

Evaluation of the secondary use of electronic health records to detect seasonal, holiday-related, and rare events related to traumatic injury and poisoning

Timothy Bergquist (✉ trberg@uw.edu)

University of Washington <https://orcid.org/0000-0001-5614-8977>

Vikas Pejaver

University of Washington

Noah Hammarlund

University of Washington

Sean D. Mooney

University of Washington

Stephen J. Mooney

University of Washington

Research article

Keywords: learning healthcare system, data science, population health, electronic health records

Posted Date: July 5th, 2019

DOI: <https://doi.org/10.21203/rs.2.11038/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published on January 13th, 2020. See the published version at <https://doi.org/10.1186/s12889-020-8153-7>.

Abstract

Background The increasing adoption of electronic health record (EHR) systems enables automated, large scale, and meaningful analysis of regional population health. We explored how EHR systems could inform surveillance of trauma-related emergency department visits arising from seasonal, holiday-related, and rare environmental events. **Methods** We analyzed temporal variation in diagnosis codes over 24 years of trauma visit data at the three hospitals in the University of Washington Medicine system in Seattle, Washington, USA. We identified seasons and days in which specific codes and categories of codes were statistically enriched, meaning that a significantly greater than average proportion of trauma visits included a given diagnosis code during that time period. **Results** We confirmed known seasonal patterns in emergency department visits for trauma. As expected, cold weather-related incidents (e.g. frostbite, snowboarding injury) were enriched in the winter, whereas fair weather-related incidents (e.g. bug bites, boating accidents, bicycle accidents) were enriched in the spring and summer. Our analysis of specific days of the year found that holidays were enriched for alcohol poisoning, assaults, and firework accidents. We also detected one time regional events such as the 2001 Nisqually earthquake and the 2006 Hanukkah Eve Windstorm. **Conclusions** Though EHR systems were developed to prioritize operational rather than analytic priorities and have consequent limitations for surveillance, our EHR enrichment analysis nonetheless re-identified expected temporal population health patterns. EHRs are potentially a valuable source of information to inform public health policy, both in retrospective analysis and in a surveillance capacity.

Background

Electronic Health Records and Meaningful Use

The past decade has seen a substantial increase in the rate of Electronic Health Record (EHR) adoption in healthcare.[1] While the primary drivers of EHR adoption have been the 2009 HITECH act and the data exchange capabilities of EHRs,[2] secondary use of EHR data to improve patient safety and health is a key benefit of large-scale adoption.[3] EHRs contain a rich set of information about patients and their health experiences, including doctor's notes, medications prescribed, and billing codes.[4] As hospitals improve data capture quality and quantity, opportunities arise for meaningful use of the data outside the clinic.

Electronic Health Records and Public Health

Public health surveillance – monitoring disease prevalence, and the conditions and behaviors that affect prevalence – is a core component of preventive medicine. Surveillance is conventionally categorized as either 'active' (wherein a health authority contacts care providers or the public to assess conditions) or 'passive' (wherein care providers are mandated to report certain conditions to the health authority).[5] For example, the Center for Disease Control's Behavior Risk Factor Surveillance System (BRFSS),[6] in which

trained interviewers contact tens of thousands of respondents by phone each year, is an active system. By contrast, the National Highway Transport Safety Administration's Fatality Analysis Reporting System, in which state transportation departments report motor vehicle crashes to a central system, is a passive system.

With the increasing adoption of EHRs, automated and scalable public health surveillance has become possible. Clinical data that is collected in routine medical care can be algorithmically processed for syndromic surveillance, a passive reporting technique wherein patient cases of a particular disease or condition relevant to population health (frequently, but not exclusively infectious disease) are automatically flagged and reported to appropriate authorities in real time. EHRs have been shown to be a reliable data source capable of facilitating syndromic surveillance.[7–11] The prevalence estimation of EHRs have also been shown to accurately reflect the known prevalence of a served region. For example, when compared to the gold standard BRFSS dataset, Klompas *et al.* found that an EHR-based diabetes prevalence detection algorithm was nearly as accurate as the BRFSS dataset.[8] Perlman *et al.* found that measures of smoking prevalence, obesity rates, hypertension, and diabetes that were derived from the EHR were as accurate as the gold standard BRFSS datasets.[12] The reliability of different conditions often differs by healthcare system, but as more sites adopt EHRs, the estimates should improve for more conditions.[13]

Previous efforts to use EHRs for public health reporting have revolved around using syndromic surveillance to electronically report cases to a data repository external to the EHR. For instance, Klompas *et al.* developed a platform for integrating EHR data for use in public health called the Electronic medical record Support for Public Health (ESP).[14] The platform enabled automated systems to pull relevant records from the EHR, and then aggregate data for visualization and analysis in an application called RiskScape.[7] A more recent example of integrating clinical data into a repository for public health surveillance was the Public Health Community Platform (PHCP), an attempt by multiple public health organizations (APHL, ASTHO, JPHIT) to standardize and develop a platform for EHR to cloud-based public health data sharing and electronic case reporting.[14, 15] While the pilot study faced several challenges, it demonstrated long-term feasibility for widespread integration between clinical practice and public health.

The EHR as a Generalizable Population Health Surveillance Platform

While syndromic surveillance typically focuses on the detection and prevalence estimation of specific conditions, electronic health record databases can act as a generalized population health surveillance

system, giving insight into previously unmonitored diseases. For instance, Melamed *et al.* showed the utility of EHRs to link diseases to seasonal trends.[16]

In this paper, we explore the utility of the EHR as a generalizable event and trend detection platform. We look at deviations from seasonal and temporal trends in medical information collected in routine clinical care, conceptualizing these deviations as events of potential interest to authorities tasked with monitoring population health. We focus specifically on traumatic injuries because they have known seasonal trends [16–18] and gold standard events by which we can validate a generalizable event detection method (e.g., we expect 4th of July to have a spike in firework accidents).

Throughout this paper, we use the term “detection” to refer to the association of statistical trauma trends with individual dates or seasons (e.g., can we “detect” winter or July 4th based on relative diagnosis code frequencies?). We look for diagnosis codes that are statistically “enriched” (a greater proportion of overall visits than would be expected due to chance alone) for different periods of time. We define a code as “enriched” when that code is significantly associated with a given period of time. For instance, we expect injuries from snow sports like skiing, snowboarding, and snowmobiling to be “enriched” in the winter months. We compare trends found to expected trends from literature and common knowledge to test the validity of this event detection technique.

Methods

Data Source

We obtained a limited data set from the UW Medicine (the University of Washington Health System) enterprise data warehouse (EDW). The EDW includes patient data from over 4.5 million patients spanning ~25 years, and representing various clinical sites across the UW Medicine system including University of Washington Medical Center, Harborview Medical Center, and Northwest Hospital and Medical Center. The EDW gathers data from over 60 sources from across these institutions including lab results, microbiology reports, demographic data, diagnosis codes, and reported allergies.

