STATE OF WASHINGTON

DEPARTMENT OF HEALTH

*P.O. Box 47811*
*Olympia, Washington 98504*

Hello,

This cover letter is meant to assist in completing a Data Sharing Agreement (DSA) between the Washington State Department of Health (DOH) and your agency. The DSA documents the conditions under which DOH shares confidential information or limited dataset(s) with other entities. Please read the entire DSA and share with other individuals who will use this data to become familiar with the data use policies.

**Two copies of the DSA should be completed.**

More specific guidelines for completing the DSA are addressed below by page number:

Page 1

- Insert your agency/organization name in the header where specified.
- Complete the table with your organization name (if CDC please specify which division).
- The business contact should be a person at your agency who will be the data steward for your agency and interact with the data most frequently.

Page 10

- Please leave the period of performance blank. DOH will complete upon receipt of the DSA. The start date will be the date of DSA receipt and will be effective for the next three years from that date.

Page 12

- The business contact should fill out the 'Information Recipient' section and provide a wet signature (as opposed to a digital signature).

Page 13: Exhibit 1

- This exhibit should be completed in full by the business contact, except for the period of performance. Please leave the period of performance blank. DOH will complete upon receipt of the DSA. The start date will be the date of DSA receipt and will be effective for the next three years from that date.

Page 18

- The business contact should fill out the 'Information Recipient' section and provide a wet signature (as opposed to a digital signature).
- Please note that copies of all papers, presentations, reports, or publications developed using data obtained under this agreement should be sent to the attention of: the Rapid Health Information NetwOrk (RHINO) program at syndromic.surveillance@doh.wa.gov.

Pages 19 – 21: Confidentiality Agreement

- The confidentiality agreement will need to be completed by the business contact as well as all individuals who will access the data. A wet signature should be provided (as opposed to a digital signature).

Page 26: Appendix C Certification of Data Disposition

- Only fill out this appendix if/when the people and organization under this DSA decide to no longer use the data. This assures that no data will be kept or used outside of the DSA and must be destroyed. At the end of the DSA's period of performance, should it not be renewed, please send this completed appendix to the following DOH Office Manager's address:

  Communicable Disease Epidemiology
  c/o Leslie Byerly
  1610 NE 150th St
  MS K17-9
  Shoreline, WA 98155

Pages 27 - 43: Appendices D & E

- Please review the small numbers publishing guidelines (Appendix D) and the RHINO data limitations and best practices for data use (Appendix E) with all staff who will access this data.

Once the two DSAs have been completed, they should be mailed to the following DOH Office Manager's address:

Communicable Disease Epidemiology
c/o Leslie Byerly
1610 NE 150th St
MS K17-9
Shoreline, WA 98155

The two DSAs will be signed by DOH and one will then be mailed back to your agency for your records.

In addition to the DSA, which applies to an organization, each individual who wishes to access the data will need to complete (1) a Data Request Form (https://www.doh.wa.gov/Portals/1/Documents/5230/420-254-RHINODataRequestForm.docx) and (2) a Confidentiality Agreement (https://www.doh.wa.gov/Portals/1/Documents/5230/420-252-ConfidentialityAgreement.pdf, also Appendix A of the DSA). The Data Request Form can be emailed to Syndromic.Surveillance@doh.wa.gov in the original .docx format so that the bottom portion can be completed by DOH staff. The Confidentiality Agreement requires a wet signature and can also be emailed to Syndromic.Surveillance@doh.wa.gov.

Please direct any questions you have in completing the DSA or accompanying documents to Syndromic.Surveillance@doh.wa.gov. Thank you!

For persons with disabilities, this document is available on request in other formats. To submit a request, please call 1-800-525-0127 (TDD/TTY call 711).

DOH 420-253
April 2020

# DATA SHARING AGREEMENT
## FOR
## CONFIDENTIAL INFORMATION OR LIMITED DATASET(S)
## BETWEEN
## STATE OF WASHINGTON
## DEPARTMENT OF HEALTH
## AND
## (INSERT AGENCY/ORGANIZATION NAME)

This Agreement documents the conditions under which the Washington State Department of Health shares confidential information or limited Dataset(s) with other entities.

**CONTACT INFORMATION FOR ENTITIES RECEIVING AND PROVIDING INFORMATION**

| | **INFORMATION RECIPIENT** | **INFORMATION PROVIDER** |
|---|---|---|
| Organization Name | | Washington State Department of Health (DOH) |
| **Business Contact Name** | | Kevin Wickersham |
| Title | | PHEP Surveillance Program Manager |
| Address | | 1610 NE 150th St. MS: K17-9<br>Shoreline, WA 98155-9701 |
| Telephone # | | 206-450-9827 |
| Email Address | | kevin.wickersham@doh.wa.gov |
| | | |
| **IT Security Contact** | | Tracy Auldredge |
| Title | | DOH Chief Information Security Officer |
| Address | | PO Box 47890<br>Olympia, WA 98504-7890 |
| Telephone # | | 360-236-4432 |
| Email Address | | security@doh.wa.gov |
| | | |
| **Privacy Contact Name** | | Judith Morton |
| Title | | Acting Privacy Officer |
| Address | | P. O. Box 47890<br>Olympia, WA 98504-7890 |
| Telephone # | | (360) 236-4221 |
| Email Address | | privacy.officer@doh.wa.gov |

## DEFINITIONS

**Authorized user** means a recipient's employees, agents, assigns, representatives, independent contractors, or other persons or entities authorized by the data recipient to access, use or disclose information through this agreement.

**Authorized user agreement** means the confidentiality agreement a recipient requires each of its Authorized Users to sign prior to gaining access to Public Health Information.

**Breach of confidentiality** means unauthorized access, use or disclosure of information received under this agreement. Disclosure may be oral or written, in any form or medium.

**Breach of security** means an action (either intentional or unintentional) that bypasses security controls or violates security policies, practices, or procedures.

**Confidential information** means information that is protected from public disclosure by law. There are many state and federal laws that make different kinds of information confidential. In Washington State, the two most common are the Public Records Act RCW 42.56, and the Healthcare Information Act, RCW 70.02.

**Data provider** means any individual or entity that provides data to the RHINO program. This includes all participating hospitals, clinics, and providers.

**Data storage** means electronic media with information recorded on it, such as CDs/DVDs, computers and similar devices.

**Data transmission** means the process of transferring information across a network from a sender (or source), to one or more destinations.

