



National Institute of Dental
and Craniofacial Research

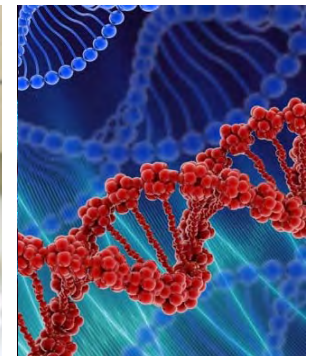
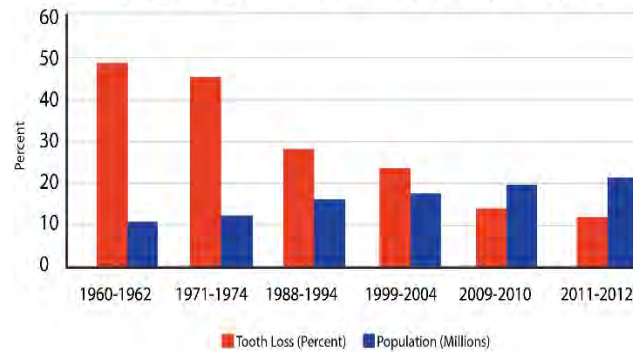
BIG DATA and Oral Health Research

Bruce A. Dye, DDS, MPH

19 April 2016



Tooth Loss in the United States among adults age 65-74 years



Informatics-BIG DATA-Oral Health Research

The Landscape Ahead

- Brief Introduction to Health Informatics
- Brief Description of what BIG DATA is
- Highlight some BIG DATA projects using publicly available data

Informatics-BIG DATA-Oral Health Research



 **KRONOS**
CREATED by marketoon studios

Workforce Innovation That Works™

KRONOS.COM

Informatics-BIG DATA-Oral Health Research

Health Informatics and Evolution of BIG DATA

- Informatics Science is one of the fastest growing research fields
- It is a multidisciplinary approach to information and knowledge management in biomedical research, clinical care, and public health
- Origins in computer science / information technology
- Evolving with more of a focus on Data Science

Informatics-BIG DATA-Oral Health Research

Health Informatics and Evolution of BIG DATA

- Informatics is Transforming the Way We Work
 - Data mining,
 - Natural language / text processing,
 - Cognitive science,
 - Clinical research,
 - Genomics / proteomics, and
 - Database management and data analytics for large amounts of data generated in health care delivery or produced through public health activities.

Informatics-BIG DATA-Oral Health Research

Health Informatics and Evolution of BIG DATA

- Recognized Core Areas of Informatics (AMIA)
 - Translational Bioinformatics
 - Clinical Research Informatics
 - Clinical Informatics
 - Consumer Health Informatics
 - Public Health Informatics

Informatics-BIG DATA-Oral Health Research

Health Informatics and Evolution of BIG DATA

- Why is BIG DATA becoming a big deal in Informatics?
 - Facilitating biomedical science
 - Promoting the quality and safety of patient care
 - Improving population health

Informatics-BIG DATA-Oral Health Research

Health Informatics and Evolution of BIG DATA

- BIG DATA can be characterized by:
 - The amount of collected data (Volume)
 - The source and type of data (Variety)
 - The speed of generated data (Velocity)
 - The quality of incoming data (Veracity)

Informatics-BIG DATA-Oral Health Research

What creates BIG DATA?

- Data Accumulation Matters (Velocity)
 - Data accumulates expansively
 - No economic trade off between retaining vs deleting data
- Data Quality Matters (Veracity)
 - More data doesn't always give us more "value"
 - Information from data becomes more valuable when the data is more reliable
 - Data-driven decision-making requires accurate and reliable data

Informatics-BIG DATA-Oral Health Research

What creates BIG DATA?

- Data Size Matters (Volume)

- At its most elementary level—BIG DATA is about bringing datasets together
- There is no critical mass of data alone needed to make it “Big.”

- Data Complexity Matters (Variety)

- You need a mix of varying types and sources of data to make it complex and large.
- Data linkages are important

Informatics-BIG DATA-Oral Health Research

What creates BIG DATA?