“Injury and poisoning” is a category of clinical affliction that includes any traumatic injury or poisoning and is coded as E-codes (E000-E999) or 800-999 codes using the ICD-9-CM diagnosis coding standard or S00-T99 or V00-Y99 codes using the ICD-10-CM coding standard, as defined in the CDC’s guidelines for traumatic injury and poisoning.[19, 20] From the EDW, we selected records of all visits between January 1, 1994 and May 2, 2017 for patients who were over the age of 18 as of May 2, 2017 and where, for each visit, at least one ICD-9-CM code or ICD-10-CM code in the “Injury and poisoning” category was recorded. For each patient record, we collected patient visit information which included de-identified patient ID,

diagnosis coding method (ICD-9-CM or ICD-10-CM), visit number identifier, admission date and time, diagnosis codes (ICD-9-CM or ICD-10-CM), and diagnosis code description. These data represent just over 3,000,000 unique trauma-related visits to the UW medical system made by over 650,000 unique individuals.

Data Cleaning

UW Medicine adopted the ICD-10-CM billing code system in mid-2015. In order to ensure we had consistent data throughout, we mapped ICD-10-CM codes to their ICD-9-CM equivalents, using the Center for Medicare and Medicaid Services (CMS) General Equivalence Mappings.^[21] Since ICD-10-CM has more detailed coding descriptions than ICD-9-CM, there is a potential for data loss when converting from ICD-10-CM to ICD-9-CM. While this may be an issue in some studies, we were more interested in the high level view of UW's patient population, and this data loss was not a major concern for this study. We used a custom tool, DxCodeHandler (<https://github.com/UWMooneyLab/DxCodeHandler>), to handle code conversion, ICD hierarchy traversal, and diagnosis code manipulation (Supplemental Materials).

Obtaining Count Data

Per our selection criteria, each patient visit included one or more ICD-9-CM or ICD-10-CM billing codes representing the billing information for the patient visit. We attributed all codes appearing in a visit to the day that visit occurred such that each day was considered a collection of independent code counts. Because codes recorded in the visit records are leaves in the ICD-9-CM hierarchy tree, we included all the ancestors of the codes in our analysis. For example, a day that had the code E880.0 (Accidental Fall on or from Escalator) would also have E880 (Accidental Fall from Stairs or Steps), E880-E888 (Accidental Falls), and E000-E999 (External Causes of Injury or Poisoning) counted on that day. In this manner, we have a set of days that have accumulated counts of diagnosis codes at different levels in the hierarchy.

Binomial Test and Hypothesis Testing

For each diagnosis code, both billable and parent codes, we tested the null hypothesis that the prevalence of each diagnosis code, when calculated against all trauma visits, was consistent across time. We tested this hypothesis using a binomial test, where we tested whether a diagnosis code is more or less prevalent in a given time period when compared to the expected prevalence if the null hypothesis were true. If a code-time period pair had a p-value less than the Bonferroni cutoff, we said that the code is enriched for that tested time period. We used an $\alpha=0.01$ when calculating the Bonferroni cut off for each experiment. We ran this test for every code that appears more than 10 times in our dataset for all four seasons and

for all 365 (non-leap year) days. For each code-time period pair, we generated a score by calculating the $-\log(p\text{-value})$ from the binomial test.

Enrichment of Seasons

To find seasonal statistical enrichment of ICD-9-CM billing codes we summed daily counts of each of the 4,582 poisoning and injury billing codes within each season. We defined Winter as December - February, Spring as March - May, Summer as June - August, and Autumn as September - November. For each season/code pair, we performed a binomial test, treating the sum of all codes in that season as the trials, and the count of the code in question for that season as the successes. The expected rate of appearance for each code in question was established by calculating its proportion of all trauma visits across all seasons and years. Thus, the p-value from this test is interpretable as the probability that these many codes or more would be seen in a given season under the null hypothesis that codes are evenly distributed across the year. We used a Bonferroni correction at $n = 18,328$ (4×4582). We also filtered out codes that appeared less than 10 times over the course of the 24-year period.

Enrichment of Dates

We used an analogous method to detect code enrichments for days of the year. Again, we computed the sum of codes occurring on each of the 365 (non-leap-day) days of the year. For each code/day pair, we performed a binomial test using the total number of codes used on that day as the number of trials, and the number of times the specific code of interest was used as the number of successes. The expected rate was derived from the baseline rate of appearance for the code of interest per day across the entire year when compared to the total number of trauma visits on that given day. We calculated a Bonferroni cutoff at $n = 1,672,430$ (4582×365). We counted codes as enriched if the p-value was less than the Bonferroni correction and the daily rate of the code was greater than the baseline expected rate of the code (we did not look at depletions). We also filtered out codes that appeared less than 10 times over the course of the 24 year dataset period.

IRB Considerations

We received an IRB non-human subjects research designation from the UW Human Subjects Research Division to construct a limited dataset for all patients from the EDW over the age of 18 under a waiver of consent. (IRB number: STUDY00000669) Data was extracted by an honest broker, the UW Medicine Research IT data services team, under a waiver of consent. No direct identifiers were given to the research team by the honest broker, and all indirect identifiers fell under the limited dataset category.

Results

Statistical Enrichment of Seasons

We detected patterns of seasonal enrichment consistent with our expectations about seasonal behavior. For example, in winter, we found enrichment of not only accidents from snow sports such as skiing and snowboarding, among others, but also cold weather-related ailments such as frostbite and hypothermia. Other codes that may be related to snow sport accidents such as head injuries, sprains, and strains were also enriched (Figure 1). To a lesser degree, spring sees more snow sport related accidents (Washington's mountains typically remain snow-covered into the late Spring or early Summer) but begins to have more fair weather activities such as off-road vehicle accidents, skateboarding accidents, and outdoor related ailments like allergies, lawn mower accidents, and sporting accidents (Figure 2). Summer sees disproportionate numbers of accidents related to outdoor activities in warm weather such as bites and stings from bugs, heat stroke, motorcycle accidents, bicycle accidents, and water transport accidents (Figure 3). While fall is the least distinctive of the seasons, it shares some trauma types with summer and has a unique enrichment for vehicle accidents (Figure 4). This may be because fall contains high traffic holidays (Thanksgiving, Labor Day) and increased levels of rain in Seattle.

Statistical Enrichment for Days of the Year

To complement our seasonal analyses, we explored enrichment of diagnosis codes for all 365 days of the year. Each date that had a code scored below the Bonferroni threshold was flagged as having possible significance. We detected 100 days that had at least one code flagged as enriched. We generated an enrichment score for each of the dates by calculating the $-\log(p\text{-value})$ of the lowest p-value for the date. The top 15 dates with the highest scoring codes are shown (Figure 5). The days in which enrichment of many codes is common are a mixture of holidays and one time events. For example, there was enrichment of codes related to fights, firework accidents, and alcohol poisoning on January 1st (Table 1). Analogously, there was a large increase in the number of firework related accidents and burns on the 4th and 5th of July as well as an increase in the number of off-road vehicle accidents and poisoning by alcohol (Table 2-3). We also observe an increase in alcohol poisoning, vehicle accidents, and an increase in possible self-harm on Christmas Eve (Table 4). For tables 1-4, we limit the reporting of codes to those that had more than 30 appearances over the 24 years of data. This reduces false positives arising from extremely rare codes that appeared during the baseline period. We also report by percent increase rather than $-\log(p)$ for better interpretability.

