NCIPC Guidelines for

Implementing the CDC Policy on Releasing and Sharing Data for Contracts

Contents

I. Purpose

II. Data Covered by These Guidelines

III. Data Not Covered by These Guidelines

IV. Considerations for Data Release and Sharing

V. Minimal Data Quality Standards for Release

VI. Minimal Data Documentation Standards for Release

VII. Data Release Responsibilities

VIII. Timelines for Release or Sharing of Data Collected with NCIPC Funds

IX. Procedures for Research and Surveillance Contracts

X. Composition and Roles of NCIPC Data Release Work Group

XI. Timelines for Implementing and Revising NCIPC Data Release Guidelines

Appendix A: Sample Language for Consent Forms

Appendix B: Responsibilities for Release or Sharing of Data (NCIPC Research and Surveillance Contracts)

Glossary

References

I. Purpose

On April 16, 2003, the Centers for Disease Control and Prevention (CDC) Policy on Releasing and Sharing Data went into effect.[1] Each CDC center, institute, and office (CIO), including the National Center for Injury Prevention and Control (NCIPC), is responsible for developing specific implementation procedures for its staff to follow. Because issues related to data release can vary from project to project, NCIPC is instituting plans for determining or arbitrating specific data release procedures for each project.

The overall purpose of NCIPC's data release and sharing implementation guidelines is to ensure that the goals of the CDC/ATSDR Policy on Releasing and Sharing Data are met. Those goals are (1) to routinely provide data to partners for appropriate public health purposes and (2) to release and share all data as soon as is feasible without compromising privacy concerns, federal and state confidentiality concerns, proprietary interests, national security interests, or law enforcement activities. These guidelines articulate the administrative and scientific procedures for NCIPC scientists and staff to follow when disseminating data or responding to requests to release or share data.

II. Data Covered by these Guidelines

These guidelines apply to new surveillance or research projects that total $500,000 in direct and indirect costs per year in which data are collected 90 days or more after April 16, 2003. This includes new projects conducted through grants, cooperative agreements, or contracts. These guidelines also cover all existing surveillance data sets for which NCIPC is the current data steward.

All data should be released as soon as is feasible without compromising privacy concerns, federal and state confidentiality concerns, proprietary interests, national security interests, or law enforcement activities. Release of data during or related to public health emergencies also involve special considerations that must be addressed on a case-by-case basis.

Guidelines cover:

- Data collected by CDC using federal resources;
- Data collected for CDC by other agencies or organizations (through procurement mechanisms such as grants, contracts, or cooperative agreements); and
- Data reported to CDC (e.g., by a state health department).[2]

Applicable definitions:

NCIPC personnel: NCIPC employees, fellows, visiting scientists, and others (e.g., on-site and off-site contractors) who are involved in designing, collecting, analyzing, reporting, or interpreting data for, in collaboration with, or on behalf of NCIPC.

Data: Scientific records which are as accurate and complete as possible.

Data release*:* Dissemination of data either for public use or through an ad hoc request that results in the data steward no longer controlling the data.

Data sharing: The process of granting certain individuals or organizations access to data that contain individually identifiable information. This access is granted with the understanding that identifiable or potentially identifiable data cannot be re-released further unless a special data-sharing agreement that governs the use and re-release of such data is agreed upon by NCIPC and the data providers.

Data Release Work Group: A work group comprising the NCIPC Associate Director for Science (ADS), the Division ADS, the director of the Office of Statistics and Programming, and the Chief Technology Officer, or their designees. This work group will oversee implementation of these guidelines and address related questions or concerns.

III. Data Not Covered by These Guidelines

These guidelines **do not** cover data shared with NCIPC but owned by other organizations (e.g., data provided to NCIPC by managed care organizations, preferred provider organizations, or technology firms for a specific research project). Such data may be covered by other policies or procedures that reflect pertinent laws, regulations, and agreements—such as the Freedom of Information Act (FOIA).

In addition, surveillance and research projects that began data collection prior to the effective date of these guidelines, April 16, 2003, are not required to comply with the data sharing and release requirements herein. However, in the event of a specific request for such data, it is strongly advised that they be shared or released to the extent possible without compromising confidentiality or other concerns.

IV. Considerations for Data Release and Sharing

There are a number of aspects to consider when determining (1) whether to release research or surveillance data and (2) how these data should be made available to data users.

Who should have access to the data? In general, NCIPC should make research and surveillance data available to the general public unless there are specific reasons not to, such as data limitations or concerns about invading the privacy of participants.

When should access to the data be limited? Some data sets (or data elements within a data set) may require restricted access because the data are of poor quality or are associated with formal confidentiality or nondisclosure protections.

