Combining statistics from two national complex surveys to estimate injury rates per hour exposed and variance by activity in the USA

Tin-Chi Lin
*Harvard T.H. Chan School of Public Health*

*Et al.*

Let us know how access to this document benefits you.

Follow this and additional works at: [https://escholarship.umassmed.edu/oapubs](https://escholarship.umassmed.edu/oapubs)

Part of the Clinical Epidemiology Commons, Community Health and Preventive Medicine Commons, and the Epidemiology Commons

Repository Citation

Creative Commons License

This work is licensed under a Creative Commons Attribution-Noncommercial 4.0 License

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in Open Access Articles by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.
Combining statistics from two national complex surveys to estimate injury rates per hour exposed and variance by activity in the USA

Tin-chi Lin,1,2 Helen R Marucci-Wellman,1 Joanna L Willetts,1 Melanye J Brennan,1 Santosh K Verma1,3

ABSTRACT

Background A common issue in descriptive injury epidemiology is that in order to calculate injury rates that account for the time spent in an activity, both injury cases and exposure time of specific activities need to be collected. In reality, few national surveys have this capacity. To address this issue, we combined statistics from two different national complex surveys as inputs for the numerator and denominator to estimate injury rate, accounting for the time spent in specific activities and included a procedure to estimate variance using the combined surveys.

Methods The 2010 National Health Interview Survey (NHIS) was used to quantify injuries, and the 2010 American Time Use Survey (ATUS) was used to quantify time of exposure to specific activities. The injury rate was estimated by dividing the average number of injuries (from NHIS) by average exposure hours (from ATUS), both measured for specific activities. The variance was calculated using the ‘delta method’, a general method for variance estimation with complex surveys.

Results Among the five types of injuries examined, ‘sport and exercise’ had the highest rate (12.64 injuries per 100 000 h), followed by ‘working around house/yard’ (6.14), driving/riding a motor vehicle (2.98), working (1.45) and sleeping/resting/eating/drinking (0.23). The results show a ranking of injury rate by activity quite different from estimates using population as the denominator.

Conclusions Our approach produces an estimate of injury risk which includes activity exposure time and may more reliably reflect the underlying injury risks, offering an alternative method for injury surveillance and research.

INTRODUCTION

Non-fatal injury posts a substantial burden on the nation’s health and safety. The US National Center for Health Statistics (NCHS) estimated that in 2012 over 30 million non-fatal injuries required medical attention;1 the lifetime medical and work-loss costs associated with non-fatal injuries totalled over $457 billion in 2013.2 There is no daily routine or activity that is genuinely free from injury risk, but if certain activities are known to have higher injury risk, appropriate task-based intervention strategies could be developed and tracked, which may prevent more injuries from occurring.

In order to calculate injury rates by activity, both injury cases, known to occur while performing specific activities, and exposure time of the same activities need to be collected. In reality, few national surveys have the capacity to collect both numerator and denominator statistics. For example, the National Health Interview Survey (NHIS) asks respondents to report the occurrence of injuries and also inquires about activities performed at the time of injury (working, driving, etc.). However, the survey does not record the exposure time corresponding to these activities except for usual weekly work hours. As such, with data from the NHIS alone it is not possible to estimate injury rates in terms of exposure hours.

In the absence of information about exposure hours in the estimation of risk, researchers typically report the frequency of injuries by activity grouping and often report the ‘injury rate per population’, that is, the number of injuries that occur in a population at risk in a given time period, to describe injury risk at the population level.1 3–6 As a basic and important epidemiological measure, this expression of injury risk is widely used. In the case of the NHIS since no exposure time in each activity is gathered, the population (estimated directly from the NHIS)4 is used as the denominator, and the rate per population by activity (or by cause) basically reflects the relative distribution of injury cases by activities (or by cause).1 3–6

However, the injury rate per population calculated this way does not take into account exposure hours in the calculation and may conceal the true picture of injury risk per unit time exposed.7 8 Moreover, because the population at risk is often treated as a constant,3–4 this expression of risk implicitly assumes that the exposure for the entire population is identical for all different types of activities, which may overstate the exposure time of activities where people typically spend less time (eg, sports and exercising) relative to the activities that people spend more time (eg, working). Thus, rate per population should not be used to compare injury risk between different activities or demographic groups where time exposed to specific activities is known to vary.