Rare Events as Case Studies

We detected enrichment of unusual codes on multiple days that did not seem linked to their respective day by either holiday or seasonal event. Upon further evaluation, we inferred that we had detected past

environmental events that showed up as single day enrichments. Feb 28, Dec 15, May 31, and Nov 8 were four of the days in the top 15 highest scoring days that followed this pattern (Figure 5). Because these enriched days fell in single years, we were able to search for news stories published on or immediately after these days to see if we could find the cause of the increase in these unusual codes.

Nisqually Earthquake

In our analysis, February 28th was shown to have an increase in earthquake related accidents, ICD-9-CM code E909.0. On February 28, 2001, there was a magnitude 6.8 earthquake centered in Western Washington.[\[22, 23\]](#) All the earthquake codes found on February 28th in our dataset were from 2001, consistent with there being very few earthquake related accidents in the EHR except during the major earthquake.

Hanukkah Eve windstorm

Our event detection method also discovered a significant increase on December 15 of the ICD-9-CM code E868.3 (accidental poisoning by carbon monoxide from incomplete combustion of other domestic fuels). Nearly all the codes were found to have been coded in 2006. The Hanukkah Eve windstorm of Dec 15, 2006 led to widespread and lengthy power outages. In the aftermath, there were news stories about the increase in carbon monoxide poisonings due to people barbecuing and running generators in their homes without ventilation.[\[24, 25\]](#) Indeed, public health authorities responded with concerns that the dangers of carbon monoxide poisoning were not widely understood in select communities.[\[26\]](#)

Industrial Accidents

We detected two other single day enrichments: May 31 with an enrichment of E891.3 (Burning caused by conflagration) and Nov 8 with an enrichment of 987.6 (Toxic effect of chlorine gas). We were able to link these two enrichments to the May 31, 2004 monorail fire in Seattle [\[27\]](#) and the November 8, 1994 chlorine spill and fire at the Coastal Dock in Ballard, WA.[\[28\]](#)

Discussion

We explored the value of UW Medicine electronic health record data for detecting public health-related environmental and seasonal causes of traumatic injury. Our analysis finds that tests for seasonal and daily enrichment of the frequency of emergency room visits for trauma detects expected events, including both seasonal trends such as winter sports-related injuries, day-specific events such as July 4th burns, and rare events such as the Nisqually earthquake.

Interesting Anomalies

Non-enriched Holidays

While most of our results confirmed expected seasonal and date-specific trends, we were surprised not to find enrichment of alcohol related injuries on St. Patrick's Day or the day following, given that St. Patrick's Day is associated with increased alcohol consumption.[29, 30] This may indicate the effectiveness of extra police patrols deployed for that day. This could also be a false negative due to the conservative nature of Bonferroni corrections.

Prior studies have examined date-related events in relation to traumatic injury. One study found that on April 20th, a date associated with celebrating marijuana consumption, there was an increase in the number of car accidents.[31] While we did not observe a statistical enrichment in car accidents, our method did identify a statistical enrichment in burns (940-949), another potential consequence of marijuana use.[32] Future work could analyze clinical notes which might allow us to identify if this enrichment is attributable to elevated marijuana use.

Enrichment of Post-surgical Complications in Winter

We also saw unexpected trends in post-surgical complications, with those terms being enriched in the winter months at the very end and beginning of the year. One hypothesis is that there is a relative increase in the number of surgeries in November and December as people schedule elective surgeries before insurance deductibles reset in the new year. An alternate hypothesis is that people defer reporting minor surgical complications until after the end-of-year holidays. We were unable to explore these hypotheses for this study because our data was limited to visits including trauma codes and did not include surgical appointments. It is also important to note that we saw a relative increase in the number of surgical complications due to lower numbers of trauma visits in the winter, and not necessarily an absolute increase in the number of post-surgical complications (Figure 6). Since codes related to post-surgical complications are less specific and are more likely to appear during trauma visits than other codes discussed thus far, the effect of this "lowered baseline" is particularly noticeable.

Unlinked Events

There were multiple dates that had significant enrichment of codes on a date where nearly all the codes came from one year. For instance, there were a large number of visits with the code 994.9 (other effect of external causes) on one of the masked days. This code is too vague to understand the common injuries of patients and, at the time of this study, we did not have access to de-identified clinical notes from which

to illicit the causes of these injuries. There was also no readily available source of news that we found to corroborate a large number of people being injured by any social or environmental event. We were not able to discern whether these dates were false positives, whether the codes were entered incorrectly, or whether there was a common event that caused these injuries. In this paper, we have masked the specific dates of these unlinked days to protect against the potential de-identification of patients since the circumstances surrounding these injuries are unknown.

Study Strengths and Limitations

Our study has several notable strengths. First, the UW Medicine system has used EHRs for a long time, affording us access to over 20 years of clinical data from a large urban health care system. Second, UW Medicine's location in Western Washington lends itself to year-round yet season-specific outdoor activities whose resulting injuries show up as specific trauma codes, including snow sports in the winter and boating in the summer. This access increased our ability to detect seasonal trauma trends.

However, our study also has limitations. First, as with any study of electronic health records, we cannot rule out biases due to site-specific coding practices or changes in practitioner knowledge of the health record system. However, we have no reason to believe errors caused by these issues would vary by season or day. Second, the UWMC is mainly a referral institution, such that many patients visit the system only for specialty services. This is mitigated in our study by the fact that we only considered trauma-related diagnosis codes and that UW Medicine is the only Level I trauma center in Washington, Alaska, Montana and Idaho. Nevertheless, further validation studies are needed to evaluate the representation of the UWMC data in the Seattle Region.

Using Electronic Health Records for Event Detection

Our method could be used in a live surveillance situation by alerting authorities and doctors when an unusual increase of cases with a particular diagnosis code show up across multiple hospitals with linked EHR systems. It could spark an investigation into what is causing the sudden increase but also could initiate public health policy development that previously would take longer to assess and carry out. While our method focused on traumatic injury, it could easily be expanded to include surveillance of all diagnosis codes. Incorporating clinical notes into the method could more accurately cluster events and better inform detected trends. Natural language processing techniques could be used to find "enriched" keywords on the detected days to add context to the detected events in a data driven automated manner.

Conclusions

In conclusion, electronic health record data hold considerable potential for public health surveillance. We explored the potential to leverage UW Medicine's enterprise data warehouse to detect seasonal, holiday, and rare events using diagnosis codes for injuries and poisonings. Our method detected many of the trends for seasons and specific dates we expected, while identifying several intriguing new enrichments. Future research should focus on improving our trend and event detection method to differentiate between one-time effects like the Nisqually earthquake, and repeat events like Independence Day. Incorporating clinical notes into a detection method could more accurately cluster events and better inform detected trends. Expanding the method to all diagnosis codes could detect new non-trauma related events. Our findings add to the growing body of literature showing that electronic health records hold considerable potential as generalizable population health surveillance platforms.

