

Paper 194-2010**Creating Dynamic Series Charts Using Base SAS®, SAS/GRAPH®, and the Annotate Facility**

Jun Ju, Sue Nowlin, John Sestito
National Institute for Occupational Safety and Health, Cincinnati, OH

ABSTRACT

The National Institute for Occupational Safety and Health (NIOSH) provides occupational health surveillance information through the systematic collection, analysis, and dissemination of this information via publications and the Internet. NIOSH is currently developing a content management system to automate the production of charts and tables for the NIOSH Worker Health Chartbook web site [<http://www.cdc.gov/niosh/docs/2004-146/>] using Base SAS®, SAS/GRAPH®, and MS SQL Server. This web site contains over 10,000 charts, updated annually, with approximately 5,000 charts per series.

Series charts are sets of charts with each graph illustrating a group of data from an input data set. They are beneficial to graphically explore data in detail. While SAS® has provided By-Group Processing along with graphic procedures to make series charts, it is labor-intensive and time-consuming to produce complex and dynamic series charts using By-Group. This paper demonstrates how to construct dynamic series charts, using SAS DATA steps to manipulate data and implement SAS graphical procedures and SAS Annotate facilities to make the graphs dynamic.

BACKGROUND

The Worker Health Chartbook of the National Institute of Occupational Safety and Health (NIOSH) is a descriptive epidemiologic reference on occupational morbidity and mortality in the United States. It presents data in an easily accessible format: more charts and graphs than text or analysis. The goal of the Chartbook web site is to be a comprehensive resource for agencies, organizations, employers, researchers, workers, and others who want to know about occupational injuries and illnesses. It accommodates charts reflecting up to 1300 industry categories, defined by North American Industry Classification System (NAICS <http://www.census.gov/eos/www/naics>), using available data in the "Injuries, Illnesses and Fatalities (IIF)" program of Bureau of Labor Statistics (BLS).

The first NIOSH Worker Health Chartbook (2004) published on the internet was developed using static HTML. The individual charts (over 400) were created through a manual process of downloading BLS data, sub-setting into excel files, producing initial charts using excel, then exporting the charts for refinement using CoreIDRAW® software. This process took over 2 years to complete and did not lend itself for easy updating of the charts when new data was available from BLS. To solve this labor intensive process and to produce a mechanism to easily manage the web site content, NIOSH developed a content management system using SAS and Microsoft SQL/server database. This system allows for the creation or modification of large numbers of series charts on an as needed basis using a much reduced amount of processing time.

INTRODUCTION

Series charts can be produced with different formats, such as histograms, plots, maps. In the NIOSH Worker Health Chartbook project, we graphed the data mainly with two types of formats: vertical bar charts and bar line charts. Charts illustrated below represent those two types.