**Direct identifier** Direct identifiers in research data or records include names; postal address information (other than town or city, state and zip code); telephone numbers, fax numbers, e-mail addresses; social security numbers; medical record numbers; health plan beneficiary numbers; account numbers; certificate/license numbers; vehicle identifiers and serial numbers, including license plant numbers; device identifiers and serial numbers; web universal resource locators (URLs); internet protocol (IP) address numbers; biometric identifiers, including finger and voice prints; and full face photographic and any comparable images.

**Disclosure** means to permit access to or release, transfer, or other communication of confidential information by any means including oral, written, or electronic means to any party except the party identified or the party that provided or created the record.

**Encryption** means the use of algorithms to encode data making it impossible to read without a specific piece of information, which is commonly referred to as a "key." Depending on the type

of information shared, encryption may be required during data transmissions, and/or data storage.

**ESSENCE** means the CDC National Syndromic Surveillance Program (NSSP) Electronic Surveillance System for the Early Notification of Community-based Epidemics (ESSENCE) platform. ESSENCE is a CDC-hosted platform which authorized users access through a web browser interface. ESSENCE contains syndromic surveillance data from Washington and other participating states, and includes analytical tools with which authorized users may interact with the data.

**Health care information** means any information, whether oral or recorded in any form or medium, which identifies or can readily be associated with the identity of a patient and directly relates to the patient's health care…." RCW 70.02.010(7)

**Health information** is any information that pertains to health behaviors, human exposure to environmental contaminants, health status, and health care. Health information includes health care information as defined by RCW 70.02.010 and health related data as defined in RCW 43.70.050.

**Health Information Exchange (HIE)** means the statewide hub which provides technical services to support the secure exchange of health information between HIE participants.

**Human research review** is the process used by institutions that conduct human subject research to ensure that:

- the rights and welfare of human subjects are adequately protected;
- the risks to human subjects are minimized, are not unreasonable, and are outweighed by the potential benefits to them or by the knowledge gained; and
- the proposed study design and methods are adequate and appropriate in light of the stated research objectives.

Research which involves human subjects or their identifiable personal records should be reviewed and approved by an institutional review board (IRB) per requirements in federal and state laws and regulations and state agency policies.

**Human subjects research**; **human subject** means a living individual about whom an investigator (whether professional or student) conducting research obtains (1) data through intervention or interaction with the individual, or (2) identifiable private information.

**Identifiable data or records** contains information that reveals or can likely associate the identity of the person or persons to whom the data or records pertain. Research data or records with direct identifiers removed, but which retain indirect identifiers, are still considered identifiable.

**Indirect identifiers** are indirect identifiers in research data or records that include all geographic identifiers smaller than a state, including street address, city, county, precinct, Zip Code, and their equivalent postal codes, except for the initial three digits of a ZIP code; all elements of dates (except year) for dates directly related to an individual, including birth date, admission date, discharge date, date of death; and all ages over 89 and all elements of dates (including year) indicative of such age, except that such age and elements may be aggregated into a single category of age 90 or older.

**Limited dataset** means a data file that includes potentially identifiable information. A limited dataset does not contain direct identifiers.

**Potentially identifiable information** means information that includes indirect identifiers which may permit linking an individual to that person's health care information. Examples of potentially identifiable information include:

- birth dates;
- admission, treatment or diagnosis dates;
- healthcare facility codes;
- other data elements that may identify an individual. These vary depending on factors such as the geographical location and the rarity of a person's health condition, age, or other characteristic.

**Publishing** means to release or make information available to the public.

**Restricted confidential information** means confidential information where especially strict handling requirements are dictated by statutes, rules, regulations or contractual agreements. Violations may result in enhanced legal sanctions.

**State holidays** Days of the week excluding weekends and state holidays; namely, New Year's Day, Martin Luther King Jr. Day, President's Day, Memorial Day, Labor Day, Independence Day, Veterans' Day, Thanksgiving day, the day after Thanksgiving day, and Christmas. Note: When January 1, July 4, November 11 or December 25 falls on Saturday, the preceding Friday is observed as the legal holiday. If these days fall on Sunday, the following Monday is the observed holiday.

**Normal business hours** are state business hours Monday through Friday from 8:00 a.m. to 5:00 p.m. except state holidays.

## GENERAL TERMS AND CONDITIONS

### I. USE OF INFORMATION

The Information Recipient agrees to strictly limit use of information obtained or created under this Agreement to the purposes stated in Exhibit I (and all other Exhibits subsequently attached to this Agreement). For example, unless the Agreement specifies to the contrary the Information Recipient agrees not to:

- Link information received under this Agreement with any other information.
- Use information received under this Agreement to identify or contact individuals.

The Information Recipient shall construe this clause to provide the maximum protection of the information that the law allows.

### II. SAFEGUARDING INFORMATION

A. CONFIDENTIALITY

Information Recipient agrees to:

- Follow DOH small numbers publishing guidelines as well as dataset specific requirements. (Appendix D and E)
- Limit access and use of the information:
  - To the minimum amount of information.
  - To the fewest people.
  - For the least amount of time required to do the work.
- Assure that all people with access to the information understand their responsibilities regarding it.
- Assure that every person (e.g., employee or agent) with access to the information signs and dates the "Use and Disclosure of Confidential Information Form" (Appendix A) before accessing the information.
  - Retain a copy of the signed and dated form for each user as long as required in Data Disposition Section.
  - Forward a copy of the signed and dated form for each user to the RHINO program at syndromic.surveillance@doh.wa.gov to obtain access credentials for new users.

The Information Recipient acknowledges the obligations in this section survive completion, cancellation, expiration or termination of this Agreement.

B. SECURITY

The Information Recipient assures that its security practices and safeguards meet Washington State Office of the Chief Information Officer (OCIO) security standard 141.10 *Securing Information Technology Assets*.

For the purposes of this Agreement, compliance with the HIPAA Security Standard and all subsequent updates meets OICIO standard 141.10 "Securing Information Technology Assets."

The Information Recipient agrees to adhere to the Data Security Requirements in Appendix B. The Information Recipient further assures that it has taken steps necessary to prevent unauthorized access, use, or modification of the information in any form.

**Note:** The DOH Chief Information Security Officer must approve any changes to this section prior to Agreement execution. IT Security Officer will send approval/denial directly to DOH Contracts Office and DOH Business Contact.

C. BREACH NOTIFICATION

The Information Recipient shall notify the DOH Chief Information Security Officer within one (1) business days of any suspected or actual breach of security or confidentiality of information covered by the Agreement.

