



# Report of the Expert Panel on Public Use Data Access and Disclosure Control

---

Jennifer Madans, PhD, Chair  
Associate Director for Science  
National Center for Health Statistics, CDC

National Children's Study Federal Advisory Committee Meeting  
April 22–23, 2008



# Expert Panel

---

'This expert panel was convened to share their experiences and expertise, and to offer recommendations on how to provide as much **access** as possible to investigators while insuring the **confidentiality** of National Children's Study participants.'

Meeting: October 11, 2007



# Expert Panel

---

Jennifer Madans, NCHS (Panel chair)

Myron Gutmann, University of Michigan

Marilyn Seastrom, NCES

Paul Sorlie, NHLBI

Alan Zaslavsky, Harvard University

Peter Scheidt, NICHD, Study Program Office

Sarah Knox, NICHD, Study Program Office

Alexa Fraser, Westat, Study Coordinating Center

Marsha Hasson, Westat, Study Coordinating Center

Nancy Weinfield, Westat, Study Coordinating Center



## Expert Panel

---

- Prior to the meeting, panel members reviewed materials provided on the Study design and plans for data access
- One day meeting to discuss plans for data access developed to date and to provide recommendations for future work
- There was a high degree of consensus regarding the Panel's recommendations



# Terminology

---

## ○ **Disclosure control**

- Extends beyond access to data and must be integrated into all study activities
- The Panel focused primarily on disclosure control as it relates to making study data available but did touch on disclosure control in other survey activities and on access to specimens
- The Panel focused on confidential data; not public use files



# Terminology

---

## ○ **Access vs. release**

- The difference is a function of where the control over the data resides
- When data are released, the stewards no longer can control the data (affects ability to fulfill promises made in the informed consent)
- Access can be provided in ways in which the steward retains control



# Terminology

---

- **Public Use files vs. use by the public**
  - Public use files -- files that do not contain confidential data which are released for use without any restriction
  - Data that are made available to the public (i.e., those who have no connection to the Study) – can contain confidential data or can be public use files
- **Confidential data vs. sensitive data**
  - Confidential data -- data that do or can identify a subject
  - Sensitive data -- refers to the content of the information
  - Not all sensitive data are confidential but since there are no absolute confidentiality protections, stewards tend to be more careful in protecting the sensitive data



# General Guidance from Panel

---

- The access plan must be an integral part of the Study design and developed early in the process – not something done 'after'
- The more complex the study, the more complex the access plan will be
- Access Issues will be:
  - More complex than expected
  - More expensive and staff intensive than expected
- The access plan should incorporate established best practices





# Components of an Access Plan

---

- **Principles** governing data access and protection of confidentiality – e.g., Is strict confidentiality protection a priority?
- **Policies**
  - Policies are contained in Study documents and in the informed consent materials
    - Who should have access to potentially identifiable data, what data should they have access to and under what conditions?
    - When will data be released/available for access?
- **Practices**
  - Mechanisms by which policy is operationalized



## Link to Informed Consent

---

- There is a particularly important link between the access plan and the informed consent process
  - Unless there is a clear legal or regulatory requirement for protecting confidentiality – the informed consent will govern data access and release
  - The informed consent provides the justification for the data access and release policy



# Link to Informed Consent

---

- **Link between the access plan and the informed consent process**
  - The informed consent should specify who has access to what and under what conditions ('insiders' and 'outsiders')
  - The wording of the informed consent is crucial – the panel repeatedly asked about the exact language in the consent materials
  - Access and release practices must be developed to meet the requirements of the informed consent



# Complications

---

- Access plans need to be developed up front **but** are living documents
  - Affected by changes in the environment
  - Affected by other technological changes
- Confidentiality risk is cumulative
  - All new releases must be considered in reference to what has already been released
- Identifiability risk increases when
  - Study sites are known
  - Sample selection criterion are based on public information (e.g., year of birth)
  - Data are from multiple sources
  - Data collection is longitudinal
  - The number of investigators is large



# Policy Issues That need to be Addressed

---

- **How will different study partners\* be treated in the Access Plan?**
  - 'Insiders' usually have full access to confidential data
    - they are closely bound to the study and the commitment to protect confidentiality
    - A subset need access to identifiable data to do their jobs
  - Need to clearly specify the nature of the access that 'outsiders' will have
    - Can vary by characteristics of the 'outsider'

\*Study Center Staff, CC staff, Study Investigators, Adjunct Study Investigators, Scientific Community; General Public



## Examples of Policy/Practice Questions Raised by the Panel

---

- How should Adjunct Study Investigators be treated?
  - Are they 'insiders'?
  - How will their role be explained in the informed consent?
  - What kind of agreements need to be developed to assure that confidentiality will be preserved? What kind of oversight arrangements are needed? Who pays the costs of maintaining the access system?



## Examples of Policy/Practice Questions Raised by the Panel

---

- Is there a distinction between the scientific community and the general public in terms of access rights? What criteria are used to differentiate between the two or within each group?
- What mechanisms should be used to provide access to the scientific community/general public? What confidentiality oversight is needed? How will use be monitored? How will users be tracked? Who will pay the costs?



# Available Access Mechanism

---

- Access mechanisms are available that offer various levels of confidentiality protection (from less to more)
  - Data Use agreements
  - Licensing
  - Remote access systems (users can access confidential data but cannot see it)
  - Data Centers
- Within each mechanisms, confidential data should be made available on a need to know basis – even for insiders\*\*\*\*





# Major Recommendations

---

- **Establish a Data Access Committee to advise the Data Steward**
  - The Data Access Committee will need to be a standing committee, as issues related to access will be both ongoing and dynamic.
  - The Data Access Committee will be needed to make evolving decisions both about who the data users are and what they will have access to.
  - The Data Access Committee should develop an access plan that addresses policies and practices (as noted above) as its first task.



# Major Recommendations

---

- **Establish a Disclosure Review Board**

- That is independent of the Publications Subcommittee
- To Oversee the actual work of disclosure control, i.e., develop strategies to carry out disclosure control and implement those strategies.
  - The Disclosure Review Board should not set policy – the setting of policy should be separated from implementation
- The Disclosure Review Board may need to review every publication for disclosure depending on access mechanism used/informed consent