocoding errors. Accurate location information is even more important since the state level database is also used to submit collisions involving a fatality to the nationwide Fatality Analysis Reporting System (FARS).

The objective of this paper is to summarize the sources of geocoding errors that occur before and after collision data is compiled into the California SWITRS database and the national FARS database. Consideration is also given to the potential errors that can arise from the use of Global Positioning System (GPS) coordinates as an alternative to geocoding. Finally, the importance of accurately geocoded collision data is discussed in the context of specific applications currently available in California.

COLLISION REPORTING AND DATA ENTRY WORKFLOW

In California, all fatal or injury involved collision reports from the California Highway Patrol (CHP) and local allied agencies must be submitted for entry into the SWITRS database. Figure 1 details the workflow from the initial collision report to the final database update *(15)*. After the collision report is finalized, all agencies must submit a paper copy to the CHP. The CHP then manually enters the report into SWITRS. Copies of reports involving collisions occurring on state highways are subsequently sent to the California Department of Transportation (Caltrans) for validation and to establish the state highway specific location information (postmile). Finally, Caltrans submits the completed data to the CHP and the collision record is finalized for the database. A separate process takes place concurrently for collisions involving a fatality. The CHP immediately sends copies of those reports to be entered into the separate national FARS database. The entire process currently takes approximately 14 to 18 months until a complete, finalized year of collision data is made available in SWITRS.

