The California Child and Youth Injury Hot Spot Project

Report for the Period 1995 to 1997

August 2000

Volume Three: Technical Guide

A Report By

Family Health Outcomes Project
Department of Family and Community Medicine
University of California, San Francisco

Sponsored by the Family Health Outcomes Project, UCSF,
with funding provided by the
California Department of Health Services,
Maternal and Child Health Branch

The California Child and Youth Injury Hot Spots Project

August, 2000

STUDY CONSULTANTS

Family Health Outcomes Project
Department of Family and Community Medicine
University of California, San Francisco

Geraldine Oliva, MD, MPH, Principal Investigator

Linda Remy, PhD, Co-Investigator and Senior Statistician

Ted Clay, MS, Programmer and Statistician

Victoria Gimeno, MD, MPH, Epidemiologist

Kristie Kooken, Research Assistant

Andy Peri, MA, Geographer

Pamela Weatherford, MA, Editor

Cindie Sedik, Administrative Analyst

Jennifer Gee, Project Administrative Assistant

STATE OF CALIFORNIA

Acknowledgements

This report reflects the efforts of the contractors and staff in the California Department of Health Services, Maternal and Child Health Branch (DHS/MCHB). Work was completed under contract 99-85045 with the University of California, San Francisco. Robert Bates, MD MPH, Medical Consultant, Adolescent Health and Injury Control, served as liaison between DHS/MCHB and the research team.

Geraldine Oliva, MD, MPH, led the research team, with day-to-day supervision by Linda Remy, PhD. Dr. Oliva and Dr. Remy shared primary responsibility for writing the *State Guide*. Dr. Remy and Ted Clay, MS shared primary responsibility for writing the *Technical Guide.* Dr Remy and Dr. Oliva shared primary responsibility for writing materials accompanying the *County Guide*.

Drs. Oliva and Remy developed the statistical reporting format. Ted Clay was responsible for most of the data processing and development of SAS programs to link, summarize, and report data. Andy Peri, MA, designed and prepared study maps. Pamela Weatherford, MA, made the statistical tables and publication-ready maps, and edited the reports. Victoria Gimeno, MD MPH, and Kristie Kooken were liaison to the counties and provided technical support as needed. Cindie Sendik and Jennifer Gee provided project support.

Suggested Citations

Oliva G, Remy L, Clay T, Peri A, Weatherford P, Gimeno V, Kooken K. *California Child and Youth Injury Hot Spots Project 1995-1997, Volume One: State Guide*, Sacramento, CA: California Department of Health Services, Maternal and Child Health Branch, August 2000.

Remy L, Oliva G, Clay T. *California Child and Youth Injury Hot Spots Project 1995-1997, Volume Two: County Guide*, Sacramento, CA: California Department of Health Services, Maternal and Child Health Branch, August 2000.

Remy L, Clay T, Oliva G. *California Child and Youth Injury Hot Spots Project 1995-1997, Volume Three: Technical Guide*, Sacramento, CA: California Department of Health Services, Maternal and Child Health Branch, August 2000.

TECHNICAL ADVISORY COMMITTEE

Robert Bates, MD MPH Medical Consultant,
MCHB, Adolescent Health and Injury Control

Gilberto Chavez, MD MPH, Chief,
DHS, Maternal Child Health Branch

Candice Diamond, Section Manager,
Healthcare Information Division, Patient Discharge Data Section
Office of Statewide Health Planning and Development

Jennifer Harper, Research Scientist,
DHS, Injury Surveillance and Epidemiology Section

Katherine Heck,
DHS, Maternal Child Health Branch

David Lawrence,
Center for Childhood Injury Prevention
San Diego State University

Michael Quinn, Research Manager,
DHS, Planning and Data Analysis Section

Don Taylor, MA, Epidemiologist,
DHS, Epidemiology and Evaluation Section

Roger Trent, PhD, Chief,
DHS, Injury Surveillance and Epidemiology Section

Lynn Walton-Haynes, DDS, MPH, Research Scientist,
DHS, Injury Surveillance and Epidemiology Section

Table of Contents

Tables

Over View

The California Child and Youth Injury Hot Spots Project

In this overview, we describe the Injury Hot Spots Project rational, and objectives. This is followed by a very brief summary of the study methods and a discussion of their limitations.

The primary purpose of the California Child and Youth Injury Hot Spots Project is to identify small area patterns of injury so severe as to result in hospitalization and/or death for children, adolescents, and young adults age 0 to 24. Specifically, hot spots identify California counties and ZIPs whose young residents were at very high risk for serious injury in 1995, 1996, and 1997. The second purpose of the Injury Hot Spots Project is to describe characteristics of the injured and the course of treatment for those who survive to hospital admission.

Project reports will help to better understand injury patterns, target injury prevention activities, and evaluate results of injury prevention efforts. The end of the overview discusses relationships among the three report volumes of the California Child and Youth Injury Hot Spots Project and identifies how to obtain study reports.

Because of underlying complexity and limitations, we recommend that study results be examined and understood in context, using all that is known about injuries in a particular health jurisdiction, other standard injury reports issued by the California Department of Health Services EPIC unit, and other information available to local injury prevention planners and relevant health and social service providers.

The Injury Hot Spots Project Rationale

Injuries are the leading cause of death and disability for children and young adults in California and nationally. Medical, legal and administrative costs for injury hospitalizations of California children under 21 were an estimated $626 million with millions more spent on rehabilitation services and therapies for injured children who survived. Effective injury prevention strategies exist, e.g. car seats, bicycle helmets, ipecac, electric outlet covers, pool fences. County Health Departments and health care providers can use these interventions to decrease rates of death and disability if they can better target their efforts. Geographic information systems can help this effort.

Project Objectives

- To develop a methodology to classify ZIP codes as to the level of risk for injuries by intent for children and youth aged 0-24
- To identify those ZIP codes with the highest injury burden
- To identify factors associated with high injury burden

Methods

Data Sources

Primary data sources were 1995-1997 hospital discharge data from the California Office of Statewide Health Planning and Development (OSHPD) and California Vital Statistics death files. County-level population estimates were obtained from the California Department of Finance. ZIP-level population estimates were obtained from Claritas. Information regarding ZIP changes came from Western Economics Research and the US Postal Service. ZIP-boundary information came from MapInfo.

Record Selection

The target population was California residents age 0 to 24 excluding newborns and conditions originating in the perinatal period. We selected all records with any principal external cause of injury (E-code) following recommended CDC injury categories, in the hospital discharge and Vital Statistics death files.

Record Linkage

Hospital discharge records for individuals were linked to identify people with one or more injuries, to assign the series of hospital admissions related to one injury episode (i.e., admission, transfer, readmission) into an episode of care, or EOC, and to link multiple EOCs for people with more than one injury episode. Hospital and death files were linked to confirm in-hospital deaths and to add cases that died before care could be provided. The resulting file was summarized to the county and ZIP-level.

E-code Discrepancies

The E-Code disagreed on the last EOC for 33% of cases with more than one record. This occurred either when multiple hospital records were linked, when a hospital record and a death record were linked, or in the case of multiple hospital records and a death record. To help readers understand what we encountered, the table below shows a sample case. This case matched exactly on all variables used to make the linkage.

Sample Multiple Injury Case

Admit Date	Admit Source	Injury	Discharge Date	Discharge Disposition
1	ER	Cut/Pierce (sui)	2	Routine
3	ER	Poisoning (int)	6	Other
20	ER	MV Pedestrian	21	LAMA
32	ER	No Injury	34	Routine
54	ER	Poisoning (int)	55	Died

Resolving E-Codes

If the hospital case did not link with the death file, we used the first injury record on the last EOC. In the sample case, we used the last injury (poisoning, intentional) and ignored previous injuries. For hospital cases linking with the death file, we made the following decision: Within 3 days of discharge, we treated both files as describing same injury and used the most specific E-code. More than 3 days after discharge, we used death E-code unless death E-code showed late effects, adverse effects, or no injury, in which case we used the hospital E-code.

Prioritizing E-Codes

Because of the E-code disagreements, we also had to develop rules to prioritize conflicts for the same injury episode of care. We developed the following hierarchy: Intentional before unintentional, other intentional before self-intentional, and a specific hierarchy within subgroupings from more to less specific.

Problems with ZIP Codes

We identified a number of problems with ZIP codes. These included: ZIPs change over time, ZIPs split into 2 or more, ZIPs consolidate, Post Office Box ZIPs nest in ZIPs with geography, ZIP numbers remain same but boundaries change, ZIP boundaries span county boundaries, ZIPs have no ZIP or nonexistent Zip. These problems had to be resolved in order to calculate rates and develop maps. We made a number of decision rules to handle these various issues.

Identifying Hot Spots

To identify hot spots, we calculated the rate per 100,000 population at the ZIP and county levels for all injury episodes, unintentional injury episodes, and intentional injury episodes. Then we calculated quartiles for number of injury episodes and rate for all ZIPs and counties relative to each other (statewide), and all ZIPs intra-county if the county had 12 or more ZIPs.

Defining Injury Hot Spots

- "Hot" spot: ZIPs and counties (and ZIPs within counties) with both number of injury episodes and rate in the 4th highest quartile.
- "Medium" spot: Rate or number in the 4th quartile and the other in the 3rd quartile
- "Warm" spot: Both rate and number in the 3rd quartile.
- All other ZIPs were considered not to be hot spots.

Study Limitations

In the sample table, notice that one person was hospitalized five times during the three-year study period for three different types of injuries, with poisoning (intentional) occurring two times. Without linkage, we would have counted one person five times for the same ZIP and would have double counted poisoning. This would have incorrectly inflated ZIP-level injury rates. Also notice that the person was discharged to another facility after the second injury, but we were unable to link the other record. A case such as this leads directly to understanding the study limitations.

At each step, we tested the reliability for all linkage decisions we made. Despite the reliability we attained, we may have linked some records -- thereby creating multiple-record sets -- that truly belonged to different people. We think this possible because a large number of injuries seemed to resolve inappropriately to single-record episodes. For example, among single-record injuries, 5,147 came in from another facility, and 4,467 were transferred to another facility, yet records from these other facilities were not found and linked. Some of these records may be among those incorrectly linked to other people, as discussed above.

We calculated ZIP-level rates using population estimates from a commercial vendor. If the ZIP-level population estimate was smaller than the actual, the rates will be incorrectly inflated. If the ZIP-level population estimate was larger than the actual, the rates will be incorrectly small. We checked with local county officials for a number of ZIP rates that seemed unlikely. When the 2000 census data are released, we may be able to ascertain the magnitude of discrepancies.

On balance, we believe "over-linking" and "under-linking" may cancel each other out. So long as we consider a record in the person-level file to be a "person" in a loose rather than exact sense, we think we have done as good a job as possible in carrying out the linkage task. The number of injuries we summarized to the county and ZIP-levels more likely reflects a "truer" number of injuries than if we had not linked. In that sense, our results may be more conservative, since a given ZIP will have fewer injuries and will be less likely to be identified as a hot spot.

However, because of the linkage method, results of this study will not agree -- at the county or state level -- with standard injury reports issued by the California Department of Health Services.

Results

The Last Injury Episode

- 150,552 California children, adolescents, and young adults age 0 to 24 were severely injured one or more times between 1995 and 1997.
- 11,275 died of their injuries, 61% of all deaths.
- About 4 of 5 deaths occurred before admission.
- The 141,892 young people admitted to hospital spent 674,594 days in care for their last injury.
- Inpatient hospital charges for the last injury was over $1.55 billion.

Injury Episode Trends

- The number of injury cases admitted to hospital or dying before reaching the hospital declined during the 3-year study period.
- The greatest decline was in the number of intentional injuries.
- Transportation-related accidents accounted for 27% of all injury episodes, falls 20%, self-inflicted (Suicide or suicide attempts) 9%, other intentional injuries (assaults) 13%.
- Two-thirds of the injured were male.
- 57% of the injured were over 15 years old.
- Injury episodes declined for those age 20 to 24.
- Injury episodes increased for children age 5 to 9.

- Non-Hispanic Whites were 41% of the injured, Hispanic 39%, African-American 11%, Asian 6%, Other 4%.

Injury Outcomes

- The number of routine dispositions (return home) decreased. The proportion of routine dispositions increased.
- The number and proportion of non-routine dispositions increased.
- The number and proportion of deaths decreased.

Injury Burden for Cases Admitted to Hospital

- Average EOC length of stay decreased from 5.2 to 4.3 days.
- The median injury EOC charge increased from $7,743 to $8,029.
- The average injury EOC charge increased from $13,331 to $14,366.
- Medi-Cal was the most frequently found anticipated payment source: 38% overall.
- Among those with multiple admissions, 82% showed Medi-Cal as the anticipated payor on at least one record.
- Lack of any type of insurance coverage was found in 16% of all injury-related records indicating the patient remained uninsured.

Multiply Admitted

- About 20% of injury victims admitted to hospital were discharged and transferred or readmitted numerous times with no or a very short break before readmission.
- Many such cases had more than one injury record.
- Psychiatric and/or substance abuse diagnoses were present on 44% of multiply-admitted vs 12% on singles.
- Including multiply admitted increased days of care to 907,522 and charges to $3.2 billion for all discharges.
- This is a 35% increase in days and a doubling of charges over the last EOC used for surveillance and mapping.
- Multiply admitted tended to cluster in a few counties.

Injury Hot Spots

- Of 1,563 California ZIPs, 127 were defined as "hot" spots, 213 "medium", and 115 "warm" for total injuries.
- These hotspots represented 29% of all California ZIPs and 53% of the State population age 0 to 24.
- 65% of serious injuries to 0 through age 24 year olds occurred in these ZIPs.

Conclusions

- Injuries cluster in more populated regions.
- Injury hot spots account for a disproportionate share of cases and costs.
- The multiply admitted account for a large percentage of overall costs.
- Linking death and hospital records and unduplicating cases leads to a better understanding of the number of individuals affected by injuries.
- Combining ZIP-level data for both number of injuries and injury rate to generate an injury burden score is more useful than either statistic alone as a method to identify high priority areas for intervention.

The Hot Spot Reports

Volume One: State Guide This volume is intended for all those interested in community safety, such as local health jurisdictions, hospitals, child advocates, and consumers. The State Guide begins with a brief description of the background, methods, and results of a study to identify California injury hotspots between 1995 and 1997. Then maps are presented that display injury patterns at the state- and county-level. An additional non-technical discussion of the study methods and results is presented after the maps. That section includes statistical tables and graphs summarizing state-level injury data.

Small area maps in *Volume One* characterize areas found to be "hot", "medium", and "warm" with respect to all injuries, unintentional injuries, and intentional injuries. All remaining areas are combined into the same category.

Volume Two: County Guide contains results for counties and ZIPs upon which State Guide maps are based. The accompanying text describes the Injury Episode Table, the Injury Episode-of-Care Table, and the Small Area Table (county- or ZIP-level) statistics developed by the California Child and Youth Injury Hot Spots Project, and suggests ways these data might be used.

Volume Three: Technical Guide*.* This volume is intended for health service researchers, health care providers, and others interested in the computing methods used to identify and flag injury hot spots. It contains a detailed description of the data and methods used to link and categorize injuries, summarize data to the small area, identify injury hot spots, and produce statistical tables and maps. This detailed presentation is intended to allow users and researchers to review and comment on the approaches taken and to encourage future improvements.

How This Differs from Other Injury Reports

This report differs in fundamental ways from other injury reports published by the State of California.

Multiple hospital discharge records for the same injury episode have been linked and summarized. This allows us to track the course of hospital care.

Vital Statistics death records have been linked with hospital discharge summaries. Linkage allows for reconciliation between these files for injuries resulting in death.

Linking and summarizing all records associated with the same injury provides a way to conservatively estimate injury rates in local communities.

Providing zip code (ZIP) level data for both number of injured children and injury rates can assist local groups interested in public safety to better understand geographic injury patterns, target prevention resources toward communities with the greatest injury burden (i.e., both high rates and high numbers), and evaluate such prevention efforts.

Using This Report

This report summarizes injury data for California's children and youth age 0 to 24 years statewide and for local health jurisdictions (58 counties, with Los Angeles divided into four regions, and three independent cities). The ZIP-level analysis compares each ZIP with all other ZIPs statewide and within each jurisdiction.[2]

The State Guide summarizes methods used to analyze the data, classify ZIPs, and presents overall statewide results. It is important to understand the statewide results in order to evaluate the meaning of regional data. Maps in this volume allow readers to visually compare their region's injury pattern with the statewide injury pattern, and to compare ZIPs within their region to each other.

State summary tables in the *State Guide* can be compared with region summary tables in *Volume Two*: *County Guide*. This enables the reader to compare characteristics of injured children in a particular region of interest to state averages. We hope this will contribute to a better understanding of injuries to California's young people and their course of hospital care.

ZIP-level tables in the *County Guide* compare a community's actual injury rates with injury rates statewide and within the region. This permits the reader to evaluate how well each community within a region safeguards its children.