Are the data and their documentation of appropriate quality for release? A major consideration is documentation. Good documentation regarding the study design, sampling scheme, or surveillance system attributes is essential for the data user to understand how to analyze the data. Additional quality considerations are as follows:

- ***Data quality*** (e.g., how complete is case ascertainment and are the data complete and accurate?);
- ***Representativeness*** (e.g., are the data representative of the population of interest?);
- ***Sensitivity*** (e.g., how well do the data measure the outcomes of interest?);
- ***Timeliness*** (e.g., how old are the data?); and
- ***Confidentiality*** and ***nondisclosure*** issues (e.g., what is the risk of linking to other data sets that have personal identifiers? is it possible for a research participant's identity to be obtained using the data?).

For further detail, please see Section V. Minimal Data Quality Standards for Release *and* Section VI. Minimal Data Documentation Standards for Release.

In what format and manner should data be released? There are various methods for releasing research or surveillance data. These include public-use data files, restricted-access data files (i.e., only eligible users are allowed to access the data), summary data files, a web-based interactive query system (e.g., WISQARS), summary tables (e.g., data grouped and reported by age or gender), fact sheets and publications. Some data may require the use of several of these methods to reach appropriate users. For instance, if the raw data cannot be released to the public due to confidentiality and nondisclosure restrictions, then it may be necessary to provide summary data to the public.

Are data-use agreements necessary? In some cases, it may be necessary to require a signature on a data-use agreement before providing users with access to the data by users. This can help protect against misuse of the data and the risk that the confidentiality of research participants will be compromised.

What are the strengths and limitations of the data? Every data set has its strengths and limitations in terms of how the data should be analyzed, reported, interpreted, and used. When data are released, users should also be given information on the intention of the research, considerations for data analysis, and other aspects about the research or surveillance system that are relevant to interpretation of findings and their application to prevention program planning, implementation, and evaluation.

V. Minimal Data Quality Standards for Release

All data that are released should meet certain minimum data quality standards. If the minimum standards are not met, careful consideration should be given as to whether or not the data should be released for data analysis by users other than NCIPC staff, grantees, or contractors. These standards are specified below:

Completeness of case ascertainment. High non-response rates or attrition in studies and surveillance systems can lead to findings that misrepresent the population of inference. In some cases, non-response and attrition can be adjusted for by appropriate statistical methods. But as a general rule, data with low response rates or high attrition rates should not be released. Exceptions can be made in instances where response rates are low or attrition rates are high due to the sensitive or challenging nature of the topic area (e.g., intimate partner violence, child maltreatment). Exceptions will also be considered in cases where data are of poor quality yet represent a unique information resource. These exceptions should be reviewed and considered by the NCIPC Data Release Work Group.

Completeness, validity, and consistency of data. Prior to their release, data should be fully edited for range and validity of data entries, duplicate records, and consistency across data elements. Quality assurance measures taken during data collection should be documented and this information should be made available to data users. Careful consideration should be given before releasing data elements that have not been fully edited or that have a high percentage of cases with missing data. Sometimes, only a partial research data set or surveillance data set should be released.

Description of original data element and created analysis variables. An important step in preparing research or surveillance data for release is to prepare the analysis file for ease of use. It may be worthwhile to create analysis variables to help data users select cases of interest and to define demographic groupings, such as age or race and ethnicity groupings. Created analysis variables should be distinguishable from

the original data elements that were collected (e.g., by being labeled with a "_C" at the end of the variable name). This gives data users the option of using the created analysis variables, verifying the created variable, or using the original variables to create their own analysis variables.

Assurance of confidentiality and nondisclosure. The data steward should carefully scrutinize the data set that is to be released to ensure the confidentiality of personal and health-related information and to avoid disclosure of participant identity in a research project or a surveillance system. Any data elements that could potentially be used to identify a participant should be removed from the data set prior to release.

VI. Minimal Data Documentation Standards for Release

Documentation should accompany the release of research and surveillance data sets. The purpose of such documentation is to provide enough details so that the data user will not analyze, use, or report information from the database in an inappropriate manner.

This documentation should include contact information for a designee that can answer questions regarding the data set as well as information on the research design or surveillance system attributes, essential considerations for data analysis, strengths and limitations of data, confidentiality restrictions, methods for calculating variance estimates (if applicable), and data file format and other characteristics. Published and unpublished references (e.g., peer-reviewed articles, websites, data collection and coding manuals, training modules) and summary tables of frequency distributions for all variables in the data set should also be provided.