Abbreviations

UWMC: University of Washington Medical Center

EHR: Electronic Health Record

ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification

ICD-10-CM: International Classification of Diseases, Tenth Revision, Clinical Modification

BRFSS: Behavioral Risk Factor Surveillance System

EDW: Enterprise Data Warehouse

Declarations

Ethics approval and consent to participate

We received an IRB non-human subjects research designation from the UW Human Subjects Research Division to construct a limited dataset for all patients from the EDW over the age of 18 under a waiver of consent. (IRB number: STUDY00000669) Data was extracted by an honest broker, the UW Medicine Research IT data services team, under a waiver of consent. No direct identifiers were given to the research team by the honest broker, and all indirect identifiers fell under the limited dataset category.

Availability of data and materials

The datasets generated and analyzed during the current study are not publically available due to privacy and HIPAA concerns surrounding the patient data.

Funding

This publication was supported by the National Center For Advancing Translational Sciences of the National Institutes of Health under Award Numbers UL1 TR002319 and U24TR002306. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Timothy Bergquist and Sean Mooney are supported by the National Institutes of Health grants R01 LM007722. Vikas Pejaver is supported by the Washington Research Foundation Fund for Innovation in Data-Intensive Discovery and the Moore/Sloan Data Science Environments Project at the University of Washington. Stephen Mooney is supported by grant K99LM012868.

Authors' contributions

TB analyzed and interpreted the longitudinal EHR data to detect historical events and was a major contributor in writing the manuscript. SJM was a major contributor in writing the manuscript as well as in the conception and design of the study. VP and NH were contributors in writing the manuscript. SDM helped conceive of the project with TB and SJM, funded, and helped oversee scientific progress. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank Drs. Adam Wilcox, Gang Luo, and Peter Tarczy-Hornoch on their feedback about methods, patient privacy, and University of Washington policy compliance.

Conflicts of Interest

The authors declare that they have no competing conflicts of interest.

References

1. Charles D, Gabriel M, Ma TSM Adoption of Electronic Health Record Systems among U.S. Non-Federal Acute Care Hospitals: 2008-2014.
2. Heisey-Grove D, Patel V (2014) Physician motivations for adoption of electronic health records. Washington, DC: Office of the National Coordinator for Health Information Technology
3. Birkhead GS, Klompas M, Shah NR (2015) Uses of Electronic Health Records for Public Health Surveillance to Advance Public Health. *Annu Rev Public Health* 36:345–359

4. Jones SS, Rudin RS, Perry T, Shekelle PG (2014) Health information technology: an updated systematic review with a focus on meaningful use. *Ann Intern Med* 160:48–54
5. Teutsch SM, Elliott Churchill R (2000) *Principles and Practice of Public Health Surveillance*. Oxford University Press
6. Pierannunzi C, Hu SS, Balluz L (2013) A systematic review of publications assessing reliability and validity of the Behavioral Risk Factor Surveillance System (BRFSS), 2004-2011. *BMC Med Res Methodol* 13:49
7. Klompas M, Murphy M, Lankiewicz J, et al (2011) Harnessing electronic health records for public health surveillance. *Online J Public Health Inform*. doi: 10.5210/ojphi.v3i3.3794
8. Klompas M, Eggleston E, McVetta J, Lazarus R, Li L, Platt R (2013) Automated detection and classification of type 1 versus type 2 diabetes using electronic health record data. *Diabetes Care* 36:914–921
9. Calderwood MS, Platt R, Hou X, Malenfant J, Haney G, Kruskal B, Lazarus R, Klompas M (2010) Real-time surveillance for tuberculosis using electronic health record data from an ambulatory practice in eastern Massachusetts. *Public Health Rep* 125:843–850
10. Elliott AF, Davidson A, Lum F, Chiang MF, Saaddine JB, Zhang X, Crews JE, Chou C-F (2012) Use of electronic health records and administrative data for public health surveillance of eye health and vision-related conditions in the United States. *Am J Ophthalmol* 154:S63–70
11. Klompas M, Haney G, Church D, Lazarus R, Hou X, Platt R (2008) Automated Identification of Acute Hepatitis B Using Electronic Medical Record Data to Facilitate Public Health Surveillance. *PLoS One* 3:e2626
12. Perlman SE, McVeigh KH, Thorpe LE, Jacobson L, Greene CM, Gwynn RC (2017) Innovations in Population Health Surveillance: Using Electronic Health Records for Chronic Disease Surveillance. *Am J Public Health* 107:853–857
13. Perlman SE, Charon Gwynn R, Greene CM, Freeman A, Chernov C, Thorpe LE (2018) NYC HANES 2013-14 and Reflections on Future Population Health Surveillance. *J Urban Health*. doi: 10.1007/s11524-018-0284-0
14. Klompas M, McVetta J, Lazarus R, et al (2012) Integrating clinical practice and public health surveillance using electronic medical record systems. *Am J Public Health* 102 Suppl 3:S325–32
15. Cooney MA, Iademarco MF, Huang M, MacKenzie WR, Davidson AJ (2018) The Public Health Community Platform, Electronic Case Reporting, and the Digital Bridge. *J Public Health Manag Pract* 24:185–189
16. Melamed RD, Khiabani H, Rabadan R (2014) Data-driven discovery of seasonally linked diseases from an Electronic Health Records system. *BMC Bioinformatics* 15 Suppl 6:S3
17. Bandi P, Silver D, Mijanovich T, Macinko J (2015) Temporal trends in motor vehicle fatalities in the United States, 1968 to 2010 - a joinpoint regression analysis. *Injury Epidemiology* 2:4
18. Rodríguez JM, Peñaloza RE, Moreno Montoya J (2015) Road Traffic Injury Trends in the City of Valledupar, Colombia. A Time Series Study from 2008 to 2012. *PLoS One* 10:e0144002

19. Fingerhut LA, Warner M (2006) The ICD-10 injury mortality diagnosis matrix. *Inj Prev* 12:24–29
20. (1997) Recommended Framework for Presenting Injury Mortality Data.
<https://www.cdc.gov/mmwr/preview/mmwrhtml/00049162.htm>. Accessed 18 Oct 2018
21. (2014) 2015-ICD-10-CM-and-GEMs.
22. (2001) CNN.com - Major U.S. quakes - February 28, 2001. CNN
23. (2018) Largest recorded earthquake in WA was 17 years ago. In: KING.
<https://www.king5.com/article/news/local/largest-recorded-earthquake-in-wa-was-17-years-ago/281-67102021>. Accessed 21 Nov 2018
24. Local News | Carbon-monoxide poisoning kills Burien man | Seattle Times Newspaper.
<http://community.seattletimes.nwsourc.com/archive/?date=20070124&slug=dige24m>. Accessed 13 Nov 2018
25. Hanukkah Eve Wind Storm ravages Western Washington beginning on December 14, 2006.
<http://www.historylink.org/File/8042>. Accessed 13 Nov 2018
26. Gulati RK, Kwan-Gett T, Hampson NB, Baer A, Shusterman D, Shandro JR, Duchin JS (2009) Carbon monoxide epidemic among immigrant populations: King County, Washington, 2006. *Am J Public Health* 99:1687–1692
27. (2004) CNN.com - Monorail train catches fire in Seattle - May 31, 2004. CNN
28. F/V Yardarm Knot Fire/Chlorine Release | IncidentNews | NOAA.
<https://incidentnews.noaa.gov/incident/7054>. Accessed 18 Dec 2018
29. Ruddell R, Thomas MO, Way LB (2005) Breaking the chain: Confronting issueless college town disturbances and riots. *J Crim Justice* 33:549–560
30. Glindemann KE, Wiegand DM, Geller ES (2007) Celebratory Drinking and Intoxication: A Contextual Influence on Alcohol Consumption. *Environ Behav* 39:352–366
31. Staples JA, Redelmeier DA (2018) The April 20 Cannabis Celebration and Fatal Traffic Crashes in the United States. *JAMA Intern Med* 178:569–572
32. Bell C, Slim J, Flaten HK, Lindberg G, Arek W, Monte AA (2015) Butane Hash Oil Burns Associated with Marijuana Liberalization in Colorado. *J Med Toxicol* 11:422–425

