

Data Analysis on a Shoestring



Data Analysis for Beginners: Web-based Data Query Systems

**Laurie Barker, MSPH
Division of Oral Health
Centers for Disease Control and Prevention**

The findings and conclusions in this presentation are those of the author and do not necessarily represent the views of the Centers for Disease Control and Prevention].

SAFER • HEALTHIER • PEOPLE™

**No, I'm not going to try to teach
you SAS in 30 minutes.**

**Presentation: Overview of
resources and issues with
available data**

**Then group/individual practice with data
query Web sites**

Learning goals...

- Answer questions with data from existing resources
- Familiar with tools for finding data online
- Understand rationale for public release data sets
- Confident to use online data query systems
- Select databases for further analysis
- Use online information to inform development of surveillance or other data collection plans

What do I mean by beginner?

- Maybe someone who has never used statistical software before
 - ◆ Maybe never wants to, either!
- Someone who has analyzed data sets, but not this particular data set
- Someone who is in a big hurry responding to a request from the governor's office (or anyone)
- Someone tasked with developing a burden document with little or no analytic support or State data

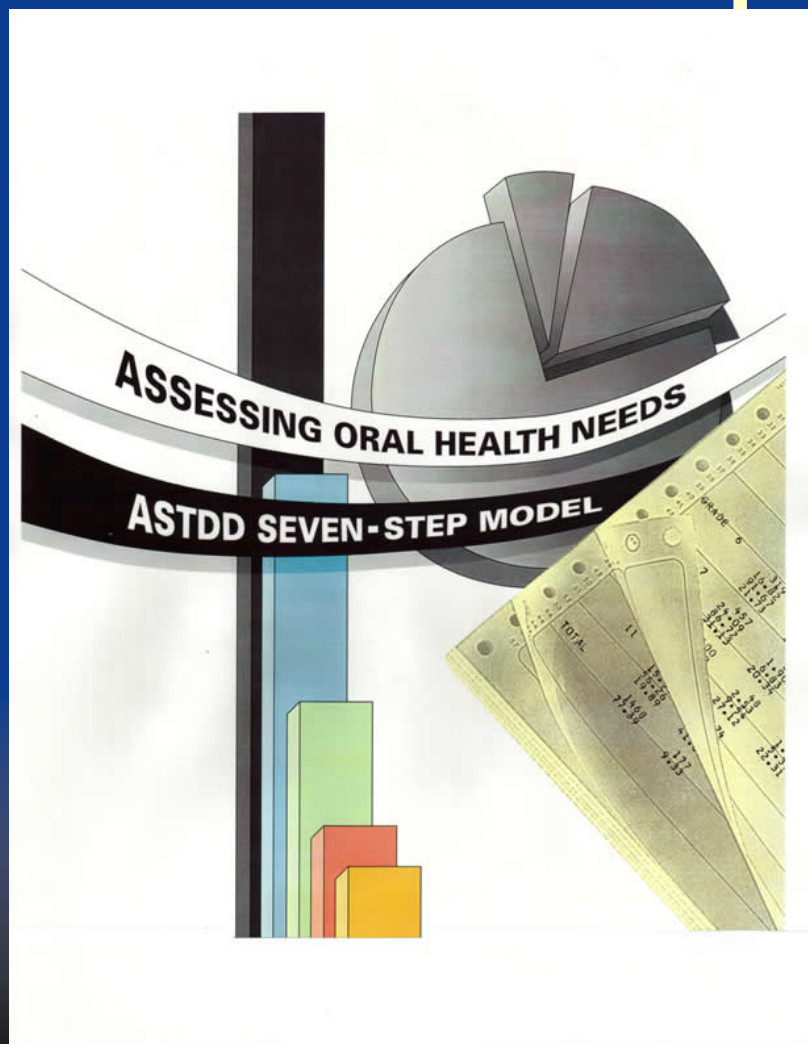
Before racing to collect new data, find out what's available already.

See handout –

Data Resource Cheat Sheet

**Lots of links in this presentation, but all are on the cheat sheet
(let me know if I missed one)**

ASTDD 7 Step Model



- Before primary data collection
 - ◆ Conduct inventory of available data
- Determine gaps in data
- Then design focused data collection to meet specific program information needs
 - ◆ Don't collect what won't be used

How or where do I find data for the inventory?

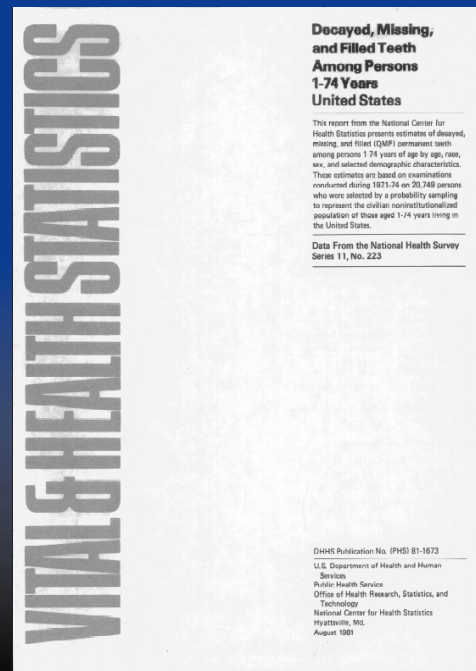
Data access of the past...



Data Tape Library,
National Center for Health Statistics



Portable computer
Year: 1982
RAM: 64K
Weight: 24.5 lb
Screen: 3 in?