Finally, for those with technical expertise who are interested in a more detailed description of the methods and analyses, refer to *Volume Three*, *Technical Guide*.

After reviewing these reports, we hope you will join others in your community and use your new knowledge to improve local understanding of injury patterns, and to better protect your community's children, adolescents, and young adults.

---

[2] Intra county or city comparisons were made only for local health jurisdictions with at least 12 ZIPs.

Obtaining Study Reports and Data

To obtain copies of these documents, contact: California Department of Health Services Maternal and Child Health Branch 714 P Street, Room 750 Sacramento, CA 95814 (916) 651-1347

The Family Health Outcomes Project can be reached at:
3333 California Street, Suite 365
San Francisco, CA 94109
Phone: 415-476-5283
Fax: 4415-502-0848
Website: http://www.ucsf.edu/fhop/

Chapter 1

Identifying Injuries Admitted to Hospital

Criteria for Identifying Injuries in Hospital Discharge Data

To conduct multi-year analyses of linked hospital outcomes, the Office of Statewide Health Planning and Development (OSHPD) provided the Family Health Outcomes Project (FHOP) with Patient Discharge Data Sets (PDDS) containing original Social Security Numbers. OSHPD gave FHOP special permission to encrypt the Social Security Numbers in these files (the resulting encrypted variable is called SSNC), using FHOP's algorithm to facilitate record linkage.[1] The original Social Security Number is dropped after encryption. We used these versions of the 1995, 1996, and 1997 PDDS to identify injured California resident children, adolescents, and young adults treated in California hospitals.

The PDDS was searched to identify all California residents from 1 day to 25 years old at discharge living in ZIP codes (ZIPs) between 90000 to 96162 hospitalized for an injury.[2] [3] Delivery and newborn records were excluded. Records lacking a valid date of birth (DOB), sex, or ZIP were not examined. The target age group extended to 24 years old. However, in the initial pull, we searched to age 25 at discharge. Someone may have been admitted for an injury at age 24 but the last injury-related discharge may have occurred when the person was 25. Additionally, out of concern that patients with certain injuries may be admitted directly to non-acute facilities such as psychiatric hospitals, we searched records from all types of care.

Records with any principal external cause of injury (E-code) were flagged unless they indicated late effects of a previous injury or adverse effects of treatment for another condition. We used software provided by the California Department of Health Services, based on recommended CDC injury categories.[4] [5] Table 1.1 summarizes the E-code grouping.

---

1 Earlier work identified that OSHPD’s record linkage number lacked the capacity to do soft matches to compensate for data recording or data entry errors. See Romano PS, Remy LL, Luft HS. (1993). Report of the California Hospital Outcomes Project: Volume 2, Technical Appendix. Sacramento, California: Office of Statewide Health Planning and Development.

2 Office of Statewide Health Planning and Development (1996). Discharge Data Tape Format Documentation.

3 The official California ZIP range begins at 90001. A number of records had the first three digits indicating the local area and the last two digits showing "00". The value 90000 indicates that the individual is believed to be a resident of Los Angeles County, but the actual ZIP of residence within the county is unknown.

4 Injury Tables, California, 1996 Deaths and Nonfatal Hospitalizations. Prepared by the Injury Surveillance and Epidemiology Section of the California Department of Health Services.

5 CDC (MMWR 1997; 46 (No. RR-14).

We also classified records as to their principal diagnosis, focusing on certain injury types. For this, we used software developed by FHOP.[6] Table 1.2 summarizes the injury grouping.

Using the above criteria, the initial condition pull identified 162,233 records during the 3-year period. Of these, 3,664 records (2.3%) showed an injury without an E-code. Table 1.3 shows correspondence between the E-code grouping and injury grouping after this step.[7]

---

6 EpiHosp Manual, Family Health Outcomes Project, 1998.

7 At this stage, when a record is selected on the basis of an injury variable, a record can have an adverse or late effect E-code. However, absent an injury, a record cannot have an adverse or late effect E-code and be selected.

Table 1.1: E-Code Grouping

Intent	Variable	Label	Codes
Unintentional	ECODE20	Cut/Pierce	E9200 - E9209
	ECODE12	Drowning/Submersion	E8300 - E8309, E8320 - E8329, E9100 - E9109
	ECODE09	Fall	E880, E8800 - E8869, E888
	ECODE14	Fire, Burn: *Fire/Flames*	E8900-E8999
	ECODE15	Fire, Burn: *Hot Object/Substance*	E9240 - E9249
	ECODE21	Firearms	E9220 - E9229
	ECODE19	Machinery	E9190 - E9199
	ECODE01	Motor Vehicle Traffic: *Occupant*	E8100, E8101, E8110, E8111, E8120, E8121, E8130, E8131, E8140, E8141, E8150, E8151, E8160, E8161, E8170, E8171, E8180, E8181, E8190, E8191
	ECODE04	Motor Vehicle Traffic: *Motorcyclist*	E8102, E8103, E8112, E8113, E8122, E8123, E8132, E8133, E8142, E8143, E8152, E8153, E8162, E8163, E8172, E8173, E8182, E8183, E8192, E8193
	ECODE02	Motor Vehicle Traffic: *Pedalcycle*	E8106, E8116, E8126, E8136, E8146, E8156, E8166, E8176, E8186, E8196
	ECODE03	Motor Vehicle Traffic: *Pedestrian*	E8107, E8117, E8127, E8137, E8147, E8157, E8167, E8177, E8187, E8197
	ECODE05	Motor Vehicle Traffic: *Other*	E8104, E8105, E8108, E8114, E8115, E8118, E8124, E8125, E8128, E8134, E8135, E8138, E8144, E8145, E8148, E8154, E8155, E8158, E8164, E8165, E8168, E8174, E8175, E8178, E8184, E8185, E8188, E8194, E8195, E8198
	ECODE10	Motor Vehicle Traffic: *Unspecified*	E8109, E8119, E8129, E8139, E8149, E8159, E8169, E8179, E8189, E8199

Table 1.1: E-Code Grouping (Continued)

Intent	Variable	Label	Codes
Unintentional (continued)	ECODE06	Pedal Cyclist, Other	E8003, E8013, E8023, E8033, E8043, E8053, E8063, E8073, E8206, E8216, E8226, E8236, E8246, E8256, E8261, E8269, E8271, E8281, E8291
	ECODE07	Pedestrian, Other	E8002, E8012, E8022, E8032, E8042, E8052, E8062, E8072, E8207, E8217, E8227, E8237, E8247, E8257, E8260, E8270, E8280, E8290
	ECODE08	Transport, Other	E8000, E8010, E8020, E8030, E8040, E8050, E8060, E8070, E8001, E8011, E8021, E8031, E8041, E8051, E8061, E8071, E8008, E8018, E8028, E8038, E8048, E8058, E8068, E8078, E8009, E8019, E8029, E8039, E8049, E8059, E8069, E8079, E8200, E8210, E8220, E8230, E8240, E8250, E8201, E8211, E8221, E8231, E8241, E8251, E8202, E8212, E8222, E8232, E8242, E8252, E8203, E8213, E8223, E8233, E8243, E8253, E8204, E8214 ,E8224, E8234, E8244, E8254, E8205, E8215, E8225, E8235, E8245, E8255, E8208, E8218, E8228, E8238, E8248, E8258, E8209, E8219, E8229, E8239, E8249, E8259, E8262 - E8268, E8272, E8273, E8274, E8275, E8276, E8277, E8278, E8279, E8282, E8283, E8284, E8285, E8286, E8287, E8288, E8289, E8292, E8293, E8294, E8295, E8296, E8297, E8298, E8299, E8310 - E8319, E8330 - E8459, E8494,E8496, E8498
	ECODE11	Natural/Environmental	E9050 - E9056, E9059, E9060 - E9064, E9065, E9069, E9000 - E9049, E9057- E9058, E9068, E907, E908, E9080 - E9089, E909, E9090 - E9099, E9280 - E9282
	ECODE13	Overexertion	E927

Table 1.1: E-Code Grouping (Continued)

Intent	Variable	Label	Codes
Unintentional (Continued)	ECODE16	Poisoning	E8500-E8699
	ECODE18	Struck by Object	E9170 - E9179, E916, E917
	ECODE17	Suffocation	E9130 - E9139, E911, E912, E913
	ECODE22	Unintentional, Other	E846 - E848, E914 - E915, E918, E9210 - E9219, E9230 - E9239, E9250 - E9269, E9288, E887, E9289
Intentional-Self-Inflicted	ECODE25	Cut/Pierce	E956
	ECODE26	Firearms	E9550 - E9554
	ECODE23	Poisoning	E950, E9500 - E9529
	ECODE24	Suffocation/Hanging	E9530 - E9539
	ECODE27	Self-Inflicted, other	E954, E9555, E9556, E9559, E9570 - E9579, E9580, E9581, E9582, E9583, E9584, E9585, E9586, E9587, E9588, E9589
Intentional-Assault	ECODE29	Battering	E9670 - E9679
	ECODE32	Cut/Pierce	E966
	ECODE28	Fight-Unarmed	E9600
	ECODE26	Firearms	E9650 - E9654
	ECODE30	Assault, Other (pdds) Other Homicide (vs)	E961, E9683, E9680, E9681, E9685, E963, E964, E9682, E9620 - E9629, E9601, E9655 - E9659, E9684, E9688, E9689
Undetermined/ Other	ECODE34	Undetermined Intent	E984, E9850 - E9854, E986, E9870 - E9879, E9881, E9882, E9887, E9885, E9886, E9883, E9800 - E9829, E9830 - E9839, E9855, E9880, E9884, E9888, E9889
	ECODE33	Legal Intervention/War	E970, E972, E973, E974, E975, E971, E978, E9900 - E9919, E992 - E994, E996, E9970 - E9972, E995, E9978, E998, E976, E9979
Late Effects	ECODE35	Late Effects	E9290 - E9295, E9298, E9299, E959, E969, E977, E989, E999

Table 1.2: FHOP Injury Grouping

Variable Name	Label	ICD-9-CM Principal Diagnosis
FX	Fracture	800, 80,1 802, 803, 804, 805, 806, 807, 808, 809, 810, 8110, 8111, 812, 813, 814, 815, 816, 817, 818, 819, 820, 821, 822, 823, 824, 825, 826, 827, 828, 8290, 8291
JNTDISL	Dislocation	830, 831, 832, 833, 834, 835, 836, 837, 838, 839
SPRNSTR	Sprains and strains	840, 841, 842, 843, 844, 845, 846, 847, 848
CONCUS	Concussion	850
INJCNS	Injury to central nervous system	851, 852, 853, 854, 950, 951, 952, 953, 954, 955, 956, 957
INJCORP	Injury to chest, abdomen and pelvis	860, 861, 862, 863, 864, 865, 866, 867, 868, 869
INJWND	Wounds	870, 871, 872, 873, 874, 875, 876, 877, 878, 879, 880, 881, 882, 883, 884, 885, 886, 887, 888, 889, 890, 891, 892, 893, 894, 895, 896, 897
INJOTH	Other injuries	900, 901, 902, 903, 904, 905, 906, 907, 908, 909, 910, 911, 912, 913, 914, 915, 916, 917, 918, 919, 920, 921, 922, 923, 924, 925, 926, 927, 928, 929, 958, 959, 990, 991, 992, 993, 9940, 9943, 9944, 9945, 9946, 9947, 9948, 9949, 995
FBOTH	Foreign body	930, 931, 932, 933, 934, 939
FBGI	Gastrointestinal foreign body	935, 936, 937, 938
INJBURN	Burns	940, 941, 942, 943, 944, 945, 946, 947, 948, 949
INJPOIS	Poisonings	960, 961, 962, 963, 964, 965, 966, 967, 968, 969, 970, 971, 972, 973, 974, 975, 976, 977, 978, 979, 980, 981, 982, 983, 984, 985, 986, 987, 988, 989
INJDRWN	Drownings	9941

Table 1.3: Correspondence between E-Code and Injury Grouping

Discharge E-Code	Grand Total		Discharge Diagnosis Codes						
	Count	Percent	Burns	CNS	Chs/Ab/Pel	Concus	Drown	Fracture	Jnt Disloc
Total	162,233	100.0	4,392	11,567	11,968	4,604	1,596	57,967	1,832
%Diagnosis Codes			2.7	7.1	7.4	2.8	1.0	35.7	1.1
Unintentional									
Cut/Pierce	4,763	2.9	2	288	143	1		271	5
Drowning/Submersion	1,713	1.1	2	10	3	2	1,528	32	2
Fall	31,583	19.5	9	2,498	804	1,420	5	21,709	479
Fire/Burn:									
Fire/Flames	1,019	0.6	826		1			24	
Hot Object/Substance	2,789	1.7	2,598			1		3	
Firearms	1,509	0.9		89	255			441	
Machinery in Operation	1,085	0.7	28	34	16	6		454	7
Motor Vehicle:									
Occupant	21,862	13.5	76	2,923	2,466	1,317	6	8,351	248
Motorcyclist	2,681	1.7	16	248	273	113		1,513	53
Pedalcycle	1,987	1.2	2	313	86	124		1,064	12
Pedestrian	7,108	4.4	22	1,068	376	333		3,656	58
Other	305	0.2	4	48	17	13		149	7
Unspecified	911	0.6	3	123	68	15		424	24
PedalCyclist, Other	4,144	2.6		405	381	333		2376	41
Pedestrian, Other	629	0.4	4	50	55	20		349	3
Transport, Other	3,453	2.1	21	292	343	174	2	1,994	52
Natural/Environmental	3,501	2.2	7	28	62	10	3	197	5
Overexertion	2,782	1.7	14	18	1			1,307	317
Poisoning	8,233	5.1	82	8	6	4		10	
Struck by Object	7,139	4.4	1	496	449	306		3,296	200
Suffocation	1,040	0.6	1	5	1		2	4	
Other	7,448	4.6	385	158	144	47	4	2,009	171
Intentional, Self Inflicted									
Cut/Pierce	1,721	1.1		20	92	1		1	
Firearms	273	0.2		72	43			85	
Poisoning	12,644	7.8	6	4	2	1		3	
Hanging/Suffocation	207	0.1	1					4	
Self-Inflicted, Other	606	0.4	39	25	15	5	1	135	1
Intentional, Assault									
Battering	1,102	0.7	17	185	35	7		290	2
Cut/Pierce	5,066	3.1		108	2,183	6		98	3
Fight	3,410	2.1	1	394	174	157		1,695	24
Firearms	9,905	6.1		417	2,707	3		2,661	2
Other	3,525	2.2	75	496	173	123	4	1,270	9
Intentional, Other									
Undetermined, Other	1,049	0.6	33	51	69	3	2	211	5
Legal/War	259	0.2	1	12	45	1		82	1
Late Effects	382	0.2	12	58	28	3		111	23
Adverse Effects	736	0.5	24	42	70	2	2	147	8
NoE-codes	3,664	2.3	80	581	382	53	37	1,541	70

Table 1.3: Correspondence between E-Code and Injury Grouping (Continued)

Discharge E-Code	Grand Total		Discharge Diagnosis Codes					
	Count	Percent	Jnttraum	Other	Poison	Spr/Strain	Wound	None
Total	162,233	100.00	301	8,781	18,152	2,389	17,950	20,734
% Diagnosis Codes			0.2	5.4	11.2	1.5	11.1	12.8
Unintentional								
Cut/Pierce	4,763	2.9	1	284	3	9	2,964	792
Drowning/Submersion	1,713	1.1		7		3	2	122
Fall	31,583	19.5	33	888	11	537	798	2,392
Fire/Burn:								
Fire/Flames	1,019	0.6		7	80		3	78
Hot Object/Substance	2,789	1.7		5	1		2	179
Firearms	1,509	0.9	1	60			575	88
Machinery in Operation	1,085	0.7		115		3	391	31
Motor Vehicle:								
Occupant	21,862	13.5	3	2,072	11	372	2,044	1,973
Motorcyclist	2,681	1.7		132		29	193	111
Pedalcycle	1,987	1.2		141		14	157	74
Pedestrian	7,108	4.4		708		49	453	385
Other	305	0.2		22		2	15	28
Unspecified	911	0.6	1	37	1	8	55	152
Pedal Cyclist,Other	4,144	2.6		136	1	30	231	210
Pedestrian,Other	629	0.4		77		1	43	27
Transport,Other	3,453	2.1	2	191	2	36	214	130
Natural/Environmental	3,501	2.2		392	558	2	930	1,307
Overexertion	2,782	1.7	30	43		744	34	274
Poisoning	8,233	5.1		90	6,790	1	6	1,236
Struck by Object	7,139	4.4	1	606	1	214	746	823
Suffocation	1,040	0.6		81	1	1	1	943
Other	7,448	4.6	10	429	10	216	654	3,211
Intentional, Self-Inflicted								
Cut/Pierce	1,721	1.1		40	5		306	1,256
Firearms	273	0.2		4	2		50	17
Poisoning	12,644	7.8		3	10,152		6	2,467
Hanging/Suffocation	207	0.1		86	1	1		114
Self-Inflicted, Other	606	0.4		21	9	7	32	316
Intentional, Assault								
Battering	1,102	0.7		199	2	1	19	345
Cut/Pierce	5,066	3.1		202			2,361	105
Fight	3,410	2.1		214	2	19	250	480
Firearms	9,905	6.1	2	466	2	1	3,293	351
Other	3,525	2.2		237	40	10	616	472
Intentional, Other								
Undetermined, Other	1,049	0.6		45	307	1	121	201
Legal/War	259	0.2		22		2	49	44
Late Effects	382	0.2	23	77	3	18	26	
Adverse Effects	736	0.5	1	321	81	6	32	
No E-codes	3,664	2.3	193	321	76	52	278	

Chapter 2

Linking Discharge Records into Episodes of Care

All records associated with one "person" were found and linked into episodes of care. A number of issues arose in the linkage that had to be resolved. Various data cleaning activities were carried out. After linkage was complete, reliability statistics were calculated.