Tables

Table 1 – Top 10 most enriched codes for January 1st.

The top 10 most enriched codes for January 1st. As expected for New Year’s Day, the most enriched codes were related to firework accidents, alcohol, and assaults. To reduce the false positive rate of the code enrichment from extremely rare codes that appeared during the baseline period, the enriched codes were only counted if

they appeared more than 10 times over the 24 year period. We also report by percent increase rather than $-\log(p)$ for better interpretability.

Dx Code	January 1st Average Code Count	Daily Average Code Count	% Increase	Description
E923.0	1.74	0.07	2469.74	Accident caused by fireworks
E923	2.39	0.22	981.77	Accident caused by explosive material
E965	1.39	0.42	235.15	Assault by firearms and explosives
854.06	1.57	0.49	217.75	Intracranial injury with loss of consciousness of unspecified duration
E922.9	1.52	0.54	180.57	Accident caused by unspecified firearm missile
E922	1.87	0.69	171.55	Accident caused by firearm and air gun missile
E860	5.39	2.01	168.36	Accidental poisoning by alcohol, not elsewhere classified
E860-E869	5.96	2.23	167.25	Accidental Poisoning By Other Substance
E860.0	5.17	1.95	165.22	Accidental poisoning by alcoholic beverages
980.8	1.57	0.61	154.88	Toxic effect of other specified alcohols

Table 2 - Top 10 enriched codes for July 4th.

The top 10 most enriched codes for July 4th. As expected for Independence Day, the most enriched codes were related to firework accidents, burns, and alcohol poisoning. To reduce the false positive rate of the code enrichment from extremely rare codes that appeared during the baseline period, the enriched codes were only counted if they appeared more than 10 times over the 24 year period. We also report by percent increase rather than $-\log(p)$ for better interpretability.

Dx Code	July 4th Average Code Count	Daily Average Code Count	% Increase	Description
E923.0	4.26	0.06	6913.3	Accident caused by fireworks
E923	4.91	0.21	2194.4	Accident caused by explosive material
E820-E825	1.70	0.69	147.0	Motor Vehicle Non-traffic Accidents
980.8	1.43	0.61	133.5	Toxic effect of other specified alcohols
948.00	2.87	1.54	85.8	Burn involving less than 10 percent of body surface with third degree burn
948.0	2.87	1.56	83.8	Burn involving less than 10 percent of body surface
948	4.09	2.24	82.5	Burns classified according to extent of body surface involved
E819.2	3.00	1.70	76.2	Motor vehicle traffic accident of unspecified nature injuring motorcyclist
851	2.09	1.22	71.1	Cerebral laceration and contusion
851.8	1.35	0.79	69.9	Other and unspecified cerebral laceration and contusion, without mention of open intracranial wound

Table 3 – Top 10 enriched codes for July 5th.

The top 10 most enriched codes for July 5th. As expected for the day after Independence Day, the most enriched codes were related to firework accidents and burns as the injured persons from July 4th continue to appear in the hospital. To reduce the false positive rate of the code enrichment from extremely rare codes that appeared during the baseline period, the enriched codes were only counted if they appeared more than 10 times over the 24 year period. We also report by percent increase rather than $-\log(p)$ for better interpretability.

Dx Code	July 5th Average Code Count	Daily Average Code Count	% Increase	Description
E923.0	7.43	0.05	14186.38	Accident caused by fireworks
E923	8.65	0.20	4144.02	Accident caused by explosive material
940	1.35	0.31	330.34	Burn confined to eye and adnexa
944.2	1.61	0.41	288.29	Blisters, epidermal loss [second degree] of hand, unspecified site
948.00	5.87	1.54	282.12	Burn involving less than 10 percent of body surface with third degree burn
948.0	5.87	1.55	278.04	Burn [any degree] involving less than 10 percent of body surface
948	7.48	2.23	235.42	Burns classified according to extent of body surface involved
944.2	4.57	1.40	225.29	Blisters, epidermal loss [second degree]
943.2	2.70	0.83	223.88	Blisters, epidermal loss [second degree]
941.2	2.74	0.86	218.77	Blisters, epidermal loss [second degree]
921.3	1.35	0.48	182.44	Contusion of eyeball

Table 4 – Top 10 enriched codes for December 24th.

The top 10 most enriched codes for December 24th. The most enriched codes were related to alcohol poisoning, injury to spleen, and injury undetermined whether accidental or purposely inflicted. To reduce the false positive rate of the code enrichment from extremely rare codes that appeared during the baseline period, the enriched codes were only counted if they appeared more than 10 times over the 24 year period. We also report by percent increase rather than $-\log(p)$ for better interpretability.

Dx Code	December 24th Average Code Count	Daily Average Code Count	% Increase	Description
865.0	1.65	0.93	77.16	Injury to Spleen without mention of open wound into cavity
980.0	4.39	2.56	71.76	Toxic effect of ethyl alcohol
865	1.65	1.01	64.33	Injury to spleen
E980-E989	2.96	1.87	57.91	Injury Undetermined Whether Accidentally Or Purposely Inflicted
E980	1.83	1.17	55.93	Poisoning by solid or liquid substances
980	4.57	3.29	38.70	Toxic effect of alcohol
E812.1	1.96	1.46	34.11	Other motor vehicle traffic accident involving collision with motor vehicle injuring passenger in motor vehicle other than motorcycle
E816	2.22	1.73	28.51	Motor vehicle traffic accident due to loss of control
E849.9	2.65	2.21	19.85	Accidents occurring in unspecified place
E819.9	5.17	4.61	12.31	Motor vehicle traffic accident of unspecified nature