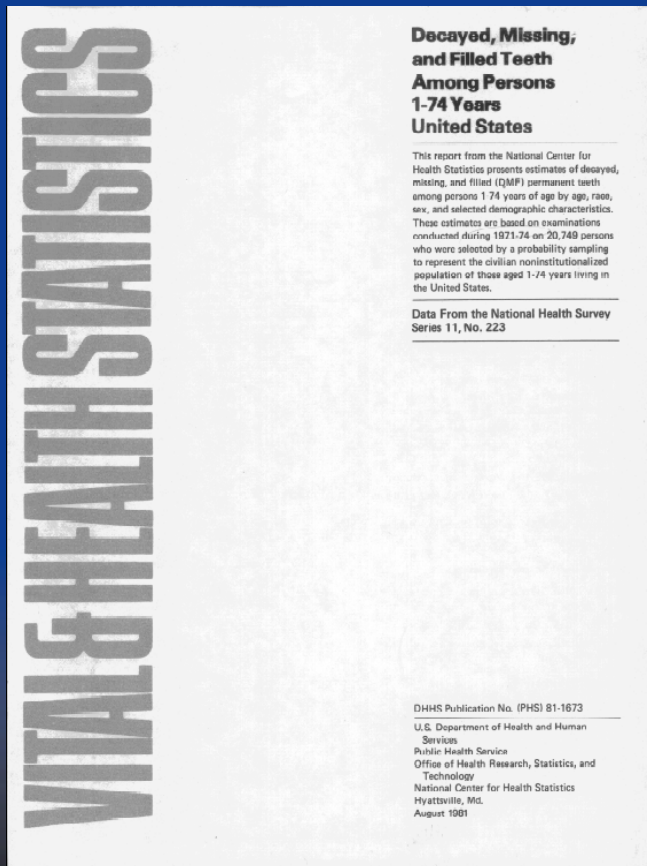
Data collected:
1971-1974
Report: 1982
Weight: 3 lb
Media: trees

Data access now...

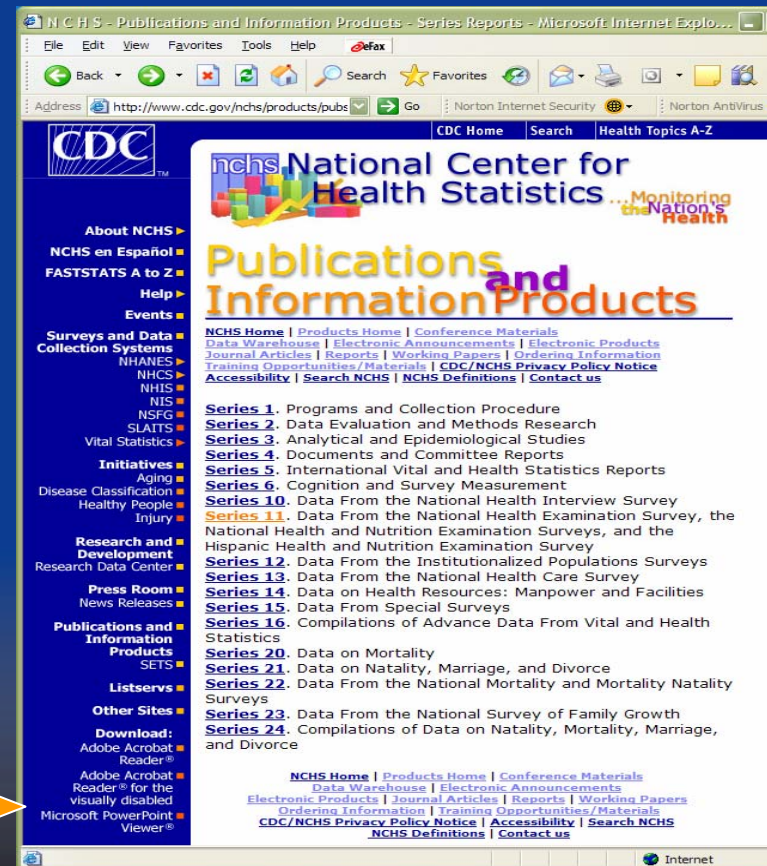
The image displays four overlapping screenshots of web browsers showing different CDC data systems:

- Top Left: National Oral Health Surveillance System (NOHSS)**
 - Address: <http://www.cdc.gov/nohss/>
 - Navigation: CDC Home, Search, Health Topics
 - Logo: CDC and ASTDD (Association of State and Territorial Dental Directors)
 - Section: National Oral Health Surveillance System (NOHSS)
 - Text: "NOHSS is a collaborative effort between CDC's Division of Oral Health and the Association of State and Territorial Dental Directors (ASTDD). NOHSS is designed to help public health programs monitor the burden of oral disease, use of the oral health care delivery system, and the status of community water fluoridation on both a state and national level. NOHSS includes..."
 - Menu: NOHSS Main Menu
 - Dental Visits
 - Teeth Cleaning
 - Complete Tooth
- Top Right: NIDCR/CDC Dental, Oral and Craniofacial Data Query System (DQS)**
 - Section: NIDCR/CDC Dental, Oral and Craniofacial Data Query System (DQS)
 - Text: "DQS is an on-line interactive data query system that makes oral health data readily available to policy makers, research professionals, program evaluators, dental practitioners, and the public. Jointly sponsored by the National Institute of Dental and Craniofacial Research (NIDCR) and the Division of Oral Health, Centers for Disease Control and Prevention (CDC), the DQS allows users to quickly retrieve nationally representative frequencies, percentages, and confidence intervals on oral health data from national health survey databases."
 - Features:
 - [DQS Features](#): Detailed information about the DQS, how to define research questions, how to submit data queries, and how to interpret results.
 - [Create a Data Query](#): Generate frequencies, percentages, and confidence intervals on oral health data using the DQS data query system.
- Bottom Left: National Survey of Children's Health Data Resource Center**
 - Address: <http://www.nschdata.org/DesktopDefault.aspx>
 - Section: National Survey of Children's Health Data Resource Center — Your Data... Your story
 - Buttons: DRC Home, What's New, Glossary, Search, Contact Us
- Bottom Right: PRAMS-Pregnancy Risk Assessment Monitoring System**
 - Address: <http://www.cdc.gov/PRAMS/index.htm>
 - Section: PRAMS-Pregnancy Risk Assessment Monitoring System | CDC Reproductive Health - A
 - Buttons: Home, About CDC, Press Room, A-Z Index, Contact Us
 - Text: Department of Health and Human Services, Centers for Disease Control and Prevention
 - Section: Pregnancy Risk Assessment Monitoring System (PRAMS): Home
- Bottom Left (MEPSnet):**
 - Address: <http://www.meps.ahrq.gov/MEPSnet/MEPSnetIntro.htm>
 - Section: MEPS Medical Expenditure Panel Survey
 - Text: Welcome to MEPSnet
- Bottom Right (HCUPnet):**
 - Address: <http://hcup.ahrq.gov/HCUPnet.asp>
 - Section: H-CUPnet
 - Text: A free, on-line query system based access to health statistics and information at the national, regional, and State level
 - Text: Welcome to HCUPnet