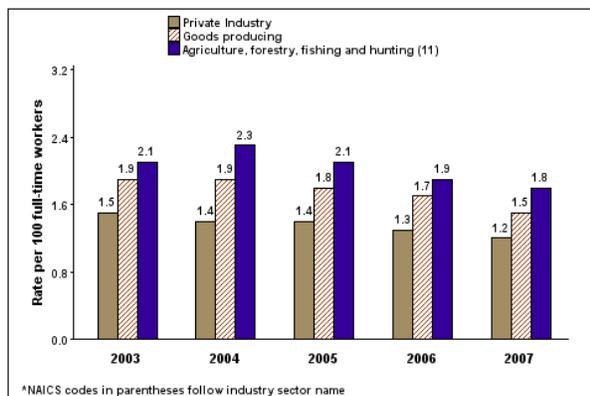


Figure 1. Bar chart

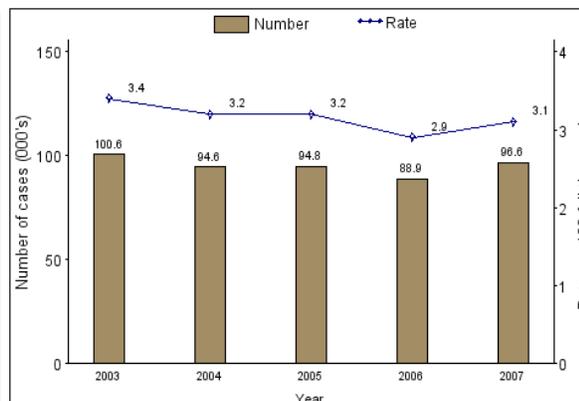


Figure 2. Bar-line Chart

The bar chart presented in Figure 1 is a sample of Bar series charts, it describes the *Days Away From Work* incidence rates of three different industrial categories (*Private Industry, Goods Producing Industries, Agriculture, Forestry, Fishing and Hunting Industry*) for year 2003-2007. Figure 2 Bar-line chart is one of the Bar-line series charts. It shows the *Total Injury and Illness* numbers and rates for industry category *Other Services* (NAICS code 81) over the years of 2003-2007. To generate these two types of series charts, the same strategy in terms of the mechanism of data manipulation and chart production was utilized, with differences in the programming details in terms of different data groups and chart types. This paper will discuss creating series bar line charts as an example to demonstrate an approach of how to create series charts.

DATA PREPARATION

The Injuries, Illnesses, and Fatalities (IIF) program “provides data on illnesses and injuries on the job and data on worker fatalities.” It contains three databases: “Census of Fatal Occupational Injuries”, “Nonfatal cases involving days away from work: selected characteristics” and “Occupational Injuries and Illnesses: Industry Data”. The database source used for demonstration in this paper is “Occupational Injuries and Illnesses: Industry Data”. (Please visit website <http://www.bls.gov/data/> for database details.)

In our approach of producing series charts, data preparation is the most essential part: data for each chart in the series needs to be assembled specially in accordance with data grouping criterion. By-group process would be an ideal procedure to apply if data is arranged in a straightforward format and there are no specific requirements for each chart. However our database contains several data tables and we need to gather information from different tables to group data properly. Instead of twisting By-group process, we utilized the following method, illustrated using SAS, to prepare the data for each chart in the series.

```

/* Allocate a temporary file */
filename temp temp;

data _null_ ; set total;
  file temp;
  put 'data rate; set all;' ;
  str= 'where naics_code in (%trim("' || trim(naics_code) || '"));';
  str2=' valr=value;' ;
  put str / str2 / ;
  put 'run;' ;
  put 'data num; set all;';
  stn= 'where naics_code in (%trim("' || trim(naics_code) || '"));';
  stn2=' valn=value;' ;
  put stn / stn2 / ;
  put 'run;' ;
  st2='%barlinecht (' || compress(naics_code) || ');' ;
  put st2;
run;

%include temp;
run;
```

```
filename temp clear;
```

As it shows above, there are two data sets involved in this SAS program, *Total* and *All*. Data set *Total* holds all the NAICS codes, and *All* contains the statistical numbers, referred as counts and incidence rates in our database, for each NAICS industry category. What the SAS code does here is looping through each *naics_code* variable value in file *Total*, preparing a temporary file to hold SAS statements for every data-grouping and charting based on each *naics_code* value. Below presents the content of the temporary file if *Naics_code* variable in *Total* data set has two values: "000000" and "GP1AAA".

```
data rate; set all;
  where naics_code in ('000000');
  valr=value;
run;
data num; set all;
  where naics_code in ('000000');
  valn=value;
run;
```

```
%barlinecht(000000);
```

```
data rate; set all;
  where naics_code in ('GP1AAA');
  valr=value;
run;
data num; set all;
  where naics_code in ('GP1AAA');
  valn=value;
run;
```

```
%barlinecht(GP1AAA);
```

DYNAMIC CHARTS CREATION

In general, dynamic charts creation can be addressed in many aspects, like graph sizes, colors, number scales, data notations, etc. As far as the series charts that we created for the Chartbook project, the Dynamic charts creation involves mainly two parts: customized axis tick values and SAS Annotate facility usage. Following are two charts from our bar line series.