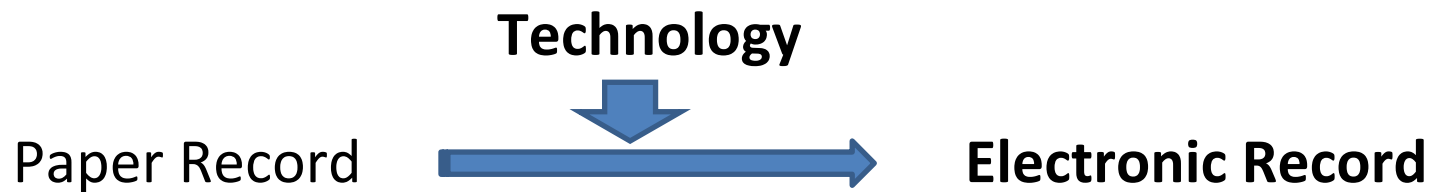
- “Big Data” begins to form when a group of data sets are brought together become so large and complex that it begins to challenge contemporary data processing and analytical approaches.

Informatics-BIG DATA-Oral Health Research

- Technology is also propelling this phenomena of “Big Data” and it’s use in research.
- Technology is allowing us to collect, process, and share data easier and faster then ever before.



Informatics-BIG DATA-Oral Health Research



- Technology and creation of Electronic Health Record (EHR).
- Big Data analytics are not about analyzing a provider's EHRs in isolation.
 - For example, to identify how many widgets were rendered
- It is about analyzing provider health recorders to identify factors that are **associated** with those widgets rendered.

Informatics-BIG DATA-Oral Health Research

- The contribution of technology to the evolution of BIG DATA
 - described well by Kenneth Cukier (TED talk).
 - “We use to look at small data to try to understand the world.”
 - “The idea is that more data doesn’t let us see more of the same but it allows us to see better, to see different, to see something new.”
- Inherent to this is that BIG DATA methods permits us to observe varying amounts of information in a way that we can extract value faster and more meaningful.

Informatics-BIG DATA-Oral Health Research

- Observation is fundamental to Epidemiology.
 - Epidemiology is the study of the distribution and determinants of diseases and conditions in man.
 - Epidemiology is essentially all about collecting data and analyzing it to reveal patterns, trends, and associations.
- BIG DATA is about observing better—identifying patterns in ways that make us better understand the world we live in.

What is “Big Data”?

BIG DATA is:

Very large complex datasets that may be analyzed computationally to reveal patterns, trends, and associations, especially relating to human health, behavior, and interactions.