Data Report Library, just online



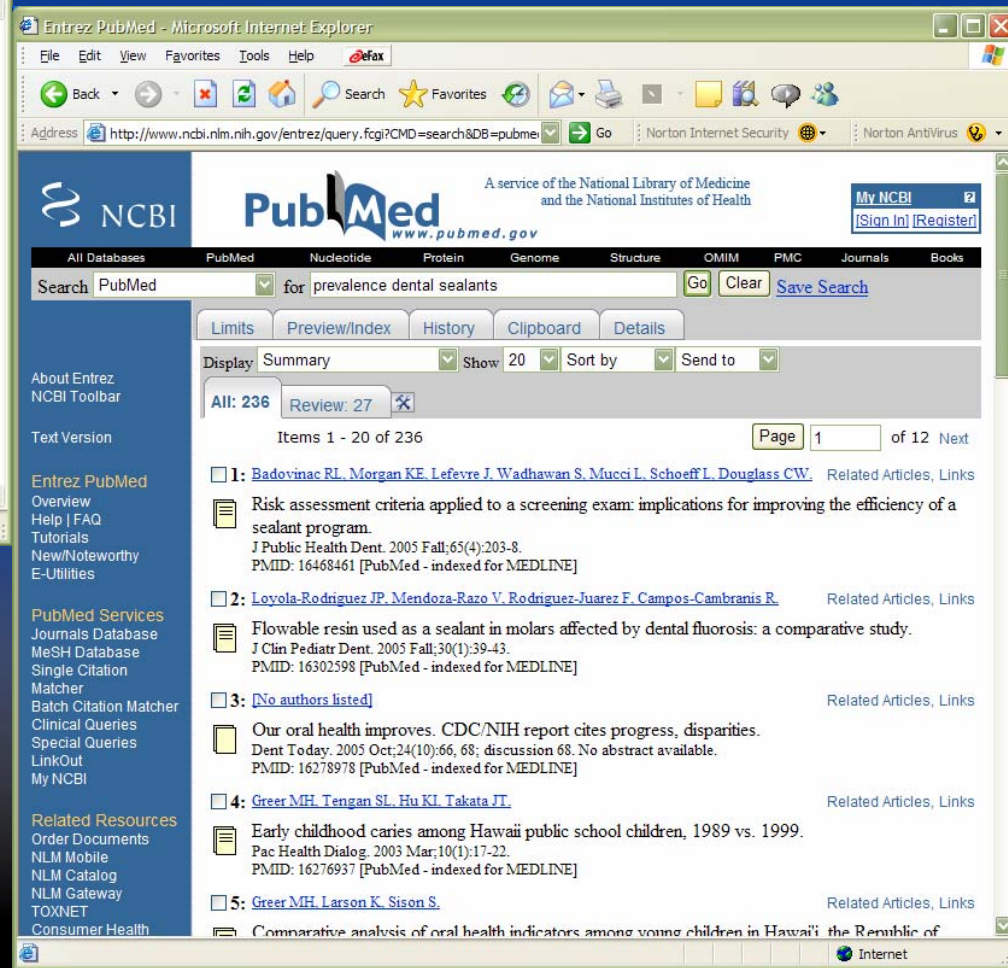
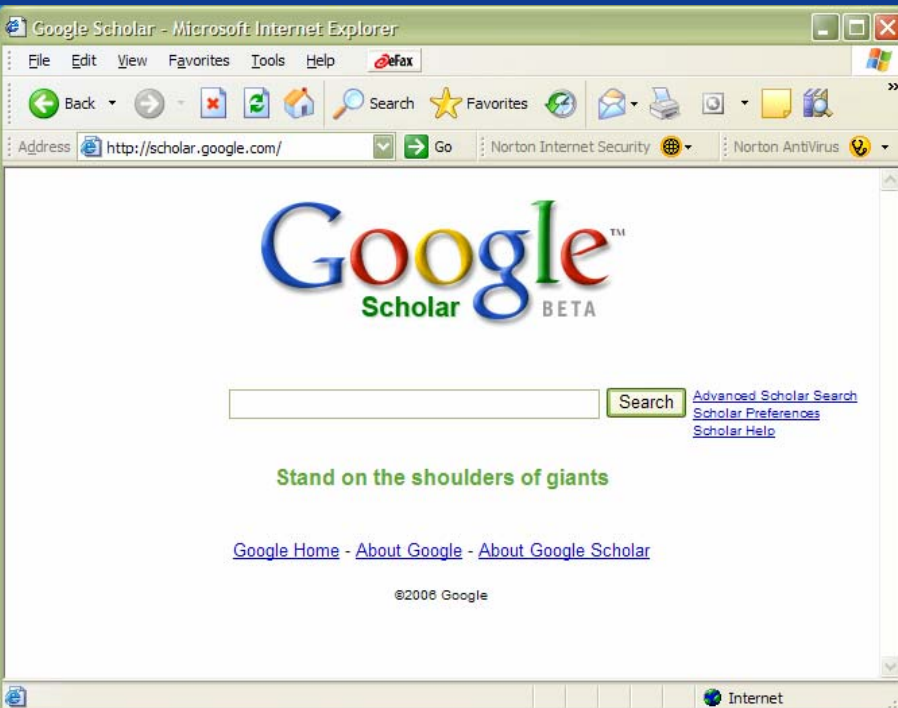
- Old reports digitized for Web access
- New reports may appear on Web before in print (or not)