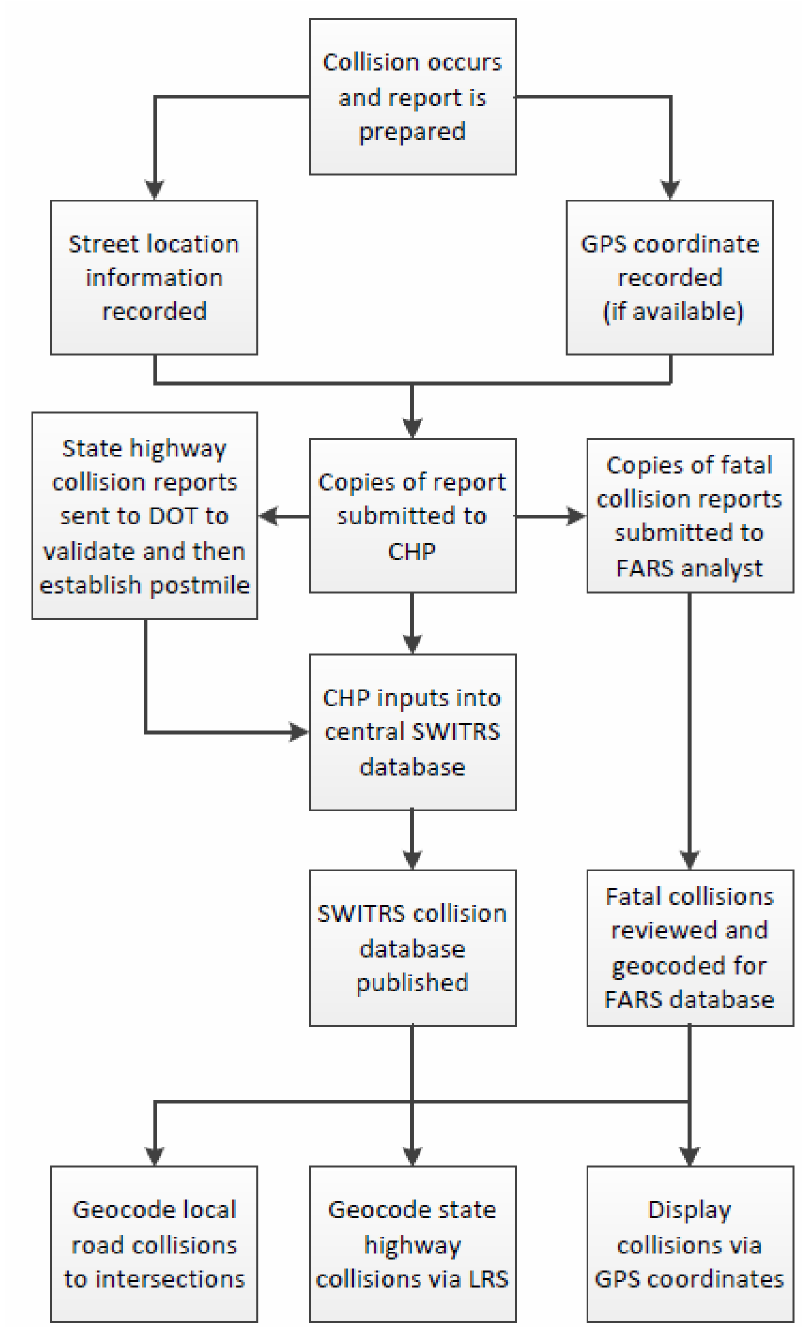

FIGURE 1 Overview of the California Collision Reporting and Data Entry Workflow.

TYPES OF LOCATION DATA ENTRY ERRORS

At any stage in the process there is potential for input error of the location data that could cause geocoding inaccuracies. Errors occurring during any of the following four steps will be examined in more detail before a discussion of other technical reasons why geocoding can fail:

- Initial traffic collision report completion
- Statewide database entry by the CHP
- State highway reference location (postmile) input by Caltrans
- Fatal collision geocoding by FARS analysts

Initial Traffic Collision Report Completion

The California collision investigation manual provides guidelines for every law enforcement agency in the state to use in completing collision reports *(16)*. Each collision report must follow established guidelines for entering detailed information on the vehicles, drivers, conditions, and various other elements, including the precise location where the collision occurred. This descriptive location information is the key component needed to successfully geocode the collision after its inclusion in the statewide database. Officers must enter the primary road name and the distance and direction from a nearby intersecting street. Table 1 shows several sample descriptive locations for collisions. When possible, a GPS coordinate location should also be included. Additional information is required for collisions that occur on state highways. This is completed after the initial report and the process is summarized in the state highway reference location input section below.

TABLE 1 Examples of Descriptive Location Available in SWITRS

Primary Road	Secondary Road	Offset Direction	Offset Distance	State Hwy	Route	Postmile
1st St	University Ave	North	40	No		
Main St	High St	-	0	No		
Rt 880	Hegenberger Rd	North	482	Yes	880	25.58

Despite the extensive guidelines provided, many geocoding errors are attributed to incorrect or invalid location information in the collision report. Geocoding can fail for one or more of the following reasons:

- Street names do not exist, contain spelling errors, or are described using unknown abbreviations
- The distance/direction offset from an intersection is invalid
- Primary and secondary streets do not actually intersect
- Incorrect city or county associated with the intersection

A special circumstance, allowed within the guidelines of the collision investigation manual, but that does not facilitate accurate geocoding, is the use of ‘identifiable landmarks’ as a secondary street. Utility poles, fire hydrants, power boxes, private roadways or other fixed objects are occasionally used in place of a secondary street *(17)*. Bigham et al. identified the occurrence of this practice in 0.4% of fatal and severe injury collisions that occurred in 2004. None of these collisions can be properly geocoded without local knowledge of the landmark and manual assignment of the coordinate location.

Including a GPS coordinate with a collision report provides a potential alternative to geocoding via the intersection. Not all police departments have the technological capability and some may face institutional barriers to including GPS coordinates, however, it is becoming a more common option. It could be assumed that using GPS would result in a more accurate coordinate location; however, that is not always the case. Bigham et al. evaluated California SWITRS data from 2009 to 2011 and found that 43% of GPS coordinates were categorized as correct, 2.5% were unknown, and the remaining 54.5% exhibited some type of discrepancy with the descriptive location *(18).* The inability to verify the true location resulted in many disparities in the descriptive location listed in the collision record. Several studies in other states have also shown mixed results for GPS coordinate accuracy. In Kentucky, earlier use of GPS resulted in correct coordinates only 50% of the time, but in more recent years approximately 92% of randomly reviewed records were found to be accurate *(19, 20)*. Meanwhile, Sarasua et al. found that approximately 80% of GPS coordinates in South Carolina collision data from 2004 to 2006 were within reasonable levels of accuracy *(21).* The sources of the GPS errors range from general operator error to adverse weather affecting measurements to improper manual entry of coordinates into the collision database *(19-23).*