Examples of BIG DATA Projects

- Using linked data for BIG DATA research projects

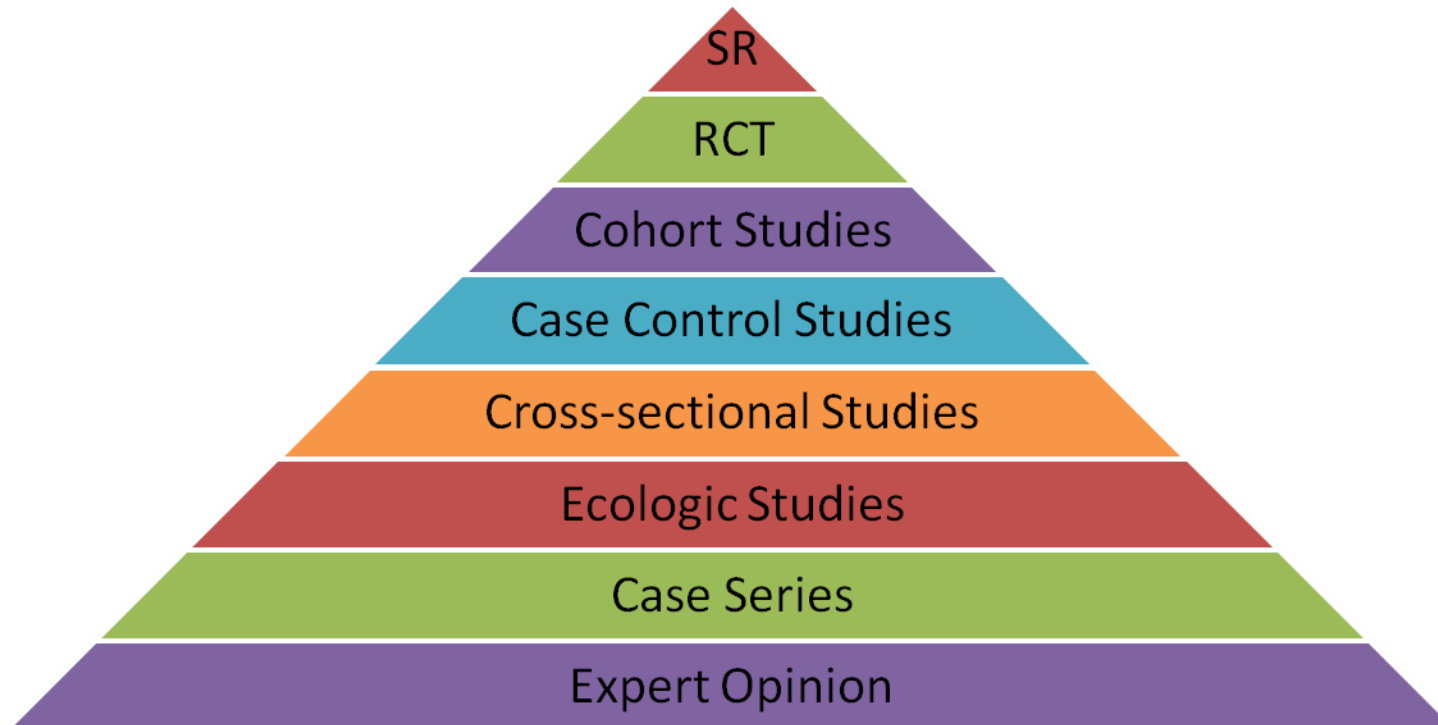


Examples of BIG DATA Projects



Examples of BIG DATA Projects

- Review Study Design Hierarchy & Levels of Evidence

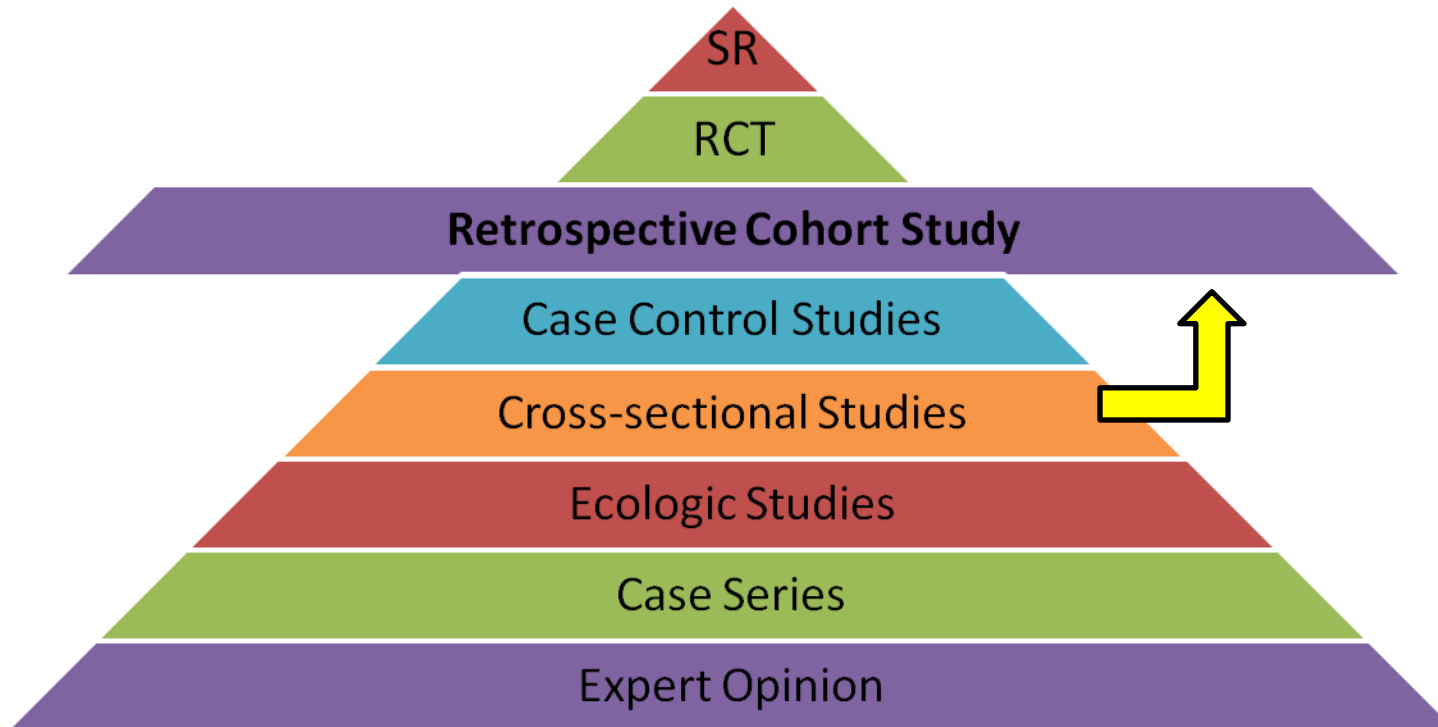


Examples of BIG DATA Projects

- What is the potential of BIG DATA?
- Observational studies on steroids
 - Where the sum has more added value than the individual data files

Examples of BIG DATA Projects

- Boosting NHANES from a good x-sectional study to a good cohort study accounting for temporal influence



Examples of BIG DATA Projects

- Example 1: NHANES and Death Records

Examples of BIG DATA Projects

- Obesity really associated with increase risk of death?
- NHANES contributes baseline information
 - Age 17-60 in 1988-1994
 - Body Measures: Obese ≥ 30 BMI
 - Lab Profiles: Metabolic Health
- NDI contributes mortality information:
 - Underlying and multiple causes of death through 2006



Examples of BIG DATA Projects

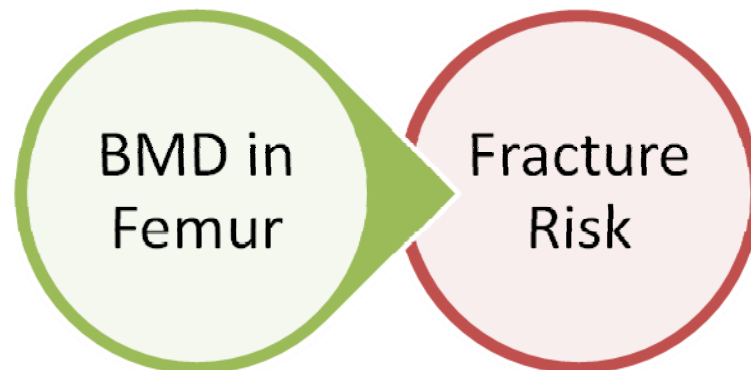
- Compared to healthy lean group
 - Risk of all-cause mortality was twice as high in Metabolic unhealthy obese group
 - No risk difference in metabolic healthy obese group
- Strengths
 - Nationally representative sample of the US population
 - Robust end point (all-cause mortality)
 - Relatively long follow-up period (~15 years)
 - Measured rather than self-reported weight and height
 - Data collection is standardized and use appropriate quality controls.

Examples of BIG DATA Projects

- Example 2: NHANES and Medicare Records

Examples of BIG DATA Projects

- Bone mineral density really associated with increase risk of fracture?
- NHANES contributes baseline information
 - Bone Mineral Density in Femur (DEXA) in 1988-1994
 - Age 65 and older
- Medicare contributes event information:
 - Incident fracture from Medicare Claims Files through 2007