<http://scholar.google.com>

<http://www.pubmed.gov>

Web versions of reports are searchable



Use published reports

- get estimates
- identify data sources
- author contact information

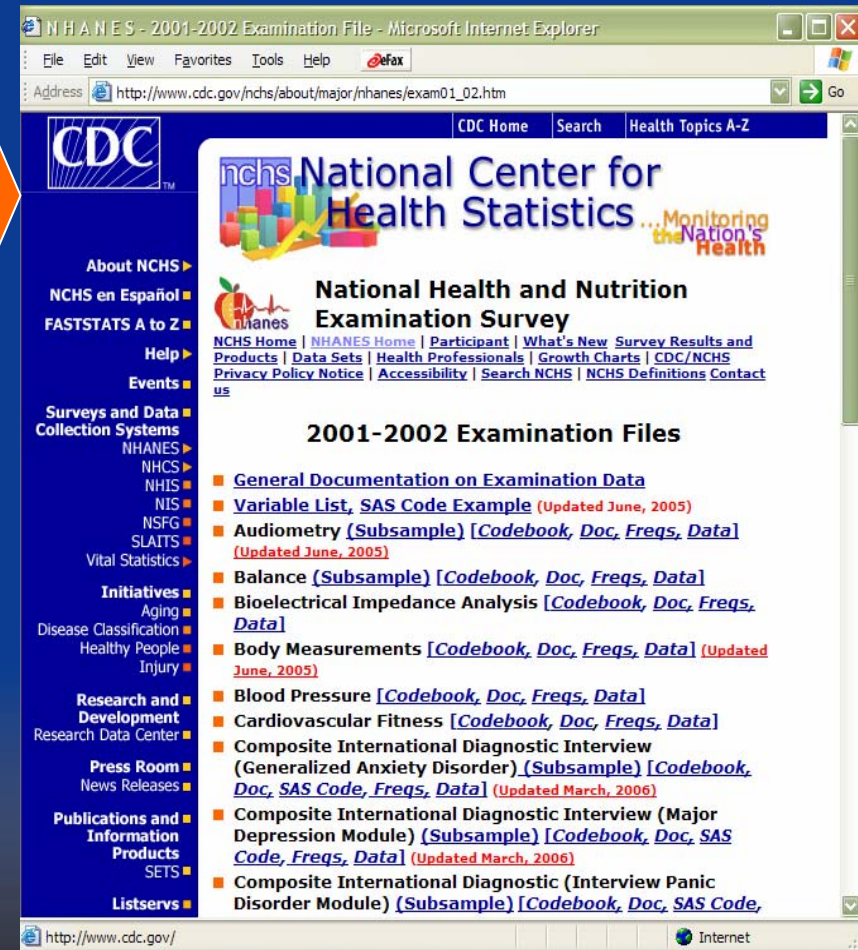
**That's great, but the reports don't have
the exact number I was looking for...**

Show me the data!

Two conflicting mandates for government data agencies

- Make all data collected available to the public, as quickly as possible (within 1 year of collection or publication)
- Protect the privacy of persons and entities providing data and confidentiality of sensitive data provided
- Solution: **Public Release data sets**
 - ◆ Tested (and sometimes altered) to reduce risk of disclosure
 - ◆ Users must report to data providing agency if user has been able to identify survey participants

Data set warehouse, just online

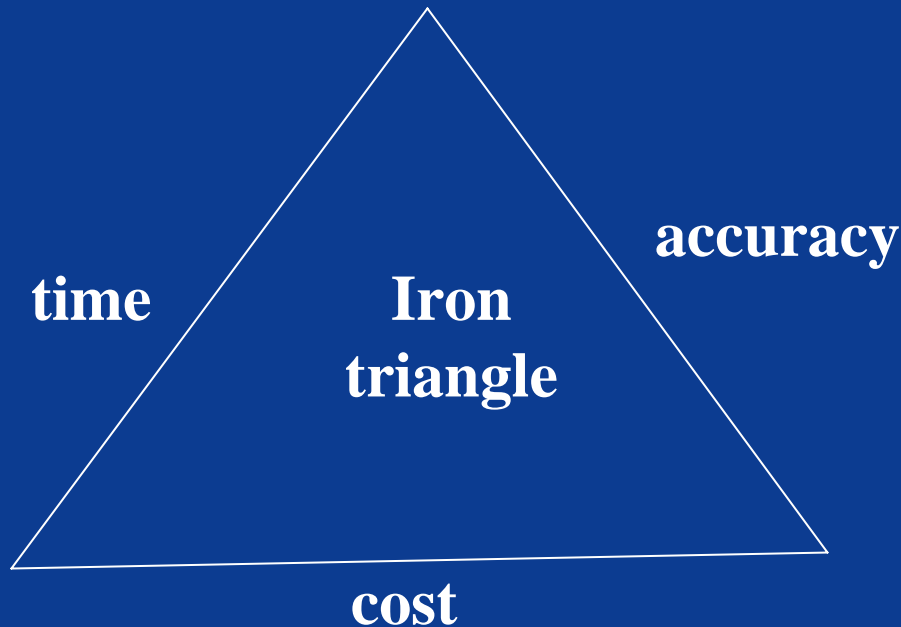


- Public release data sets available by download from Web sites
- Still requires user to know SAS/SUDAAN or other software to analyze
- Makes a lot of data more readily available to analysts
- Still requires special software to analyze properly
- Some data only available through Research Data Center, or other data use arrangement

Downloading data tips

- Check the size of the file
 - ◆ Dialup may be too slow
- Get the documentation, too
 - ◆ Check for changes to analytic guidelines
 - ◆ Read to see content and collection methods
- If you download data, note the date downloaded and the file date (date posted to Web site)
- Check the date of the file, and check for updates
 - ◆ Data may change, data sets withdrawn

Why would a data set change?



- Faster production means potential sacrifice to accuracy of 1st release
- Possible changes as issues discovered
- Change less likely the longer data set has been available

Feedback from data users

Likelihood of change



Managing data files downloaded from Web sites

- Think about file structure in advance
- Protect the original files
 - ◆ Don't assume the same file will be available from the Web site, exactly the same as when first downloaded
- Create “working files”
- Save work regularly
- Back up files regularly
 - ◆ Some office networks have back-up built in
 - ◆ But if it's critical, create your own back-up
 - ◆ Store back-up copies in separate places
 - ★ If the building burns down, will both the original and back up be gone?