Statewide Database Entry by the CHP

Ideally, data entry errors would be limited to the initial collision report. However, due to the California requirement to submit paper reports to CHP, regardless of whether they have been entered electronically, there is ample room for transcription error. With over 200,000 injury involved collisions and approximately 275,000 voluntarily reported property damage only collisions each year, it is a significant task for the CHP to manually key the reports into SWITRS. This task is subject to human error, as the street names, offset distances or directions and GPS coordinates can all be incorrectly entered. Any error could compromise the ability to accurately geocode the collision.

State Highway Reference Location (postmile) Input by Caltrans

Caltrans is responsible for maintaining the state highway system and addressing infrastructure safety concerns at collision locations. The state highway system is based on a postmile measuring system originally designed in 1964 to catalogue highway infrastructure and related events. The postmile system measures the distance along highways in each county, and following new construction, the measurement values are adjusted using realignments. All events and infrastructure data (e.g., pavement types, constructions zones, collisions and other information related to the highway) require a postmile value. Therefore, to properly assign collisions to the system, an extra processing step is required before they can be entered into the SWITRS database. State highway collision reports are submitted to Caltrans, where personnel must interpret the descriptive location in the collision report and identify the county, state route number, direction of travel, re-alignment type, and postmile measurement. Clearly, this process is also subject to human error, including the following error types:

- Incorrect postmile within allowed range for a highway (not easily detectable)
- Invalid postmile value outside of established ranges for a highway
- Non-existent highway number or direction
- Incorrect county

Despite the potential errors, using the postmile system is more appropriate for geocoding state highway collisions, as discussed in more detail in the geocoding processes section.

Fatal Collision Geocoding by FARS Analysts

FARS is a publicly available national collision database that includes detailed information on all collisions involving a fatality. In 2005, FARS began including latitude and longitude coordinates to facilitate mapping and spatial analysis of the data. To generate the coordinates, FARS analysts must calculate the coordinate location when entering each crash into the database. The coordinate location is derived from custom software that allows the analyst to search on a map via several methods and click on the determined point.

This process is also subject to human error and as a result, the coordinate location is incorrect for some records. However, due to the nature of the descriptive location fields available in FARS, the accuracy of the coordinate location cannot be directly validated. Instead, this process requires matching the collision record to the one in the state collision database to retrieve some of the original location fields. Table 2 shows the available fields for both data sources. Bigham and Husby analyzed the GPS coordinates of FARS data in California by matching a random sample of collision records to the SWITRS record based on dates, times, street intersections, postmile values, and other fields *(24).* After finding the SWITRS record, the actual primary and secondary streets with offset distance and direction could be associated with the FARS record. This allowed the coordinate location to be reviewed and revealed many potential inconsistencies. Approximately 10% of collisions appeared to be placed more than 300 feet from the descriptive location and were deemed incorrect. Another 8% to 10% of collisions could not be verified using the available descriptive location, raising some concerns about their accuracy. Although the descriptive locations are not always accurate, the findings suggested that up to 20% of FARS coordinate locations were potentially incorrect.