Examples of BIG DATA Projects

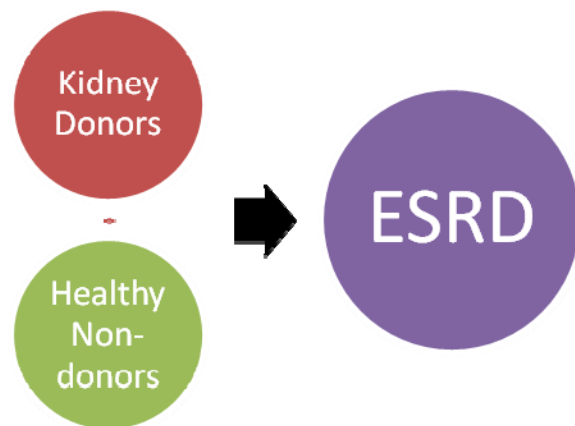
- Compared to healthy lean group
 - For each SD decrease in BMD: risk of fractured doubled
 - Age, sex, R/E was still associated with increase risk of fracture even after controlling for femur BMD
- Strengths
 - Nationally representative sample of the US population
 - Clinically diagnosed end point (ICD-9)
 - At least a decade-long follow-up period
 - Data collection was standardized and used appropriate quality controls.

Examples of BIG DATA Projects

- Example 3: NHANES, USRDS and CMS Data

Examples of BIG DATA Projects

- Greater risk of ESRD after live kidney donation?
- US Renal Data System
 - Registry of all kidney donations (~96,000 from 1994-2011)
 - Identify the SES, BMI, Smoking and Blood Pressure in Cases
- CMS information identifies ESRD
 - CMS Medical Evidence Form 2728—Certification of ESRD
- NHANES information creates healthy non-donor group:
 - Data 1988-1994 used to create population control
 - Replacement Matching use SES, BMI, Smoking and Blood Pressure information to build a control group of ~96,000 people



Examples of BIG DATA Projects

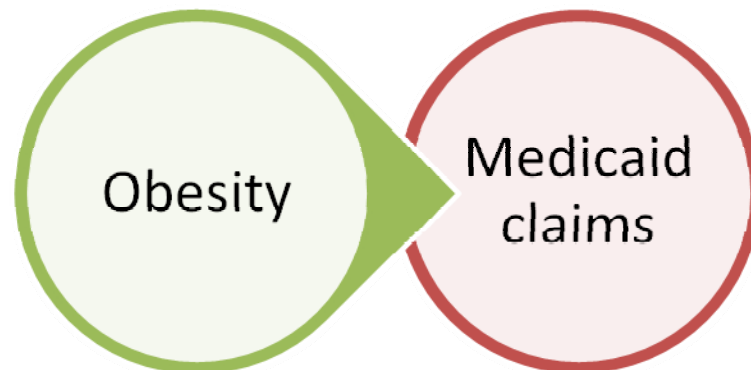
- Estimated Risk of ESRD at 15 Years of DOnation
 - 30.8 / 10,000 for donors
 - 3.9 / 10,000 for healthy non-donors
- Strengths
 - Inclusion of all kidney donors in 20 year period
 - Use of data from people not selected based on existing health condition to create healthy non-donor pool
 - Availability of clinically diagnosed end points
 - A ~15 year follow-up period
 - Data collection was standardized and used appropriate quality controls.

Examples of BIG DATA Projects

- Example 4: NHANES and Medicaid Claims Data

Examples of BIG DATA Projects

- What is the econometric cost of obesity to Medicaid?
- NHANES contributes baseline information
 - Age 20+ in 1999-2004
 - Body Measures: Obese ≥ 30 BMI
- Medicaid Analytic eXtract (MAX) data contributes:
 - Amount of FFS costs paid through Medicaid same year



Examples of BIG DATA Projects

- Average FFS costs paid by Medicaid in same year as obesity was determined during 1999-2004
 - \$95.02 for non-obese Medicaid receiptents
 - \$225.72 for obese Medicaid receiptants
- Strengths
 - Nationally representative sample of the US population
 - Measured rather than self-reported weight and height
 - Actual costs paid based on claims data
 - Data collection was standardized and used appropriate quality controls.

Informatics-BIG DATA-Oral Health Research

- NHANES and BIG DATA
 - Information from data becomes more valuable when the data is more reliable
 - With NHANES—we have Veracity
 - It's about observing better—identifying patterns in ways that make us better understand the world we live in
 - With Linkage—we obtain Volume, Variety and Velocity
 - Facilitating Data-driven decision-making

Informatics-BIG DATA-Oral Health Research

Thanks!