Data file structure – one I use for NHANES

- NHANES 1999-2000
(similar folder for 2001-2002)

- ◆ Original files

- ★ Codebooks and documentation
- ★ XPT files
 - Demo
 - OHXDENT
 - ...

- ★ SASV8 files

- ◆ Working files

- NHANES 1999-2002

- ◆ Analysis files

- ★ SASV8 files
- ★ SASV6 files
- ★ My own documentation, variable crosswalks, notes

- ◆ Project-specific

- ★ SAS programs
- ★ Output data sets
- ★ Project documentation

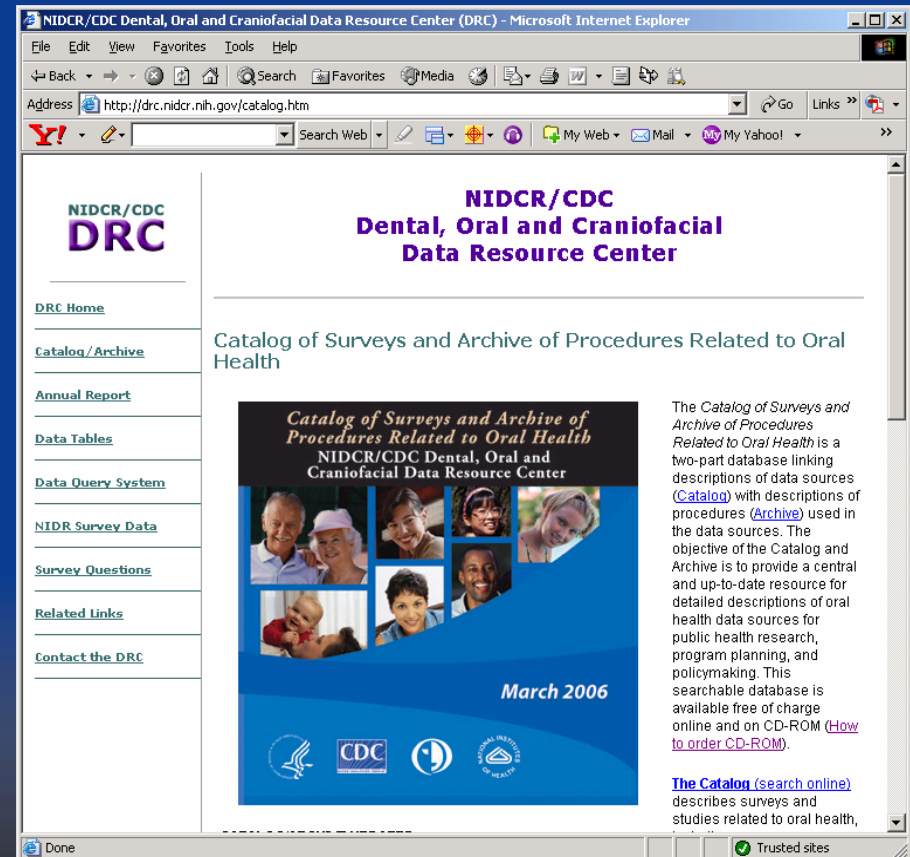
**Well, that's nice, but I don't have SAS
or whatever, and I don't really have
time or money to get the software and
do the analysis from scratch.**

OR

**I'd really like to know more about these
data sets before I get into it and learn
that it doesn't have what I was really
looking for.**

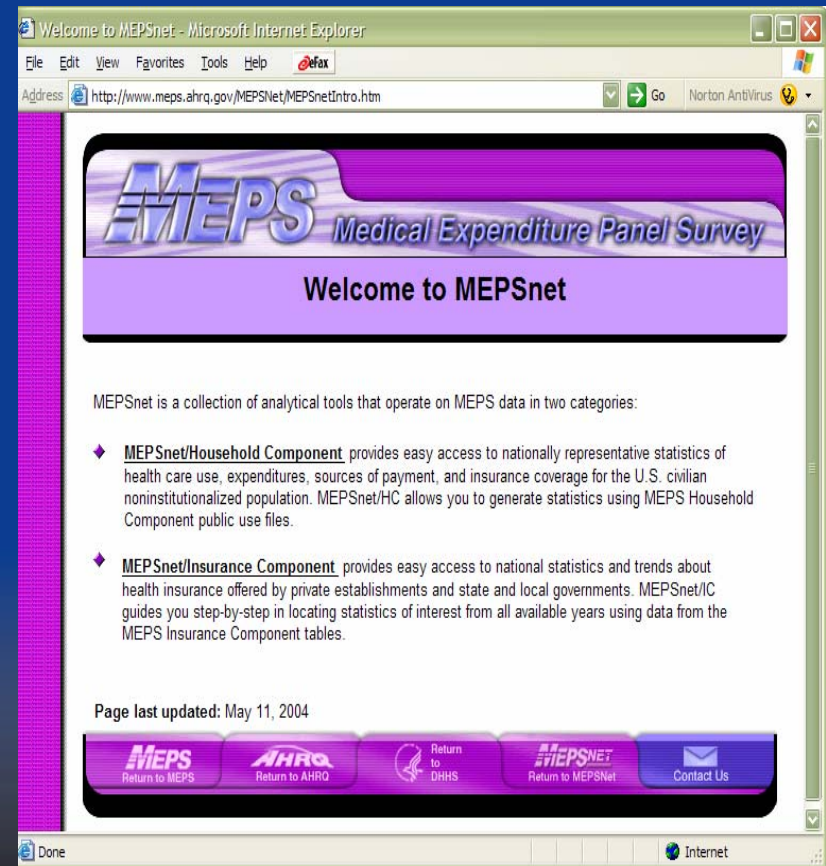
A tool to find surveys with oral health information of use to you...