TABLE 2 Comparative Examples of Descriptive Location Available in SWITRS vs FARS

SWITRS						
Primary Road	**Secondary Road**	**Offset Direction**	**Offset Distance**	**State Hwy**	**Route**	**Postmile**
1st St	University Ave	North	40	No		
Main St	High St	-	0	No		
Rt 880	Hegenberger Rd	North	482	Yes	880	25.58
FARS						
Primary Road	**Secondary Road**	**Milepost**				
1st St	-	-				
Main St	High St	-				
I-880	Hegenberger Rd	26				

Perhaps more troublesome is the inability to accurately evaluate the locations, and the lack of a feedback loop to the state database. FARS data is derived from the same reports used to compile the SWITRS collision database, yet represents a completely separate dataset. Any geocoding efforts applied to the FARS data are not integrated back into the SWITRS dataset. After FARS is published, only the given coordinates can be used since there is inadequate information for any subsequent geocoding process.

GEOCODING PROCESS ERRORS

After the SWITRS data is compiled and published, any data entry errors will be incorporated into the geocoding process. In the absence of high quality GPS coordinates from the data entry phase, collisions must be geocoded before they can be utilized for spatial analyses. The geocoding process is not an exact science, however, and even if the primary and secondary road descriptions are completely accurate, new errors can be introduced. Collisions may simply not be geocoded or they may be located inaccurately. Bigham et al. outlined a large-scale process used to geocode California collision data. Two different types of processes were employed depending on whether a collision occurred on a state highway or a local road *(17, 25)*. Each of these processes and their contribution to geocoding errors are described below. In addition, the process to display GPS coordinates is briefly outlined since this process could also introduce errors.

Geocoding Local Road Collisions to Intersections

Collisions occurring on local roads are geocoded via standard intersection based geocoding methods using a variety of mature software platforms and services. If the data is invalid, the software may be able to interpret the correct location in some instances, but other times it will fail to locate the site or locate the site incorrectly. However, even in the case of a perfectly coded intersection, there may be naming discrepancies in the street network. For example, the street network may only recognize older street names, preventing the record from being geocoded. The spatial quality of the street network can also adversely affect the positional accuracy *(26)*. For example, if a digital street is not spatially referenced correctly, a geocoded collision could be placed a significant distance from the actual location.

Another factor is that many local road collisions do not occur directly at an intersection, but instead mid-block between intersections. Custom developed software programs are required to offset the proper distance and direction from the intersection and while this process adds precision, it may cause further errors. For example, the collision may be offset in the incorrect direction or the distance may be inaccurate.

Geocoding State Highway Collisions Via a Linear Referencing System

Collisions occurring on state highways, especially on freeways and near interchanges, cannot be effectively geocoded using standard intersection geocoding processes. For example, if a collision occurs at the ‘intersection’ of two freeways, this could be referenced to eight or more different locations by the geocoding process. To overcome this problem, a linear referencing system (LRS) can be used to locate the collision based on the highway number, direction of travel, and postmile value. An LRS interpolates locations along a linear feature based on relative measurements. In the case of a highway feature, an event such as a collision that has a known postmile value can be measured by its distance from a recognized reference marker such as a ramp or intersection. The biggest advantage of using a highway LRS is the ability to apply the assigned postmile value for collisions that are ignored by the intersection geocoding process. Without an LRS, only a small fraction of highway collisions could be properly geocoded.

An improperly calibrated LRS, however, can present further barriers to geocoding regardless of whether the route and postmile associated with the collision are correct. The development of an LRS, especially based on historical postmile reference markers, requires a significant level of effort and is susceptible to errors. For example, if the postmile value of a ramp exit on a freeway is entered incorrectly into an LRS, all measurements near that ramp will be inaccurate after calibration. Bigham and Kang present the process used to develop an LRS for the California state highway system to geocode SWITRS collision data *(25)*. Their work outlines common errors that can affect the quality of an LRS:

- Incorrect postmile marker placement (non-sequential order)
- Incorrect postmile marker placement (not on route)
- Only one known postmile marker on route
- Incorrect measures of accumulation
- No postmile markers near the end of a route

If any of those errors are present, an LRS will be unable to correctly locate collisions occurring near the locations of the calibration errors. An LRS is also subject to the same spatial quality considerations of the street network as intersection based geocoding.