Episode of Care

An episode of care (EOC) is the period initiated by patient presentation with a diagnosis of a clinical condition and concluded when the condition is resolved.[8] Time horizon (also referred to as the period of the study) is a critical issue for evaluating the personal, social, and economic burden of a particular clinical condition. For example, the California Hospital Outcomes Project defined an EOC for a heart attack patient as the period from first admission to an acute care hospital for a heart attack to final discharge within 30 days, including transfers or readmissions to another hospital.[9]

In the case of injuries, we are uncertain as to the time horizon. Depending on the nature and severity of the injury, some patients may be admitted repeatedly. For example, multiple admissions for the same injury episode may be needed to complete the planned course of treatment, to treat late effects of the injury, or to treat adverse effects of prior injury-related admissions. Additionally, transfers from one facility to another as a result of an injury are common.

It also may be difficult to identify the EOC because of problems associated with coding hospital discharge records. Because of coding errors or omissions in the discharge abstract record, earlier or later hospitalizations associated with the same episode may not show the injury status. For example, the first hospital admitting the patient may not have recorded the injury or its mechanism. For this reason, coding rules permit a receiving hospital to indicate an E-code if they do not know that the transferring hospital recorded the injury. These rules may result in both facilities showing an E-code for the same injury EOC.

Failure to identify the series of admissions associated with an injury will underestimate the personal, social, and economic burden. It also will result in over-estimating the true number of injuries.

---

8 Schulman KA, Yabroff KR, Kong J, Gold KF, et al. (1999) A claims data approach to defining an episode of care. Health Services Research, June 1999, 34(2): 603:621.

9 Romano PS, Remy LL, Luft HS. (1993). Report of the California outcomes project: Volume 2, Technical Appendix. Sacramento, California: Office of Statewide Health Planning and Development.

Record Linkage

To obtain a non-duplicated count of injuries resulting in hospital admission, we began by linking records using the methodology and software developed by the California Hospital Outcomes Project.[10] The methodology requires (1) the encrypted SSNC and/or (2) a combination variable created out of the DOB, sex, and ZIP of residence. This secondary variable is used to soft match SSNC, to confirm a linkage, or to match records lacking a SSNC to one having a SSNC. After linkage is made, a new variable is created (SSNC2) to identify records treated as coming from the same person.

By the end of the process, linked sets were an exact or soft match on SSNC (or had picked up an SSNC if the record did not have one), DOB, and ZIP, and an exact match on sex. Two SSNC were considered to soft match if 7 of 9 digits agreed in any position.

After linkage, 120,818 records resolved into one-record injury episodes, and 91,727 records had 2 or more records for 29,129 SSNC2, a total of 149,947 SSNC2. Records in the multiples file had been linked but had not yet been assigned into EOC. Thus, subsequent pulls against the master files had identified 50,312 additional records potentially associated with the initial injury records.

We also found and made a separate file of 3,950 records in the study age range that indicated the patient had died but which PDDS coders had not identified as an injury case. These records were used in later steps to merge the PDDS with the California Vital Statistics (VS) death certificates file. Specifically, these records were used to link records that VS coders indicated were injured but which PDDS coders did not indicate were injured.

Issues that Arose in Linking Episodes of Care

Table 2.1 illustrates seven common issues that arose in the multiple records file. The examples are from the first 75 records in that file. These consisted of 31 SSNC2 sets of 2 to 5 records, all of which resolved into one of the issues in Table 2.1.

In the examples, SSNC = SSNC2 and DOB and sex were the same on these record sets. That is, we had an exact match on these key variables. ZIP of residence was identical on all but four sets. On these four sets, the 3-digit ZIP code (ZIP3) was identical on all records, and digit 4 and/or 5 differed.

The examples are as follows:

1. The classic episode of care. In this example, a patient is admitted through the emergency room for a motor vehicle occupant injury, transferred to another facility on day 7, received in the other facility on day 7, and discharged on day 46. Failure

---

[10] Romano PS, Remy LL, Luft HS. (1996) Report of the California Hospital Outcomes Project: Acute myocardial infarction, Volume 2, Technical Appendix. Sacramento, California: Office of Statewide Health Planning and Development.

to link the episode would under-estimate injury-related hospitalization days and total charges.

2. More than one record E-coded. This is a classic episode but with both records E-coded. Failure to link records would have incorrectly increased the number of injuries for motorcycle accidents.

3. Index record E-coded, short interval to readmission. In this example, observe the 2-day interval to readmission. Notice that the second admission also was from the emergency room. Was this a readmission due to premature discharge for the injury or for another illness that may or may not be injury-related?

4. Both records E-coded with same injury, short interval to readmission. In this example, we have a 3-day readmission interval. The second had a routine admission source with routine discharge for a 1-day stay. Is this a premature discharge, planned readmission for routine follow-up treatment, or admission for another illness?

5. Index record E-coded, long interval to readmission, possible records not linked. This patient was discharged to another facility on day 24, but did not show up again until day 89 with late effects, transferred on day 115 to another facility and discharged on day 117. Records 2 and 3 correctly showed late effects of the initial injury.

6. Different E-codes, interval to readmission. In this particular example, the index admission appears to have been in another facility whose record was not found. The patient stayed 3 days and was readmitted routinely a week later for a 4-day stay. Is this a second injury, planned readmission for follow-up treatment to the original injury, or the result of premature discharge for the original injury?

7. Four or more admissions, varying intervals between admissions. This involved 33,921 records (37% of multiples) for 4,927 SSNC2 (17% of multiple SSNC2) with 4 to 103 records each. These sets were the most problematic and fell into two primary categories: those with isolated "injuries" incidental to treatment for a condition such as cancer or HIV, and those with mental illness or substance abuse diagnoses plus one or more injury records. Table 2.2 shows the ICD-9-CM codes used to group these diagnoses.

Cancer or HIV Diagnoses. A subset of SSNC2 (543 SSNC2, 0.36%) had records indicating cancer (N = 411, 0.28%) or HIV (N = 139, 0.10%) diagnoses. During the 3-year study period, 11% had 1 record, 14% had 2, and 75% had 3 to 72 records, for a total of 4,695 records, with 615 records indicating an injury. Mental illness diagnoses associated with late-state AIDS (N = 134) occurred late if at all in multi-record sets. Substance abuse diagnoses were found for 72 SSNC2, including 44 with a dual diagnosis. The median age was 13, and 75% were below age 20. In virtually all records, cancer or HIV was the principal diagnosis and the principal procedure was related to treatment for that diagnosis (i.e., radiation, transfusions).

A manual review was conducted of all multi-record sets for cancer and HIV patients with 4 or more records. A very small group, primarily over age 21 also with mental health or substance abuse diagnoses, seemed to have been "injured" within the intent of this study (i.e., suicide attempts, fall, motor vehicle). However, when examined over the entire group, most "injuries" seemed to be miscoded adverse effects or unintended consequences following treatment for the primary diagnosis. After linkage with the death file, all patients with cancer or HIV diagnoses on any record were dropped from the study.

Mental Illness or Substance Abuse Diagnoses. A total of 27,311 SSNC2 (18%) had mental illness or substance abuse diagnoses on their records, as either a primary or secondary diagnosis. This included 17,586 with a mental illness diagnosis, and 15,315 with a substance abuse diagnosis. Of these, 5,590 were dually diagnosed. During the 3-year study period, 50% had one record, 25% had 2 records, and 25% had 3 to 66 records, for a total of 54,231 records, of which 36,550 were injury-related. With a median age of 19, 95% of these patients were age 12 years or older.

A manual review was conducted of all multi-record sets with 4 or more records and 2 or more injury records. Most cases had different types of injuries spread over multiple episodes of care. The multiply-injured were more likely to show up in the same county in the same or multiple ZIPs within that county.

To have retained multiple injury episodes involving SSNC2 with mental illness or substance abuse diagnoses would have inflated the injury rate for a given ZIP, particularly in case 7, in that one individual would appear multiple times. That is, a given ZIP injury rate would be based incorrectly on many admissions for one person, rather than one injury for each of many people.

ZIPS (N = 795) in 14 counties accounted for 11,740 (85%) of these unusual records. In these counties, 208 ZIPs had 20 to 103 records each, and accounted for 7,359 records associated with 1,925 unusual SSNC. Although each unusual SSNC2 by definition had a minimum of 4 records in the larger database, the average number of records within target counties and target ZIPs varies. These unusual SSNC2 tended to have high mobility within a small geographic area, occasionally crossing over into adjacent counties. The rest of their records will be in other counties and/or ZIPs. In this study, we assigned the county to be that which occurred the most frequently, and the ZIP to be that ZIP which occurred the most frequently within the assigned county.

Table 2.3 displays the 14 counties and ZIPs associated with these unusual SSNC2. This table identifies the number of discharge records, and number of unusual SSNC2 countywide and within the target ZIPs.

These counties and hospitals treating most residents in these ZIPs may wish to consider reviewing local applications of discharge abstract coding rules and/or examine local efforts to coordinate mental health services for adolescents and young adults. Repetitive admissions may indicate a breakdown in overall care coordination. These may include failure to identify post-discharge needs, communicate those needs to outpatient providers or caregivers, or educate patients about medication regimens.[11]

Additional Data Cleaning

Before proceeding to create the single record EOC file from the multiples file, we needed to eliminate redundant records, correct bad linkages, identify various types of problematic cases, and find a county of residence for records lacking one. Finally, because we were interested in making a better estimate for the social burden of injuries, we developed an algorithm to replace charges for records lacking one. This problem is created because Kaiser hospitals and children's hospitals are exempt from reporting their charges to OSHPD.

Correcting Linkage Problems

Redundant records are created when 2 or more records exist for the same person for the same admission to the same hospital. We identified 37 redundant record sets, including one with two death records. After review of the sets, one record in the set was discarded.

Bad linkages can occur under a variety of circumstances. We identified one linkage that had to be corrected. We broke up the SSNC2 set by reassigning a new SSNC2 to the single record that did not belong.

We checked for two types of particularly problematic situations that would need to be fixed in order to improve results when we linked to the death file. First, we identified 119 SSNC2 with both male and female sex. We broke up these SSNC2 by adding the suffix "M" or "F" to the appropriate identifier. For a given SSNC2, we removed any records that were found after a patient died.

Because of redundant records and separating SSNC2 due to bad linkages or differing sex, we created new SSNC2 sets, including some with no injuries, and eliminated 262 records.

Assigning County of Residence

We found 4,098 records from 260 ZIPS that did not show a county of residence, a variable critical to estimating both county and ZIP-level injury rates.[12] Most of these ZIPs had other records in the source master files (all discharges, all ages) containing the county of residence. Therefore, we returned to the masters and made a dataset containing all ZIP5s, ZIP3s, and their associated county, saving the county most frequently recorded for each ZIP5. Then we merged this file with the

---

[11] Wray NP, Petersen NJ, Souchek J, Ashton, C., et al. The hospital multistay rate as an indicator of quality of care. Health Services Research, Aug 1999, 34(3):777-790.

[12] County of residence is assigned by OSHPD based on the most current information available from the USPS. About 40% of ZIPs span county boundaries. For these ZIPS, OSHPD assigns the ZIP to the county with the most discharges. Personal communication, Candace Diamond, OSHPD, August 31, 1999. Over the study period, we found one ZIP split into two counties (91361, Ventura (1995) and Los Angeles (1996 and 1997)). We assigned all records for this ZIP to Los Angeles County.

injury files to impute a county of residence for ZIPs missing them. If no other ZIP5 record existed, we assigned the county with the most ZIP3.

Modifying Charges

Because we wished to estimate the social burden of childhood illness, we modified charges for records lacking them (n = 14,036), as follows.

First, all records admitted and discharged on the same day, (length of stay (LOS) = 0 days) were changed to a LOS of 1 day. Because OSHPD coding rules require charges to be reported for the year, the LOS upper range was truncated at 365.

Next, we focused on records with charges (n = 202,459). We divided total charges by LOS to obtain charges per day. After a series of diagnostic runs, we defined a charge inlier as a record with a charge less than $100 per day and a charge outlier as one with a charge greater than $50,000 per day. Following another series of diagnostic runs, inlier records in mental health DRGs (426, 427, 428, 429, 430, 435, 436) were recalculated by multiplying $50 times the LOS. Other inlier records were recalculated by multiplying $100 times the LOS. Finally, after another series of diagnostic runs, outlier records were recoded as follows: If the total charge was more than $1 million and LOS was between 8 and 20 days, we replaced the total charge by multiplying $50,000 times the LOS. Otherwise, we truncated the total charge on the record at $1.5 million.

Using the file with truncated charges, we obtained the value of the 50%ile for charges per day within each DRG and year. We multiplied this value times LOS to impute total charges where it was missing. Next, for cases still missing a charge, we used the DRG median, and finally the year median over all DRGs.

In summarizing charges to the SSNC2, ZIP, and county levels, we used the variable based on the truncated and replaced values. We truncated the lower and upper bounds of charges and used the median rather than the mean to replace records missing charges because charges are not distributed normally and we wanted to be conservative when summarizing charges.

Making the Episode-of-Care File

Given data issues that had been identified, we decided to use only the last injury episode for a given SSNC2. This had several advantages. First, by definition for cases resolving to one record, that was the last episode. Second, in the case of an SSNC2 with multiple records, using the last injury episode got us closer in time to the last hospital outcome record for the patient. In the case of patients who died, this would be a decided advantage in merging with the Vital Statistics death certificate file. At this stage, we dropped all multi-record injury episodes that began when the patient was 25 years old (1,413 SSNC2). To summarize the available information in the multiples file and to maximize the likelihood of finding patients who had died, we created a file with one record per SSNC2 (n = 29,129) containing the following data:

- Over all records:

  Total number of records, days, and charges over all admissions.

  Number of records with an injury indication; number of injury-related records, days, and charges. Injury-related records were defined as date neighbors (records 0 to 1 day apart containing an indication of an injury on one or more records), and extended through the last late effect or adverse effect record found before a subsequent non-injury-related admission.

  Count over all diagnosis fields the number of records with substance abuse, mental illness, cancer, or HIV diagnoses; number of records with adverse effects or late effects diagnoses.

  E-code most often found, and the MDC and injury variable found on the record with that E-code.

  DOB and race/ethnicity most often found.

  Payor source most often found. Number of times the patient was uninsured and number of times Medicaid was the anticipated payor.

  ZIP and county of residence most often found.

- From the first record with a valid E-code or valid injury in the last injury-related EOC:

  Sex, DOB, admission date, race/ethnicity, anticipated payor, hospital facility identifier, ZIP of residence, facility ZIP, county of residence, county of hospital.

  The E-code, injury, and major diagnostic category associated with that record.

- From all records in the last injury-related EOC:

  Length of stay and charges, number of diagnoses, number of procedures and classification of procedures,[13] adverse or late effects, and complications of care.[13] [14]

- From the last record in the last injury-related EOC:

  Discharge date, disposition, ZIP and county of residence.

---

[13] Elixhauser, A. (1996). Clinical Classifications for Health Policy Research, Version 2: Software and user's guide. (AHCPR Pub. No. 96-0046.) Healthcare Cost and Utilization Project (HCUP-3) Research Note 2. Rockville, MD: Agency for Health Care Policy and Research.

[14] American Nurses Association. (1997). Implementing Nursing's Report Card. A Study of RN Staffing, Length of Stay and Patient Outcomes. Washington, DC: American Nurses Publishing.

- From the last record:

  ZIP and county of residence, DOB, admission date, and discharge date on the last record to facilitate linkage with the death file.

