

The Dead Go Live! Injury Mortality Data via a Web-based Application

Steve James, Centers for Disease Control and Prevention, Atlanta, GA

ABSTRACT

How many teenagers died from gunshot wounds last year? What's the leading cause of death for people in my age group? While death and dying are not normally favorite topics at the dinner table, there are many who take great interest in the answers to questions like these. Public health scientists, researchers and policy makers use this information to determine intervention strategies to reduce injury in the United States. Until recently, most requests to the CDC for data such as this were fulfilled through static web pages and from ad-hoc requests. In March, 2000, the Centers for Disease Control and Prevention unveiled a web-based application to provide these data via the internet. This paper will describe the application and some of the design considerations used when creating the application.

INTRODUCTION

For years, the National Center for Injury Prevention and Control (NCIPC) at CDC would get numerous requests from public health officials, researchers and concerned citizens for data related to injuries. Up until a few years ago, all requests were handled with ad-hoc queries and a few standardized reports. In 1994 some static web pages replicating the hardcopy reports were available through the web, but ad-hoc requests continued.

In an effort to reduce the number of ad-hoc requests while providing timely and informative data on injury mortality, NCIPC developed a web application to provide this information via the internet. Named WISQARS™ for Web-based Injury Statistics Query and Reporting System, the system was released to the public on March 1, 2000. Within the first three months, over 50,000 requests have been processed. As a result, more people have been able to get more information on injury mortality than ever before, while at the same time nearly eliminating the need to service ad-hoc requests. By virtually every measure, WISQARS™ has been a roaring success.

To see the application, visit <http://www.cdc.gov/ncipc/wisqars>.

DESIGN CHARACTERISTICS

There were several factors that made significant impact on the design of the system. The first was that the data that we would be presenting were static. The mortality data that's used in the Injury Center comes from the National Center for Health Statistics (NCHS). They collect, clean and organize the data that they collect from state death certificate data. And once the data are released, they are rarely changed. That meant the structure that was set up for the data would be stable and could be optimized for retrieval. Additional data would be added only annually, and the existing data need not change.

Secondly, the CDC corporately has a policy that web applications be available to the largest number of people, regardless of the technology that they have. So if we provided access to people using a Java applet for example, we also had to provide similar access to people whose browser's didn't support Java. We also had to be conscious of people with disabilities who were using browser-readers to visit our pages. Thus, we were unable to use products like Web/AF to develop a user interface and were restricted to simple HTML.

And finally we had to do it as cheaply as possible. While we had some management support in the development of our project, we didn't have associated funds to purchase a lot of development tools. So while setting up a data warehouse with MDDb and associated software may have been the better long-range solution, we had to use what we had and hope that once we had shown some success with the mortality data that funding would become available to expand the scope of the data presented.

THE WISQARS™ SYSTEM

After several months of haggling over exactly what the system would and wouldn't do, a design was finalized and development of the system began. We were fortunate in that we were able to borrow space on a UNIX server that had SAS/IntrNet™ so we did not have to outlay any cash for development tools.

The design we came up with involved simple HTML pages that the user would enter query parameters, and SAS® on the back end using SAS/IntrNet™. We developed what was essentially our own MDDb for the data, optimizing it for retrieval and minimizing the processing that would have to be done to fulfill the request.

WISQARS™ provides mortality data in three report formats:

- 1) Injury Mortality Reports, which can be used to determine injury deaths and death rates for specific external causes of injuries,
- 2) Leading Causes of Death Reports, which can be used to determine the number of injury-related deaths relative to the number of other leading causes of death in the United States or in individual states, and
- 3) Years of Potential Life Lost (YPLL) which is used to measure premature mortality.

All reports (with the exception of YPLL reports which do not group anything by age) are available by year, age, race, sex, Hispanic origin, region and state. Reports can be requested by standard 5-year age groupings (e.g., 0-4 years or 5-9 years) or a custom-defined range (e.g., single year-of-age, 13-19 years or 6-60 years). Reports can be generated to look at age-adjusted death rates across years. Race categories are White, Black, American Indian/Alaskan Native, Asian and Pacific Islander, and other (all non-white and non-black and may include other races not listed). Injury Mortality Reports can be requested by mechanism/cause and manner of death/intent of injury in a specific state by sex, age and race). For the Leading Cause of Death Reports, there is an option to request the frequencies of deaths by specific mechanisms/causes of injury for unintentional injuries, homicides, and suicides within specific age groups or overall ages.

INJURY MORTALITY REPORTS

These reports provide mortality statistics for a specific cause of injury. The cause of injury is specified by mechanism/cause and manner/intent. An example would be Unintentional Motor Vehicle Traffic deaths. The mechanism/cause is Motor Vehicle Traffic and the manner/intent would be Unintentional. The user has the ability to choose up to four different "by" variables that will vary the output appropriately. Additionally there is a "text only" output option that reduces the amount of HTML that the output has so that the user can more easily transfer the data to other software packages.

The output is context sensitive in that it will include or not include footnotes that are appropriate to the type of output. For example, if the report includes a value of Hispanic Origin, a footnote will appear describing some of the limitations of the reporting of that value.