Displaying Collisions Via GPS Coordinates

If GPS coordinates are included for a collision in either SWITRS or FARS, these coordinates can be directly imported into a GIS system. The aforementioned GPS accuracy issues notwithstanding, there is less possibility for translation error since it does not involve a street geocoding process. However, errors can be introduced during the import into a GIS if the latitude and longitude coordinates are transposed or an incorrect or undefined coordinate system is applied to the original coordinates. These errors should be immediately visible, however and can be corrected before any spatial analysis of the collisions is completed.

DISCUSSION

This paper outlines a range of reasons involved in the failure of geocoding collision data. Pinpointing the exact cause of a geocoding error for any given record is not always possible, but understanding the factors responsible for errors can provide important context for ongoing SWITRS geocoding efforts in California. Bigham et al. geocoded 86% of collisions on local roads and determined that 97% of the locations were accurate *(17)*. It was more difficult to evaluate and quantify the overall accuracy of collisions on state highways geocoded via LRS. Following additional research, errors involving the original LRS were identified and corrected, resulting in improved location accuracy *(25)*.

Manual reviews of the data have continued in an iterative process to improve geocoding accuracy since the geocoded data forms the foundation of the Transportation Injury Mapping System (TIMS). TIMS is a web-based system that provides collision data querying and mapping and other tools for traffic related research, policy, and planning in California *(11)*. Since its release in 2011, SWITRS fatal and injury collisions have been geocoded each year and made available through the TIMS site. The raw collision data is retrieved from the CHP, undergoes the geocoding process, and is then made available for direct file download or through the mapping applications available in TIMS. Since TIMS is open to the public and makes the collision data more accessible than the raw non-geocoded SWITRS data, it is widely used throughout the state by a variety of agencies. This underscores the need for high levels of location accuracy since, in many cases, especially in large state or regional analyses, users may not consider geocoding accuracy. The geocoded data is also one of several data sources being used to help inform the state’s distribution of traffic safety related funding.

Several programs exist in California through which local agencies can apply for funds to implement traffic safety countermeasures in their community or region. These programs are typically federally funded programs that are administered in the state by Caltrans. Several examples include the Active Transportation Program (ATP), Safe Routes to School program (SRTS), and the Highway Safety Improvement Program (HSIP) *(27-29)*. These programs are moving to a data-driven application process to maximize potential benefits. The HSIP has the most specific requirements, which necessitate that applicants show evidence of the effectiveness of a proposed safety countermeasure through a benefit/cost calculation. The dollar cost of the countermeasure is compared with the predicted dollars of safety benefits resulting from the decreased number and severity of collisions *(30)*. The benefits must be larger than the costs in order for the project application to even be reviewed. Agencies must also submit evidence that collisions occurring in a particular location can be reduced by a countermeasure, which requires geocoded collision data. Geocoding errors could potentially cause invalid applications or increased efforts on the part of applying agencies to review collision locations prior to completing applications.

While there is a lack of specific research on how collision data geocoding errors affect traffic safety spatial analyses, the subject has been frequently explored in health related research for other data sources that utilize street level geocoding. The impact of geocoding errors for health data typically depends on the type of spatial analyses and the size of the distances under consideration, with smaller scale focused analyses more likely to be affected *(31)*. For example, calculating collision hotspots at individual intersections would be more severely affected by geocoding errors than if an entire neighborhood was analyzed. Another key finding applicable to collision data has shown that geocoding for rural areas tends to be of lower quality than for urban areas due to the differences in the street network quality *(32, 33)*. This is an important consideration given the disproportionate number of fatal and severe injury collisions that occur on rural roadways *(34)*. Overall, researchers acknowledge the significance of geocoding errors and understand the need for better methods to clarify and quantify the impacts in health analysis *(35)*. Collision data and traffic safety analyses are no exception and it is clear that collision geocoding errors can lead to invalid conclusions.