## III. RE-DISCLOSURE OF INFORMATION

Information Recipient agrees to not disclose in any manner all or part of the information identified in this Agreement except as the law requires, this Agreement permits, or with specific prior written permission by the Secretary of the Department of Health.

If the Information Recipient must comply with state or federal public record disclosure laws, and receives a records request where all or part of the information subject to this Agreement is responsive to the request: the Information Recipient will notify the DOH Privacy Officer of the request ten (10) business days prior to disclosing to the requestor. The notice must:

- Be in writing;
- Include a copy of the request or some other writing that shows the:
  - Date the Information Recipient received the request; and
  - The DOH records that the Information Recipient believes are responsive to the request and the identity of the requestor, if known.

## IV. ATTRIBUTION REGARDING INFORMATION

Information Recipient agrees to cite “Washington State Department of Health Rapid Health Information NetwoOrk (RHINO)” or other citation as specified, as the source of the information subject of this Agreement in all text, tables and references in reports, presentations and scientific papers.

Information Recipient agrees to cite its organizational name as the source of interpretations, calculations or manipulations of the information subject of this Agreement.

## V. OTHER PROVISIONS

With the exception of agreements with British Columbia for sharing health information, all data must be stored within the United States.

## VI. AGREEMENT ALTERATIONS AND AMENDMENTS

This Agreement may be amended by mutual agreement of the parties. Such amendments shall not be binding unless they are in writing and signed by personnel authorized to bind each of the parties.

## VII. CAUSE FOR IMMEDIATE TERMINATION

The Information Recipient acknowledges that unauthorized use or disclosure of the data/information or any other violation of sections II or III, and appendices A or B, may result in the immediate termination of this Agreement.

## VIII. CONFLICT OF INTEREST

The DOH may, by written notice to the Information Recipient:

Terminate the right of the Information Recipient to proceed under this Agreement if it is found, after due notice and examination by the Contracting Office that gratuities in the form of entertainment, gifts or otherwise were offered or given by the Information Recipient, or an agency or representative of the Information Recipient, to any officer or employee of the DOH, with a view towards securing this Agreement or securing favorable treatment with respect to the awarding or amending or the making of any determination with respect to this Agreement.

In the event this Agreement is terminated as provided in the paragraph above, the DOH shall be entitled to pursue the same remedies against the Information Recipient as it could pursue in the event of a breach of the Agreement by the Information Recipient. The rights and remedies of the DOH provided for in this section are in addition to any other rights and remedies provided by law. Any determination made by the Contracting Office under this clause shall be an issue and may be reviewed as provided in the "disputes" clause of this Agreement.

## IX. DISPUTES

Except as otherwise provided in this Agreement, when a genuine dispute arises between the DOH and the Information Recipient and it cannot be resolved, either party may submit a request for a dispute resolution to the Contracts and Procurement Unit. The parties agree that this resolution process shall precede any action in a judicial and quasi-judicial tribunal. A party's request for a dispute resolution must:

- Be in writing and state the disputed issues, and
- State the relative positions of the parties, and
- State the information recipient's name, address, and his/her department agreement number, and
- Be mailed to the DOH contracts and procurement unit, P. O. Box 47905, Olympia, WA 98504-7905 within thirty (30) calendar days after the party could reasonably be expected to have knowledge of the issue which he/she now disputes.

This dispute resolution process constitutes the sole administrative remedy available under this Agreement.

## X. EXPOSURE TO DOH BUSINESS INFORMATION NOT OTHERWISE PROTECTED BY LAW AND UNRELATED TO CONTRACT WORK

During the course of this contract, the information recipient may inadvertently become aware of information unrelated to this agreement. Information recipient will treat such information respectfully, recognizing DOH relies on public trust to conduct its work. This information may be hand written, typed, electronic, or verbal, and come from a variety of sources.

## XI. GOVERNANCE

This Agreement is entered into pursuant to and under the authority granted by the laws of the state of Washington and any applicable federal laws. The provisions of this Agreement shall be construed to conform to those laws.

In the event of an inconsistency in the terms of this Agreement, or between its terms and any applicable statute or rule, the inconsistency shall be resolved by giving precedence in the following order:

- Applicable Washington state and federal statutes and rules;

- Any other provisions of the Agreement, including materials incorporated by reference.

## XII. HOLD HARMLESS

Each party to this Agreement shall be solely responsible for the acts and omissions of its own officers, employees, and agents in the performance of this Agreement. Neither party to this Agreement will be responsible for the acts and omissions of entities or individuals not party to this Agreement. DOH and the Information Recipient shall cooperate in the defense of tort lawsuits, when possible.

## XIII. LIMITATION OF AUTHORITY

Only the Authorized Signatory for DOH shall have the express, implied, or apparent authority to alter, amend, modify, or waive any clause or condition of this Agreement on behalf of the DOH. No alteration, modification, or waiver of any clause or condition of this Agreement is effective or binding unless made in writing and signed by the Authorized Signatory for DOH.

## XIV. RIGHT OF INSPECTION

The Information Recipient shall provide the DOH and other authorized entities the right of access to its facilities at all reasonable times, in order to monitor and evaluate performance, compliance, and/or quality assurance under this Agreement on behalf of the DOH.

## XV. SEVERABILITY

If any term or condition of this Agreement is held invalid, such invalidity shall not affect the validity of the other terms or conditions of this Agreement, provided, however, that the remaining terms and conditions can still fairly be given effect.

## XVI. SURVIVORSHIP

The terms and conditions contained in this Agreement which by their sense and context, are intended to survive the completion, cancellation, termination, or expiration of the Agreement shall survive.

## XVII. TERMINATION

Either party may terminate this Agreement upon 30 days prior written notification to the other party. If this Agreement is so terminated, the parties shall be liable only for performance rendered or costs incurred in accordance with the terms of this Agreement prior to the effective date of termination.

**XVIII. WAIVER OF DEFAULT**

This Agreement, or any term or condition, may be modified only by a written amendment signed by the Information Provider and the Information Recipient. Either party may propose an amendment.

Failure or delay on the part of either party to exercise any right, power, privilege or remedy provided under this Agreement shall not constitute a waiver. No provision of this Agreement may be waived by either party except in writing signed by the Information Provider or the Information Recipient.

**XIX. ALL WRITINGS CONTAINED HEREIN**

This Agreement and attached Exhibit(s) contains all the terms and conditions agreed upon by the parties. No other understandings, oral or otherwise, regarding the subject matter of this Agreement and attached Exhibit(s) shall be deemed to exist or to bind any of the parties hereto.