Figures

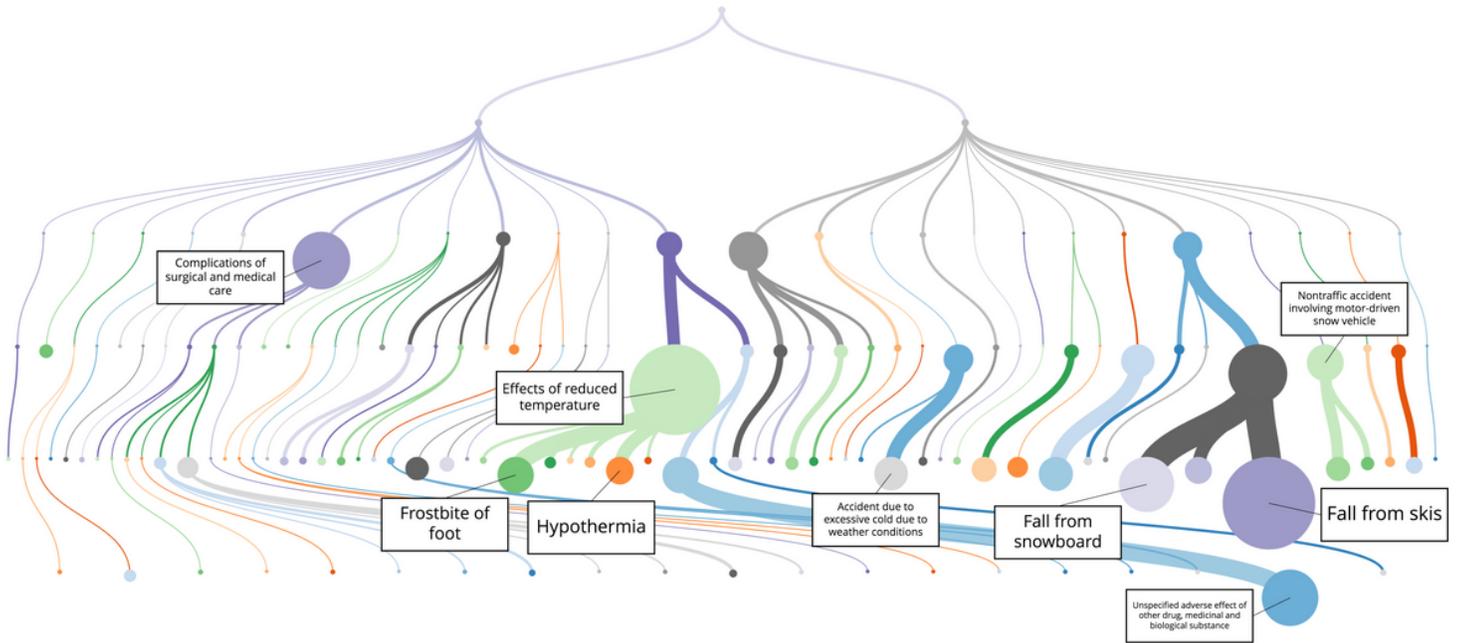


Figure 1

Hierarchical visualization of the enriched diagnosis codes in winter. The enriched ICD-9 codes for the winter months across 24 years of data. The radius of the individual nodes is proportional to the significance of the enrichment of that node's ICD-9 code in the winter. The depth of the node in the visualized hierarchy is equal to the depth of the code in the ICD-9 hierarchy. Cold weather and snow related injuries are enriched (Fall from snowboard, Hypothermia, Frostbite, etc.) Colors are meant solely for visual clarity as are the widths of the paths. The path widths are calculated from the smallest attached node.

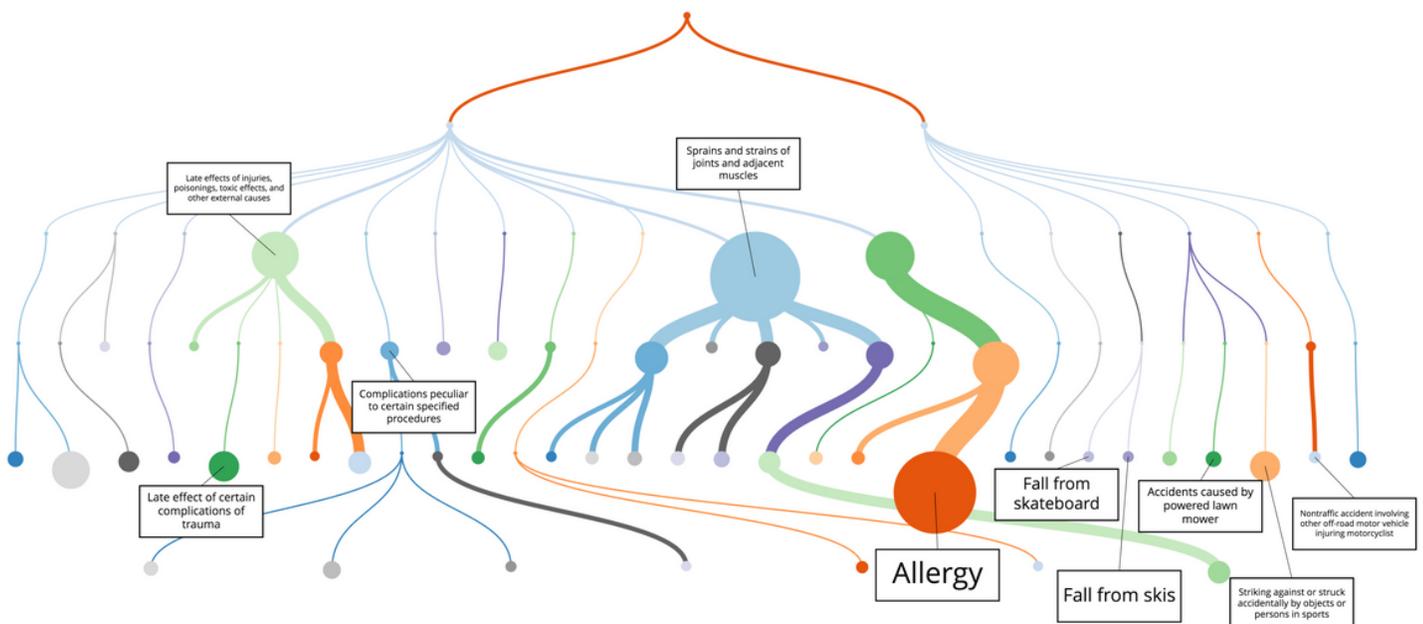


Figure 2

Hierarchical visualization of the enriched diagnosis codes in spring. The enriched ICD-9 codes for the spring months across 24 years of data. The size of the individual nodes is proportional to the significance of the enrichment of that node's ICD-9 code in the spring. The depth of the node in the visualized hierarchy is equal to the depth of the code in the ICD-9 hierarchy. "Fall from skis" is enriched as people get the last runs of the season in. Fair weather related activities and ailments are on the rise such as allergies, fall from skateboard, lawn mower accidents, and sporting accidents. Colors are meant solely for visual clarity as are the widths of the paths. The path widths are calculated from the smallest attached node.