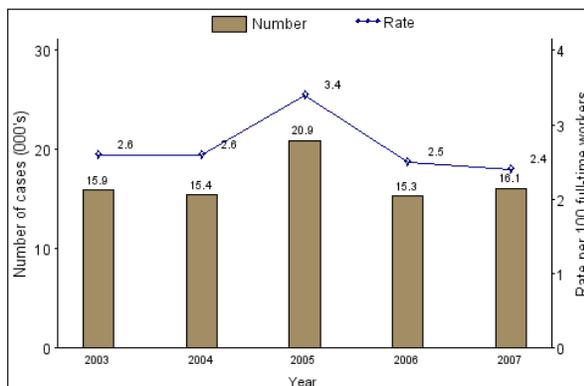


Figure 3. Bar-line chart

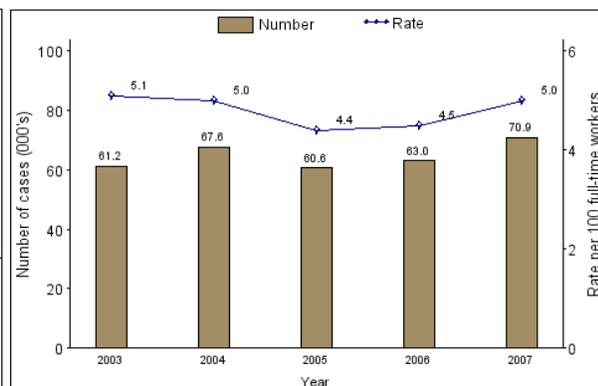


Figure 4. Bar-line chart

The charts demonstrate that data ranges and scale values for response axes are customized with "nice" tick numbers. For instance, Figure 3 bar-line chart has 0-30 range for left response axis with 10 as the scale value even though the actual data ranges from 15.3-20.9; the left response axis in Figure 4 chart has 0-100 value range with scale value 20, and the actual values are between 60.6-70.9. By default, SAS graph procedures pick a value range based on the actual data for the response axes, and scale those axes with *n* equal intervals. That could cause the tick marks on the axes to have "ugly" labels with complicated decimal numbers. In order to simplify and "beautify" the response axes, we adopted the algorithm, from the book *Graphics Gems*, authored by Andrew S. Glassner, as illustrated in SAS Language below:

```
* Find a "nice" number approximately equal to x;
```

```

* Round the number if round=1, else take ceiling if round=0;
%macro nicenum(x, round);
%global nicenum_val;
data _null_;
  x = symget('x');
  round = symget('round');
  exponent = floor(log10(x));
  f = x / (10**exponent);

  if round then do;
    if f < 1.5 then nf = 1;
    else if f < 3 then nf = 2;
    else if f < 7 then nf = 5;
    else nf = 10;
  end;
  else do;
    if f <= 1 then nf = 1;
    else if f <= 2 then nf = 2;
    else if f <= 5 then nf = 5;
    else nf = 10;
  end;
  nice = nf * (10**exponent);
  call symput('nicenum_val', nice);
run;
%mend;

%global graph_min;
%global graph_max;
%global nfrac;
%global tick_size;

%macro graphminmax(ntick=5, min=0, max=);
data _null_;
  min = symget('min');
  max = symget('max');
  diff = max - min;
  call symput('diff', diff);
run;
%nicenum(&diff, 0);
%let range = &nicenum_val;
data _null_;
  tick_size = symget('range') / (symget('ntick') - 1);
  call symput('tick_size', tick_size);
run;
%nicenum(&tick_size, 1);
%let tick_size = &nicenum_val;
data _null_;
  min = symget('min');
  max = symget('max');
  tick_size = symget('tick_size');
  graph_min = tick_size * floor(min/tick_size);
  graph_max = tick_size * ceil(max/tick_size);
  nfrac = max(-floor(log10(tick_size)), 0);
  call symput('graph_min', graph_min);
  call symput('graph_max', graph_max);
  call symput('nfrac', nfrac);
run;
%mend;

```

While it is certainly not the only algorithm to build dynamic axes in the graph, we chose this algorithm because it can form axes scales with nice and simple decimal numbers.