A further refinement of the expression of injury risk is warranted given the availability of adequate denominator statistics. To estimate injury rates by activity that take into account exposure time, we combine statistics from two national complex surveys for the numerator and denominator as inputs for the numerator and denominator of rate calculation. We adopt a method to estimate the variance of the injury rates that takes into account the complex survey designs. Using different national surveys for the numerator and denominator of rate...
calculation is an approach first employed to assess the association between smoking and lung cancer,9 and has been applied to other topics, particularly cancer mortality.10–13

Some studies have examined the injury rate of certain activities in the USA such as transportation fatalities16 or fatal workplace injuries17 similarly using different data sources for the numerator and denominator. However, these studies are concerned about specific injuries or specific subpopulations; no previous research has attempted to examine a wide spectrum of activities in which people are engaged on a daily basis. We add to the literature by estimating injury rates of all possible activities (known as person-time incidence rate or incidence density rate) for which no national estimates have been produced. We also demonstrate the ability to approximate the variance of the rates that integrates the complex survey designs of both numerator and denominator data.

**METHODS**

**Data**

We used the 2010 US adult population to illustrate the procedures. The numerator data, or injury outcomes, came from the 2010 NHIS18; the denominator data, or exposure time data, came from the 2010 American Time Use Survey (ATUS).19 The NHIS is a cross-sectional household interview survey that has been monitoring the US population’s health since 1957, where the target universe of the NHIS is the civilian noninstitutionalised population.18 The NHIS uses multistage sampling that involves stratification, clustering and oversampling of specific population subgroups. The survey is administered by the US Census Bureau under a contractual agreement with the NCHS. We used NHIS data for 2010, extracted from the Integrated Health Interview Series database,20 and restricted analyses to those aged 18+. There were 27 157 adult observations in our analysis, and the response rate in 2010 was 79.5%.18

For the main portion of the interview, ‘Family Core’, all members of the household 18 years of age and over who are at home at the time of the interview are invited to participate and to respond for themselves.18 For adults not available for interview, information is provided by another adult family member in the household. The Family Core includes an injury and poisoning section, in which the respondents report each injury or poisoning episode that was severe enough to seek medical treatment 3 months prior to the interview. Additional supporting information on the nature and circumstances of the injury is also gathered. We excluded the poisoning episodes, and only used injury episodes that occurred within a 6-week recall period (as opposed to 3 months) because studies4,21 showed that recall bias increases with the time elapsed between injury and interview, especially after 6 weeks.

The ATUS measures the amount of time people spend performing various activities; the survey started in 2003 and has been collected annually. The target universe is composed of the civilian, non-institutionalised population that are at least 15 years of age residing in occupied households in the USA.19 The ATUS is a stratified, three-stage sample where an eligible person is randomly selected from the household to conduct the interview and to participate in a 24-h time-use diary that documents the time spent in specific activities that are then coded using a standardised lexicon.22 The ATUS is conducted by the US Census Bureau for the US Bureau of Labor Statistics. We used ATUS data for 2010, extracted from the ATUS-X database,23 and restricted analyses for persons 18+. There were 16 679 adult observations in our analysis and the response rate of the ATUS in 2010 was 56.9%.19

The NHIS inquires about activities that were being performed at the time of injury after the respondent describes the circumstances leading to the incident; there are 11 categorical options.24 We estimated the rate of injury for the following five categories (as worded in the NHIS questionnaire):24 (1) working at a paid job, (2) sports and exercise, (3) sleep, resting, eating, drinking, (4) working around the house/yard and (5) driving/riding in a motor vehicle. These activities were selected because the ATUS has recorded time use in a 24-h period in much a more detailed manner, which could be appropriately collapsed to allow matching of hours exposed corresponding to each of those five categories in the NHIS. The NHIS activities and corresponding collapsing of ATUS activities are described in table 1.

**Calculation of injury rate per hour exposed and variance using statistics from different surveys for the numerator and denominator**

We let $Y$ be the number of injury episodes of US adults in 2010 and related to a specific activity, and let $X$ be the population’s total hours of exposure to the specific activity in 2010 that is also related to the same activity as the injury episodes. The rate per hour exposed is simply the ratio of $Y$ to $X$, which we estimated using the ratio of the corresponding sample means $\bar{Y}$ and $\bar{X}$,25 assuming both surveys generalise to the same population:

$$R = \frac{Y}{X}$$

(1)

The sample means $\bar{Y}$ and $\bar{X}$ were calculated individually, incorporating the survey-appropriate sampling weights. Both $\bar{Y}$ and $\bar{X}$ were annualised, because in the calculation of a rate the numerator and denominator need to have the same calendar period time, which in our study was the year of 2010. Since we have restricted the recall period of injury episodes to 6 weeks.
(42 days), the annualised estimate $\bar{Y}$ was calculated by multiplying the weighted 6-week averages by 365/42, with the assumption that the injury pattern over a 6-week period represents that of a year. The time-diary data of ATUS were collected for a single day and weekend days were over-represented (50% of the sample data). The weighting variables were constructed to reduce the impact of over-representation of weekend days so that the weighted estimates represent a ‘typical day’ in a calendar quarter. The annualised estimates were obtained by multiplying the weighted daily averages by 365, assuming that the ‘typical day’ in the calendar quarter (after weighting) represents the time-use pattern throughout the year.