- DRC Catalog
- Searchable data base of meta data about surveys with oral health data
- Includes links to data Web sites
- Contact information for each survey
- Literature search for each survey, updated every 6 months



Online Data Query Systems

- Provide Web-based tools for data analysis
- Requires no knowledge of statistical software required
- Not usually designed for advanced, multivariate modeling
- Do provide tools to create estimates within groups (stratified tables)
- Lower burden on user to protect data



Not all data are on the Web

- Web based systems are a good starting place for an inventory
- Further inquiry into state-held data, and data held by other organizations

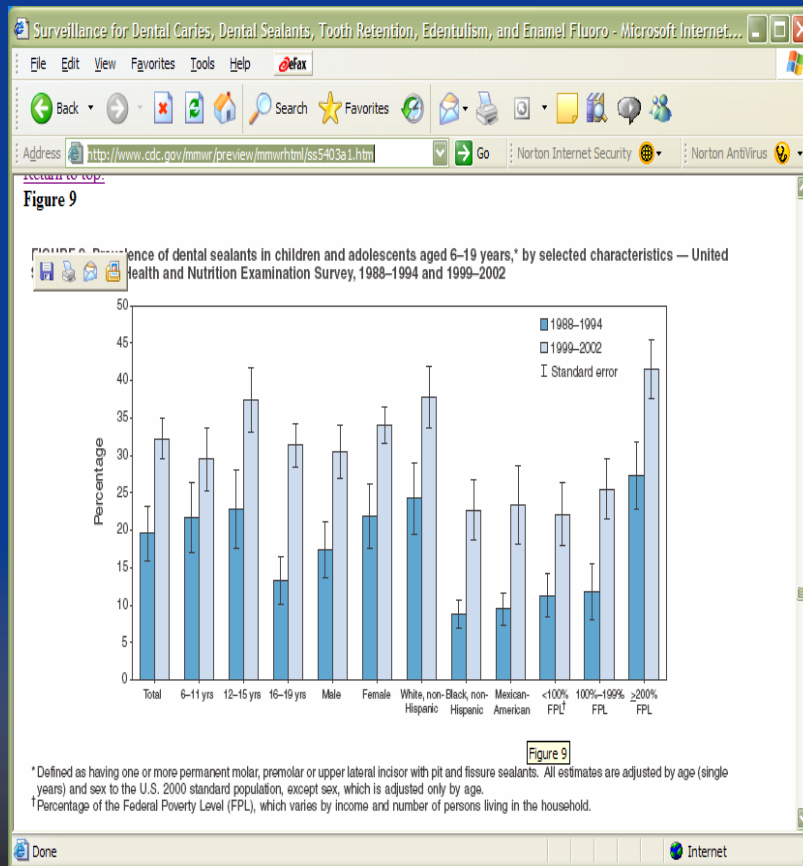
You can't always get what you want...

**But if you try sometimes
you just might find
you get what you need...**

What questions can I answer using online data query systems?

- How many children in my county have sealants by tooth type, grouped by severity of early childhood caries?
 - ◆ Ok – probably not, unless state has system on Web
- How many children in the US and in my State have dental sealants?
 - ◆ How different is my county likely to be from the U.S. average?

Prevalence of dental sealants United States and individual states



Dental Sealants - Overview - Microsoft Internet Explorer

Address: <http://www2.cdc.gov/nohss/IndicatorV.asp?Indicator=1>

National Oral Health Surveillance System
[Home](#) | [Glossary](#) | [Related Links](#) | [Contact Us](#)

Dental Sealants

Percentage of 3rd Grade students with dental sealants on at least one permanent molar tooth

See also information on [State Oral Health Surveys](#)

Overview by State

Sort list by: State [Percent with Sealants](#)

State	School Year	Percent with Sealants ¹	Response Rate ² (%)	Percent eligible for free and reduced-cost lunch program ³		
				Sample Schools ⁴	Students ⁵	State
Arkansas	2001-2002	% 24.4 CI (21.3-27.5) N 815	86	55	NR	45
Colorado	2003-2004	% 36.0 CI (33.9-38.2) N 2031	68	34	NR	33
Delaware	2001-2002	% 34.1 CI (29.9-38.2) N 1032	43	37	41	40
Idaho	2000-2001	% 52.1 CI (48.8-55.5) N 3126	71	50	NR	39
Kansas	2003-2004	% 33.1 CI (30.2-36.0) N 3375	32	NR	NR	NR
Maine	1998-1999	% 47.6	51	NR	31	32

This system was developed with the collaboration of the Centers for Disease Control and Prevention (CDC) and the Association of State and Territorial Dental Directors (ASTDD).

It's easier to criticize than to be correct: anticipate questions

- **In our sealant example**
 - ◆ **Age group differences**
 - ★ NHANES – children 6-11
 - ★ NOHSS Basic Screening Survey (BSS) – 3rd grade
 - ◆ **Sampling differences**
 - ★ NHANES household
 - ★ NOHSS BSS – school-based
- **Would those differences change the decision made on the basis of these data?**

Data progression

Good data routinely collected, analyzed, reported, and used for monitoring trends and decision-making



**Surveillance
System**

Better data



Weak, or not quite what
you wanted, data



Photo by Sgt. 1st Class Doug Sample, USA

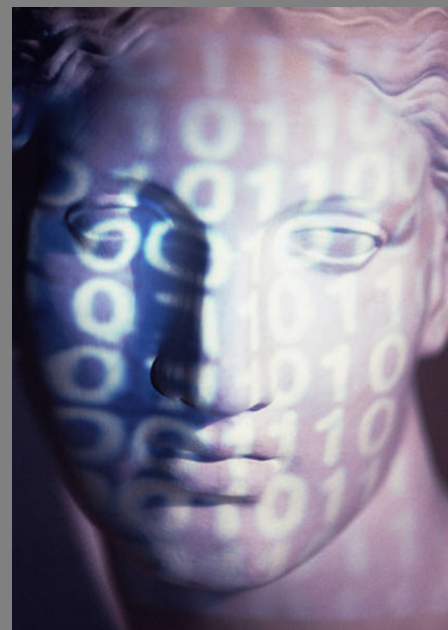
No data



Using Web systems as a starting point

- Developing information pieces with data
 - ◆ Burden document
 - ◆ Short pieces
- Planning surveillance activities
 - ◆ Choose indicators
 - ◆ Design sample
 - ◆ Plan analysis