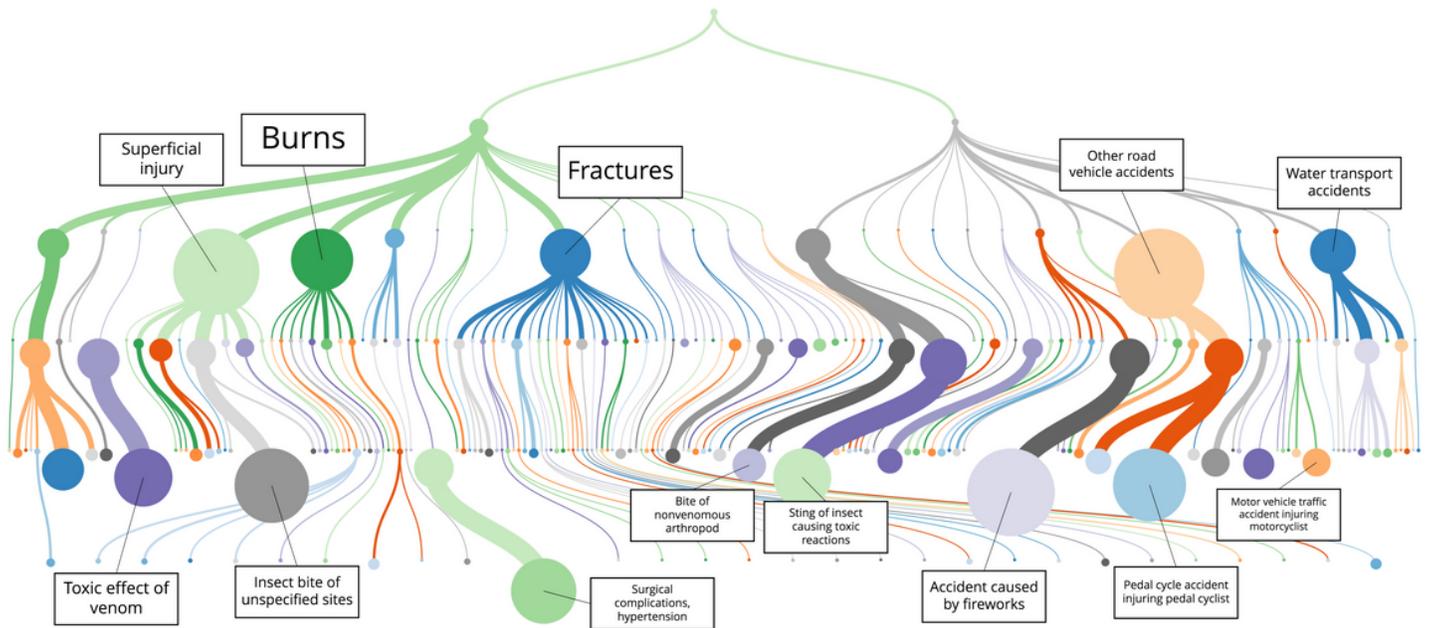


Figure 3

Hierarchical visualization of the enriched diagnosis codes in summer. The enriched ICD-9 codes for the summer months across 24 years of data. The size of the individual nodes is proportional to the significance of the enrichment of that node's ICD-9 code in the summer. The depth of the node in the visualized hierarchy is equal to the depth of the code in the ICD-9 hierarchy. Fair weather and heat injuries are enriched (Bug bites, boating accidents, motorcycle accidents, etc.). Colors are meant solely for visual clarity as are the widths of the paths. The path widths are calculated from the smallest attached node.

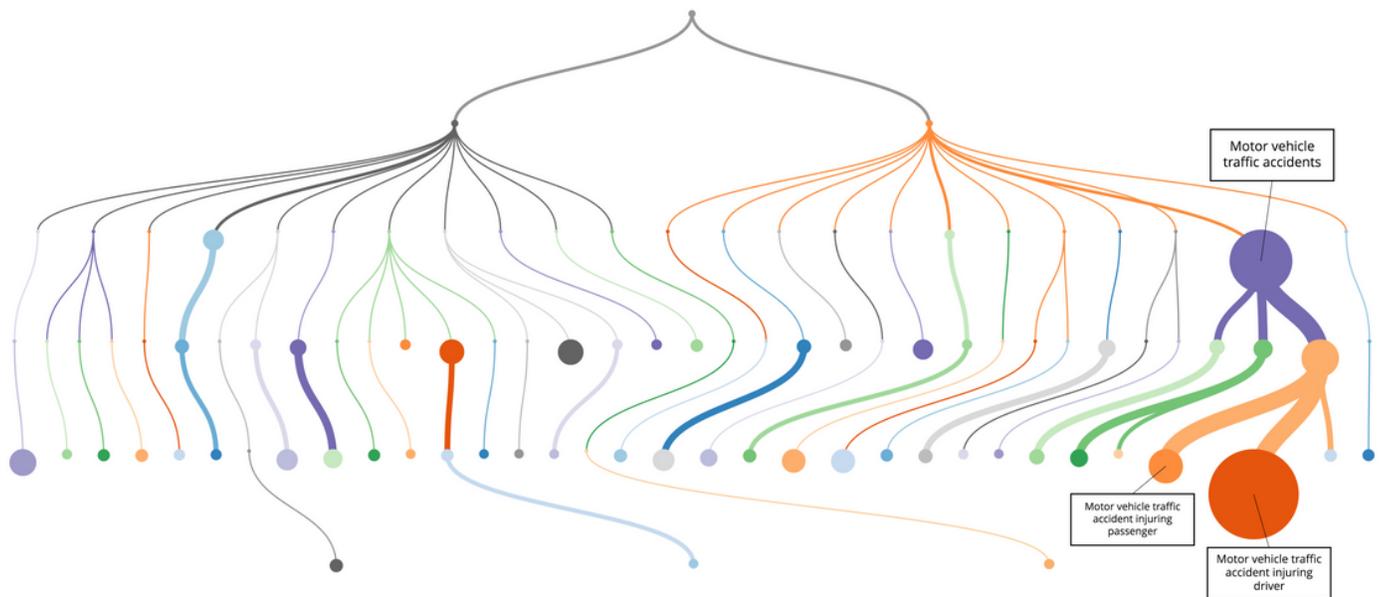


Figure 4

Hierarchical visualization of the enriched diagnosis codes in fall. The enriched ICD-9 codes for the fall months across 24 years of data. The size of the individual nodes is proportional to the significance of the enrichment that node's ICD-9 code in the fall. The depth of the node in the visualized hierarchy is equal to the depth of the code in the ICD-9 hierarchy. Motor Vehicle Accidents are enriched as fall contains multiple high traffic holidays like Labor Day weekend, Veteran's Day, and Thanksgiving. Colors are meant solely for visual clarity as are the widths of the paths. The path widths are calculated from the smallest attached node.