PDDS Linkage Reliability

SSNCs were present for about 50% of injury records for patients age 0 to 14, 62% for age 15 to 19, and 74% for age 20 to 24. There was a slight trend that did not vary by age for the percent of present SSNCs to increase about 2 to 3 percent between 1995 and 1997. SSNC were found for about 60% of injury records for Non-Hispanic White patients, 70% for African-American patients, 66% for Asian, and 55% for Hispanic and Others. There was no important age by race interaction with respect to the presence or absence of the SSNC. Probably due to their younger age, injury patients were less likely to have a SSNC than older patients admitted for heart attack or congestive heart failure.[15, 16]

Based on the availability of the SSNC, we are more certain that we correctly linked records associated with episodes of care as age increased and for Black injury victims. Episodes for younger children and for those of Hispanic and Other race/ethnicity were more likely to resolve (perhaps incorrectly) into single record episodes.

In the linkage algorithm, linkages based on soft-matches occurred infrequently. Among multi-record SSNC2, all linked records were, by definition, in full agreement on sex. We compared information found on the first injury record in the last injury EOC to the information most frequently found and last found over all records for the multi-record SSNC2 sets during the 3-year study period. Differences may be due to incorrect linkage, typographic errors, or in the case of ZIP, moving from one community to another. Agreement above .75 may be taken to represent excellent agreement beyond chance.[17] Among multi-record SSNC2 sets:

- Dates of birth agreement: first:last 98.5%, first:most 98.9%, last:most 99.0%.
- County of residence agreement: first:last 97.2%, first:most 97.8%, last:most 97.6%.
- 3-Digit ZIP agreement: first:last 94.1%, first:most 95.4%, most:last 94.9%.
- 5-digit ZIP agreement: first:last 89.0%, first:most 91.5%, most:last 90.6%.

---

[15] Romano PS, Remy LL, Luft HS. (1993). Report of the California outcomes project: Volume 2, Technical Appendix. Sacramento, California: Office of Statewide Health Planning and Development.

[16] Alexander M, Grumbach K, Remy L, Rowell R, Massie B. (1999) Congestive heart failure hospitalizations and survival in California: Patterns according to race/ethnicity. American Heart Journal, 1999;137:919-27.

[17] Fleiss JL. (1981) Statistical Methods for Rates and Proportions, Second Edition. New York: John Wiley & Sons.

- Race/Ethnicity agreement: first:most 89.9%. Most disagreement was among cases where one record showed "Other Race", but other records for the SSNC2 indicated a different race/ethnicity, which agreed 92.7%.

- After excluding cases with "Other Race" on any record, kappa coefficients for first:most were: Non-Hispanic White, K = .885, CI = .879-.890; African-American, K = .896, CI = .888-.904; Hispanic, K = .871, CI = .865-.877; Asian, K = .873, CI = .861-.886.

- SSNC agreement among those with an SSNC on at least one record: first:last 93.3%, first:final 96.6%, first:assigned 98.0%, first:most 98.9%.

We also compared anticipated payor before linkage for all records, after linkage for the first record in the last injury EOC, and among the multi-record set.

- Agreement on payment source: first:most 82.2%.

- Medicaid was the most frequently found payment source: 38% overall before linkage, and 35% after linkage for the first record in the last injury EOC. Among the multi-record set, 82% showed Medicaid as the anticipated payor on at least one record.

- Apparent failure to qualify for any type of insurance coverage resulted in 16% of all records indicating the patient remained uninsured at discharge. After linkage, 19% remained uninsured. Among the multi-record set, 22% were uninsured at least once. This included 10% persistently uninsured for 2 or more admissions.

On the basis of these results, we believe we linked records for injury patients with an acceptable degree of reliability. To the extent that we were unable to correctly link multiple records for the same EOC, population rates will be higher. Overall, however, ZIP-level rates will be lower (i.e., more conservative) than if we had not linked.

Table 2.1: Issues in Multiple Record SSNC2

Issue Type	Admission Date	Discharge Date	Admission Source	Disposition	E-code	Note
1	1	7	ER	Other Facility	MV Occupant	Classic EOC
	7	46	Other Facility	Routine	None	
2	1	3	ER	Other Facility	Motorcycle	Both records E-coded.
	3	15	Other Facility	Routine	Motorcycle	
3	1	2	ER	Routine	MV Occupant	2-day interval to readmit. Premature discharge or late effects?
	4	6	ER	Routine	No Ecode	
4	1	4	ER	Routine	MV Pedalcycle	Both records E-coded. 3-day interval to readmit. Premature discharge? 2nd injury? Planned readmission?
	7	7	Routine	Routine	MV Pedalcycle	
5	1	24	ER	Other Facility	Firearms (assault)	Discharged to other facility, no next record. 2-month lag to next record, 2 records comprise an EOC. Both records correctly recoded as late effects.
	89	115	Other Facility	Other Facility	Late Effects	
	115	117	Other Facility	Routine	Late Effects	
6	1	4	Other Facility	Routine	MV Pedestrian	Both records E-coded. 6-day interval to readmit. Premature discharge? Late effects? Second injury? Planned readmit?
	10	14	Routine	Routine	Unintentional, Other	
7	1	2	ER	Routine	Cut/Pierce (suicide)	Multiple different or no injury records. Note transfer to other facility, but no other facility record follows
	3	8	ER	Other Facility	Poisoning	
	20	21	ER	LAMA	MV Pedestrian	
	32	34	ER	Routine	No Ecode	
	54	55	ER	Died	Poisoning	

Table 2.2: ICD-9-CM Classifications for Cancer, HIV, Mental Illness, Substance Abuse

Variable Name	Label	ICD-9-CM Principal or Any Secondary Diagnosis
IMMUNE	HIV	042, 043, 044, 07953, 279, 79571, V08
CANCER	Cancer	140, 141, 142, 143, 144, 145, 146, 147 148, 149, 150, 151, 152, 153, 155, 156 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 230, 231, 232, 2330, 2332, 2333, 2334, 2335, 2336, 2337, 2338 2339, 234
LEUKALL	Acute lymphocytic leukemia	2040
LEUKOTH	Other leukemias	2041, 2042, 2048, 2049, 205, 206, 207, 208
NEOPBEN	Benign neoplasm	210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229
CACERV	Cervical cancer	2331
NEOPOTH	Other neoplasm	235, 23,6 237, 238, 239, 7842
ALCDRUG	Alcohol drug disorders	29,1 292, 303, 304, 305, 7903, V586,
PSYCHOT	Other psychotic disorders	293, 294, 295, 2960, 2961, 2969, 297, 298, 299, 301, 302, 306, 310, 316, V40, V663, V665, V673
PSYDEPR	Depression	2962, 2963, 2964, 2965, 2966, 2967, 2968, 3004, 3090, 3091, 311
PSYNURO	Other mental disease	3000, 3001, 3002, 3003, 3005, 3006, 3007, 3008, 3009, 3070, 3072, 3073, 3074, 3076, 3077, 3078, 3079, 308
NUTEATD	Eating disorders	3071, 3075
PSYADJ	Adjustment reaction	3092, 3093, 3094, 3095, 3096, 3097, 3098, 3099
PSYCOND	Conduct disorder	312
PSYEMOT	Emotional disturbance	313

Table 2.3: Counties and ZIPs with Many Multi-Record SSNC2

County	N of ZIPs	N of Recs	Unique SSNC2	Avg. Recs Per SSNC2	Recorded ZIP	N of Disc.	N of SSNC2
Alameda	12	352	88	4.0	94536	28	5
					94538	24	7
					94541	22	10
					94546	20	9
					94587	24	10
					94601	37	15
					94602	42	19
					94603	35	11
					94605	40	13
					94607	27	8
					94621	26	8
					94704	27	5
Contra Costa	7	247	49	5.0	94509	20	3
					94521	37	7
					94553	32	7
					94565	31	6
					94801	44	7
					94804	41	13
					94806	42	13
Fresno	9	296	50	5.9	93701	24	9
					93702	32	11
					93703	24	3
					93704	27	6
					93705	28	4
					93721	26	6
					93722	78	9
					93726	22	5
					93727	35	9
Kern	6	190	36	5.3	93268	30	3
					93305	25	8
					93306	42	9
					93307	27	9
					93308	35	8
					93309	31	9
Los Angeles	73	2630	471	5.6	90000	20	9
					90001	24	8
					90002	39	11
					90004	57	9

Table 2.3: Counties and ZIPs with Many Multi-Record SSNC2 (Continued)

County	N of ZIPs	N of Recs	Unique SSNC2	Avg. Recs Per SSNC2	Recorded ZIP	N of Disc.	N of SSNC2
Los Angeles (continued)					90011	75	16
					90018	43	15
					90019	43	11
					90022	49	12
					90023	39	11
					90026	37	12
					90028	33	7
					90029	26	7
					90034	28	10
					90037	39	9
					90042	27	7
					90043	32	9
					90044	46	11
					90047	55	10
					90057	34	11
					90059	48	9
					90062	26	8
					90063	22	5
					90065	32	5
					90069	27	2
					90201	51	9
					90242	26	8
					90247	43	9
					90250	47	11
					90255	26	5
					90262	26	8
					90272	28	3
					90280	54	14
					90303	20	4
					90638	31	4
					90640	46	9
					90650	110	22
					90701	26	2
					90706	46	15
					90723	28	11
					90731	22	6
					90732	21	4
					90744	31	6
					90804	27	8
					90805	48	12
					90806	26	6
					90807	27	5

Table 2.3: Counties and ZIPs with Many Multi-Record SSNC2

County	N of ZIPs	N of Recs	Unique SSNC2	Avg. Recs Per SSNC2	Recorded ZIP	N of Disc.	N of SSNC2
Los Angeles (continued)					90813	37	15
					91001	34	7
					91103	27	8
					91306	22	6
					91311	24	5
					91331	65	18
					91335	32	7
					91342	64	15
					91343	28	9
					91351	23	7
					91384	27	2
					91401	24	6
					91402	34	8
					91405	47	11
					91423	25	6
					91605	26	7
					91732	84	23
					91733	29	5
					91744	27	8
					91766	32	13
					91767	40	15
					91770	25	11
					91780	22	7
					93534	29	9
					93535	35	8
					93536	25	5
					93550	32	10
Orange	12	367	65	5.6	90620	29	4
					92647	26	8
					92648	22	3
					92649	26	2
					92667	28	7
					92677	32	8
					92680	20	3
					92683	43	9
					92692	20	3
					92707	22	4
					92804	72	11
					92805	27	10
Riverside	13	482	97	5.0	91720	27	8

Table 2.3: Counties and ZIPs with Many Multi-Record SSNC2 (Continued)

County	N of ZIPs	N of Recs	Unique SSNC2	Avg. Recs Per SSNC2	Recorded ZIP	N of Disc.	N of SSNC2
Riverside (continued)					91760	20	4
					92201	36	9
					92220	36	7
					92223	20	4
					92240	27	6
					92501	81	20
					92503	75	28
					92506	22	7
					92507	29	10
					92509	56	19
					92544	22	8
					92553	31	11
Sacramento	12	366	79	4.6	95608	20	8
					95610	21	4
					95628	28	6
					95660	48	9
					95758	28	7
					95817	36	11
					95820	43	9
					95821	21	6
					95823	48	18
					95825	27	11
					95826	24	7
					95828	22	7
San Bernardino	21	790	170	4.6	91710	29	10
					91730	28	10
					91761	20	5
					92307	25	6
					92316	23	8
					92324	31	14
					92335	69	19
					92336	39	8
					92345	27	11
					92346	52	17
					92354	34	10
					92373	30	9
					92374	50	11
					92376	81	24
					92392	57	12
					92399	27	6

Table 2.3: Counties and ZIPs with Many Multi-Record SSNC2

County	N of ZIPs	N of Recs	Unique SSNC2	Avg. Recs Per SSNC2	Recorded ZIP	N of Disc.	N of SSNC2
San Bernardino (continued)					92404	33	12
					92405	25	11
					92407	34	12
					92410	42	17
					92411	34	10
San Diego	25	945	165	5.7	91910	37	10
					91911	83	17
					91950	53	11
					92019	20	4
					92020	47	9
					92021	41	6
					92026	38	8
					92027	24	8
					92028	28	8
					92040	32	8
					92054	32	10
					92056	43	7
					92064	41	9
					92101	25	11
					92102	42	7
					92103	21	17
					92104	31	14
					92105	74	17
					92110	35	16
					92113	54	13
					92114	28	12
					92116	20	8
					92123	45	13
					92124	24	2
					92139	27	9
San Francisco	4	170	42	4.0	94103	33	8
					94110	69	23
					94112	34	8
					94124	34	12
Santa Clara	3	79	17	4.6	94040	20	5
					95124	22	5
					95127	37	8
Stanislaus	5	211	41	5.1	95307	23	6

Table 2.3: Counties and ZIPs with Many Multi-Record SSNC2 (Continued)

County	N of ZIPs	N of Recs	Unique SSNC2	Avg. Recs Per SSNC2	Recorded ZIP	N of Disc.	N of SSNC2
Stanislaus (continued)					95350	59	15
					95351	50	15
					95355	44	14
					95380	35	6
Ventura	6	234	39	6.0	93001	43	10
					93003	43	9
					93023	52	4
					93030	51	17
					93035	21	3
					93065	24	5

Chapter 3

Linking Hospitalization and Death Records

Annual injury reports typically summarize severe non-fatal injuries resulting in hospitalization separately from fatal injuries.[18] *The goal in linking hospitalization and death records was to obtain a more complete picture of the long-term outcomes of injury. The linkage method may result in a conservative zip-level injury rate. We were willing to accept some level of discrepancies in making matches. The resulting rate will be lower than if we had rejected those discrepancies. In this chapter, we describe methods developed to make the linkages, the rationale behind those methods, and the results. We also describe results of statistical tests to examine reliability of linkages made.*

The PDDS Death Files

In preparing to merge the injury data with the Vital Statistics (VS) death certificate files, we made a file of all patient discharge data set (PDDS) records for California residents age 1 day (excluding newborns) to 25 years at discharge who died in a licensed California inpatient facility in 1995, 1996, and 1997. Three SSNC (encrypted social security numbers) had to be reassigned. After removing redundant records, the resulting file had 3,950 observations.

We created this file because we wanted to protect against the possibility that VS coders may have flagged as injured some cases that PDDS coders had not. While we did not want to search through all PDDS records, we did want to search among patients that PDDS coders said had died.

We concatenated data for injury patients who died (N = 2,607) with this file for a total of 6,557 in-patient deaths.

The Vital Statistics Death Certificate File

The 1995, 1996, and 1997 Vital Statistics (VS) death certificate files were searched to identify all California residents age 1 day to 25 years who died in California, excluding infants with congenital deformities and conditions originating in the perinatal period.[19] In making the VS file, we wanted to protect against the possibility that PDDS coders had flagged as injured some cases that VS coders had not.

All records were classified as to their injury status, again using software provided by the California Department of Health Services.[20] The SSNC was encrypted with

[18] See, for example, Injury Tables, California, 1996 Deaths and Nonfatal Hospitalizations. Prepared by the Injury Surveillance and Epidemiology Section of the California Department of Health Services.

[19] 1989-1997 California Death File Documentation, Vital Statistics Statistical Master File, California Department Of Health Services, Center of Health Statistics.

[20] Because of the way the VS file is structured, it is not possible to identify the injury type or other potential injuries by summarizing the principal diagnosis (i.e., broken bones, concussion), or to identify the affected body system. The PDDS file has separate fields for the principal diagnosis, E-code, and major diagnostic

the same algorithm used to encrypt the PDDS. We found 6 pairs of exact duplicate records in the VS file, and dropped one record of the duplicate pair. Two records with the same SSNC had to be reassigned because they obviously were different people. The resulting file had 19,515 observations: 11,923 (61.1%) had an injury, 15,630 (80.1%) had a SSNC and 9,964 (51.6%) had both.

Table 3.1 summarizes the VS records by death location. Of the injury deaths, 6,025 were reported to have occurred out-of-hospital. An additional 74 injury victims were dead on arrival or the location of their death was unknown. Death in the emergency room or an outpatient department was reported for 3,635 injury victims. Inpatient deaths were reported for 5,880 people (2,189 injury victims, and 3,691 for all other causes).

The PDDS file had identified 6,557 in-hospital deaths from all causes (2,621 injury victims, and 3,950 for all other causes). In addition to the primary purpose of this linkage, i.e., to obtain a non-duplicated count of injury deaths, reconciling differences between the PDDS and VS files became a sub-focus of investigation.

In a series of tests, we identified 153 ZIPs with 4,421 injuries that did not appear to be valid. More than 95% of bogus ZIPs were in the VS file. Since ZIP is a key component of the linkage algorithm, we believe the presence of so many bogus ZIPs on so many VS injury records affected our ability to make linkages and affected reliability statistics we report in subsequent sections. Later in this chapter and also in Chapter Four, we describe the method used to handle these ZIPs.

We identified 295 VS records that lacked a date of birth (DOB). Of these, 279 were coded as 0 years old, including 137 injury victims. Of these injury victims, 86.1% were male, 81% were Hispanic, 96.4% lacked a SSNC, 91.2% died out-of-hospital, and most died in rural counties or counties bordering Mexico. The most frequently occurring injury causes were drowning, natural/environmental, various types of assault and suicide. We concluded these were unlikely to be infants. Since we did not know their age, we discarded all VS records lacking a DOB with an age of 0 when we made the "person-level" file.