**XX. PERIOD OF PERFORMANCE**

This **Agreement** shall be effective from_____through_____.

## SPECIAL TERMS AND CONDITIONS

**XXI. The information recipient shall:**

a. Only utilize the information obtained through this agreement for purposes of public health and/or healthcare practice, which do not constitute research activities as defined in RCW 42.48.010. Information may be obtained and used for research purposes only after approval by an Institutional Review Board (IRB) and execution of a Confidentiality Agreement for the research project.
b. Take all reasonable steps to prevent unauthorized access to the ESSENCE platform and any data obtained through this agreement which may be considered private or confidential under state or federal law.
c. Not publish or otherwise disclose any data which may directly or indirectly identify an individual, except as allowed by law within the confines of a public health investigation. Furthermore, the information recipient shall not publish the identity of a data provider (hospital, clinic, or provider) except with the consent of the data provider.

d. Not attempt to determine the identity of persons whose information is included in the data set or use the data in any manner that identifies individuals or their families, except to investigate events of potential public health importance (e.g., notifiable conditions, outbreaks).
e. Not attempt to obtain additional information about a patient or their visit from a patient's electronic medical record except for purposes agreed upon by the data provider (hospital, clinic, or provider) and the information recipient.
f. Except as required by state or federal law, not provide or otherwise utilize data obtained through this agreement for purposes of regulatory action or law enforcement against a data provider (hospital, clinic, or provider) or individual.

**XXII. The Information Recipient may:**

a. Adhering to the DOH Small Numbers Publishing Guidelines (Appendix D) and RHINO Data Best Practices (Appendix E), and without including direct or indirect identifiers, publish, redisclose, or release aggregated data in order to protect public health.
b. Link data obtained through this Agreement with data from other sources, in order to identify, characterize, and/or solve a health problem, or evaluate the success of a health program. Any linked dataset containing data elements obtained through this agreement are subject to the terms of this Agreement, similar agreements governing linked datasets, and all state and federal laws that govern any included datasets.
c. Use data obtained through this Agreement to follow up on specific visits in order to investigate events of potential public health importance (e.g., notifiable conditions, outbreaks). In support of such an investigation, data obtained through this Agreement may be shared with health officials on a "need to know" basis, sharing the fewest number of data elements with the fewest number of individuals, for the least amount of time necessary.

**IN WITNESS WHEREOF, the parties have executed this Agreement as of the date of last signature below.**

| **INFORMATION PROVIDER** | **INFORMATION RECIPIENT** |
| --- | --- |
| State of Washington Department of Health | _____ |
| _____<br>Signature | _____<br>Signature |
| _____<br>Print Name | _____<br>Print Name |
| _____<br>Date | _____<br>Date |

# EXHIBIT I

## 1. PURPOSE AND JUSTIFICATION FOR SHARING THE DATA

Provide a detailed description of the purpose and justification for sharing the data, including specifics on how the data will be used. (**to be completed by Information Recipient**)

Is the purpose of this agreement for human subjects research that requires Institutional Review Board (IRB) approval?

☐ Yes ☐ No

If yes, has an IRB review and approval been received? If yes, please provide copy of approval. If No, attach exception letter.

☐ Yes ☐ No

## 2. PERIOD OF PERFORMANCE

This **Exhibit** shall have the same period of performance as the **Agreement** unless otherwise noted below:

Exhibit ___ shall be effective from_____through_____.

## 3. DESCRIPTION OF DATA

Information Provider will make available the following information under this Agreement (Include the name of the database and a list of all the data elements being provided):

The Information Provider will provide access to the CDC National Syndromic Surveillance Program (NSSP) Electronic Surveillance System for the Early Notification of Community-based Epidemics (ESSENCE) platform for a limited number of authorized users employed or contracted by the Information Recipient. User accounts will be established and managed by the Information Provider.

Authorized users will, upon execution of this Agreement and receipt of signed confidentiality agreements (Appendix A) from each authorized user, have access to the complete dataset contained within ESSENCE for the Information Recipient's jurisdiction. For example, an authorized user employed by a local health jurisdiction (LHJ) will have access to all ESSENCE

data reported by facilities located in that jurisdiction, and all ESSENCE data for residents of that jurisdiction. An authorized user employed by a hospital will have access only to data from that hospital.

Authorized users have the ability to interact with and analyze the data within the ESSENCE platform. Additionally, authorized users have the ability download partial or complete datasets from the platform for additional analysis outside of the ESSENCE platform.

Data elements which may be found in ESSENCE for each record (visit) include:

- Facility name
- Facility type
- Admission reason code
- Patient's chief complaint(s) – original and processed entries
- Patients discharge diagnosis(es)
- Patient's Date of Birth
- Patient's age
- Visit/Admission date and time
- Discharge date and time
- Date and time of death (if applicable)
- Patient's medical record number
- Zip code city, county, and state of patient residence
- Discharge disposition
- Patient's sex
- Patient's race
- Patient's ethnicity
- Facility zip code
- Procedure code
- Initial Temperature
- Initial ED acuity assessment
- Onset date
- Clinical Impression
- Problem list
- Medication list
- Initial pulse oximetry
- Initial systolic and diastolic blood pressures
- Height
- Weight
- Body mass Index
- Pregnancy status

- Smoking status
- Travel history
- Visit type
- Mode of arrival
- Clinical Impression
- Triage notes
- Insurance coverage
- Insurance company ID
- Discharge instructions
- Various administrative and system data elements

It is important to note that, while the above listed data elements may exist in the ESSENCE platform, the elements included for each individual record may vary. This is a result of variances in data submission among facilities.

The information described in this section is:

☒ Restricted Confidential Information (Category 4)
☐ Confidential Information (Category 3)
☐ Potentially identifiable information (Category 3)
☐ Internal [public information requiring authorized access] (Category 2)
☐ Public Information (Category 1)

Any reference to data/information in this Agreement shall be the data/information as described in this Exhibit.