NCIPC staff recommend providing the following documentation:

1. Name, address, phone number, and e-mail address of contact person;
2. Data access procedures and other pertinent documentation;
3. Data-use agreement (if applicable);
4. Background on study design;
5. Period of the study;
6. Sampling methods;
7. Data collection procedures, protocols, and coding guidelines;
8. Description of created analysis variables and how they were constructed;
9. Limitations (or caveats) and strengths of the data;
10. Confidentiality restrictions (e.g., suppressed data, omitted records);
11. Data dictionary (i.e., table of contents, codebook of data elements, labels, data values, sample weights, coding conventions, record layout, and field lengths);
12. Data set names and attributes (e.g., number of observations, number of variables, record length) and available formats (e.g., ASCII, SAS, DBF, SPSS);
13. Statistical procedures relevant to data analysis (e.g., examples of computer programs for calculating standard errors of estimates);
14. References; and
15. Summary tables of all variables with variable names, labels, and frequency counts.

Other information about the research or surveillance system. Unpublished documents (e.g., study design, data collection and coding manuals, data collection instruments, training modules) should also be made available to data users. These background documents will need to go through a clearance process similar to NCIPC Institutional Review Board (IRB) clearance for release of data from contracts and cooperative agreements. Unpublished documents should be well organized, clearly written,

and self-explanatory. These documents can be made available as pdf and html files at the data-release website or at other designated locations.

VII. Data Release Responsibilities

A data steward must be assigned for each research or surveillance project that is funded or conducted by NCIPC. This steward has overall responsibility for meeting CDC data release and sharing requirements for that project's data set(s). Usually the grantee serves as the data steward for data sets resulting from grants and cooperative agreements. For cooperative agreements, NCIPC staff may participate in decisions related to data release or sharing. For research contracts, the NCIPC lead scientist or technical monitor will usually serve as the data steward in collaboration with the project officer. Appendix B, Responsibilities for Release or Sharing of Data, reviews responsibility for tasks that should be completed prior to releasing or sharing data sets.

Investigators are expected to provide information on the availability of data sets on their institution's website. Investigators may also choose to advertise availability of data sets through newsletters and other communication channels.

VIII. Timelines for Release or Sharing of Data Collected with NCIPC Funds

NCIPC recognizes that investigators on research and surveillance projects have a legitimate interest in benefiting from their investment of time and effort in designing the research and collecting data. Accordingly, NCIPC policies aim to balance the first and continuing use of original data by the original investigators with the need to share data for broader public health use.

Data and pertinent documentation must be released through a public-use data set or shared with restrictions after the data have been evaluated for quality and shared with any partners in the original data collection effort. In general, release should occur as soon as possible after acceptance of the main findings for publication. If data from epidemiologic, surveillance, or longitudinal studies are collected across multiple time periods, data from each collection phase must be released or shared following the timeline noted above. Requests for exceptions from these timelines should be addressed to the division ADS.

IX. Procedures for Research and Surveillance Contracts

When proposing research and surveillance contracts, NCIPC staff must provide the following information to their division ADS or, for offices, the office director:

1. A description of the proposed data release or sharing plan;
2. Consent forms that address data release or sharing plans;
3. Justification for not sharing or releasing data, as appropriate;
4. Timelines;
5. The name of the project data steward;
6. Plans for assessing data quality (see *Section V. Minimal Data Quality Standards for Release*); and
7. Plans for assuring proper documentation (see *Section VI. Minimal Data Documentation Standards for Release*).

These plans should be submitted at the same time and using the same clearance chain as research protocols submitted for IRB approval. Forms will be provided by the NCIPC ADS

Office. Plans for data release or sharing must be addressed in all consent forms sent to the IRB for approval. The Division ADS, or for offices, the Office Director, must approve and clear data release or sharing plans for research and surveillance contracts before data collection can begin. Documentation of approved data release or sharing plans will be maintained by the ADS office with final clearance authority.

The Division ADS will monitor all ongoing research contracts annually to ensure compliance with these guidelines for data release and sharing. Lead scientists are expected to provide updates on implementation of data release and sharing plans at the time they submit IRB continuation reports for each research project. Forms will be provided by the ADS office. At a minimum, yearly reports should include updates on plan implementation, documentation of quality checks, and venues used for data release and archiving. A format for data sharing and release implementation plans will be forthcoming. Documentation of approved data release or sharing plans will be maintained by the ADS office.

X. Composition and Roles of NCIPC Data Release Work Group

An NCIPC Data Release Work Group has been formed to assist in implementation of these guidelines. The work group will address questions and concerns as well as requests for exemptions. Any such requests will be addressed to the assigned project officer.