The variance of $R$, $\text{VAR}(R)$, was approximated using the ‘delta method’, which is essentially a first-order Taylor series expansion. The delta method has been used extensively in statistics and can be applied to different sampling designs. (StataCorp, LP. Stata Survey Data Reference Manual, Release 13. College Station, Texas: Stata Press Publication 2013.) This method is the default approach for variance estimation with complex surveys as their data. Chapter 9 of Sampling: Design and Analysis (Lohr) describes the method in detail. The delta method shows that the variance can be approximated by a combination of $\text{VAR}(\bar{Y}), \text{VAR}(\bar{X}), \bar{Y}$ and $\bar{X}$ (online supplementary material provides the details of the derivation):

$$
\text{VAR}(\bar{Y}/\bar{X}) \approx \frac{\text{VAR}(\bar{Y})}{\bar{X}^2} + \frac{\bar{Y}^2 \cdot \text{VAR}(\bar{X})}{\bar{X}^4}
$$

(2)

If the data source for the numerator or denominator is a complex survey, the calculation of $\text{VAR}(\bar{Y}), \text{VAR}(\bar{X}), \bar{Y}$ and $\bar{X}$ needs to include the survey designs so that the design information is appropriately reflected in $\text{VAR}(\bar{Y}/\bar{X})$. Operationally, we did this by using the Stata command ‘svy: mean’ (Stata Corporation LP 2013) to calculate $\bar{X}, \bar{Y}, \text{VAR}(\bar{X})$ and $\text{VAR}(\bar{Y})$ individually, and substituted them into equation (2) to obtain $\text{VAR}(\bar{Y}/\bar{X})$. The SE($\bar{Y}/\bar{X}$) was obtained by taking the square root of $\text{VAR}(\bar{Y}/\bar{X})$. Stata V13 was used for all analyses.

Comparing the results with rate-per-population estimates and those using the NHIS data alone

To illustrate the difference between different expressions of injury risk, we estimated injury rate per population and contrasted the results with rate per hour exposed. Similar to the method used to calculate rate per hour exposed, the numerator was the number of injury episodes. Converse to the rate per hour approach was that the adult population (estimated directly from the NHIS) was used as the denominator. The NCHS has constantly provided annualised injury risk estimates, which we followed this convention.

RESULTS

There were 12 679 adult observations in the 2010 ATUS, and 27 157 in the 2010 NHIS. The demographic profile (weighted) of the two samples was similar. The mean age was 46.2% and 46.3%, and the proportion of females versus males was 51.7% and 51.6% for the NHIS and ATUS, respectively. The left panel of table 2 summarises the average numbers of injuries an adult would sustain over a year (ie, the $\bar{Y}$ in equation 1) as well as the SEs (ie, the square root of $\text{VAR}(\bar{Y})$ in equation 2). The right panel summarises the average activity exposure time over a year (ie, the $\bar{X}$ in equation 1). The calculation of $\bar{Y}, \text{SE}(\bar{Y}), \bar{X}$ and $\text{SE}(\bar{X})$ took into account appropriate survey designs.

Table 3 reports the estimated injury rates per hour of activity exposed. To compare the injury risk associated with different activities, the rates were normalised to ‘number of injury episodes per 100 000 h’ for each type of activity.

US adults sustained 1.45 injuries per 100 000 work hours (table 3). The comparable figure was 12.64, 0.23, 6.14 and 2.98 injuries per 100 000 h for the other four categories, respectively.

Table 4 describes injury rates per 1000 persons (population) for the same five activity types. The injury rate-per-hour approach depicts a ranking of injury risk quite different from the rate-per-population approach. For example, the injury rate per hour exposed in a sport activity was 4.24 times as high as the injury rate per hour exposed to driving (table 3). By contrast, the rate per population in a sport activity was only 1.08 times as high as rate per population exposed to driving (table 4).

DISCUSSION

We combined statistics from two different complex national surveys for the numerator and denominator to estimate injury rates per hour exposed by activity, and approximated the variance using the delta method, a commonly used method for variance estimation with complex surveys. Researchers in other fields have combined statistics from two surveys to generate ratio estimates, although this approach has not been accomplished to estimate injury rates by activity groupings. By combining these two complex surveys and incorporating the time spent in specific activities, we identified a different rank order of injury risk compared with using a per person risk metric.