Linking PDDS and VS Files

Linkage between the PDDS and VS files was done in a series of steps. First, for SSNC2 with multiple PDDS records, we had to determine which version of the PDDS variable (first, last, most, see Chapter Two) would maximize linkage. Using that information, we implemented a series of steps to link cases that VS coders and OSHPD coders said had died, regardless of injury status. Then we linked injury cases PDDS coders said had lived to cases that VS coders said had died regardless of injury status.

---

category. The VS file strips the E from the E-code and uses this stripped value in lieu of the ICD-9-CM principal diagnosis descriptor. With this coding, the VS file has no possibility of further describing the injury type or the affected body system.

Selecting the "Best" Linkage Variables

Earlier we reported linkage reliability for multi-record SSNC2 within the PDDS. This included concordance between SSNC2 compared with SSNC, among SSNC assigned to the same SSNC2, DOB, race/ethnicity, ZIP of residence, and county of residence. There, agreement generally was highest comparing the "most" variable to others. However, this did not mean necessarily that we would make the most linkages using the "most" variables. For example, the last record for the SSNC2 might provide the most temporally relevant information for linking the VS file.

To determine which SSNC variation might offer the greatest likelihood of making a match, we selected from the linked multi-record file those cases that died (N = 409), keeping those with a SSNC. When merging by the SSNC variation, we required agreement as to the sex of the person to make the link with the VS file. Table 3.2 shows that the SSNC most often found in the PDDS (N = 313) linked to the most VS records (N = 232).

We used the SSNC most often found to calculate agreement between other relevant VS and PDDS variables used to refine and validate linkages. Table 3.3 summarizes the result of efforts to find the best linkage variables based on DOB, ZIP, and county of residence. We printed all cases where a PDDS DOB version disagreed with the VS DOB (N = 7). All DOB disagreements would be accepted using the linkage algorithm for soft matches.

Table 3.4 shows the kappa statistic and confidence interval for the VS race/ethnicity and the PDDS race/ethnicity on the first record in the last injury EOC and the race/ethnicity most often found. Most disagreement was among cases where one or another file showed "Other Race". After excluding "Other Race" cases, kappa statistics were best using race/ethnicity most often found.

Only about 50% of linked test cases agreed as to the injury causing death. Table 3.5 lists those cases from the first 50 in the test file where the PDDS E-code disagreed with the VS E-code. DOB agreed exactly on all but 2 of the 50 cases, and those two cases soft-matched. The most common disagreement was that PDDS coders said the person had been injured, but the injury was not recorded in the VS file. The second most common disagreement was between the E-code most often found and the VS E-code, as in records 5, 6, 7, 14, 18, and 40. This suggests a history of multiple injuries, resulting finally in death after the last hospital discharge. Records 18, 23, 44, and 49 indicate disagreement as to the type of injury killing the individual, and disagreement as to intent.

Based on these results, we decided to use the following multi-record EOC file variables when merging with the VS file:

- SSNC, DOB and race/ethnicity most often found,
- ZIP5, ZIP3, and county of residence on the very last record,
- the last valid E-code variable.

Linking the PDDS Death File and the VS File

The primary goal of this step is to link cases PDDS coders said had died with cases in the VS file. We began by separating the PDDS and VS files based on the availability of a SSNC (PDDS: no SSNC, N = 2,855; SSNC, N = 3,702; VS: no SSNC, 3,885; SSNC, 15,630). We developed a multi-step method using different combinations of SSNC, exact or soft matches of DOB or DOD, ZIP5 or ZIP3. Once we had linked using the various combinations, we calculated an agreement score between other variables in the PDDS and VS files not used to make the link. Other comparison variables included the above plus race/ethnicity, injury status (four variations), injury date compared with admission date, DOD compared with discharge date, and county of residence. The agreement score had a possible maximum of 10, depending on variable(s) used to make the link, and sex always was required to agree. At each stage, two raters reviewed printouts on all cases and agreed on acceptance criteria. These criteria were hard-coded into the program and executed, non-accepted cases were returned to their original unmatched source, and the next step would be executed.

In a sequence of eight steps, we linked 4,904 death records with an average agreement score of 8.4 (range 8.0 to 9.8 plus 100% agreement on sex). Overall, 4,685 (95.5%) linkages agreed as to whether an injury had or had not occurred. PDDS and VS coders: agreed that 2,686 cases were not injured, agreed that 1,999 cases (1,792 singles, 207 multiples) were injured (but not necessarily on injury mechanism), and disagreed on 219 cases as to whether the case was an injury. This latter category included 156 PDDS cases (36 singles, 120 multiples) and 63 VS cases.

Among linked records agreeing as to whether an injury had or had not occurred, we found 440 cases (9.0%) in the PDDS file that VS coders said had died in other than an inpatient setting. This included 396 cases that VS coders said died in the emergency room or outpatient, 2 cases reported as dead on arrival, 16 unknown, and 26 thought to have died outside of a hospital setting.

We reviewed summary statistics and printed and reviewed 454 cases disagreeing as to where a patient had died. Of cases disagreeing on site of death, the PDDS identified 379 (83.5%) as entering through the emergency room, and 70.5% were from the single record file. Homicide, suicide, and transportation-related accidents accounted for 69.2% of the cases. Matched pairs disagreeing on site of death agreed: exactly on sex (100%), DOB (91.9%), date of death (92.3%), county of residence (95.4%, ZIP5 (68.0%), ZIP3 (90.1%), and race/ethnicity (84.2%). To make the ZIP-level file, we kept all cases in this group.

We also reviewed summary statistics and printed and reviewed cases disagreeing as to whether an injury had occurred. These linked cases agreed exactly on sex (100%), DOB (95.6%), date of death (97.3%), county of residence (94.1%, ZIP5 (77.0%), ZIP3 (89.2%), and race/ethnicity (84.2%). Of 156 PDDS cases disagreeing, no VS record had an injury. Of 63 VS cases disagreeing, 27 were late effects of injuries occurring before 1995, 15 were adverse effects of treatment, and the remaining 21 appeared to be legitimate injuries in the study period.

This left 14,611 VS records unlinked and 1,653 PDDS death records unlinked. Of PDDS death records still unlinked, 452 were coded as injured (single = 370, multiple = 82). Of VS records remaining unlinked, 9,861 were coded as injured and 4,750 were coded as not injured. Of the VS injured: 6,019 were coded as having died out-of-hospital with 1,161 dying at home, 66 dead on arrival or unknown, 3,296 dying in the emergency room or outpatient, and 480 dying inpatient.

Linking the PDDS File of Injury Survivors and the VS File

In the above step, we linked cases where both PDDS and VS coders agreed the person died. The next step was to find PDDS injury patients discharged alive (N = 147,340) who may have died after discharge. Because VS coders and PDDS coders may have disagreed as to injury status, we used all remaining records in the VS file (N = 14,611). We were <u>not</u> seeking to link cases that VS coders said were injured but PDDS coders said were not.

For this step, we used only PDDS injury survivors with a SSNC (N = 94,883; VS N = 11,367) and only did a sequence of four algorithms. We linked 247 injury cases discharged alive who later died. The VS file identified the death location as follows: out-of-hospital, 150 (60.7%); unknown, 1 (0.4%); emergency room or outpatient, 77 (31.2%) and inpatient, 19 (7.7%). Of these, the VS file reported an injury on 159 linked sets (64.4%).

Linkage Reliability

To check linkage reliability, we selected cases where either matching algorithm (PDDS discharged dead, PDDS discharged alive) had identified an injury on either side of the linkage (PDDS or VS) (N = 2,441). Table 3.6 summarizes results of the linkage reliability tests for key linkage variables.

We achieved an exact match on SSNC for 93.5% of linkages, and a soft match on 97.6%. Of 960 cases that were not an exact SSNC match, 99.9% soft-matched on DOB and 93.9% soft matched on SSNC.

DOB matched exactly on 92% of linkages, and soft-matched on 98.5%. Of 195 cases not exactly matched on DOB, 159 were soft matches, and all were an exact or soft match on SSNC.

We achieved an exact ZIP match on 72.3% of linkages, and a 3-digit match on 93%. County of residence agreed exactly on 95.1%. An exact ZIP match was difficult to achieve because of the large number of bogus ZIPs in the VS file.

Cases agreed exactly on the E-code for 67.3% of linkages. For 29 cases, the VS file showed late effects of an injury while the PDDS file showed no E-code. These 29 cases were an exact or soft-match on SSNC and DOB.

VS and PDDS disagreed as to whether an injury had occurred on 253 cases. SSNC and DOB were exact or soft matches on 96.9% of these cases, and by definition, sex agreed 100%.

Table 3.7 presents Kappa statistics and upper and lower 95% confidence intervals for race/ethnicity. When records with Other Race/Ethnicity were included, neither the kappa statistic nor the upper confidence interval included 0.80. When cases with Other/Race ethnicity were excluded, reliability improved for Non-Hispanic White and Hispanic All Race. Under either scenario, the kappa and/or the confidence interval was greater than 0.80 for African-Americans or Asians.

Creating the "Person-Level" File

The "person-level" file was constructed in three steps. First, we had multiple variations of the same variables that had to be resolved, depending on whether the PDDS cases resolved to a single or multiple records, and whether the PDDS records linked with VS records. Second, we had to bring in non-linked VS records. Finally, we reassigned injury classifications that were inconsistent between the PDDS and VS, and removed "people" who did not meet study criteria.

Decision Rules for Linked Cases

For linked cases, we had to decide what information to retain in the person-level injury analysis file. For example, for multi-record SSNC2, we had versions of the same variables (first, last, most, etc), and if they linked, equivalent VS information. For the linked single-record injury EOC and for deaths PDDS coders had coded as not injured but that were linked with VS records, we had equivalent information on the PDDS record and the VS record. Given these various scenarios, we made the following assignments to make the final variables:

- ZIP. If the ZIP was not bogus, we used in descending order: The ZIP most often found, the ZIP on the first record in the last EOC, the ZIP last found, the ZIP on the death record if there was a linkage. If all available ZIPs were bogus, we used the first one found by searching in the same order.

- Race/Ethnicity. If the case came from the multiples file, we used the race/ethnicity most often found. If other race/ethnicity was indicated and there was a death linkage, we used race/ethnicity on the death record.

- If the case came from the multiples file, we used the DOB, county of residence and SSNC most often found, and assigned the first valid E-code, injury variable, and MDC found in the last injury episode.

- We used the injury date found in the VS file if the case linked. Otherwise, we used the admission date found on the first record in the last EOC as the injury date.

- If the case had died and had not linked, we assigned date of death to be the discharge date on the last record found for the person.

- If the case had been discharged alive in the last record in the PDDS file but VS linkage was made, we changed the disposition to "Dead" and used date of death from the VS file.

- E-Code. We reassigned linked VS and PDDS records that disagreed as to the cause of injury, as follows:

If the PDDS said they were discharged within 3 days of the death, we treated both the PDDS and VS files as if they were describing the same injury, and used the most specific E-code. This was defined by prioritizing E-codes: Intentional before unintentional, assault before suicide, and a specific hierarchy within these subgroupings from more to less specific. Example of disagreements are shown in Table 3.5. Given these examples, the algorithm would make the following assignments:

- Select "Firearms (assault)" over "Motor Vehicle Occupant" (SSNC 5).
- Select "Firearms (suicide)" over "Poisoning (suicide)" (SSNC 6).
- Select "Motor Vehicle, unspecified" over "Struck by object (unintentional)" (SSNC 23)
- Select "Firearms (assault)" over "Firearms (unintentional)" (SSNC 44).
- Select "Poisoning (unintentional)" over "Undetermined (other)" (SSNC 49).

If the VS death occurred more than 3 days after the PDDS discharge date, we used the VS E-code unless the VS E-code showed late effects, adverse effects, or had no injury indicated, in which case we used the PDDS E-code.

Because we had reassigned the E-code variable, we then had to re-calculate whether the injury was intentional or unintentional.

Decision Rules for Non-Matched Injury VS Records

- We assigned a disposition of "Dead".
- If the injury date was missing, we assigned date of death to injury date.

Removal of Cases

Having done everything we could to qualify a case for inclusion, we removed all cases (N = 9,131) meeting the following criteria:

1. **After linkage, the patient was older than 24 years on the date of the last injury episode (N = 7,013).**

   We calculated age at injury using the recorded or assigned admission date and DOB. If the age at injury was older than 24 years, the case was removed.

2. **The injured person did not have a DOB and had an age in years of 0 (N = 137).**

   As discussed earlier, we reviewed records meeting this criteria and concluded they were unlikely to be infants. Since we did not know their age, we could not determine whether these cases met the age criteria for the study.

3. **The injury occurred before 1995 (N = 592).**

   We removed all cases with an injury date before January 1, 1995 as outside the study window.

4. **The ZIP of residence was a military Post Office Box (N = 73).**

   We identified a small number of Post Office Box ZIPS for military personnel stationed out of the United States. San Francisco was coded as the county of residence, but these ZIPS were not geographically in California. These people apparently were injured in or near the military community and returned to California for care.

5. **Any diagnosis record associated with the SSNC2 indicated the patient had cancer (N = 402) or HIV (N = 139).**

   As discussed in Chapter Two, these did not seem to be injury patients within the intent of this study.

6. **The Principal E-code indicated adverse effects of treatment (N = 583) or late effects of injury (N = 192)**

   If after linking, no E-code or injury variable diagnoses could be found except adverse or late effects, the case was discarded. Adverse effects as the only principal cause of injury was outside the purpose of this study. For late effects, we presumed the injury was on another record, and that these were failures of the linkage algorithm, or that the injury occurred before 1995.

The Analysis File

We began the linkage process with 162,233 PDDS injury records (including 2,621 records with death as the disposition) and 11,923 VS injury records (with 2,189 in-patient deaths). After dropping cases that did not meet study criteria, the person-level analysis file contained 150,552 all-cause injury cases, with 110,778 unintentional and 37,168 intentional injury cases. *Volume One, State Guide* summarizes characteristics of injury cases at the state level and *Volume Two, County Guide* presents the same information at the county level.

The final analysis file contained 28,273 PDDS cases linking two to 103 PDDS records, 113,602 PDDS cases resolving into one record episodes, 17 deaths PDDS coders said were not injured that linked with the VS as injured, and 8,660 VS injury cases we did not link.

After linkage, death was the outcome for 11,275 cases (7.5%). Table 3.8 shows the distribution of patients who died as a result of an injury, by death location and linkage status.

We linked 2,212 VS records with PDDS cases. VS and PDDS coders agreed 1,651 patients (74.6%) died in-hospital. We were unable to link 403 cases PDDS coders said died in-hospital and 398 cases VS coders said died in-hospital. No unlinked VS case had a SSNC or date of birth, making linkage impossible. Of 3,413 patients VS coders said died in the emergency room or outpatient, we linked 401 (11.7%). Of 68 patients VS coders were unsure as to the site of death, we linked 8 (11.8%). Of 5,324 cases VS coders thought died out-of-hospital, we linked 135 (2.5%).

Discussion

We began linkage using the methodology developed by the California Hospital Outcomes Project for patients with acute myocardial infarction.[21] Our task was made more difficult by several issues. First, we did not set a 30-day limit to end the episode of care, as we had no way to know how long it might take to resolve the long-term effects of injury. Second, we found a sizeable group on the PDDS side with multiple injury records that could not be resolved using the California Hospital Outcomes methodology. Sorting out issues associated with these records was complex.

Third, because of the age group we were studying and because so many injury victims die before reaching a hospital, many records on both sides (PDDS and VS) lacked a SSNC. By contrast, most AMI patients have a SSNC, and the California Hospital Outcomes Project identified that records lacking SSNCs differ systematically from patients with SSNCs. We have made the same finding for injury cases. [22]

Finally adding to the linkage complexity was the discovery of so many VS records with bogus ZIPs. We do not believe this has been previously reported, and this fact doubtless has limited previous PDDS/VS linkage efforts. For example, OSHPD began linking the PDDS and VS files in 1996, and its algorithm required an exact match on 5-digit ZIP.[23]

Despite added complexities of our task, reliability statistics tended to meet or exceed those reported by the California Hospital Outcomes Project, both within the PDDS and between the PDDS and VS files. For example, in OSHPD's 1996 report based solely on PDDS records, race differed on linked record sets for 11.3% and date of birth discrepancies occurred in 5.5%. Among linked injury cases, race differed within the PDDS set for 10.1% of linked record sets and between the PDDS and VS files for 11.2% of linked record sets. Depending on the version used,

---

[21] Romano PS, Remy LL, Luft HS. (1996) Report of the California Hospital Outcomes Project: Acute myocardial infarction, Volume 2, Technical Appendix. Sacramento, California: Office of Statewide Health Planning and Development.