**4. STATUTORY AUTHORITY TO SHARE INFORMATION**

**DOH statutory authority** to obtain and disclose the confidential information or limited Dataset(s) identified in this Exhibit to the Information Recipient:

**RCW 43.20.050 – Powers and duties of state board of health**
**RCW 43.70.050 – Collection, use, and accessibility of health-related data**
**RCW 43.70.130 – Powers and duties of the secretary**
**RCW 70.02.050 – Disclosure without patient's authorization**
**RCW 70.02.220 - Sexually transmitted diseases—Permitted and mandatory disclosures**
**RCW 70.02.230 - Mental health services, confidentiality of records—Permitted disclosures**
**RCW 70.02.240 - Mental health services—Minors—Permitted disclosures**
**RCW 43.70.057 - Hospital emergency room patient care information—Data collection, maintenance, analysis, and dissemination—Rules**

## 5. ACCESS TO INFORMATION

METHOD OF ACCESS/TRANSFER

☐ DOH Web Application (indicate application name):
☐ Washington State Secure File Transfer Service (sft.wa.gov)
☐ Encrypted CD/DVD or other storage device
☐ Health Information Exchange (HIE)**
☒ Other: Authorized users will access the data through the CDC NSSP ESSENCE platform

**Note: DOH Chief Information Security Officer must approve prior to Agreement execution. DOH Chief Information Security Officer will send approval/denial directly to DOH Contracts Office and DOH Business Contact.

FREQUENCY OF ACCESS/TRANSFER

☐ One time: DOH shall deliver information by _____ (insert date)
☐ Repetitive: frequency or dates _____ (insert dates if applicable)
☒ As available within the period of performance stated in Section 2.

## 6. REIMBURSEMENT TO DOH

Payment for services to create and provide the information is based on the actual expenses DOH incurs, including charges for research assistance when applicable.

Billing Procedure

- Information Recipient agrees to pay DOH by check or account transfer within 30 calendar days of receiving the DOH invoice.

- Upon expiration of the Agreement, any payment not already made shall be submitted within 30 days after the expiration date or the end of the fiscal year, which is earlier.

Charges for the services to create and provide the information are:

☐ $_____

☒ No charge.

**7. DATA DISPOSITION**

Unless otherwise directed in writing by the DOH Business Contact, at the end of this Agreement, or at the discretion and direction of DOH, the Information Recipient shall:

☐ Immediately destroy all copies of any data provided under this Agreement after it has been used for the purposes specified in the Agreement . Acceptable methods of destruction are described in Appendix B. Upon completion, the Information Recipient shall submit the attached Certification of Data Disposition (Appendix C) to the DOH Business Contact.

☐ Immediately return all copies of any data provided under this Agreement to the DOH Business Contact after the data has been used for the purposes specified in the Agreement, along with the attached Certification of Data Disposition (Appendix C)

☐ Retain the data for the purposes stated herein for a period of time not to exceed _____ (*e.g., one year, etc.*), after which Information Recipient shall destroy the data (as described below) and submit the attached Certification of Data Disposition (Appendix C) to the DOH Business Contact.

☒ Other (Describe): Authorized users have the ability to download (copy) partial or complete datasets from the platform. Upon request by DOH program staff, at the end of the Agreement term, or when no longer needed, the Information Recipient shall destroy all copies of any data provided under this Agreement. Acceptable methods of destruction are described in Appendix B.

**8. RIGHTS IN INFORMATION**

Information Recipient agrees to provide, if requested, copies of any research papers or reports prepared as a result of access to DOH information under this Agreement for DOH review prior to publishing or distributing.

In no event shall the Information Provider be liable for any damages, including, without limitation, damages resulting from lost information or lost profits or revenue, the costs of recovering such Information, the costs of substitute information, claims by third parties or for other similar costs, or any special, incidental, or consequential damages, arising out of the use of the information. The accuracy or reliability of the Information is not guaranteed or warranted in any way and the Information Provider(s) disclaim liability of any kind whatsoever, including, without limitation, liability for quality, performance, merchantability and fitness for a particular purpose arising out of the use, or inability to use the information.

☒ If checked, please submit the following:

- Copies of all papers, presentations, reports, or publications developed using data obtained under this agreement to the attention of: the RHINO program at syndromic.surveillance@doh.wa.gov.

**9. ALL WRITINGS CONTAINED HEREIN**

This Agreement and attached Exhibit(s) contains all the terms and conditions agreed upon by the parties. No other understandings, oral or otherwise, regarding the subject matter of this Agreement and attached Exhibit(s) shall be deemed to exist or to bind any of the parties hereto.

**IN WITNESS WHEREOF, the parties have executed this Exhibit as of the date of last signature below.**

| **INFORMATION PROVIDER** | **INFORMATION RECIPIENT** |
|---|---|
| State of Washington Department of Health | _____ |
| _____ | _____ |
| Signature | Signature |
| _____ | _____ |
| Print Name | Print Name |
| _____ | _____ |
| Date | Date |

**APPENDIX A**

**USE AND DISCLOSURE OF CONFIDENTIAL INFORMATION**

People with access to the information must sign and date the "Use and Disclosure of Confidential Information Form" (Appendix A) before accessing the information.

The Information Recipient must retain a copy of the signed and dated form for each user as long as required in Data Disposition Section.

The Information Recipient must forward a copy of the signed and dated form for each user to the RHINO program at syndromic.surveillance@doh.wa.gov to obtain access credentials for new users.

People with access to confidential information are responsible for understanding and following the laws, policies, procedures, and practices governing it. Below are key elements:

A. CONFIDENTIAL INFORMATION
Confidential information is information federal and state law protects from public disclosure. Examples of confidential information are social security numbers, and healthcare information that is identifiable to a specific person under RCW 70.02. The general public disclosure law identifying exemptions is RCW 42.56.

B. ACCESS AND USE OF CONFIDENTIAL INFORMATION
1. Access to confidential information must be limited to people whose work specifically requires that access to the information.
2. Use of confidential information is limited to purposes specified elsewhere in this Agreement.

C. DISCLOSURE OF CONFIDENTIAL INFORMATION
1. An Information Recipient may disclose an individual's confidential information received or created under this Agreement to that individual or that individual's personal representative consistent with law.
2. An Information Recipient may disclose an individual's confidential information, received or created under this Agreement only as permitted under the ***Re-Disclosure of Information*** section of the Agreement, and as state and federal laws allow.

D. CONSEQUENCES OF UNAUTHORIZED USE OR DISCLOSURE
An Information Recipient's unauthorized use or disclosure of confidential information is the basis for the Information Provider immediately terminating the Agreement. The Information Recipient may also be subject to administrative, civil and criminal penalties identified in law.

E. ADDITIONAL DATA USE RESTRICTIONS: ESSENCE User Code of Conduct

System Monitoring —As an authorized user, you understand and acknowledge that your use of this system will be monitored for system management and to ensure protection against unauthorized access or use. Unauthorized access or use may subject a user to administrative, civil, criminal, or other adverse action to the extent allowed by law.