Informatics-BIG DATA-Oral Health Research

The screenshot shows the CDC website for the National Health and Nutrition Examination Survey (NHANES). The browser address bar shows <http://www.cdc.gov/nchs/nhanes.htm>. The page header includes the CDC logo and the text "Centers for Disease Control and Prevention" and "CDC 24/7: Saving Lives. Protecting People.™". A navigation bar contains an "A-Z Index" and letters A through Z. The main content area is titled "National Health and Nutrition Examination Survey". On the left, a sidebar menu lists various options, with "Questionnaires, Datasets and Related Documentation" circled in red and a red arrow pointing to it. The main content area includes a breadcrumb trail "NCHS Home > Surveys and Data Collection Systems", social media icons for Facebook, Twitter, and a plus sign, and a heading "National Health and Nutrition Examination Survey". Below this heading is a paragraph describing the survey: "The National Health and Nutrition Examination Survey (NHANES) is a program of studies designed to assess the health and nutritional status of adults and children in the United States. The survey is unique in that it combines interviews and physical examinations." There are also sections for "Selected Participants", "Information for Health Professionals", and "NHANES National Youth Fitness Survey". On the right side, there is a "What's New" section with sub-sections for "Publications" and "Data Release", each containing a list of recent items with links and dates.

CDC Home
Centers for Disease Control and Prevention
CDC 24/7: Saving Lives. Protecting People.™

A-Z Index A B C D E F G H I J K L M N O P Q R S T U V W X Y Z #

National Health and Nutrition Examination Survey

NCHS Home > Surveys and Data Collection Systems

Facebook Twitter +

National Health and Nutrition Examination Survey

The National Health and Nutrition Examination Survey (NHANES) is a program of studies designed to assess the health and nutritional status of adults and children in the United States. The survey is unique in that it combines interviews and physical examinations.

Selected Participants
Have you been selected to take part in the National Health and Nutrition Examination Survey?

Information for Health Professionals
Learn about participant involvement and benefits with the National Health and Nutrition Examination Survey

NHANES National Youth Fitness Survey
The NHANES National Youth Fitness Survey (NNYFS) was a one year survey conducted in 2012. The NNYFS collected nationally representative data on physical activity and fitness levels of children and adolescents in the

What's New

Publications

- Nut Consumption Among U.S. Youth, 2009-2012
- Winter 2015-16 Newsletter [PDF - 187 KB]

Data Release

- Taste & Smell (CSX_H 2013-2014)
- Apolipoprotein B (Subsample) (ApoB_H 2013-2014)
- Sleep Disorders (SLQ_H 2013-2014)
- Taste & Smell (CSQ_H 2013-2014)

[More »](#)

Informatics-BIG DATA-Oral Health Research

http://www.cdc.gov/nchs/nhanes/nhanes_questionnaires.htm

NHANES - Questionnaires, ...

File Edit View Favorites Tools Help

CDC Home
Centers for Disease Control and Prevention
CDC 24/7: Saving Lives. Protecting People.™

A-Z Index A B C D E F G H I J K L M N O P Q R S T U V W X Y Z #

National Health and Nutrition Examination Survey

NCHS Home > Surveys and Data Collection Systems > National Health and Nutrition Examination Survey

f t +

Questionnaires, Datasets, and Related Documentation

- Continuous NHANES Data, Questionnaires and Related Documentation
 - Search Continuous NHANES Variables
 - NHANES 2015-2016
 - NHANES 2013-2014
 - NHANES 2011-2012
 - NHANES 2009-2010
 - NHANES 2007-2008
 - NHANES 2005-2006
 - NHANES 2003-2004
 - NHANES 2001-2002
 - NHANES 1999-2000
- Prior to 1999 Data, Questionnaires, and Related Documentation
- Other NHANES Data
- Survey Methods

Informatics-BIG DATA-Oral Health Research

The screenshot shows a web browser window with the URL http://www.cdc.gov/nchs/nhanes/other_nhanes_data.htm. The page is titled "National Health and Nutrition Examination Survey" and features a sidebar with navigation links. The main content area includes a breadcrumb trail: "NCHS Home > Surveys and Data Collection Systems > National Health and Nutrition Examination Survey > Questionnaires, Datasets, and Related Documentation". Below this, there are social media icons for Facebook, Twitter, and a plus sign. The section "Other NHANES Data" contains a list of links: "Genetic Data Repository", "NHANES I Epidemiologic Followup Study", "NHANES Linked Data Files", and "NHANES National Youth Fitness Survey (NNYFS)". The link "NHANES Linked Data Files" is circled in red, and a red arrow points to it from the right.