[22] Romano, PS, Luft, HS, Rainwater JA, Zach AP. Report on Heart Attack 1991-1993, Volume 2: Technical Guide. Sacramento, CA: California Office of Statewide Health Planning and Development, December 1997.

[23] Data File Development. Unpublished document dated 10/23/98 provided by Andye Zach of OSHPD.

date of birth agreed exactly between 98% and 99% of the time within the PDDS, and from 92.0 to 98.5% between the PDDS and VS files.

Despite the reliability we attained in the linkage, we may have linked some records -- thereby creating multiple-record sets -- that truly belonged to different people. We think this possible because a large number of injuries seemed to resolve inappropriately to single-record episodes. For example, among single-record injuries, 5,147 came in from another facility, and 4,467 were transferred to another facility, yet records from these other facilities were not found and linked. Some of these records may be among those incorrectly linked to other people, as discussed above. However, since the records were not a good enough match to make a linkage, the records inappropriately remained single.

On balance, we believe "over-linking" and "under-linking" may cancel each other out. So long as we consider a record in the person-level file to be a "person" in a loose rather than exact sense, we think we have done as good a job as possible in carrying out the linkage task. The number of injuries we summarized to the county and ZIP-levels more likely reflects a "truer" number of injuries than if we had not linked, and in that sense may be more conservative, since a given ZIP will have fewer injuries and will be less likely to be identified as a hot spot.

Table 3.1: VS Deaths 1995-1997 Location by Injury Status

Death Location	Not Injured		Injured		Total	
	N	Col%	N	Col%	N	Col%
Out of Hospital	1,944	25.6	6,025	50.5	7,969	40.8
DOA/Unknown	47	0.6	74	0.6	121	0.6
ER/Outpatient	1,910	25.2	3,635	30.5	5,545	28.4
Inpatient	3,691	48.6	2,189	18.4	5,880	30.1
Total/Row%	7,592	38.9	11,923	61.1	19,515	100.0

Table 3.2: PDDS Deaths from Multiples File: Linkage to VS file based on SSNC Version (N = 409)

SSNC Variation	Found N	Found %	Linked N	Linked %
Found first	251	61.4	207	82.5
Found on last EOC	253	61.9	214	84.6
Found most	313	76.5	232	74.1
Found on death record	251	61.4	208	82.9

Table 3.3: PDDS Deaths from Multiples File: PDDS Agreement with VS Across Variations of Key Linkage Variables

PDDS Variable	PDDS Variation Agreeing with VS data (N = 232)					
	First		Last		Most	
	N	Agree %	N	Agree %	N	Agree %
Date of Birth	225	97.0	226	97.4	225	97.0
ZIP (5-digit)	173	74.6	185	79.7	173	74.6
ZIP (3-digit)	212	91.4	217	93.5	206	88.8
Residence County	224	96.6	222	95.7	224	96.6

Table 3.4: PDDS Deaths from Multiples File: Agreement with VS for Race/Ethnicity First and Most Often Found

Race/Ethnicity	Including Other Race/Ethnicity (N = 232)				Excluding Other Race/Ethnicity (N = 220)			
	First		Most		First		Most	
	Kappa	95% CI	Kappa	95% CI	Kappa	95% CI	Kappa	95% CI
Non-Hispanic White	0.81	0.73-0.89	0.83	0.75-0.91	0.85	0.77-0.92	0.87	0.80-0.94
Hispanic-All Race	0.82	0.74-0.89	0.80	0.72-0.88	0.83	0.75-0.91	0.84	0.77-0.91
African-American	0.90	0.83-0.98	0.94	0.87-1.00	0.92	0.85-0.99	0.94	0.87-1.00
Asian	0.92	0.84-0.99	0.87	0.78-0.97	0.96	0.89-1.00	0.93	0.86-1.00

Table 3.5: PDDS Deaths from Multiples File: Examples of Disagreement with VS as to Injury

SSNC	First E-Code	Last Valid E-Code	E-Code Most Often	E-Code VS File
2	Drowning/Submersion	Drowning/Submersion	Drowning/Submersion	No ecode
3	Struck by Object (unintentional)	Struck by Object (unintentional)	Struck by Object (unintentional)	No ecode
5	Firearms (assault)	Firearms (assault)	MV Occupant	Firearms (assault)
6	Firearms (suicide)	Firearms (suicide)	Poisoning (suicide)	Firearms (suicide)
7	Firearms (assault)	Firearms (assault)	MV Pedestrian	Firearms (assault)
8	Undetermined Other	Undetermined Other	Undetermined Other	Firearms (assault)
9	Other (unintentional)	Other (unintentional)	Other (unintentional)	No ecode
11	Firearms (assault)	Firearms (assault)	MV Occupant	Firearms (assault)
12	Poisoning (unintentional)	Poisoning (unintentional)	Poisoning (unintentional)	No ecode
13	Other (assault)	Other (assault)	Other (assault)	Battering
14	Hanging/Suffocation (suicide)	Hanging/Suffocation (suicide)	Poisoning(suicide)	Hanging/Suffocation (suicide)
15	Poisoning (suicide)	Poisoning (suicide)	Poisoning (suicide)	No ecode
18	Cut/Pierce (assault)	Cut/Pierce (assault)	Fight (assault)	No ecode
19	Poisoning (unintentional)	Poisoning (unintentional)	Poisoning (unintentional)	No ecode
23	Firearms(assault)	Firearms(assault)	Struck by Object (unintentional)	MV Unspecified
25	Fall	Fall	Fall	No ecode
32	Fall	Fall	Fall	No ecode
33	Fall	Fall	Fall	No ecode
35	Suffocation (unintentional)	Suffocation (unintentional)	Suffocation (unintentional)	No ecode
36	MV Occupant	MV Occupant	MV Occupant	No ecode
37	Fall	Fall	Fall	No ecode
40	Firearms (suicide)	Firearms (suicide)	Poisoning (suicide)	Firearms (suicide)
41	Suicide(other)	Suicide(other)	Suicide(other)	No ecode
43	MV Occupant	MV Occupant	MV Occupant	Late Effects
44	Firearms (unintentional)	Firearms (unintentional)	Firearms (unintentional)	Firearms (assault)
45	Other(unintentional)	Other(unintentional)	Other(unintentional)	No ecode
49	Undetermined Other	Undetermined Other	Undetermined Other	Poisoning (unintentional)
50	Undetermined Other	Undetermined Other	Undetermined Other	Poisoning (unintentional)

Table 3.6: PDDS Agreement with VS for Exact and Soft-match Linkage Variables

PDDS Variable	PDDS Agreeing with VS (N = 2,441)			
	Exact Match		Soft Match	
	N	Agree %	N	Agree %
Sex	2,441	100.0		
SSNC	2,283	93.5	2,382	97.6
Date of Birth	2,246	92.0	2,405	98.5
ZIP	1,765	72.3	2,269	93.0
Residence County	2,321	95.1		
E-code (Exact)	1,642	67.3		
Any Injury	2,188	89.6		

Table 3.7: PDDS Agreement with VS for Race/Ethnicity Including and Excluding Other Race/Ethnicity

Race/Ethnicity	Including Other Race/Ethnicity			
	Yes (N = 2,441)		No (N = 2,281)	
	Kappa	95% CI	Kappa	95% CI
Non - Hispanic White	0.74	0.72-0.77	0.78	0.75-0.81
Hispanic - All Race	0.76	0.73-0.78	0.80	0.77-0.82
African - American	0.96	0.88-0.93	0.93	0.90-0.95
Asian	0.79	0.75-0.84	0.86	0.82-0.90

Table 3.8: Injury Deaths 1995-1997 Location by Linkage Status

Death Location			Not Linked				Linked			
	Total		PDDS		VS		PDDS Dead		PDDS Live	
	N	Col%	N	Row%	N	Row%	N	Row%	N	Row%
Out of Hospital	5,324	47.2			5,190	97.5	6	0.1	128	2.4
DOA/Unknown	68	0.6			60	88.2	7	10.3	1	1.5
ER/Outpatient	3,413	30.3			3,012	88.3	329	9.6	72	2.1
Inpatient	2,470	21.9	403	16.3	398	16.1	1,651	66.8	18	0.7
Total/Row%	11,275	100.0	403	3.6	8,660	76.8	1,993	17.7	219	1.9

Chapter 4

Injury Summary Files

We earlier developed a multi-step method to find the geographic space for a given year, and then to adjust that geographic space for population change.[24] In this chapter, we describe work to expand that methodology across multiple years. We began by identifying every ZIP that existed in 1995 and reconciling inconsistencies among data sources as to the county or counties associated with that ZIP in that year. Next, we summarized the episode-of-care (EOC) file, containing injuries from 1995 through 1997, to the ZIP associated with the injury irrespective of year. Then we classified ZIP-level records as to whether the ZIP had a population in 1998, had injuries, was bogus, or had consolidated between 1995 and 1998. We estimated the 1996 population of every ZIP that could map. Having resolved all issues associated with ZIPs, we put all ZIP-level files together, calculated injury rates, and identified hot spots. We then wrote programs to produce within-county or within-region ZIP-level reports. Finally, we created a file to use as input for making maps and produced the maps.

Data Sources

To summarize injury EOC data to the ZIP and county level, we used data from multiple vendors:

- A MapInfo file contained the 1995 ZIP, the enclosing ZIP if a post office, county (actually a 1st, 2nd and 3rd county, for ZIPs split across county boundaries), area size, the name of the community associated with the ZIP and other fields (N = 2,657).[25] This file included ZIPs starting with "0" for various parks and forests, ZIPs known to have no population such as Post Office Box (POB) ZIPs, and ZIPs with a population and physical geography but receiving mail only at post office boxes. ZIPs in these latter two categories are shown as enclosed by other ZIPs.[26]

- A file from GeoLytics had ZIP and county boundaries for 1995. Working from the CD and its accompanying software, all California ZIPs with a value 90000 or

24 Oliva G, Remy L, Clay T, Greene J, Guendelman S. The effect of increasing social burden on patterns of hospitalization for children age 0-4 in California. Manuscript in preparation. Presented at APHA Annual Meeting 1999, Chicago, Il Paper in progress.

25 MapInfo (1996). Troy, NY. The source of the 1995 5-Digit ZIP data is cited as under license to MapInfo by Geographic Data Technologies, Lynn, NH.

26 This latter situation can cause some misunderstandings when viewing maps. For example, the Town of Ross, California has both area and population but receives all its mail at a post office located in downtown Ross. The MapInfo file shows the Ross ZIP as enclosed by its geographic neighbor, the Town of San Anselmo. Thus, Ross will not show on the map.

greater were exported to a MapInfo format file (N = 1,575).[27] The GeoLytics and MapInfo files contained the same list of 1,575 ZIPs with a physical area that could be mapped. These ZIPs became the target list to which all others were linked.

- Claritas provided a ZIP-level file estimating population by race, sex and age groups for 1998.[28] This file contained 1,889 ZIPs. It did not have a variable associating a ZIP with a county.

- Another Claritas file cross-referenced PO ZIPs into their enclosing population ZIPs. It contained 2,845 ZIPs (2,845 minus 1,889, or 956 POB and national park ZIPs), and lacked a county variable. All 1,889 parent ZIPs in this file are in the Claritas population file.

- Western Economics Research provided a file to identify ZIP change information since June 1991.[29] This file identified whether a changed ZIP is a POB. We understand it does NOT reflect ZIP boundary shifts that affect the ZIP used by a certain number of addresses. Boundary shifts between 1995 and 1998 might require combining statistics for ZIPs, but we do not have this data.

- We used a 1990 census file to identify the county or counties (a maximum of 5) that ZIPs existing in 1990 were nested within or spanned, and the portion of the total ZIP population associated with each county.[30]

- DOF county population estimates are released yearly by the California Department of Finance.[31] These contain estimated county-level population by sex, race/ethnicity, and age in years.

- Data elements that hospitals forward to OSHPD include the ZIP of the patient's residence but do not include the patient's county of residence or any other geographic information. OSHPD annually purchases a dataset from the United States Postal Service (USPS), and then uses that to assign every ZIP to one

---

27 CensusCD+Maps, Version 2.1. (1998) GeoLytics, Inc. East Brunswick, NJ. According to the GeoLytics documentation, everything came from Tiger 95 files except for the ZIP boundaries, which came from Geographic Data Technology, Inc. of Lebanon, NH.

28 FHOP commissioned Claritas to make this file according to FHOP specifications for age and race/ethnicity categories by ZIP. Population estimates were as of mid-year 1998. Claritas, Boston MA.

29 Western Economics Research. W.E.R, The Information Connection, Inc. P. O. Box 107, Mill City, OR 97360. Website: http://www.teleport.com/~wer/zip_reference.html#change.

30 1990 Census of Population and Housing Summary Tape File 3B, ZIP Code Areas 7,8,9. U. S. Department of Commerce, Bureau of the Census, Data User Services Division, Washington, DC 20233.

31 State of California, Department of Finance, Race/Ethnic Population with Age and Sex Detail, 1970-2040.Sacramento, CA, December 1998.

and only one county.[32] The methodology we use makes the same assumption.[33]

Clarifying ZIP Relationships Within and Among Counties

In nested groups typically used for contextual analysis (i.e., individual, in family, in neighborhood, in city, in county), a group at one level of observation is contained within a group at the next higher level.[34] Unfortunately, however, this ideal picture changes when the ZIP is the nesting unit. In any given year and over time, the geographic location (boundaries) of ZIPs may be either wholly nested or may overlap other geographic boundaries. Further, to maximize mail delivery efficiency, the United States Postal Service (USPS) consolidates or splits ZIPs as a function of population change and also changes ZIP geographic boundaries. That is, a given ZIP number may exist from year to year, but area, boundaries, and population a ZIP number encompasses may change over time. This causes substantial methodologic issues that must be addressed when one wishes to study and map information gathered over time.

In the present instance, MapInfo had assigned ZIPs it believed existed in 1995 to a maximum of three counties, and had identified ZIPs nested within larger geographic ZIPs. Using information from the United States Postal Service (USPS), OSHPD had annually assigned every ZIP to one and only one county. The 1990 census had assigned ZIPs and population to one or more counties (FIPS codes, a maximum of 5) based on boundaries existing at that time. Given vagaries of the USPS, a ZIP could be within one county in one year and another the next. Western Economic Research maintained information for all ZIP changes between the 1990 census and 1999, but this file had no boundary information.

Given various combinations of competing and missing information, our task was to assign each injury in the patient discharge data set (PDDS) file to a mappable ZIP. This methodologic challenge was exacerbated by bogus ZIPs, i.e., ZIPs recorded in the PDDS and Vital Statistics (VS) files that no available source identified as ever having existed.

We created a file from the master PDDS files of all discharges for 1995 through 1997, regardless of patient age or diagnosis, finding every county associated with every ZIP recorded in the PDDS over that time. Working from a map provided by

32 For a discussion of the method and the consequences in terms of county-level error, see Appendix F, Methodology for Assigning Patients to a County of Residence Using the Patient's ZIP Code. Office of Statewide Health Planning and Development (1996). Discharge Data Tape Format Documentation.

33 The USPS assigns one county name for each current ZIP to indicate the general geographic area served by the ZIP in that year. The county name identifies the location of the USPS station assigned to deliver the mail to that ZIP. Because some ZIP geographic boundaries span two or more counties, this can cause some confusion on some maps. The name of a city physically located in County A (from MapInfo) will be attached to a ZIP that spans the boundaries of both County A and County B. Residents of both County A and County B will see on their respective county maps the name of the city physically located in County A associated with the ZIP that spans both counties.

34 Boyd LH and Iversen GR. (1979) Contextual Analysis: Concepts and Statistical Techniques. Belmont, CA: Wadsworth.

OSHPD,[35] we divided Los Angeles county into four sub-regions, using the Health Service Area (HSA) and Health Facilities Planning Area (HFPA) variables in the PDDS files. HSAs and HFPAs are geographic areas designated by OSHPD.[36] When summarizing data to the region, we made 63 records, one for each county and four for the sub-regions of Los Angeles County. In addition, Berkeley (in Alameda County), Pasadena and Long Beach (both in Los Angeles County) have their own city departments of public health. These departments provided a list of ZIPs they monitor. We grouped these ZIPs to prepare reports and maps for these cities.

A guiding decision was that the 1995 ZIP geography as reflected by the Geolytics ZIP boundaries would be the bottom line to assign ZIPs to counties for mapping. Our goal was to identify (1) ZIPs whose boundaries were totally outside the county or sub-region, and (2) ZIPs that crossed county borders, and of those, the county containing the most ZIP area.

However, in order to do regional roll-ups of ZIP-level injury data, we also had to assign one and only one region to estimate ZIP-level population changes between 1998 (the year of the ZIP-level population estimates) and 1996 (the year we were using to calculate population rates). To do this, we used information from Geolytics, MapInfo, the 1990 Census, and the PDDS regions file. We initially assumed that the county containing the largest area of a ZIP would contain the largest share of population. We compared the 1990 Census ZIP population as distributed among counties with the list of counties MapInfo assigned to a ZIP and the county OSHPD assigned to a ZIP. If there was a conflict, we assigned the ZIP to the county the 1990 census said had the largest proportion of the ZIP population.