As SAS Annotate facility extends the ability of graphic procedures to make customized charts, for bar line charts, we utilized the Annotate facility to draw the legend, bars and lines, and label numbers incorporated with the GPLOT procedure. As to bar-line charts creation, SAS has provided the GBARLINE procedure, starting in SAS 9.1, and the statistical procedure SGLOT (new in SAS 9.2) to create bar-line charts. However, those two procedures have not been developed to integrate with SAS annotate facility well, hence have limitations. Below contains a portion of SAS code that was implemented to create dynamic bar line charts in the Chartbook project.

```

/* draw legend */
data anno_1;

```

```

        length function color $ 8 style $ 12 text $ 30;
        retain xsys ysys hsys '3';
* use move function;
        function='move'; x=41; y=94; output;
* use bar function to draw the bar with designated color;
        function='bar'; x=35; y=98; color="#A28D64"; line=1; style='solid';
            style='solid'; output;
* following code lines are to draw the bar outlines;
        function='move'; x=41; y=94; output;
        function='bar'; x=35; y=98; color=black; line=1; style='empty'; output;
        function='move'; x=35; y=94; output;
        function='draw'; color=black; line=1; size=0.2; x=41; y=94; output;
        function='move'; x=41; y=94; output;
* label the bar "Number";
function='label'; style="arial" ; position='6'; xsys='3';
        x=42; ysys='3'; y=97;
        text='Number'; size=3.8;
        color='black'; output;

run;

/* draw vertical bars */
data plotlabel11; set chart;
        length color $8;
        retain xsys ysys '2';
        %bar(years-&barwidth, &mintickl, years+&barwidth, valn, &barcolor, 0, solid);
run;

/* draw bar outlines */
data plotlabel12; set chart;
        retain xsys ysys '2';
        function='move'; x=years-&barwidth; y=&mintickl; output;
        function='draw'; color='black'; line=1; size=1;
            x=years-&barwidth; y=valn; output;
        function='move'; x=years-&barwidth; y=valn; output; function='draw';
            color='black'; line=1; size=1; x=years+&barwidth;
            y=valn; output;
        function='move'; x=years+&barwidth; y=valn; output;
        function='draw'; color='black'; line=1; size=1;
            x=years+&barwidth; y=&mintickl; output;
run;

/* Label counts */
data plotlabel13; set chart;
        length color $ 8;
        retain function 'label' color 'black' when 'a'
            xsys ysys '2' position '2' size 2.8 hsys '3';
        text=left(put(valn, 12.1));
        if valn=. then text='(-)';
        x=years;
        if valn=. then valn=0;
        y=valn+(valn*0.005);
run;

/* label rates */
data plotlabel2; set chart;
        length color $ 8;
        retain function 'label' color 'black' when 'a'
            xsys ysys '2' position '2' size 2.8 hsys '3';
        text=left(put(valr, 12.1));
        if valr=. then text='(-)';
        x=years+(&barwidth+0.125);
        if valr=. then valr=0;
        y=valr+(valr*0.005);
run;

.....

/* Incooperate the annotated data set with GPLOT procedure */
proc gplot data=chart anno=anno_1;
        plot1 valn*years/overlay noframe anno=plotlabel1 vaxis=axis1 haxis=axis3;
        plot2 valr*years/noframe vaxis=axis2 anno=plotlabel2 legend=legend2;

```

```
run;
;
```

CONTENT MANAGEMENT SYSTEM

Besides chart production, the content management system for the Chartbook project also comprises recording the metadata of each chart/table, namely: chart/table name, description, filename path/location, chart title, and data year range into a SQL database. A Microsoft Access application is also built as a user interface to manage the database (Figure 7 displays the main form of the application). Figure 5 below illustrates the major components of the content management system:

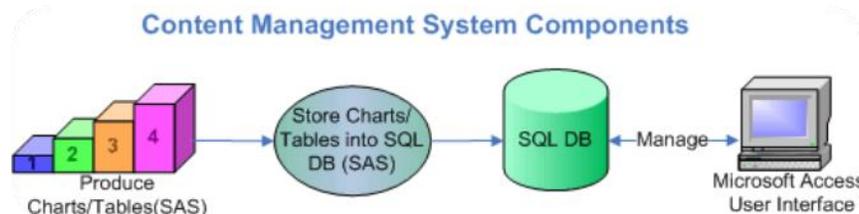


Figure 5. Content Management System Components

In this system, SAS fulfilled the tasks of producing charts/tables and storing the metadata of every chart/table into the database. The SAS program of loading data into the SQL database is demonstrated below:

* Note: The IGNORE_READ_ONLY_COLUMNS option is used because the destination table has an auto-numbered primary key;

```
libname in 'current file location';
libname out odbc dsn=SQL_Server_DSN ignore_read_only_columns=yes;

proc append base=out.DestinationTable
            data=in.TableToBeLoad force;
run;
quit;
```

CONCLUSION

Series charts can be as brief as one chart if the whole data set is considered as one group. They also can be presented as many as the number of records in the table if every record is distinguished out by certain criteria. Creating Series charts is an efficient approach to overview data from different perspectives, and it can be achieved through different routes, depending on data features. This paper covers only one of the methods of creating series charts, which fits with the industrial occupational injuries and illnesses data from BLS. Further work is needed and different methods will be explored when different data features are handled.

NIOSH disclaimer: The findings and conclusions in this paper have not been formally disseminated by the National Institute for Occupational Safety and Health and should not be construed to represent any agency determination or policy.

REFERENCES

- NIOSH. (2004), "Worker Health Chartbook". Available: <http://www.cdc.gov/niosh/docs/2004-146/>
- Bureau of Labor Statistics, "Fatal Occupational Injuries in the United States, 1995-1999", Report 965
- Andrew S. Glassner. (1993), "Graphics Gems", p61-63
- SAS Institute Inc. (2004), "SAS/GRAPH 9.1 Reference".
- SAS Institute Inc. (2009). "SAS/GRAPH 9.2: Statistical Graphics Procedures Guide". Also Available at: <http://support.sas.com/documentation/cdl/en/grstatproc/61948/HTML/default/titlepage.htm>
- Luke Castellanos, Nick Spanos (2004) "Creating Ranking Charts Using SAS/GRAPH and the Annotate Facility", SUGI Paper 035-29

David Mink, David J. Pasta (2006) "Improving Your Graphics Using SAS/GRAPH Annotate Facility", SUGI Paper 085-31

Darrell Massengill (2005) "Tips and Tricks: Using SAS/GRAPH Effectively", SUGI Paper 90-30

Rita Tsang (2005) "Creating bar charts and scatter plots using the SAS GPLOT procedure. Both on the same page? It is easier than you think!" SUGI Paper 181-30

ACKNOWLEDGMENTS

Special thanks to David Wall for finding the "nice" graph label algorithm and translating it into SAS. We are also grateful to SAS Support group who has provided great help during our SAS programming to create high-quality charts.

CONTACT INFORMATION

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

Jun Ju
National Institute of Occupational Safety and Health
Cincinnati, OH 45226
(513) 841-4354
Email: jj@cdc.gov

Sue Nowlin
National Institute of Occupational Safety and Health
4676 Columbia Pkwy, MS-R4
Cincinnati, OH 45226
(513) 841-4467
Email: snowlin@cdc.gov

John Sestito
National Institute of Occupational Safety and Health
4676 Columbia Pkwy, MS-R4
Cincinnati, OH 45226
(513) 841-4184
Email: Jsestito@cdc.gov

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

Appendix:

Worker Health e-Chartbook

Worker Health Chartbook

The Worker Health Chartbook 2008 is a descriptive epidemiologic reference on occupational morbidity and mortality in the United States. A resource for agencies, organizations, employers, researchers, workers, and others who need to know about occupational injuries and illnesses, the Chartbook includes more than 400 figures and tables describing the magnitude, distribution, and trends of the Nation's occupational injuries, illnesses, and fatalities.