Death rates based on fewer than 20 deaths are considered unstable, and traffic-lighting is used to identify those numbers which based on too-few deaths..

LEADING CAUSES OF DEATH REPORTS

The Leading Causes of Death Reports (LCD) reports information on both injury and non-injury related deaths. It allows a user to see the impact of injury-related deaths on total mortality. It is displayed in the form of a chart with columns for different age groups and Totals. The user can select the same demographic information that he can using the Injury Mortality Reports, but does not select a specific injury category.

The user can vary the output in one of two ways. The first is that the user can change the number of causes reported for each age group. It's possible that a particular cause of death may not fall in the top 10 rankings for an age group. A user can select additional rankings to see where the ranking of that particular injury falls. For instance, Homicide is not normally in the top ten causes of death overall. Using a higher ranking than 10 will likely show it. Secondly the user can determine which age groups are used to display the information: either one of three pre-defined formats or a user-defined format.

Additionally the LCD reports allow a user to find out the specific injuries that make up each of the displayed injury categories. Each injury category is a link that executes another program that produces a bar-graph showing the relative contribution of each type of injury along with the number of deaths and the percentage.

Here a technical trick was used that allowed the bar chart be produced more easily. Rather than using SAS®/Graph with Annotate to produce the chart, simple HTML was used. A one pixel .GIF file was created and it's display dimensions were manipulated using the HEIGHT= and WIDTH= parameters of IMG tag. By reading the output of a PROC FREQ with a DATA STEP, one could easily write out the appropriate code to produce the bar charts in an HTML table. For more information on this technique, see paper #406, *Web-Application Bar Charts without SAS/Graph*.

YEARS OF POTENTIAL LIFE LOST (YPLL)

Years of Potential Life Lost (YPLL) reports are a way to measure premature mortality. Like the LCD reports, they include information on Injury as well as non-Injury mortality. YPLL is calculated by subtracting the age of death of the decedent from a standard age, such as 75, and then summing these for each cause of death.

Suppose, in a simplistic example, if we were going to compare the Year of Potential Life Lost before age 75, and we had 400 deaths of children age 2 years old of Unintentional Motor Vehicle, and we had 1000 adults age 73 die of Heart Disease. The YPLL of Unintentional Motor Vehicle deaths would be $(75-2)*400 = 29,200$, and for Heart Disease the YPLL would be $(75-73)*1000 = 2,000$. So while many more people died of Heart Disease than from Motor Vehicle injuries, the YPLL of Motor Vehicles is much higher.

The user can select the same demographic information as the Injury Mortality Reports, except no age selection is allowed that would restrict the data to certain age groups. It is possible to change the standard age that's used to calculate the YPLL though. The user can choose to get a list of the top categories, which would include the injury categories of Unintentional, Homicide and Legal Intervention, and Suicide, or the user can select a specific cause of injury and determine the YPLL for that category alone.

LESSONS LEARNED

In building a system such as this, one of the key factors is having a good design. Repairing a flaw in the design phase is significantly cheaper and easier than in later phases. Putting the effort into the design helps ensure a good system. This goes against what many would prefer, since they are interested in starting the coding as soon as possible. It's best to put that off as long as possible until the design is finalized and any changes can be made more cheaply.

While we were able to allow people inside of CDC to use the system prior to its release to the entire Internet, we were concerned about system load. We really didn't have any idea of how many people

would come to visit the site; would it be 1, 10, 50, 5000 people per day. One of the problems with load testing like this is that you don't know how many will hit the site. It's like inviting people to a party but having no idea of how many will come. And it's not like you can call up 1,000 of your friends and ask them to visit your site on a particular day.

Having management support is another critical factor in the success or failure of the system. In part, you have to decide what the cost of failure is. If someone comes to your site and has a bad experience, what is the impact? If you're a business, then you potentially lose customers and sales. If you're a government agency, you weaken credibility. One problem with the web is that you don't necessarily know who your users are. They could be having a bad experience and you'd never know it. Management support and involvement is important here because of the impact that the web application can have on your business without notice.

One of the things that we designed into the system was a logging function. Essentially what we did was store all of the query parameters that a user entered into a SAS® data set for later analysis. This has been very useful in determining the number of requests, types of requests, etc. without costly web statistics software. We can determine if users are having trouble by seeing how often they put invalid query parameters in.

CONCLUSION

Building a web application has had dramatic effect at NCIPC. Not only are more people getting information on injury than ever before, there's less effort used in getting them that information. Virtually all the ad-hoc requests have been eliminated. With the lack of ad-hoc requests the work can be more easily managed and be more proactive rather than reactive. This makes for a smoother work environment and allows newer technology to be incorporated easier, which is more rewarding to the IT staff.

REFERENCES

SAS, Web/AF and SAS/IntrNet are registered trademarks or trademarks of SAS Institute Inc., Cary, NC

WISQARS is a registered trademark of the National Center for Injury Prevention and Control of Centers for Disease Control and Prevention

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Steve James
Centers for Disease Control and Prevention
Atlanta, GA 30341
Work Phone: 770-488-4269 Fax: 770-488-1665
Email: sjames@cdc.gov