Warnings, Alerts, and Anomalies —Syndromic surveillance systems emphasize the use of statistical alerting algorithms to help users determine where to focus additional attention. Time series visualization and statistical alerts alone are generally insufficient for issuing public alerts or warnings. Users typically "drill down" to these data to assess the distribution of affected emergency department (ED) visits (or other events captured by the syndromic surveillance system) and may use additional variables such as person, place, or time and other clinical assessments. Analyses may include quality checks to confirm data are complete and accurate.

To that end, users are expected to respect the role of state and local jurisdictions and their respective authority related to public health matters within their jurisdiction by

- Consulting a jurisdiction whose data you intend to access and use (including jurisdictions within your own) to discuss a finding or interpretation of these data before issuing a public statement or warning, taking public health action, or seeking further information from data providers within the other jurisdiction when that action includes disclosure of information derived in part or in whole from the other jurisdiction's data*.

- Informing those who use your data about significant anomalies already understood or under investigation to prevent duplication of effort and unnecessary queries. This includes anomalies due to artifacts (like exercises or batched data) and those due to real local events.

Data Sharing —the design of the BioSense** Platform ensures that all sites contribute data toward national syndromic surveillance (with limited details at aggregate levels) while also allowing jurisdictions to control whether and how much data are shared at local and state levels. Users are expected to act responsibly by

- Assuming the risk and liability of any of their use or misuse of the BioSense Platform or data produced, including use that complies with third-party rights (i.e., downstream Data Use Agreements).

- Sharing data with other authorized users in accord with applicable agreements and laws.

- Ensuring that the use of these data is in accord with acceptable practices for ensuring the protection, confidentiality, and integrity of contents.

- Making NO attempt to identify individuals represented in these data or data sources except as part of an authorized public health investigation follow-up and to the extent allowed by applicable law.

- Making NO attempt to use these data where prohibited by local, state, or federal law or regulation.

- Keeping usernames and passwords confidential; this system is intended for authorized users only.

Violation of Code of Conduct may result in CDC disallowing access to the BioSense Platform and associated data and tools within. By accepting this code of conduct, you acknowledge that you are an authorized user of the BioSense Platform and have read and understand the BioSense Platform Code of Conduct.

*Cross-jurisdictional consultation and coordination are strongly encouraged, to assist in the interpretation of data and gain further information to inform effective public health action. While beneficial, this should not prevent a jurisdiction from exercising their authority to protect public health.

**BioSense and ESSENCE are used interchangeably

Print Name: _____

Signature: _____

Date:_____

Email Address:_____

Phone Number:_____

**APPENDIX B**

**DATA SECURITY REQUIREMENTS**

**Protection of Data**

The Information Recipient agrees to store information received under this Agreement (the data) within the United States on one or more of the following media, and to protect it as described below:

**A. Passwords**

1. Passwords must always be encrypted. When stored outside of the authentication mechanism, passwords must be in a secured environment that is separate from the data and protected in the same manner as the data. For example passwords stored on mobile devices or portable storage devices must be protected as described under section *F. Data storage on mobile devices or portable storage media*.

2. Complex Passwords are:

   - At least 8 characters in length .
   - Contain at least three of the following character classes: uppercase letters, lowercase letters, numerals, special characters.
   - Do not contain the user's name, user ID or any form of their full name.
   - Do not consist of a single complete dictionary word, but can include a passphrase.
   - Changed at least every 120 days.

**B. Hard disk drives –** Data stored on workstation hard disks:

1. The data must be encrypted as described under section *F. Data storage on mobile devices or portable storage media*. Encryption is not required when Potentially Identifiable Information is stored temporarily on local workstation hard disks. Temporary storage is thirty (30) days or less.

2. Access to the data is restricted to authorized users by requiring logon to the local workstation using a unique user ID and Complex Password, or other authentication mechanisms which provide equal or greater security, such as biometrics or smart cards. Accounts must lock after 5 unsuccessful access attempts and remain locked for at least 15 minutes, or require administrator reset.

**C. Network server and storage area networks (SAN)**

1. Access to the data is restricted to authorized users through the use of access control lists which will grant access only after the authorized user has authenticated to the network.

2. Authentication must occur using a unique user ID and Complex Password, or other authentication mechanisms which provide equal or greater security, such as biometrics or smart cards. Accounts must lock after 5 unsuccessful access attempts, and remain locked for at least 15 minutes, or require administrator reset.

3. The data are located in a secured computer area, which is accessible only by authorized personnel with access controlled through use of a key, card key, or comparable mechanism.

4. If the servers or storage area networks are not located in a secured computer area **or** if the data is classified as Confidential or Restricted it must be encrypted as described under *F. Data storage on mobile devices or portable storage media*.

**D. Optical discs (CDs or DVDs)**

1. Optical discs containing the data must be encrypted as described under *F. Data storage on mobile devices or portable storage media*.

2. When not in use for the purpose of this Agreement, such discs must be locked in a drawer, cabinet or other physically secured container to which only authorized users have the key, combination or mechanism required to access the contents of the container.

**E. Access over the Internet or the State Governmental Network (SGN).**

1. When the data is transmitted between DOH and the Information Recipient, access is controlled by the DOH, who will issue authentication credentials.

2. Information Recipient will notify DOH immediately whenever:

    a) An authorized person in possession of such credentials is terminated or otherwise leaves the employ of the Information Recipient;

    b) Whenever a person's duties change such that the person no longer requires access to perform work for this Contract.

3. The data must not be transferred or accessed over the Internet by the Information Recipient in any other manner unless specifically authorized within the terms of the Agreement.

a) If so authorized the data must be encrypted during transmissions using a key length of at least 128 bits. Industry standard mechanisms and algorithms, such as those validated by the National Institute of Standards and Technology (NIST) are required.

b) Authentication must occur using a unique user ID and Complex Password (of at least 10 characters). When the data is classified as Confidential or Restricted, authentication requires secure encryption protocols and multi-factor authentication mechanisms, such as hardware or software tokens, smart cards, digital certificates or biometrics.

c) Accounts must lock after 5 unsuccessful access attempts, and remain locked for at least 15 minutes, or require administrator reset.

**F. Data storage on mobile devices or portable storage media**

1. Examples of mobile devices are: smart phones, tablets, laptops, notebook or netbook computers, and personal media players.

2. Examples of portable storage media are: flash memory devices (e.g. USB flash drives), and portable hard disks.

3. The data must not be stored by the Information Recipient on mobile devices or portable storage media unless specifically authorized within the terms of this Agreement. If so authorized:

   a) The devices/media must be encrypted with a key length of at least 128 bits, using industry standard mechanisms validated by the National Institute of Standards and Technologies (NIST).