Table of Contents

- Chapter 1 - Labor Force Data and Forecasts
- Chapter 2 - Overview of Worker Health
- Chapter 3 - Agriculture, Forestry, Fishing, and Hunting industry sector (11)
- Chapter 4 - Mining industry sector (21)
- Chapter 5 - Construction industry sector (23)
- Chapter 6 - Manufacturing industry sector (31-33)
- Chapter 7 - Wholesale and Retail trade industry sectors (42, 44-45)
- Chapter 8 - Transportation, Warehousing, and Utilities industry sectors (48-49, 22)
- Chapter 9 - Health care and Social Assistance industry sector (62)
- Chapter 10 - Services industry sectors (51-56; 61; 71-72; 81)
- Chapter 11 - Fatal injuries and Musculoskeletal disorders

Index / Search

Last modified: July 10, 2008 16:22

Figure 6. Sample NIOSH Worker Health Chartbook page

Categories

Category Tree

- root - Root of all chapters
 - Part 1 - Overview of Worker Health
 - Chapter 1 - Fatal Injury
 - Chapter 2 - Nonfatal injuries and illnesses
 - Part 2 - Labor Force Data and Forecasts
 - Chapter 3 - Age
 - Chapter 4 - Gender
 - Chapter 5 - Race or Ethnicity
 - Chapter 6 - Occupation
 - Chapter 7 - Industry
 - Part 3 - Industry Sectors
 - Chapter 8 - Agriculture, Forestry, Fishing, and Hunting industry sector
 - Chapter 9 - Mining industry sector (21)
 - Chapter 10 - Construction industry sector (23)
 - Chapter 11 - Manufacturing industry sector (31-33)
 - Chapter 12 - Wholesale and Retail trade industry sectors (42, 44-45)
 - Chapter 13 - Transportation, Warehousing, and Utilities industry sector
 - Chapter 14 - Health care and Social Assistance industry sector (62)
 - Chapter 15 - Services industry sectors (51-56; 61; 71-72; 81)
 - Part 4 - Mortality and morbidity outcomes
 - Chapter 16 - Fatal occupational injuries
 - Chapter 17 - Nonfatal Injuries and illnesses involving days away from work
 - Part 5 - State
 - Part 6 - Economics
 - Part 7 - General Population Health Risk and Status

Category Details

ID (no edit): 14

Chapter: Chapter 3

Title: Age

Table Location:

Sort Order: 1

Last Updated: 3/4/2008 12:51:11 PM

NORA Sector: U.S. Workforce

NORA Subsector:

NORA Cross-sector: All

Web Content (Introductory material): How did employed U.S. workers differ by age?

Parent category: Part 2

Data Item:

Sort ord:	Description
1	U.S. workers aged 25-54 accounted for 99.7 million workers in 2006, o 1986, 1996, 2006, and projected 2016. Since 1986, the age distributio