XI. Timelines for Implementing and Revising NCIPC Data Release Guidelines

Beginning in FY04, all funding announcements for research and surveillance projects with at least $500,000 in direct and indirect costs (regardless of funding mechanism) will require the submission and approval of a data release plan before projects are funded.

Appendix A. Sample Language for Consent Forms

The intent to share or release data must also be addressed in any documents generated for providing informed consent in research or other human subjects activities. The following is proposed language to add to the confidentiality section of a consent document or verbal script.

> "Research or other data collected by NCIPC is routinely shared or released to partners or the general public in a reasonably timely fashion. Before any data are shared, all personal identifying information will be removed. Parties obtaining data that contain sensitive or potentially identifiable information may be required to sign a special agreement to further ensure confidentiality of the data."

Appendix B: Responsibilities for Release or Sharing of Data

NCIPC Research and Surveillance Contracts

Tasks	Responsible Party
1. Manage Process/Data Steward	Contractor/Lead Division/Office
2. Determine Type of Release of Data	Lead Division/Office
3. Request Exemptions	Lead Division/Office
4. Determine Where Data Set will be Housed	Lead Division/Office
5. Review and Approval of Data Sharing Plan	Division ADS/Office Director *
6. Assume Cost for Preparing Data Set	**
7. Perform Quality Checks	**
8. Prepare Documentation	**
9. Review and Clearance of Documentation	Division ADS/Office Director *
10. Prepare Data Set	**
11. Advertise Availability of Data Set	**
12. Track Publications and Uses of Data ***	**
13. Monitor Compliance with CDC/NCIPC Policies	Project Officer/Technical Monitor/Lead Scientist

* At the discretion of the Project Officer or Division ADS, may also be reviewed by the NCIPC Data Release Work Group
** Final responsibility dependent upon contract language
*** Discretionary activity

Glossary

Archiving data:	Storage of data for secondary use, with or without an expiration date and with or without the informed consent of those who provided private or confidential information.
Confidentiality:	The expectation that information entrusted to CDC will not be divulged to others in ways that are inconsistent with the conditions agreed to when the information was originally disclosed.
Controlled access:	A system that allows researchers restricted use of data that cannot be released either to the public or under special-use agreements. Under this system, researchers can use the data but cannot take possession of it.
Data:	Scientific records which are as accurate and complete as possible.
Data release:	Dissemination of data either for public use or through an ad hoc request that results in the data steward no longer controlling the data. This does not include the release of data under FOIA.
Data sharing:	The process of granting certain individuals or organizations access to data that contain individually identifiable information with the understanding that identifiable or potentially identifiable data cannot be re-released further unless a special data-sharing agreement that governs the use and re-release of the data is agreed upon by CDC and the data providers.
Data steward:	The CDC employee, grantee, or contractor responsible for explaining CDC's data policy to staff and users, developing and maintaining data systems, evaluating and approving requests for access to data, and monitoring compliance with CDC policy.
Disclosure control:	Procedures used to limit the risk that information about an individual will be disclosed. These procedures may be administrative (e.g., granting authorized access), technical (e.g., setting up passwords), or statistical (e.g., cell suppression, aggregation, perturbation, and top- or bottom-coding). Statistical procedures are most commonly used to prepare a public-use data set or a data set that is linkable to another released data set.

Identifiable data:	Data which can be used to establish individual identity, either “directly,” using items such as name, address, or unique identifying number or “indirectly” by linking data about a case or individual with other information that uniquely identifies them.
Microdata:	Data files or records on an individual person or facility.
Privacy rights:	The right of people to withhold information about themselves from the knowledge of others.
Proprietary:	Produced or collected in such a way that exclusive rights may apply.
Public health emergency:	The occurrence or imminent threat of an adverse health event caused by epidemic or pandemic disease, infectious agent, biologic or chemical toxin, environmental disaster, or any agent that poses a real and substantial risk for a significant number of human fatalities or cases of permanent or long-term disability.
Public-use data:	Data that are available to anyone.
Restricted data:	Data that are shared only in a limited way because greater dissemination could have a negative effect, for example on national security.
Security:	Any mechanisms (administrative, technical, or physical) by which privacy and confidentiality policies are set up in computer or telecommunications systems.

References

1. Centers for Disease Control and Prevention Policy on Releasing and Sharing Data. Available at: URL: www.cdc.gov/od/ads/pol-385.htm.

2. CDC Council of State and Territorial Epidemiologists (CSTE) Intergovernmental Data Release Guidelines Working Group Report. CDC-ATSDR-CSTE Data Release Guidelines for Re-release of State Data. Draft Report, June 2002.