We compared our estimates for work-related injuries with those using the 2010 NHIS data alone; the results were close. The NHIS asks ‘usual work hours per week’ (but not time use...
Methodology

Table 3  Injury rate per 100 000 h (number of injuries per 100 000 exposure hours) of US adults by activities performed at the time of injury, using 2010 NHIS (numerator) and ATUS (denominator)

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>Work</th>
<th>Sport and exercise</th>
<th>Sleeping, resting, eating, drinking</th>
<th>Working around house/yard</th>
<th>Driving/riding a motor vehicle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate</td>
<td>1.45</td>
<td>12.64</td>
<td>0.23</td>
<td>6.14</td>
<td>2.98</td>
<td></td>
</tr>
<tr>
<td>SE</td>
<td>0.22</td>
<td>2.23</td>
<td>0.06</td>
<td>0.75</td>
<td>0.49</td>
<td></td>
</tr>
<tr>
<td>95% CI</td>
<td>1.01 to 1.89</td>
<td>8.28 to 17.01</td>
<td>0.13 to 0.34</td>
<td>4.66 to 7.61</td>
<td>2.03 to 3.94</td>
<td></td>
</tr>
</tbody>
</table>

ATUS, American Time Use Survey; NHIS, National Health Interview Survey.

Table 4  Annualised injury rate per 1000 population (number of injury episodes per 1000 persons) of US adults by activities performed at the time of injury, using 2010 NHIS for the numerator and denominator

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>Work</th>
<th>Sport and exercise</th>
<th>Sleeping, resting, eating, drinking</th>
<th>Working around house/yard</th>
<th>Driving, riding a motor vehicle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate</td>
<td>27.68</td>
<td>12.86</td>
<td>8.36</td>
<td>28.17</td>
<td>11.96</td>
<td></td>
</tr>
<tr>
<td>SE</td>
<td>4.28</td>
<td>2.21</td>
<td>1.97</td>
<td>3.42</td>
<td>1.95</td>
<td></td>
</tr>
<tr>
<td>95% CI</td>
<td>19.29 to 36.06</td>
<td>8.52 to 17.19</td>
<td>4.50 to 12.22</td>
<td>21.46 to 34.87</td>
<td>8.14 to 15.78</td>
<td></td>
</tr>
</tbody>
</table>

NHIS, National Health Interview Survey.
injury risk and may be used to compare across different types of injuries or demographic groups. This procedure also overcomes data availability constraints frequently encountered by injury researchers, making it possible to estimate injury risk and CIs even when no single survey has collected information on both injury cases and activity exposure time.

Key messages

What is already known on the subject

- Few surveys collect both injury cases and exposure time, making it difficult to generate exposure time-based estimates for injury risk of common day-to-day activities.
- In the absence of exposure time data, researchers often report the number of injuries that occur in a population at risk in a given time period, or ‘injury rate per population.’
- The rate-per-population approach reflects the magnitude of injury burden of a population, but may conceal the true picture of injury risk per unit time exposed.

What is this study adds

- To estimate injury rates by activity, the denominator is the number of people engaged in an activity at any given time, and the numerator is the number of injuries that occur in that activity.
- To calculate the variance of the rates, we use the ‘delta method.’
- Our results demonstrate a different ranking of injury rate by activity as compared to ranked estimates using population as the denominator.
- Our approach enables one to estimate injury risk and variance by combining injury cases and activity exposure time from separate surveys.

Acknowledgements The authors appreciate the thoughtful comments of Dr Elyssa Besen and Dr David Lombardi.

Contributors HRM-W developed the concept of the study; TL implemented the study design with contributions from HRM-W, JLV and SKV. TL performed the analysis and drafted the first version of the paper. JLV, SKV, HRM-W and TL contributed to the interpretation of the data. All five authors have performed a critical review of the manuscript and have approved this final version for publication.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

REFERENCES

33 Frazis H, Stewart J. What can we do to estimate data we will tell us about the cost of living? Monthly Lab Rev 2004;127:3.
Combining statistics from two national complex surveys to estimate injury rates per hour exposed and variance by activity in the USA

Tin-chi Lin, Helen R Marucci-Wellman, Joanna L Willetts, Melanye J Brennan and Santosh K Verma

Inj Prev  published online April 4, 2016

Updated information and services can be found at:
http://injuryprevention.bmj.com/content/early/2016/04/04/injuryprev-2015-041885

These include:

References
This article cites 20 articles, 11 of which you can access for free at:
http://injuryprevention.bmj.com/content/early/2016/04/04/injuryprev-2015-041885#BIBL

Open Access
This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

Email alerting service
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Topic Collections
Articles on similar topics can be found in the following collections
Open access (61)

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/