Delete This Category

Figure 7. Microsoft Access interface of Content Management Database

Sample Barline series charts

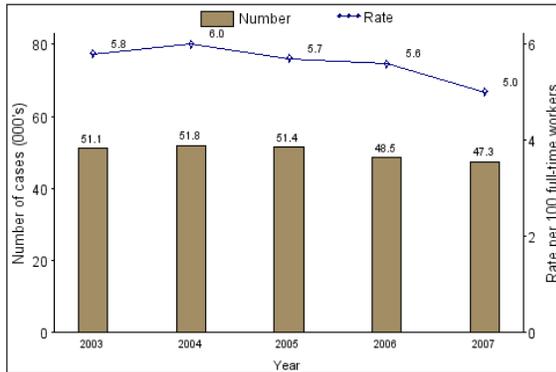


Figure 8. Oc_Inj_nr_11.gif

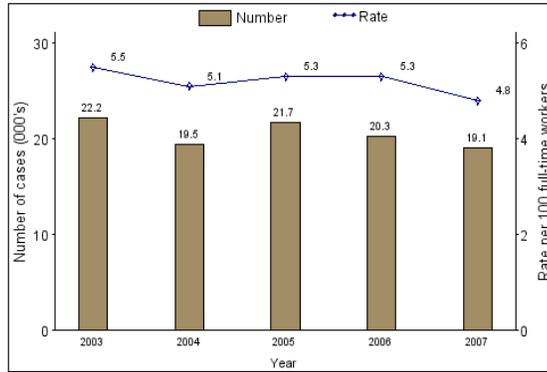


Figure 9. Oc_Inj_nr_111.gif

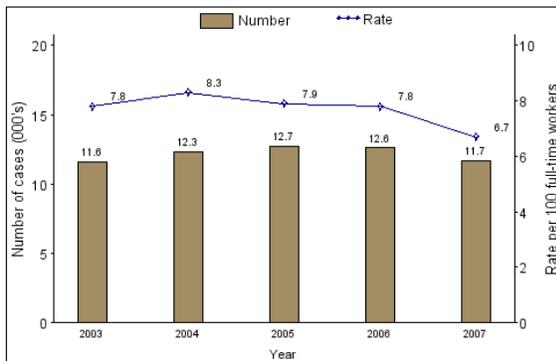


Figure 10. Oc_Inj_nr_112.gif

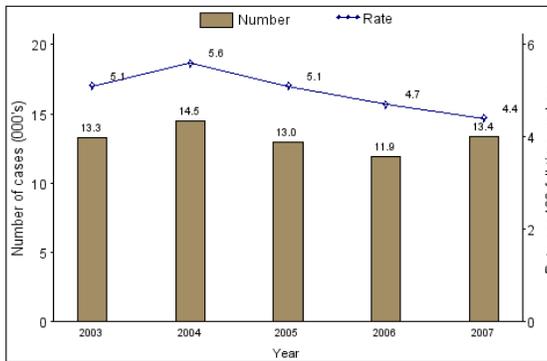


Figure 11. Oc_Inj_nr_115.gif

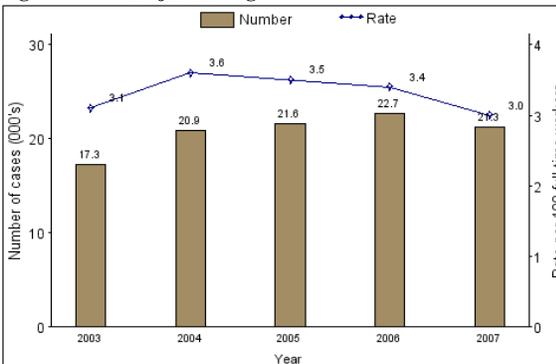


Figure 12. Oc_Inj_nr_21.gif

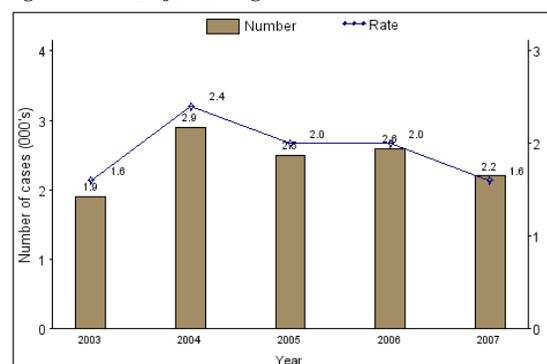


Figure 13. Oc_Inj_nr_211.gif

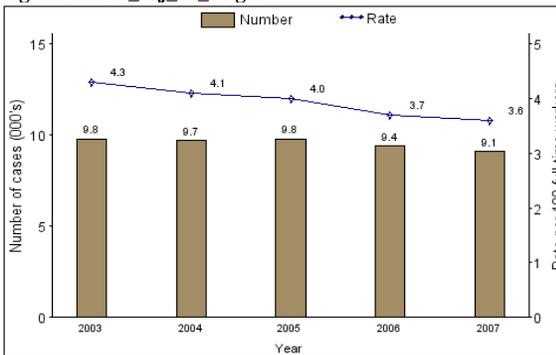


Figure 14. Oc_Inj_nr_212.gif

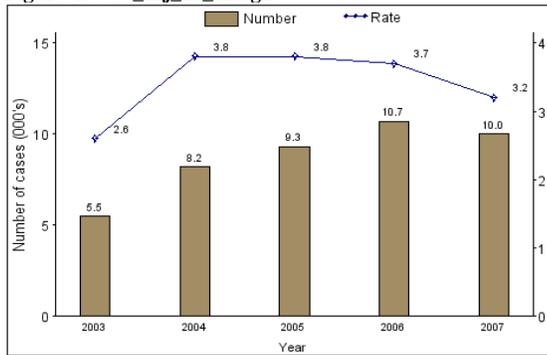


Figure 15. Oc_Inj_nr_213.gif