THE BURDEN OF ORAL DISEASE



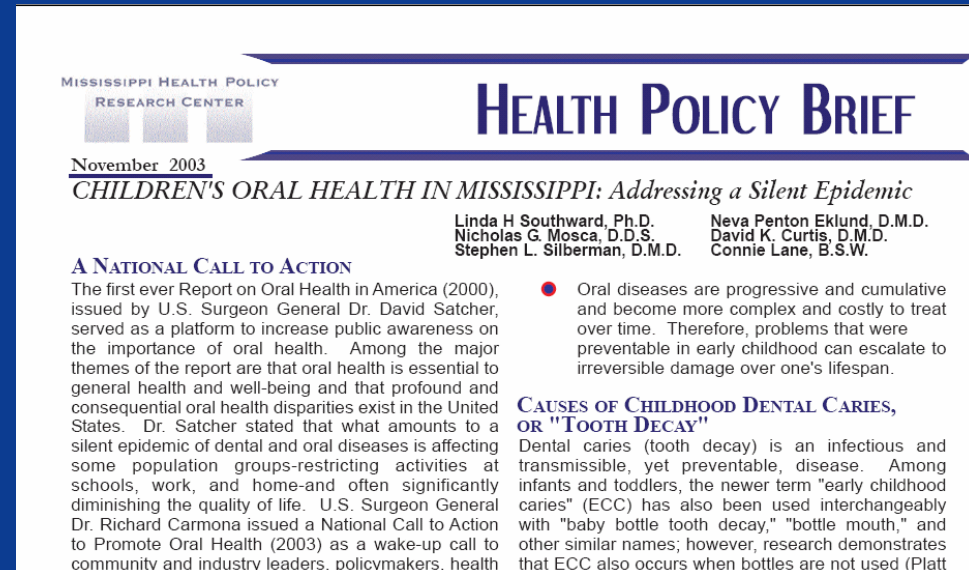
A Tool for Creating State documents

Burden Document is big...

- It can serve to illustrate gaps in data
- Can document the data inventory, if additional state sources are added that may not be reflected in the tool
- Mostly, serves as a reference from which to prepare shorter, focused information pieces
- Can help focus next efforts for analysis, data collection
- HP2010 indicators and more

Example: Mississippi policy brief

- Used information from range of sources:
 - ◆ Major reports
 - ◆ National and state data sources
 - ◆ Epidemiologic and programmatic
 - ◆ Partner sources



Some data used in this report were available from online sources.

Others were obtained and analyzed by university partner.

Burden Document indicators: Disease

- ◆ **Caries** (NHANES, NOHSS, State data)
- ◆ **Tooth loss** (NHANES, BRFSS, State data)
- ◆ **Periodontal diseases** (NHANES, state data)
- ◆ **Oral cancer** (CDC/NCI US Cancer statistics)

- ◆ **Special populations**
 - ★ Children, adults, women, disabled
- ◆ **Disparities**
- ◆ **Social impact**
- ◆ **Costs** (CMS, Medicaid, state data, MEPS)
 - ★ Direct and indirect

Burden Document indicators: Risk and protective factors

- ◆ **Water fluoridation** (CDC/WFRS, State)
- ◆ **Topical fluoride and supplements**
(Synopses, state)
- ◆ **Sealants** (NHANES, NOHSS,)
- ◆ **Preventive visits**
- ◆ **Cancer screening**
- ◆ **Tobacco control** (CDC)
- ◆ **Oral health education** (Synopses)

Burden Document indicators: Provision of services

- Dental Workforce and Capacity
- Dental Workforce Diversity
- Use of Dental Services
 - ◆ General population
 - ◆ Schoolchildren
 - ◆ Pregnant women
- Dental Medicaid and State Children's Health Insurance Programs
- Community and Migrant Health Centers and other State, County, and Local Programs

Burden Document Tool

Links to Data Sources

Table III. Proportion of Adults* with Untreated Dental Caries, by Selected Age Groups and Demographic Characteristics

	Age 35–44 Years		Age 65–74 Years	
	United States ^a (%)	<STATE> ^f (%)	United States (%)	<STATE> ^f (%)
<i>Healthy People 2010 Target</i>	15	15		
TOTAL	26		19	
Race or Ethnicity				
American Indian or	68 ^b			

Table III Sources:

Aged 35–54 years

U.S. Department of Health and Human Services. *Healthy People 2010, Progress Review, 2000*. Available at www.cdc.gov/nchs/ppt/hpdata2010/focusareas/fa21.xls.

<These data will be updated in 2006>

Aged 65–74 years

Health, United States, 2004. Table 80, p. 267. Available at www.cdc.gov/nchs/data/hs/hs04.pdf.

Burden Document Tool: Data Release Calendar

Appendix b. Data Release Calendar for Oral Health Indicators

Data Source	Website Location	Frequency of Release	Past Release Date	Future Release Dates	Indicator
Synopses of State and Territorial Dental Public Health Programs	http://www2.cdc.gov/nccdphp/doh/synopses/index.asp or www.cdc.gov/oralhealth or www.astdd.org	Annual collection, 6 month lag	FY2004 (2005 Synopses, available from ASTDD June 2005, from Web late 2005)	<ul style="list-style-type: none"> FY2005 (2006 Synopses, available from ASTDD June 2006, from Web late 2006) FY2006 (2007 Synopses, available from ASTDD June 2007, from Web late 2007) 	<p>System for recording and referring infants and children with cleft lip and cleft palate, all</p> <p>Oral health surveillance system, all</p> <p>Tribal, state, and local dental programs with a public health trained director, all</p>
Behavioral Risk Factor Surveillance System (BRFSS)	http://www.cdc.gov/brfss		2004 data released June 16, 2005	<ul style="list-style-type: none"> 2004 – released June 16, 2005 2005 – optional module only, expected June 2006 2006 – expected June 2007 2007 – optional module only, expected June 2008 	Percentage of people who had their teeth cleaned within the past year, visited dentist in past year, and complete tooth loss ages 18 or higher.
Health Resources and Services Administration (HRSA), Bureau of Primary Health Care (BPHC)	http://datawarehouse.hrsa.gov/				Community-based health centers and local health departments with oral health components, all