      - Encryption keys must be stored in a secured environment that is separate from the data and protected in the same manner as the data.

   b) Access to the devices/media is controlled with a user ID and a Complex Password (of at least 6 characters), or a stronger authentication method such as biometrics.

   c) The devices/media must be set to automatically wipe or be rendered unusable after no more than 10 failed access attempts.

   d) The devices/media must be locked whenever they are left unattended and set to lock automatically after an inactivity activity period of 3 minutes or less.

   e) The data must not be stored in the Cloud. This includes backups.

   f) The devices/ media must be physically protected by:

      - Storing them in a secured and locked environment when not in use;

- Using check-in/check-out procedures when they are shared; and
- Taking frequent inventories.

4. When passwords and/or encryption keys are stored on mobile devices or portable storage media they must be encrypted and protected as described in this section.

**G. Backup Media**

The data may be backed up as part of Information Recipient's normal backup process provided that the process includes secure storage and transport, and the data is encrypted as described under *F. Data storage on mobile devices or portable storage media.*

**H. Paper documents**

Paper records that contain data classified as Confidential or Restricted must be protected by storing the records in a secure area which is only accessible to authorized personnel. When not in use, such records is stored in a locked container, such as a file cabinet, locking drawer, or safe, to which only authorized persons have access.

**I. Data Segregation**

1. The data must be segregated or otherwise distinguishable from all other data. This is to ensure that when no longer needed by the Information Recipient, all of the data can be identified for return or destruction. It also aids in determining whether the data has or may have been compromised in the event of a security breach.

2. When it is not feasible or practical to segregate the data from other data, then ***all*** commingled data is protected as described in this Exhibit.

**J. Data Disposition**

If data destruction is required by the Agreement, the data must be destroyed using the guidelines listed within OCIO Policy 141.10, para 8.3 (Media Handling and Disposal) and NIST 800-88 (Guidelines for Media Sanitization).

**K. Notification of Compromise or Potential Compromise**

The compromise or potential compromise of the data is reported to DOH as required in Section II.C.

**APPENDIX C**

**CERTIFICATION OF DATA DISPOSITION**

Date of Disposition _____

☐ All copies of any Datasets related to agreement DOH#_____ have been deleted from all data storage systems. These data storage systems continue to be used for the storage of confidential data and are physically and logically secured to prevent any future access to stored information. Before transfer or surplus, all data will be eradicated from these data storage systems to effectively prevent any future access to previously stored information.

☐ All copies of any Datasets related to agreement DOH#_____ have been eradicated from all data storage systems to effectively prevent any future access to the previously stored information.

☐ All materials and computer media containing any data related to agreement DOH #_____ have been physically destroyed to prevent any future use of the materials and media.

☐ All paper copies of the information related to agreement DOH #_____ have been destroyed on-site by cross cut shredding.

☐ All copies of any Datasets related to agreement DOH #_____ that have not been disposed of in a manner described above, have been returned to DOH.

☐ Other

The data recipient hereby certifies, by signature below, that the data disposition requirements as provided in agreement DOH # _____, Appendix B, item J Disposition of Information, have been fulfilled as indicated above.

_____ _____

Signature of data recipient Date

**APPENDIX D**

**DOH SMALL NUMBERS PUBLISHING GUIDELINES**

These guidelines pertain only to the public release (publishing) of information. These guidelines do not pertain to information sharing within an agency or with partner agencies in the context of public health and/or healthcare operations.

***Suppression Rules:***

- Aggregate data so that the need for suppression is minimal. Suppress all non-zero counts which are less than ten.
- Suppress rates or proportions derived from those suppressed counts.
- Assure that suppressed cells cannot be recalculated through subtraction, by using secondary suppression as necessary.

***Exceptions to the Suppression Rules:***

DOH Small Numbers Publishing Guidelines allow for case-by-case exceptions in certain circumstances, so that the public may receive information when public concern is elevated and/or protective actions are warranted. Two examples of such situations are:

- In a cluster investigation, intense public interest often combines with very small numbers of cases. In order to be responsive to the community and allay fear, the Data Recipient may decide it is important to make an exception to the small numbers publishing standard while still protecting privacy.
- Similarly, in a public health emergency such as a communicable disease outbreak or other all-hazards incident, case counts may be released when the numbers are very small. This should be done in the context of an imminent public health threat, such as person to person spread of disease, where immediate action is indicated to protect public health.

When releasing small numbers to the public in the context of the above exceptions, DOH recommends limiting the amount of information shared in order to protect the identity of the person(s) involved. In these cases, DOH recommends reporting only the person's gender, decade of age, and county of residence. For minors, ages should be reported as <18.

For further guidance, please refer to Appendix F for *Draft* DOH Agency Standards and Guidelines for Working with Small Numbers. This document contains recommendations and best practices for protecting the privacy of Washington residents when presenting data to the public.

**APPENDIX E**

**RHINO DATA LIMITATIONS AND BEST PRACTICES FOR DATA USE**

Appendix E includes guidance regarding the limitations of the dataset and recommends best practices for the use of these data. This dataset is unique, as it is rapid, minimally processed and cleaned, and always preliminary. Due to these factors, the limitations must be well understood, and the data must be handled appropriately.

**Data etiquette:**

- Before releasing data originating from another jurisdiction, contact the jurisdiction to obtain approval for data release, invite collaboration, and to ensure that you have interpreted the data correctly.

**Limitations of the data:**

- Data drop-outs are common: Data are frequently missing for brief (1-2 days) and sometimes longer timeframes (weeks to months).
- Data are highly variable:
    - Across facilities, electronic health record vendors, healthcare organizations, and facility types (e.g., ambulatory, ED, inpatient). Differences may include data format, value sets used, variables included, and quality of data reported.
    - Over time. New facilities come online and drop off over time, and occasionally facilities do not report data for a span of days to weeks. Trends may be affected by changes in systems used to track records, facility workflow, business processes, or policies. In addition, electronic health record updates/modifications may impact data. We are often not informed of these changes or their effects.
- Data are always preliminary.
    - Data are updated over time as information becomes available. There is no way to tell when a visit record is "complete". Most records are complete within 1-2 weeks of the visit date, but some records may have updates months later.