CONCLUSION

The state of California recognizes the need to improve their crash data collection systems and processes. In 2011, the Office of Traffic Safety (OTS) requested that the National Highway Traffic Safety Administration (NHTSA) conduct a traffic records assessment to determine whether the current system fulfilled the need to ultimately improve the safety of the roadways *(15)*. A major recommendation of the assessment was to improve crash data systems in the state by reducing redundancy and by implementing a fully electronic data collection and submission process. The recommendation is being addressed as part of a long term strategy to upgrade SWITRS to enable the database to accept electronic reports from all allied agencies and establish error checking procedures to review the consistency of the reports. However, it is also important that the ability to spatially review the street locations and GPS coordinates is also implemented in data collection technologies used at the collision site. This will ensure improved geocoding accuracy and eventually the use of GPS could eliminate the need for geocoding altogether.

ACKNOWLEDGEMENTS

Special thanks to Grace Felschundneff for general editing and revisions.

REFERENCES

1. National Highway Traffic Safety Administration, U.S. Department of Transportation. *Traffic Safety Facts 2012 Data*, DOT HS 811 856, 2014.

2. Majka, K., et al., *Use of Geocoded FARS data to analyze fatal motorcycle crashes.* National Highway Traffic Safety Administration. Paper 07-0287.

3. Schneider, R., et al., Association Between Roadway Intersection Characteristics and Pedestrian Crash Risk in Alameda County, California. *Transportation Research Record: Journal of the Transportation Research Board, No. 2198(-1)*, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 41-51.

4. Moudon, A., et al. Risk of Pedestrian Collision Occurrence: Case Control Study of Collision Locations on State Routes in King County and Seattle, Washington. *Transportation Research Record: Journal of the Transportation Research Board, No. 2073(-1)*, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 25-38.

5. Lascala, E. A., D. Gerber, and P. J. Gruenewald. Demographic and environmental correlates of pedestrian injury collisions: a spatial analysis. *Accid Anal Prev*, 2000. Vol. 32(5), 2000, pp. 651-8.

6. Wier, M., et al., An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning. *Accident Analysis & Prevention*, Vol. 41(1), 2009, pp. 137-145.

7. Chakravarthy, B., et al., The Relationship of Pedestrian Injuries to Socioeconomic Characteristics in a Large Southern California County. *Traffic Injury Prevention*, Vol. 11(5), 2010, pp. 508-513.

8. Fischer, K., I. Sternfeld, and D. Melnick. Impact of population density on collision rates in a rapidly developing rural, exurban area of Los Angeles County. *Injury Prevention*, Vol 19, 2012, pp. 85-91.

9. National Highway Traffic Safety Administration. *Fatality Analysis Reporting System Encyclopedia*. http://www-fars.nhtsa.dot.gov/Main/index.aspx. Accessed Jun. 23, 2014.

10. Center for Advanced Infrastructure and Transportation. *Plan4Safety*. Rutgers University. http://cait.rutgers.edu/tsrc/plan4safety. Accessed Jun. 23, 2014.

11. Safe Transportation Research and Education Center. *Transportation Injury Mapping System (TIMS)*. University of California, Berkeley. http://www.tims.berkeley.edu. Accessed Jun. 25, 2014.

12. DiMaggio, C. and G. Li, Effectiveness of a Safe Routes to School Program in Preventing School-Aged Pedestrian Injury. *Pediatrics*, pVol.131(2), 2013, pp. 290-296.

13. Ragland, D. R., et al., Ten years later: Examining the long-term impact of the California Safe Routes to School program. Transportation Research Board, Transportation Research Board of the National Academies, Washington, D.C., 2013.

14. Stewart, O., A. V. Moudon, and C. Claybrooke, *Multistate Evaluation of Safe Routes to School Programs. American Journal of Health Promotion*, Vol. 28(sp3), 2013, pp. S89-S96.

15. Benac, J. D., et al., *State of California Traffic Records Assessment*, National Highway Traffic Safety Administration. 2011.