Web-based Data Query Systems

- NIDCR/CDC Data Query System
- National Oral Health Surveillance System
- Behavioral Risk Factor Surveillance System
- Youth Risk Behavior Surveillance System
- Medical Expenditure Panel Survey
- National Survey of Children's Health

- Start at CDC home page
- Go to NOHSS
- Caries Experience
- Some states have data for grades other than 3rd
- And, there is untreated decay, and sealants.
- But, state data only for this indicator
- So let's look for national data

State Oral Health Surveys - Vermont - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Mail

Address <http://www2.cdc.gov/nohss/OHSurveyDetailV.asp?StateID=50> Go Links

Y! Search Web My Web Mail

State Oral Health Surveys

Vermont
2002-2003 School Year

[Back to State Oral Health Surveys](#)

Data Submitted 2/27/2003

Grade	N	Adjusted for Non-response	% (95% CI)	Sealants	Caries Experience	Untreated Decay
K	NC	No	% CI		NC NC	NC NC
		Yes	% CI		NC NC	NC NC
1st	417	No	% CI	12.5 (9.5-16.1)	35.7 (31.2-40.6)	17.5 (14.1-21.6)
		Yes	% CI	12.5 (7.9-17.1)	35.5 (28.8-42.1)	17.2 (11.9-22.6)
2nd	412	No	% CI	42.7 (37.9-47.7)	40.0 (35.3-45.0)	13.8 (10.7-17.6)
		Yes	% CI	42.3 (33.9-50.6)	40.1 (34.3-45.9)	14.2 (10.1-18.4)
3rd	409	No	% CI	66.3 (61.4-70.8)	44.3 (39.4-49.2)	16.1 (12.8-20.1)
		Yes	% CI	66.1 (59.4-72.9)	45.1 (39.0-51.1)	16.2 (12.9-19.5)

% Percentage
CI 95% Confidence Interval
NC Not Collected

This system was developed with the collaboration of the Centers for Disease Control and Prevention (CDC) and the Association of State and Territorial Dental Directors (ASTDD).

ASTDD

Dental Visits
Teeth Cleaning
Complete Tooth Loss
Fluoridation Status
Dental Sealants
Caries Experience
Untreated Tooth Decay
State Profiles
Data Sources
Related Links
Frequently Asked Questions

Done Local intranet

Let's try google

Search: Caries
Experience
Surveillance

August 2005
MMWR

Caries
experience and
untreated decay
by primary or
permanent teeth,
sealants

Healthy People,
Data, Data2010

TABLE: Microsoft Internet Explorer

Address: http://wonder.cdc.gov/scripts/broker.exe

21-01b Dental caries experience - Primary or permanent teeth - Children (aged 6 to 8 years)

	#1	1988-94	52%	---	---	---	---	1\51%	---	---	---	42%
Race and ethnicity												
American Indian or Alaska Native only	#2	1999-00	a,2,3\91%	NA	NA	NA	---	---	---	---	---	42%
Asian or Pacific Islander only	#3	1988-94	3\DNA	---	---	---	---	1\DNA	---	---	---	42%
Asian only	#4	1993-94	3,4\90%	---	---	---	---	---	---	---	---	42%
Native Hawaiian or Other Pacific Islander only	#5	1999	3,5\79%	NA	NA	---	---	---	---	---	---	42%
Black or African American only	#6	1988-94	a,3\DNA	---	---	---	---	1\DNA	---	---	---	42%
White only	#7	1988-94	a,3\DNA	---	---	---	---	1\DNA	---	---	---	42%
2 or more races	#8	1988-94	DNC	---	---	---	---	1\DSU	---	---	---	42%
American Indian or Alaska Native; White	#9	1988-94	DNC	---	---	---	---	1\DSU	---	---	---	42%
Black or African American; White	#10	1988-94	DNC	---	---	---	---	1\DSU	---	---	---	42%
Hispanic or Latino	#11	1988-94	DSU	---	---	---	---	1\DSU	---	---	---	42%
Mexican American	#12	1988-94	a\64%	---	---	---	---	1\67%	---	---	---	42%
Not Hispanic or Latino	#13	1988-94	3\DNA	---	---	---	---	1\DNA	---	---	---	42%
Black or African American only, not Hisp/Latino	#14	1988-94	3\49%	---	---	---	---	1\56%	---	---	---	42%
White only, not Hispanic or Latino	#15	1988-94	3\49%	---	---	---	---	1\46%	---	---	---	42%
Gender												
Female	#16	1988-94	54%	---	---	---	---	1\48%	---	---	---	42%
Male	#17	1988-94	50%	---	---	---	---	1\53%	---	---	---	42%
Education level (head of household)												
Less than high school	#18	1988-94	65%	---	---	---	---	1\DNA	---	---	---	42%
High school graduate	#19	1988-94	52%	---	---	---	---	1\DNA	---	---	---	42%
At least some college	#20	1988-94	43%	---	---	---	---	1\DNA	---	---	---	42%
Disability status												
Persons with disabilities	#21	1988-94	DNA	---	---	---	---	1\DNA	---	---	---	42%
Persons without disabilities	#22	1988-94	DNA	---	---	---	---	1\DNA	---	---	---	42%
Select populations												
3rd grade students	#23	1988-94	60%	---	---	---	---	1\DNA	---	---	---	42%

Done - Focus area:21-Oral Health

Local intranet

Com F-10

Prob Don

National Center for Health Statistics

Division of Health Promotion Statistics

3311 Toledo Road
Hyattsville, MD

DATA2010 Home
Data by [data source](#)

* 21.*^Oral Health has been selected *

Local intranet

Future systems

- NHANES system in development by NCHS
- NHIS system in development by NCHS
- Extension of NIDCR/CDC Data Query System
- NOHSS workgroup meeting to review existing indicators, choose additional, evaluate NOHSS

Questions

- Web Query Practice

Web-based Data Query Systems

- NIDCR/CDC Data Query System
- National Oral Health Surveillance System
- Behavioral Risk Factor Surveillance System
- Youth Risk Behavior Surveillance System
- Medical Expenditure Panel Survey
- National Survey of Children's Health