Working in MapInfo, we linked the CensusCD boundary file with the file assigning county population priority. Within ZIPs, counties were coded to indicate their presumed dominance in terms of population. Working one county at a time, we found all ZIPs thought to be associated with that county and created a color-coded map to look at the size of the polygon within and without the county. ZIPs with discrepancies between the census information and geographic area were noted.

The starting assumption was that the larger geographic area would also contain the larger population, but this did not always hold. For example, we found a large ZIP, predominantly rural in its area (dairy farms) with its geographic area almost entirely in Marin. MapInfo had assigned the primary county as Marin, apparently on the basis of geography. However, the census indicated the population was predominantly in Sonoma. This was verified by asking people familiar with the region. Without detecting problems such as this, Marin, with its smaller population, would show a higher injury rate for this ZIP, when the injuries in fact were associated with a town physically located in Sonoma County, which has a much

---

[35] Health Facility Planning Areas - 1986. Map of health facilities planning area boundaries. Exact boundary descriptions can be obtained from the Division of Health Planning, Research and Data, of the Office of Statewide Health Planning and Development.

[36] Health Facility Planning Areas, January 1983. P-800-2, Revised 07/18/83, pursuant to Section 90811, Division 7, Title 22 of the California Administrative Code. Office of Statewide Health Planning and Development.

larger population. We also identified ZIPs whose physical boundaries were located wholly outside their county or sub-region boundaries. These inconsistencies were corrected in assigning counties to ZIPs.

We ended up with a file of 2,622 real ZIPs, including 1,575 with boundaries that could be mapped. All POB, and enclosing ZIPs were identified. Each ZIP was assigned to one of those counties for population estimation purposes and to a maximum of three counties for mapping purposes. At this point, we also created a file containing only those 1,575 ZIPs that could be mapped.

Resolving Problem ZIPs

In preparing to create a ZIP-level file summarizing injuries, we identified four ZIP problems we had to resolve. The following describes the type of problem encountered and a general description of how these were resolved:

1. **Bogus ZIP** records (1) identified a county where the injury occurred but lacked a ZIP, (b) were a ZIP ending in "00", or (c) did not match any ZIP in any file of real ZIPs. When working with the discharge-level PDDS file and the VS file, we initially found 4,421 injury records from 131 bogus ZIPs.

   There was no important relationship between the presence of a bogus ZIP and the sex, race, or county of residence. Individuals aged 15 to 24 were more likely than younger individuals to have bogus ZIPs. Records with bogus ZIPS were more likely than records with legitimate ZIPS to be clustered in the following: motor vehicle, falls, drowning, homicide, suicide.

   The VS file was the source for 94.6% of bogus ZIP records.[37] Because these accounted for 41% of VS injury records, they could not be discarded. We handled most of these in making the EOC file. Specifically, if a VS record linked with a PDDS record, we assigned legitimate ZIPs in the PDDS, using the following order to assign: the most frequently used ZIP, or if that was bogus, the last real ZIP, or if that was bogus, the first ZIP on the condition record in the last injury EOC. All remaining records with bogus ZIPs or records with county but no ZIP were from the VS file.

   We identify these ZIPs in the ZIP-level tables. They are used to calculate the county-level injury rate, but are not otherwise used.

2. **No population ZIPs** had injuries, existed in 1995 but had no population allocated through the cross walking of 1998 ZIP information back to 1995. They were handled in the same way as bogus ZIPs.

3. **Old Split ZIPs** existed before 1995 but were split into more than one ZIP by 1995. This may result from patients continuing to give their old ZIP when admitted to a hospital or because patient information was entered into the computer system before the splits occurred and the hospital had not updated

[37] We strongly recommend that the Department of Vital Statistics work with local county jurisdictions to improve the quality of the ZIP variable in the death file.

the address. These injuries were assigned to the 1995 ZIPs into which the older ZIP had been split.

4. **Consolidated ZIPs** existed separately in 1995 but were consolidated into a single ZIP in the 1998 population file. Hence we do not have separate population counts for the ZIPs. While not a problem with the injury records (numerators), it is a problem derived from using 1998 data for population estimates (denominators). We resolved this as follows. The 1998 population for the consolidated ZIP was split into two or more 1995 ZIPs, in proportion to the total number of injuries found in the 1995 ZIPs. We added these together and treated them as if they were one.

ZIP-Level Summary of Injury EOC

Working with the EOC file, we created variables to use as numerators when we calculated ZIP-level rates. Using the ZIP in the EOC file, we summarized injuries by all combinations of intent (all, intentional, and unintentional), sex, race/ethnicity (White or Other Race)[38] and age category (0 to 4, 5 to 9, 10 to 14, 15 to 19, 20 to 24). We also carried along other information needed to correctly classify ZIPs.

The ZIP-level summary of EOC records was assigned to one of three files. One consisted of 11 counties with 30 injury EOC and no ZIP. A second contained 26 ZIPs ending in "00" with 206 injuries. The third had 2,233 ZIPs with 150,316 injuries. The latter file included ZIPs that would end up being valid and others not yet classed as split or consolidated, other bogus ZIPs, or ZIPs with no associated population.

Table 4.1 identifies counties associated with EOCs completely missing a ZIP. Table 4.2 identifies the "00" bogus ZIPs and the number of injuries associated with them.

Cross-Walking Injury, Map, and Population ZIPs

The next step was to assign all ZIP-level records into six files, depending on various combinations of whether they had injuries, did or did not have population, or were bogus, split, or consolidated. On the population side, we created one file with 1,875 ZIPs having a 1998 population, another with 10 ZIPs with population in 1998 that linked to multiple 1995 mapped ZIPs, and 14 ZIPs having no 1998 population. The sum of ZIPs on the population side (1,899) is larger than the number of population records from Claritas (N = 1,889). This is caused by an overlap between records having a 1998 population and records that mapped to multiple 1995 mapped ZIPs.

On the injury side, we identified 2,204 legitimate ZIPS, 66 that linked to multiple 1995 mapped ZIPs, and 29 additional bogus ZIPs. The sum of ZIPs with injuries (2,299) is larger than the number of injury records that went into the crosswalk

[38] The Claritas file contained race/ethnic breakdowns within age categories. However, because of severe small numbers problems, it would be inappropriate to calculate separate ZIP-level estimates for each race/ethnic group, i.e., Non-Hispanic White, Hispanic, Black, Asian, and Other race/ethnicity.

(2,233). This is an overlap between records that link to legitimate ZIPs based on injury and records that link to multiple 1995 mapped ZIPs.

On both sides, ZIPs that linked to multiple mapped ZIPs were a problem. We did not know where to put the ZIP for maps, which ZIP to use to estimate 1996 population, or where to assign the ZIP to generate county-level reports. We resolved differences within and between the multiples files, and identified the basis for multiple mappings. The resulting file contained 32 mappable ZIPs. Table 4.3 identifies the consolidated ZIP, reason it consolidated, source ZIPs, 1998 population (if any), number of injuries (if any), and consolidation date.

Resolving Bogus ZIPs

At this point, we had identified all bogus and consolidated ZIPs. Now we put the various files together, resulting in 92 ZIPs that remained bogus with 417 injuries. These included 15 that remained ambivalent as to location, 11 blank, 8 discontinued, 26 ending in "00", 29 that did not match any known ZIP, and 3 with injuries but no population. We applied a county or sub-region to the records missing this information.

Estimating 1996 ZIP-Level Population

We estimated 1996 ZIP-level population by the following method. First, we merged the cross-walk file of ZIPs with population and the file of mappable ZIPs, as described above. Then, the population data was summarized into the corresponding 1995 mappable ZIPs. For each county and demographic stratum (combination of sex, race, age category), we calculated the percent of population each ZIP contributed to the total county population in 1998. Thus, within each county and demographic stratum, we calculated the percentage that each ZIP contributed, known as the ZIP percentages.

The 1996 DOF county data was summarized into the same demographic strata. DOF county totals for each stratum were multiplied by the ZIP percentages to yield 1996 ZIP-level population estimates within each stratum.

This methodology makes two assumptions. First, similar to OSHPD, ZIPs are treated as if they are entirely in one county, and if split across two or more counties, the secondary county does not contain a significant percent of the ZIP population. Second, the geographic (ZIP) population distribution within each demographic stratum is the same in 1998 as 1996.

For example, if a county had three ZIPs, and 1998 Claritas data estimated the proportion of White Females aged 5-9 to be 35% in ZIP 1, 20% in ZIP 2 and 45% in ZIP 3, those same percentages were assumed to hold in 1996. Thus if the DOF county data had 1,000 White Females aged 5-9 in 1996, we would allocate 350 to ZIP 1, 200 to ZIP 2 and 450 to ZIP 3.

The California population age 0 to 24 in the 1998 Claritas file is 11,601,712. The total California population age 0 to 24 in the 1998 DOF file is 12,437,256, or 7% higher. We used the Claritas data only to estimate within-county ZIP-level population distributions. We used the 1996 DOF county population files as the source for the mid-interval denominator, because the state uses DOF estimates to calculate state and county injury rates. Thus the sum of the 1996 ZIP-level population within the county equals the DOF estimated county population.

Identifying Hot Spots

We now had put together all available information regarding ZIP-level population and injuries. We began with 1,575 ZIPs with boundaries, excluding state and national parks. We estimated a 1996 population and calculated rates for 1,563 ZIPs. Three ZIPs with population and injuries had consolidated into one of the 1,563 ZIPs we used for rates.[39] Two ZIPs did not have 1998 population but had injuries and had consolidated.[40] These ZIPs also are among the 1,563 mappable ZIPs. Seven mappable ZIPs had no population and no injuries.[41] To identify injury hotspots (all, unintentional, and intentional) we:

- Calculated the rate per 100,000 population at the ZIP and county level
- Calculated quartiles for number of injuries and the rate per 100,000 population for all ZIPs in the state relative to each other, and for all counties in the state relative to each other
- Calculated quartiles for the number of injuries and the rate per 100,000 population for all ZIPs within counties relative to each other
- If the county had 12 or more ZIPs, we ranked ZIPs within the county.

Tables 4.4 through 4.9 display relationships between quartiles (all, unintentional, and intentional) for number of injuries and rate for injuries at the ZIP and region. For each intent and level (ZIP, county), the tables show the number, total population, average population, average ZIP area, average crowding, average number of injuries, and average rate per 100,000 population. Underlying distributions for these variables differ depending on intent.

The goal of this research was to bring attention to those areas where children and young adults were disproportionately burdened by injury. By identifying these areas, local health planning bodies will be better able to target injury prevention activities and evaluate the effectiveness of those activities.

One way to target injury prevention would be to focus on areas with the highest number of injuries. As the summary tables indicate, the number of injuries is strongly affected by the underlying population estimate. It also appears to be influenced by geographic size (area) of the community and by crowding (population divided by area). Most injuries occur in densely populated urban areas. Focusing solely on numbers of injuries would emphasize the problems of cities and minimize problems of sparsely populated rural areas.

---

39 ZIP 95667 consolidated to 95643. ZIP 92332 consolidated to 92304. ZIP 93254 consolidated to 93214.

40 ZIP 93214 consolidated to 92680. ZIP 93217 consolidated to 93203.

41 ZIPs 90506, 91371, 92096, 92280, 93043, 94128, 95721.

Another way to examine the phenomenon is to use population-based injury rates. However, rates are calculated using the estimated underlying population. A community with a "small" population but a high rate may have a problem with the underlying population estimate or truly may have a problem protecting its young population. So confidence intervals around the rate estimates also are important, as are standardized ratios and their confidence intervals.

We did a series of diagnostic analyses to understand relationships between estimated population, rates, and standardized ratio confidence intervals. We focused on ZIPs falling within the third quartile on number and rate (N3R3), fourth quartile on number and third on rate (N4R3), and second or third quartile on number and fourth on rate (N2R4, N3R4) for all, unintentional, and intentional injuries. Focusing on all injuries (Table 4.4):

- The N3R3 cell contained 115 ZIPs, had a total estimated population of 833,939, ZIP population ($\overline{X}$ = 7,315) below the state average, area ($\overline{X}$ = 64 square miles) below the state average, crowding ($\overline{X}$ = 1,364) above the state average, injuries ($\overline{X}$ = 93) below state average, and injury rate ($\overline{X}$ = 423) just above state average. No lower confidence interval for any standardized ratio was greater than 100 for any injury classification in this cell. Thus this cell is distinguished primarily by having a number of injuries above the ZIP median.

- The N4R3 cell contained 156 ZIPS for all injuries, contained the highest estimated population (3,077,567), the highest ZIP population ($\overline{X}$ = 19,602), second smallest area ($\overline{X}$ = 41), highest crowding ($\overline{X}$ = 3,194), second highest number of injuries ($\overline{X}$ = 256), and an injury rate ($\overline{X}$ = 435) above the state average. The lower confidence interval for the standardized ratio was greater than 100 for 19 of 156 total injury ZIPs, 23 of 160 unintentional injury ZIPs, and none of 135 intentional ZIPs. This cell is distinguished primarily by having a number of injuries in the top quartile of the state, although some ZIPs within it also have higher than expected rates.

- The N3R4 cell contained 58 ZIPs for all injuries, a total population of 324,586 with ZIP population ($\overline{X}$ = 5,410) below state average, area ($\overline{X}$ = 77) just below the state average, crowding ($\overline{X}$ = 1,463) above the state average, number of injuries ($\overline{X}$ = 98) just above the state average, and fourth highest rate ($\overline{X}$ = 608) in the state. The lower confidence interval for the standardized ratio was greater than 100 for 42 of 58 total injury ZIPs, 45 of 62 unintentional injury ZIPs, and 15 of 66 intentional ZIPs. This cell is distinguished by having a number of injuries above the state median, and with a larger number of ZIPS having higher than expected rates.

- The N2R4 cell contained 85 ZIPs for all injuries, the fourth lowest population (total estimated = 105,100), fifth lowest ZIP population ($\overline{X}$ = 1,251), third largest average area ($\overline{X}$ = 146), second lowest crowding ($\overline{X}$ = 120), fifth lowest number of injuries ($\overline{X}$ = 25), and second highest injury rate ($\overline{X}$ = 672). ZIP-level population estimates ranged from a low of 144 to a high of 2,617. The lower confidence interval for the standardized ratio was greater than 100 for 29 of 85 total injury ZIPs, 45 of 97 unintentional injury ZIPs, and 1 of 69 intentional ZIPs.

Of the 29 ZIPs in the N2R4 cell with the standardized ratio greater than 100, 10 had rates above 1,000, and 15 had populations below 1,000. We checked the ZIP-level tables to see how many injuries had been assigned into the affected ZIPs, thinking that might have affected the rate numerator. Eleven of 29 ZIPs received one or less. None were among the consolidated ZIPs. We contacted local public health departments in the parent counties to get some sense of the population estimate. There was a general sense that the population estimates were low. As a result, we did not flag ZIPs in this cell or the N1R4 cell.

Using this information, ZIPs were classified as hot, medium, or warm, as follows:

- If the unit of analysis (county, ZIP statewide, or ZIP within county) was in cell N4R4, it was considered to be a "Hot Spot". On the sample maps, hot spots are indicated with the color red.
- If the unit of analysis was in N3R4 or N4R3, it was considered to be a "Medium Spot". On the maps, medium spots are indicated with the color orange.
- Finally, "Warm Spots" were flagged as those ZIPs in N3R3. These are colored yellow on the maps.
- All other ZIPs were considered not to be hot spots. They are light gray on the map.

Of 1,563 mappable ZIPs, the following summarizes all, unintentional, and intentional injuries:

- All Injuries: 127 ZIPs were "hot", 214 "medium", and 115 "warm" for total injuries. These hotspots were 29.1% of all California ZIPs, 53.1% of the State population in this age group lived in these communities, and 65.1% of injuries to California's children, adolescents, and young adults through age 24 happened there.
- Unintentional Injuries: 98 ZIPs were "hot", 222 "medium", and 101 "warm" for unintentional injuries. These hotspots were 26.9% of all California ZIPs, 46.9% of the State population in this age group lived in these communities, and 57.4% of unintentional injuries to California's children, adolescents, and young adults through age 24 happened there.
- Intentional Injuries: 225 ZIPs were "hot", 201 "medium", and 142 "warm" for intentional injuries. These hotspots were 36.3% of all California ZIPs, 65.6% of the State population in this age group lived in these communities, and 82.9% of intentional injuries to California's children, adolescents, and young adults through age 24 happened there.

We did a similar analysis of county-level injuries, expressed as regions with Los Angeles divided into four regions. Of 61 regions, the following summarizes all, unintentional, and intentional injuries:

- All Injuries: 5 regions were "hot", 8 "medium", and 4 "warm" for total injuries. Of the state population age 0 to 24, 56.3% lived in these regions, and 62.4% of injuries in this age group happened in these regions.