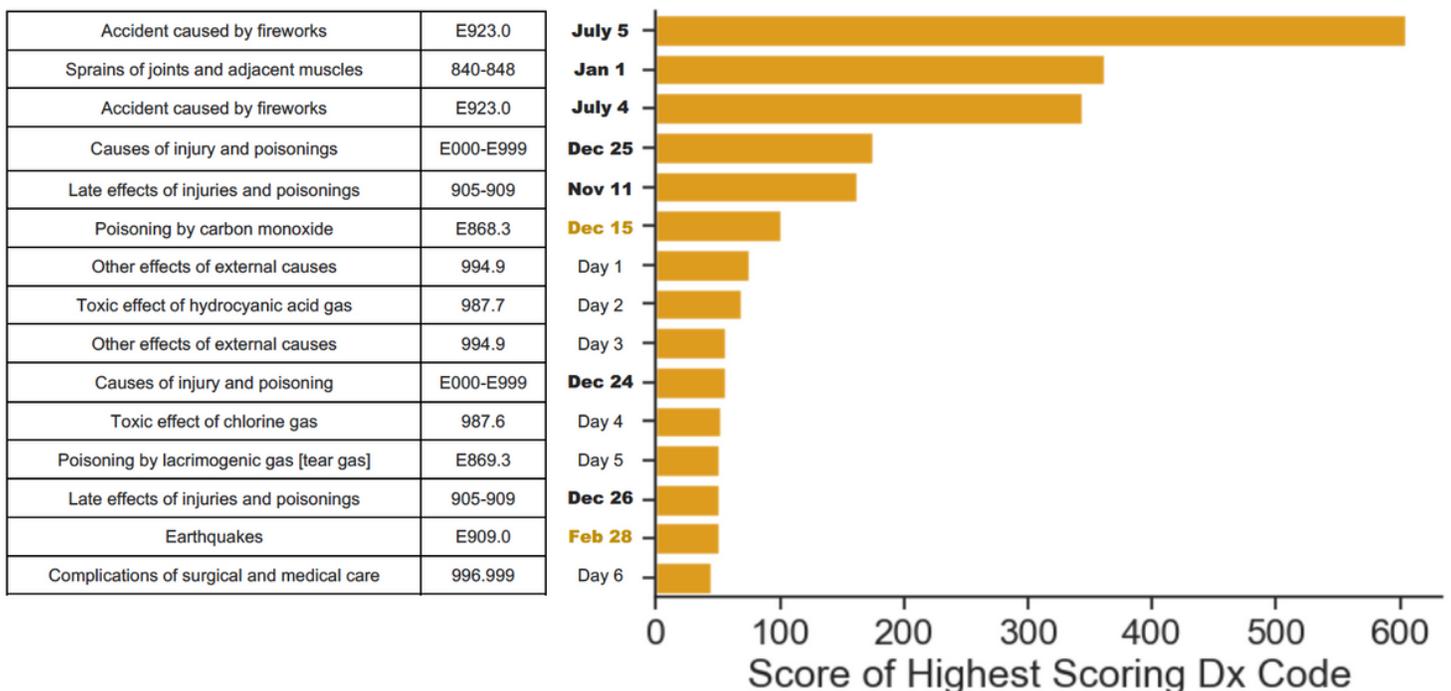


Figure 5

The top 15 highest scoring days of the year. The top 15 days with the highest scoring diagnosis codes. Each of the codes in the table are the most enriched codes on each of the days in the date column. The black bolded dates are either holidays or are dates that surround a holiday. The orange bolded dates are associated with known rare events that clearly explain the enrichment of their codes, namely the Nisqually Earthquake on Feb 28, 2001 and the Hanukkah Eve Windstorm on Dec 15, 2006. The other dates have unusual patterns of enriched codes such as chlorine gas poisoning and tear gas poisoning, but we could not find a readily available explanation to confirm some holiday, environmental, or social event on these days. Since these events appear to have happened on a single day in a single year and look to be associated with specific events, we have masked the dates due to the unknown specificity of these events and potential for identification of individuals involved in these events.

Monthly Percentage Deviation from the Annual Monthly Average

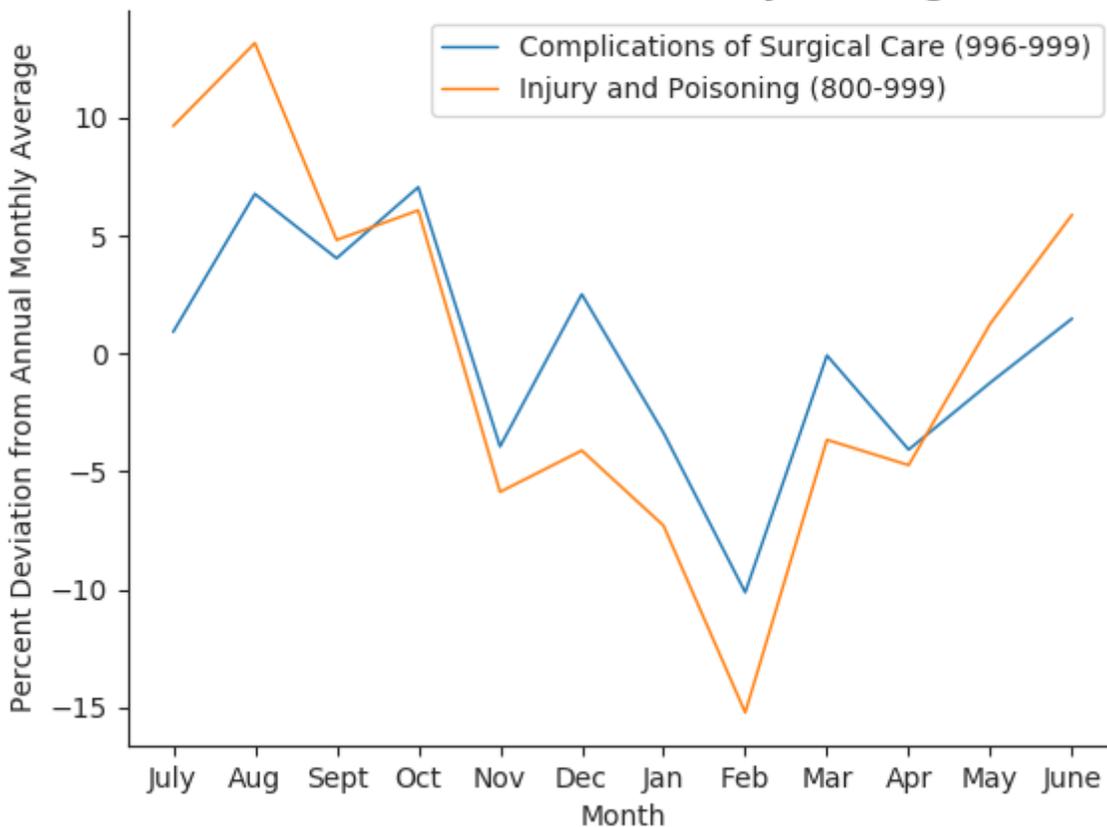


Figure 6

Comparison of the code count trend differences between 996-999 and 800-999. The percent deviation from the annual monthly average code count for both the Complications of Surgical Care (996-999) diagnosis family and the broad category of Injury and Poisoning (800-999). By calculating the average monthly code count for each family and the percent deviation per month from that expected average, we see that both code families follow a similar seasonal pattern of increase in the summer and decrease in the winter in terms of raw code count. While they follow the same pattern, Complications of Surgical Care doesn't decrease as much in the winter, and actually has a spike in December, which is why our method

picks up this diagnosis family as enriched in the winter. Since the number of trauma visits is used to establish a baseline expected rate of each code count, our method is detecting relative enrichment and not absolute enrichment.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplement1.docx](#)