

The Numbers Behind the Maps

Addressing Data Challenges

- Good health mapping requires good data
- Government often has it, but it's hard to get
- Its especially difficult if it's data about individuals, and contains individual identifiers and/or discreet spatial information

Addressing Data Challenges

Why do we want Government data?

- Large, comprehensive data sets
 - Vital Records cover all Californians
 - Health Planning (OSHPD) data covers all CA hospitalizations
 - Medi-Cal covers a large segment of the population, and provides a reasonably comprehensive set of services

Addressing Data Challenges

Why do we want Government data?

- These types of data files assist in presenting
 - Snapshots of current health status covering large numbers of people
 - Trends in the health status of individuals over time, coupled with treatment received
 - Health status trends over time in a geographic region

Addressing Data Challenges

Some Examples

- Maternal Mortality Studies
- Low birth weight and premature births
- Pre-natal care
- Differences in post-hospitalization survival rates by region
- Differences in death rates by cause of death and by region

Addressing Data Challenges

Trends in health research

- Use comprehensive data sets, but focus on ever smaller geographic areas
 - “Think globally, act locally” concept.
 - Major issues are seen by looking broadly at a population
 - However, costs of responding are lowered if the response has a tight focus, and
 - Major changes in health are often effected at the individual level

Addressing Data Challenges

Some Examples of Targeting

- Asthma in children, by neighborhood
- Differential rates in Caesarean sections by geographic area and by hospital
- Violence
- Injuries, including auto accidents
- Health care costs
- Services to the elderly

Addressing Data Challenges

Why is it so difficult to get government data?

- Privacy and Confidentiality
 - Brief discussion of background
 - Freedom from embarrassment from unwanted disclosure
 - Protection from adverse effects stemming from health status release
 - Losses of job opportunities
 - Lost ability to obtain insurance
 - Social stigmas applied
 - Etc.

Addressing Data Challenges

Why is it so difficult to get government data?

- Improper uses of health research
 - Immoral “active” or “passive” experimentation with human subjects
 - Corporate policies that “red-line” geographic areas or classifications of people for insurance exclusion purposes
 - Exclusion of individuals or families from access to care

Addressing Data Challenges

Why is it so difficult to get government data?

- Historical problems have resulted in increasing protections for individuals
 - Most of these protections result in making access to the data more difficult
 - Most of these protections are encoded in law that specifically address rights to privacy, but contain terms difficult to apply concerning proper releases of data.
 - The safest route for a governmental data steward is to not release anything

Addressing Data Challenges

Why is it so difficult to get government data?

- Principal laws and processes governing access to CA health data
 - Health and Safety Code
 - Welfare and Institutions Code
 - The Committee for the Protection of Human Subjects
 - Departmental Privacy and Security Officers
 - HIPAA (brief overview, but some discussion of the “minimum data necessary” concept, especially when applied data field by data field)

Addressing Data Challenges

Why is it so difficult to get government data?

- **Governmental Program Administrators**
 - Who is ultimately responsible for data releases?
 - What benefits do they receive? (Few)
 - What risks do they take (Many)
 - What needs to happen? (Provide more benefits (psychic works), but essentially lower the personal risks) See final chapter for how

Addressing Data Challenges

Why is it so difficult to USE government data?

- Data used for research is often collected for a different purpose
 - Like billing insurers or paying provider bills
 - Or producing a required legislative report that no one reads any longer
- This results in obscure data fields, or in misunderstood data labels
- Program changes can cause undetected changes in the data when new collection standards are issued without changing the labels

Addressing Data Challenges

Why is it so difficult to USE government data?

- Doing the conceptual “crosswalk” from “Administrative Data” to health data for research is challenging

Addressing Data Challenges

Special cases in using government data

- Combining data sets from different collection systems
- Most often involves different Departments
- Often involves coordinating different legal codes
- May need legal staff involvement
 - (oh my)

Addressing Data Challenges

Special cases in using government data

- Nevertheless, combining data sets can add tremendous descriptive power
- Can add geographical mapping components to a data set without geo-coded fields
- But, it requires the presence of unique identifiers in each data set to be able to perform the match

Addressing Data Challenges

Special cases in using government data

- Releasing identifiers make governmental administrators very nervous
- Releasing identifiers make governmental administrators VERY nervous
- Releasing identifiers make governmental administrators VERY NERVOUS

Addressing Data Challenges

Why are governmental administrators nervous?

- Large penalties for inappropriate releases
 - Financial
 - Career
- Some things the administrator thinks about
 - What identifiers are present?
 - Can any of this data be matched with another data set to produce identifiers?
 - Is the population in a geographic area so sparse that an individual may be identified by seemingly innocent demographic data”

Addressing Data Challenges

Why are governmental administrators nervous?

- How do I reduce my risk?
- What is an “acceptable level of risk” to my Department or Agency?
- Is this documented to protect me?
- What does “due diligence” demand when reviewing this release?
- How can I properly “vet” the researcher and his/her institution?

Addressing Data Challenges

Why are governmental administrators nervous?

- How can I prevent a researcher from unauthorized activity, whether the activity is accidental or by design.
- This leads to a question: How can I handle this without releasing the data at all?

Addressing Data Challenges

Looking to the Future

- Government creation of matched data sets
 - Match with identifiers and de-identify before release
 - Means the researcher must rely on the government matching accuracy
- Create access without transfer
 - Allow transfer only of de-identified data

Addressing Data Challenges

Looking to the Future

- Standardize how researchers are “vetted”
 - Delineate the various “risks” and have a standardized scale to assess an overall score
- Standardize the rules that separate a confidential database from a public database
 - HIPAA goes a long way here, but it is sometimes difficult to apply the regulation

Addressing Data Challenges

Looking to the Future

- Standardize a technique that will render a confidential dataset into an acceptable “public” GIS dataset
 - Geo-locations would be displaced sufficiently so they cannot be used to locate or identify an individual
 - Geo-locations would not unduly compromise statistical accuracy

Addressing Data Challenges

Looking to the Future

- Ideally, this would be relatively sophisticated software, but easy to use
- Would be commonly available to all governmental agencies at acceptable cost
- Would have published information for researchers to use to fully understand the level of precision afforded
- Would result in a HIPAA “safe harbor” designation