- Unintentional Injuries: 2 regions were "hot", 8 "medium", and 4 "warm" for unintentional injuries. Of the State population age 0 to 24, 44.2% lived in these regions, and 50.4% of unintentional injuries in this age group happened in these regions.

- Intentional Injuries: 12 regions were "hot", 6 "medium", and 7 "warm" for intentional injuries. Of the State population age 0 to 24, 87.9% lived in these regions, and 92.6% of intentional injuries in this age group happened in these regions.

Preparing Injury Reports

Using the ZIP-level file containing rates, we output comma-delimited ASCII files, within counties and at the state level, summarizing detailed statistical information used to classify hot spots. We created template Excel files to import the ASCII files into Excel. The resulting Excel files were used as publication-ready tables for the study report. Detailed statistics used to classify counties and ZIPs within counties are in *Volume Two, County Guide*. The Excel state and county statistical files are available from the counties. The only difference between the hard-copy report and the electronic files is that the electronic files contain information to several decimals.

We also prepared reports at the state and county level summarizing the characteristics of the injuries. In order to do these reports, we needed the county or sub-region within county to which ZIPs had resolved. To get this information, we merged the EOC file, the bogus ZIPs file, and the file of ZIPs that were okay. At the state level, we prepared three reports: one summarizing characteristics of all injuries, characteristics of injury patients surviving to admission and their clinical course of care, and characteristics of all injuries by race/ethnicity within age groups. We also prepared county-level summaries similar to the state-level report of all injuries and injuries surviving to admission.

The above information was again output to comma-delimited ASCII files. We created a template Excel file to import the ASCII files into Excel. The resulting Excel files were used to prepare publication-ready tables for the study report. The state-level tables are in *Volume One, State Guide*. The county tables are in *Volume Two, County Guide*. The Excel state and county files are available electronically upon request. The only difference between the hard-copy report and the electronic files is that the electronic files contain information to several decimals.

Mapping Hot Spots

Using the ZIP-level file with rates, we output comma-delimited ASCII files containing the ZIP, county or sub-region, and hot spot classifications. We read this into Excel, imported into MapInfo, and linked with the boundary information.

We created files containing county or sub-region boundaries and the location of Federal and state highways. The legend classifying hotspots was created. ZIP boundaries were imported into each county file and into state-level files. Within county maps, ZIPs extending outside county boundaries were trimmed to show only the within-county portion.

At this point, we created six county maps showing each ZIP classified as to its statewide comparison for all, unintentional, and unintentional injuries, and each ZIP classified within the county on the same dimensions. The maps were exported in the Word Metafile Format (WMF).

We made a Microsoft Word file template with two columns (state-wide comparison and intra-county comparison) and three rows (all, unintentional, intentional injury). We imported each map into the appropriate cell, changed the page title as appropriate, and saved each set of six county maps as a separate county file. If the county had fewer than 12 mappable ZIPs, no intra-county comparisons are made.

State-level maps were made following the same procedure. In these we compare injuries at the county-level statewide and ZIP-level statewide within county boundaries.

These maps are in reproduced in *Volume One, State Guide*. The map files are available electronically upon request.

Table 4.1: Counties Having Injuries With No ZIP

County	Injuries		
	Total	Unintentional	Intentional
Alameda	1	0	1
Central Los Angeles	15	5	9
Mendocino	1	1	0
Monterey	2	1	1
Orange	1	1	0
Riverside	1	1	0
Sacramento	1	1	0
San Bernardino	1	1	0
San Diego	5	3	2
Santa Clara	1	1	0
Santa Cruz	1	1	0
Total	30	16	13

Table 4.2: ZIPs ending in "00", the County and Number of Injuries

ZIP	County	All	Unintentional	Intentional
90000	Central Los Angeles	142	74	68
90200	Central Los Angeles	2	1	1
90700	Southeast Los Angeles	1	1	0
90800	West Los Angeles	1	0	1
91000	West Los Angeles	2	2	0
91100	Southeast Los Angeles	3	1	1
91600	Northeast Los Angeles	2	2	0
91700	San Bernardino	3	3	0
92000	San Diego	3	1	2
92100	San Diego	1	1	0
92500	Riverside	3	2	1
92600	Orange	1	1	0
92800	Orange	2	2	0
93000	Ventura	1	0	1
93300	Kern	1	1	0
93600	Fresno	1	1	0
93700	Fresno	1	1	0
94000	San Mateo	2	1	1
94100	San Francisco	7	3	4
94500	Alameda	2	1	1
94600	Alameda	13	10	1
95000	Santa Cruz	3	2	1
95100	Santa Clara	2	1	1
95400	Sonoma	2	2	0
95600	Placer	1	1	0
95800	Sacramento	4	3	1
	Total	206	118	85

Table 4.3: ZIPS Affected by Consolidation

Consolidated		Source ZIPs					
Reason	ZIP	ZIP1	ZIP 2	ZIP 3	Population	Number of Injuries	Effective Date
Population	92304	92304	92332		45	2	10/3/98
Denominator	92782	92680	92714		7255	7	7/1/96
	93203	93203	93217		6214	72	8/1/97
	93254	93214	93254		314	2	
	95667	95643	95667		10778	115	
Discontinued	92013	91913	91914	91915		2	1/1/91
	92330	92530	92532			9	7/1/91
	92343	92543	92545			6	7/1/91
	92362	92562	92563			5	7/1/91
	92370	92570	92571			10	7/1/91
	92383	92582	92583			4	7/1/91
	92390	92590	92591	92592		5	7/1/91
	92409	92408	92410			4	3/1/94
MapInfo/	90055	90014	90015			10	
Claritas	90070	90005	90010			3	
Conflict	90079	90014	90015			1	
	90224	90220	90221			4	
	90313	90232	90301			1	
	91225	91204	91205			2	
	91385	91335	91355			1	
	92174	92102	92114			3	
	92186	92101	92110			2	
	92554	92553	92555			4	
	92589	92590	92591			2	
	93613	93611	93612			2	
	95406	95401	95404			2	
MapInfo/	95426	95451	95461			5	5/18/94
USPS	95671	95630	95673			23	
Conflict							
Able to	92078	91978	92069			1	7/1/99
Ignore	92313	92313	92555			44	7/1/91
	92337	92337	92557			67	7/1/91
	95887	95691	95826			1	

Table 4.4: State-wide ZIP-Level Summary of All Injuries

Injury Type	State Quartile on Number	State Quartile on Rate Q1	Q2	Q3	Q4	Row Total or Row Average
Number	Q1	172	59	47	121	399
of	Q2	128	98	72	85	383
ZIPs	Q3	80	137	115	58	391
	Q4	12	96	156	127	388
	Column Total	392	390	390	391	1,563
Total	Q1	128,508	32,330	20,083	23,160	204,081
Population	Q2	586,076	304,356	172,356	105,100	1,167,890
	Q3	875,955	1,289,311	833,939	324,586	3,305,381
	Q4	253,004	1,822,335	3,077,567	2,123,093	7,276,000
	Column Total	1,843,546	3,448,330	4,103,946	2,575,940	11,971,762
Average	Q1	743	539	427	190	508
Population	Q2	4,543	3,138	2,394	1,251	3,057
	Q3	11,230	9,276	7,315	5,410	8,501
	Q4	21,084	19,387	19,602	16,985	18,753
	Column Average	4,073	8,842	10,523	6,588	7,659
Average	Q1	155	171	106	99	135
ZIP	Q2	65	71	92	146	90
Area	Q3	73	53	64	77	64
	Q4	50	60	41	32	43
	Column Average	106	77	65	84	83
Average	Q1	453	182	141	51	254
Crowding	Q2	866	663	487	120	579
(Pop/Area)	Q3	1,232	1,153	1,374	1,630	1,307
	Q4	1,005	1,563	3.194	3,104	2,702
	Column Average	761	980	1,794	1,284	1,204
Average	Q1	3	5	6	5	4
Number of	Q2	30	31	30	25	29
Injuries	Q3	85	92	93	99	92
	Q4	168	197	256	326	261
	Column Averagel	33	89	136	126	96
Average	Q1	125	337	438	829	269
Rateper	Q2	221	329	423	672	319
100,000	Q3	251	330	423	608	360
Population	Q4	265	339	435	640	465
	Column Average	235	335	432	639	418

Table 4.5: State-wide ZIP-Level Summary of Unintentional Injuries

Injury Type	State Quartile on Number	State Quartile on Rate Q1	Q2	Q3	Q4	Row Total or Row Average
Number of ZIPs	Q1	166	48	57	135	406
	Q2	117	90	72	97	376
	Q3	91	137	101	62	391
	Q4	18	114	160	98	390
	Column Total	392	389	390	392	1,563
Total Population	Q1	132,788	30,625	25,805	28,979	218,196
	Q2	544,029	289,144	175,800	128,714	1,137,686
	Q3	1,015,778	1,225,781	729,151`	324,820	3,295,530
	Q4	372,882	2,392,275	3,002,489	1,152,703	7,320,350
	Column Total	2,065,476	3,937,825	3,933,245	2,035,216	11,971,762
Average Population	Q1	800	638	453	215	508
	Q2	4,650	3,213	2,442	1,327	3,057
	Q3	11,162	8,947	7,219	5,239	8,501
	Q4	20,716	20.985	18,766	15,844	18,119
	Column Average	5,269	10,123	10,085	5,192	7,659
Average ZIP Area	Q1	146	156	146	103	133
	Q2	74	64	93	136	91
	Q3	69	45	65	86	62
	Q4	46	47	44	42	45
	Column Average	102	64	73	94	83
Average Crowding (Pop/Area)	Q1	485	75	149	69	251
	Q2	1,031	767	473	139	631
	Q3	1,374	1,428	1,247	1,317	1,351
	Q4	1,392	2,641	2,797	2,462	2,603
	Column Average	896	1,464	1,580	882	1,204
Average Number of Injuries	Q1	2	5	4	4	3
	Q2	24	25	24	21	23
	Q3	65	70	72	73	70
	Q4	128	168	186	225	188
	Column Average	29	80	100	74	71
Average Rate per 100,000 Population	Q1	90	257	333	641	215
	Q2	170	257	327	527	257
	Q3	195	259	330	463	275
	Q4	206	266	330	473	333
	Column Average	184	264	330	477	308

Table 4.6: State-wide ZIP-Level Summary of Intentional Injuries

Injury Type	State Quartile on Number	State Quartile on Rate Q1	Q2	Q3	Q4	Row Total or Row Average
Number of ZIPs	Q1	320	27	22	30	399
	Q2	66	170	92	69	397
	Q3	5	165	142	66	378
	Q4	0	29	135	225	391
	Column Total	391	391	391	390	1,563
Total Population	Q1	207,077	18,861	8,631	4,956	239,525
	Q2	362,811	676,950	208,009	57,351	1,305,120
	Q3	76,825	1,858,402	1,095,633	328,161	3,359,021
	Q4	0	633,638	2,554,728	3,879,730	7,068,095
	Column Total	646,714	3,187,850	3,867,000	4,270,198	11,971,762
Average Population	Q1	647	699	392	165	600
	Q2	5,492	3,982	2,261	831	3,287
	Q3	15,365	11,263	7,716	4,972	8,886
	Q4	0	21,850	18,924	17,243	18,170
	Column Average	1,654	8,135	9,890	10,949	7,659
Average ZIP Area	Q1	127	129	168	119	129
	Q2	78	74	101	151	94
	Q3	67	79	74	63	74
	Q4	0	80	44	20	33
	Column Average	118	81	75	58	83
Average Crowding (Pop/Area)	Q1	283	45	146	88	244
	Q2	727	474	446	250	471
	Q3	525	999	1,390	1,467	1,221
	Q4	0	1,205	2,029	3,678	2,921
	Column Average	361	712	1,326	2,421	1,204
Average Number of Injuries	Q1	0.1	1	1	1	0.3
	Q2	4	6	6	4	5
	Q3	14	18	20	22	19
	Q4	0	39	51	87	71
	Column Average	1	13	26	55	24
Average Rate per 100,000 Population	Q1	7	48	85	202	17
	Q2	25	47	83	167	52
	Q3	31	53	85	146	72
	Q4	0	59	90	168	130
	Column Average	20	53	88	167	103

Table 4.7: Summary of Region Statistics: All Injuries

Injury Type	State Quartile on Number	State Quartile on Rate				Row Total or Row Average
		Q1	Q2	Q3	Q4	
Number	Q1	4	4	1	5	14
of	Q2	3	4	4	4	15
Regions	Q3	5	4	4	2	15
	Q4	2	4	6	5	17
	Column Total	14	16	15	16	61
Total	Q1	30,441	30,094	7,461	28,112	96,108
Population	Q2	180,759	131,719	75,222	111,935	499,635
	Q3	692,058	427,985	451,927	108,435	1,680,405
	Q4	823,696	2,694,349	3,347,737	2,829,832	9,695,614
	Column Total	1,726,954	3,284,147	3,882,347	3,078,314	11,971,762
Average	Q1	7,610	7,524	7,461	5,622	6,865
Population	Q2	60,253	32,930	18,806	27,984	33,309
	Q3	138,412	106,996	112,982	54,218	112,027
	Q4	411,848	673,587	557,956	565,996	570,330
	Column Average	123,354	205,259	258,823	192,395	196,258
Average	Q1	57	83	88	93	79
Number of	Q2	548	361	238	407	378
Injuries	Q3	1,308	1,160	1,412	791	1,227
	Q4	4,047	7,651	7,211	8,680	7,374
	Column Average	1,179	2,314	3,330	2,942	2,468
Average	Q1	249	369	393	549	385
Rate per	Q2	303	365	421	484	378
100,000	Q3	315	361	417	486	365
Population	Q4	328	379	431	511	431
	Column Average	319	376	429	510	419

Table 4.8: Summary of Region Statistics: Unintentional Injuries

Injury Type	State Quartile on Number	State Quartile on Rate Q1	Q2	Q3	Q4	Row Total or Row Average
Number of Regions	Q1	4	3	2	5	14
	Q2	3	2	5	5	15
	Q3	6	3	4	2	15
	Q4	1	8	6	2	17
	Column Total	14	16	17	14	61
Total Population	Q1	39,279	10,433	18,284	28,112	96,108
	Q2	180,759	57,850	137,234	123,792	499,635
	Q3	829,651	290,392	451,927	108,435	1,680,405
	Q4	557,008	4,413,714	3,643,889	1,081,003	9,695,614
	Column Total	1,606,697	4,772,389	4,251,334	1,341,342	11,971,762
Average Population	Q1	9,820	3,478	9,142	5,622	6,865
	Q2	60,253	28,925	27,447	24,758	33,309
	Q3	138,275	96,797	112,982	54,218	112,027
	Q4	557,008	551,714	607,315	540,502	570,330
	Column Average	114,764	298,274	250,078	95,810	196,258
Average Number of Injuries	Q1	63	30	87	77	64
	Q2	429	263	268	296	309
	Q3	988	813	1,122	658	945
	Q4	3,921	4,645	6,240	6,279	5,357
	Column Average	813	2,514	2,555	1,124	1,816
Average Rate per 100,000 Population	Q1	212	291	317	454	311
	Q2	237	303	326	399	309
	Q3	238	280	331	404	281
	Q4	235	281	342	387	313
	Column Average	236	281	341	391	308

Table 4.9: Summary Region Statistics: Intentional Injuries

Injury Type	State Quartile on Number	State Quartile on Rate Q1	Q2	Q3	Q4	Row Total or Row Average
Number of Regions	Q1	7	5	1	1	14
	Q2	6	4	3	2	15
	Q3	1	5	7	2	15
	Q4	0	1	4	12	17
	Column Total	14	15	15	17	61
Total Population	Q1	58,455	26,655	9,893	5,890	100,893
	Q2	224,710	143,142	86,744	24,746	479,342
	Q3	73,855	519,295	794,218	308,545	1,695,913
	Q4	0	266,688	2,837,110	6,591,816	9,695,614
	Column Total	357,020	955,780	3,727,965	6,930,997	11,971,762
Average Population	Q1	8,351	5,331	9,893	5,890	7,207
	Q2	37,452	35,786	28,915	12,373	31,956
	Q3	73,855	103,859	113,460	154,273	113,061
	Q4	0	266,688	709,277	549,318	570,330
	Column Average	25,501	63,719	248,531	407,706	196,258
Average Number of Injuries	Q1	9	10	28	17	12
	Q2	57	64	72	43	60
	Q3	115	203	265	479	263
	Q4	0	561	1,805	2,032	1,892
	Column Average	37	126	621	1,497	609
Average Rate per 100,000 Population	Q1	37	64	94	96	53
	Q2	51	60	83	116	63
	Q3	52	65	78	103	78
	Q4	0	70	85	123	111
	Column Average	49	66	83	122	103