16. *Collision Investigation Manual*, *Volume HPM 110.5*, 2003, California Highway Patrol: Sacramento, CA.

17. Bigham, J.M., et al., Geocoding police collision report data from California: a comprehensive approach. *Int J Health Geogr*, Vol. 8, 2009, pp. 72.

18. Bigham, J.M. and Oum, S., Evaluation of the accuracy of Global Positioning System coordinates for collision locations in California. Working paper. University of California, Berkeley. 2014.

19. Green, E. R. and K. R. Agent, *Evaluation of the Accuracy of GPS as a Method of Locating Traffic Collisions*, 2004, Kentucky Transportation Center: Lexington, KY.

20. Green, E. R. and K. R. Agent. *Evaluation of the Locations of Kentucky's Traffic Crash Data*. Indianapolis, IN. 2011.

21. Sarasua, W., J. Ogle, and K. Geoghegan, *Use of Global Positioning System to Identify Crash Locations in South Carolina.* Transportation Research Record: Journal of the Transportation Research Board, *No. 2064*: *No. 2073*, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 43-50.

22. Sando, T., et al., *Quantification of the Accuracy of Low Priced GPS Receivers for Crash Location.* Journal of the Transportation Research Forum, Transportation Research Board of the National Academies, Washington, D.C., Vol. 44(2), 2005, pp. 19-32.

23. Ogle, J. H., *Technologies for Improving Safety Data*, 2007, Transportation Research Board of the National Academies: Washington, D.C., Transportation Research Board of the National Academies, Washington, D.C., 2007.

24. Bigham, J. M. and H. M. Husby. *A review of the spatial accuracy of FARS coordinate locations in California*, in *138th American Public Health Association Conference*. Denver, CO. 2010.

25. Bigham, J. M. and S. K. Kang, Building a Highway Linear Referencing System from Preexisting Reference Marker Measurements for Transportation Data Management. *URISA Journal*, Vol. 25(1), 2013, pp. 29-37.

26. Frizzelle, B., et al., The importance of accurate road data for spatial applications in public health: customizing a road network. *Int J Health Geogr*, Vol. 8(1), 2009, pp. 24.

27. California Department of Transportation. *Active Transportation Program (ATP)*. 2014. http://www.dot.ca.gov/hq/LocalPrograms/atp. Accessed May 19, 2014.

28. California Department of Transportation. *Safe Routes to School Program (SRTS)*. 2014. http://www.dot.ca.gov/hq/LocalPrograms/saferoutes/saferoutes.htm. Accessed Jun. 24, 2014.

29. California Department of Transportation. *Highway Safety Improvement Program (HSIP)*. 2014. http://dot.ca.gov/hq/LocalPrograms/hsip.html. Accessed May 18, 2014.

30. California Department of Transportation. *Local roadway safety - A manual for California's local road owners. Version 1.1.* Safe Tranportation Research and Education Center, University of California, Berkeley, 2013.

31. Zandbergen, P.A., et al., Error propagation models to examine the effects of geocoding quality on spatial analysis of individual-level datasets. *Spat Spatiotemporal Epidemiol*, Vol. 3(1), 2012, pp. 69-82.

32. Cayo, M.R. and T.O. Talbot, Positional error in automated geocoding of residential addresses. *Int J Health Geogr*, Vol. 2(1), 2003, pp. 10.

33. Zimmerman, D. and J. Li, The effects of local street network characteristics on the positional accuracy of automated geocoding for geographic health studies. *Int J Health Geogr*, Vol. 9(1), 2010, pp. 1-11.

34. Federal Highway Administration. *Implementing the High Risk Rural Roads Program*. http://safety.fhwa.dot.gov/local_rural/training/fhwasa10012/chap_2.cfm. Accessed Jun. 23, 2014.

35. Jacquez, G.M., A research agenda: Does geocoding positional error matter in health GIS studies? *Spatial and Spatio-temporal Epidemiology*, 2012. Vol. 3(1): p. 7-16.