CDC Home
Centers for Disease Control and Prevention
CDC 24/7: Saving Lives. Protecting People.™

A-Z Index [A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#) [X](#) [Y](#) [Z](#) <#>

National Health and Nutrition Examination Survey

NCHS Home > Surveys and Data Collection Systems > National Health and Nutrition Examination Survey
> Questionnaires, Datasets, and Related Documentation

[f](#) [t](#) [+](#)

Other NHANES Data

- Genetic Data Repository
- NHANES I Epidemiologic Followup Study
- **NHANES Linked Data Files**
- NHANES National Youth Fitness Survey (NNYFS)

National Health and Nutrition Examination Survey

About NHANES

What's New

Questionnaires, Datasets, and Related Documentation

Survey Methods and Analytic Guidelines

Search Variables

All Continuous NHANES

NHANES 2015-2016

NHANES 2013-2014

NHANES 2011-2012

NHANES 2009-2010

NHANES 2007-2008

Informatics-BIG DATA-Oral Health Research

http://www.cdc.gov/nchs/data_access/data_linkage_activities.htm

Data Access - Data Linkage ...

File Edit View Favorites Tools Help

History Data

- End Stage Renal Disease Data
- Housing and Urban Development
- Interactive Tables and Databases
- National Death Index
- Public-Use Data Files

Related Sites

- NCHS Data Visualization Gallery
- Research Data Center

Important Information

- Data Release Policy
- Data User Agreement

files enable researchers to examine the factors that influence disability, chronic disease, health care utilization, morbidity, and mortality. NCHS is currently linking various NCHS surveys with death certificate records from the [National Death Index \(NDI\)](#), enrollment and claims data from the [Centers for Medicare & Medicaid Services \(CMS\)](#), Retirement, Survivor, and Disability Insurance (RSDI) and Supplemental Security Income (SSI) benefit data from the [Social Security Administration \(SSA\)](#), End Stage Renal Disease (ESRD) data obtained from the [United States Renal Data System \(USRDS\)](#), and administrative records from the Department of Housing and Urban Development(HUD). [HUD-User website.](#)

Linked Data Resources by Topic Area.

- Mortality data
- Medicare Enrollment and Claims data (CMS)
- Medicaid/CHIP Enrollment and Claims data (CMS)
- Social Security Benefit History data (SSA)
- End Stage Renal Disease (USRDS)
- Administrative Records of housing assistance programs from the Department of Housing and Urban Development HUD

New

- Release of
- Hospitalizat Experience Fee-for-ser Over
- Linkage of to Administ CMS
- Characteris Managed 2003-2005
- Updated Pu Files with F Available

Upcoming

- Updated lin Medicaid S

NIDCR Training Opportunities Big Data Research

Residency in Dental Public Health and Oral Health Informatics (**Emerging**)

- Partnership with National Library of Medicine (NLM)
- For dentists with an interest in informatics science
- An MPH or equivalent public health-related degree required
- 2-Year Full-time Residency / Fellowship
- Mentors and Faculty come from NLM and NIDCR

NIDCR Training Opportunities Big Data Research

Residency in Dental Public Health and Oral Health Informatics (**Emerging**)

- Informatics Research Area Strength: **Public Health**
 - Focus on “Big Data” projects
 - Other Potential Research Areas: **Bioinformatics** and **Clinical Research**
- Certificate in DPH
- NIDCR Residency in DPH is CODA accredited
 - Meets education qualifications for graduates to challenge DPH certifying exam
- Certificate in Health Informatics
- Goal is to begin application process Summer 2016

NIDCR Training Opportunities Big Data Research

Questions about Residency in Dental Public Health and Oral Health Informatics:

- Bruce Dye
 - NIDCR Residency Program in Dental Public Health
 - bruce.dye@nih.gov