**Best practices:**

- It is critical for all users to have a thorough understanding of the data.
    - Monitor the quality of the data to understand gaps in reporting and changes in reporting over time. Sometimes trends are artifacts of data quality issues, not reflective of the underlying health of the population or healthcare utilization.
    - Consider alternative explanations for any trends you observe. Reach out to colleagues who are familiar working with syndromic surveillance data to understand what limitations you should be aware of when analyzing the data. Ask them if the trends you are observing make sense. Cross-check the trends against other types of data sources when feasible.
    - Perform a literature review (published pee-reviewed manuscripts as well as ISDS conference abstracts) to understand contexts in which syndromic surveillance

data have (or have not) proven useful. Have others reported success when evaluating the condition(s) you are monitoring? What limitations should you be aware of?
    - Consider whether the syndrome definition you are using is calibrated optimally for the question you are trying to answer (i.e., do you need a sensitive or specific definition), and evaluate several different definitions to understand impact of changing the definition on the results. In addition, evaluate options for validating the syndrome definitions you've selected by comparing syndromic data with a gold standard (e.g., chart review, coded diagnoses, other data sources).
    - Know which facilities are included in your data, and when they started reporting to the system.
    - Look for data drop-offs and changes in healthcare utilization. Find out when facilities that are included in your data set switched to a new EMR. Determine how you will account for missing data, and how you will account for secular trends (day-of-week effects, seasonal trends, etc.).
    - Get to know the pattern of reporting (e.g., how soon are complete data available for the previous 24 hours? Do facilities report on the weekend? Are some data delayed?) so you know whether you are working with incomplete or complete data. Also, understanding that certain data elements may be delayed (e.g., diagnoses) will help you understand the patterns you see.
    - Understand the types of facilities (e.g., ED, inpatient, outpatient, ambulatory).
    - Know the format of diagnoses (e.g., ICD9? ICD10? Single? Multiple? Primary? Secondary? Do diagnoses reflect current visit info, or do they also pull from the patients' histories?).
    - Know the format of chief complaint (e.g., single term, standardized or free-text with potential for lots of variability, patient's own words vs. clinician evaluation vs. front desk staff entry).
    - Know what optional data elements are included and may be of potential value (e.g., triage notes, clinical impression) and the completeness of data elements of interest.
- Check your data again!
    - Once you know your data, you can establish a routine of checking your data for any unexpected changes in the data.
    - Examine the data for all-visit counts by facility, and within each facility for completeness of reporting by age, sex, chief complaints, etc.
- Use both counts and percentages.
    - When building a query, check data counts to make sure they are at the magnitude expected (e.g., has there been a change in total visit counts? Are certain facilities missing when you query for a certain syndrome? Do changes in visit counts reflect when facilities started reporting or dropped out? How do counts vary by age group or other factors?). Then, review trends using percentages which normalize data.

- Respect existing relationships that have been forged with data providers, or establish and maintain relationships if they do not already exist.
    - If you need to follow up on trends or data quality concerns, contact the lead(s) who have established relationships with the data providers. If no such relationships exist, create those connections.
    - Work towards identifying and understanding data anomalies (e.g., changes in business practices or hospital policies) by reaching out to the data providers. This will help you understand and investigate trends in your data and allow you to follow up on specific records of interest more quickly. In addition, this shows the facilities that we are using their data and how valuable it is to us!
- View this dataset as a tool in the toolbox rather than a stand-alone
    - Syndromic data is not curated or cleaned. It is made available as it is created/received. As a result, it can be noisy and occasionally lead to inaccurate conclusions. There are misspellings, data entry errors, and missing data. Where syndromic data can be helpful is:
        - generating hypotheses
        - strengthening information gathered from other sources
        - investigating rumors

      It is not typically recommended that policies, public health interventions, or press releases be based <u>SOLELY</u> on syndromic data.

**APPENDIX F**

## ***DRAFT*** **DOH AGENCY STANDARDS AND GUIDELINES FOR WORKING WITH SMALL NUMBERS**

**Revision Date: Draft May 23, 2017**
**Primary Contact: Cathy Wasserman, PhD, MPH, State Epidemiologist for Non-Infectious Conditions**
**Secondary Contact: Eric Ossiander, PhD**


Summary DOH Data Presentation for the Public – Small Number Standard
Summary DOH Data Presentation for the Public – Reliability Recommendation
Summary Graphic
What is new, and how does this affect public health assessment?
Scope of the "Standards and Guidelines for Working with Small Numbers"
Why are small numbers a concern in public health assessment?
What constitutes a breach of confidentiality?
Why do we question the reliability of statistics based on small numbers?
Why do we have guidelines and standards?
Guidelines for Working with Small Numbers
General Considerations
Assessing Confidentiality Issues
Know the identifiers
Examine numerator size for each cell
Consider the proportion of the population sampled
Consider the nature of the information
How to Reduce Risk of Confidentiality Breach
General approach
Aggregation
Cell
Omission of stratification variables
Group identification
Exceptions to the suppression rule
Additional Concerns
Recommendations to Protect Confidentiality
Assessing and Addressing Statistical Issues
Relative standard error
Increase numerator size for rare events
Include confidence intervals
Bias from a known undercount
Recommendations to Address Statistical Issues
Glossary
Resources
Relevant Policies, Laws and Regulations
Appendix 1: Detailed example of disclosure risk
Appendix 2: Washington Tracking Network rule-based use of suppression and aggregation

## Summary DOH Data Presentation for the Public – Small Numbers Standard

This document presents standards for the presentation of data to the public. These standards are designed to protect the confidentiality of personal data and the privacy of Washington residents.

These standards require DOH staff who are preparing data for public presentation to:

1. When possible, aggregate data so that the need for suppression is minimal.
2. Suppress all non-zero counts which are less than ten, unless they are in a category labeled "unknown."
3. Suppress rates or proportions derived from those suppressed counts.
4. Assure that suppressed cells cannot be recalculated through subtraction, by using secondary suppression as necessary.
5. Survey data
    a. Surveys in which 80% or more of the eligible population is surveyed should be treated as above (i.e., as non-survey data).
    b. Surveys in which less than 80% of the eligible population is surveyed:
        i. If the respondents are equally weighted, then survey estimates based on fewer than 10 respondents should be "top-coded" (estimates of less than 5% or greater than 95% should be presented as 0-5% or 95-100%).
        ii. If the respondents are unequally weighted, so that cell sample sizes cannot be directly calculated from the weighted survey estimates, then there is no suppression requirement for the weighted survey estimates.

Department of Health Agency Standards for Reporting Data with Small Numbers

## Exceptions Summary

These standards are minimum requirements which are expected to be followed when releasing confidential data to the public, with limited exceptions. DOH programs may choose more stringent rules as program-specific standard practice, but they may not use less stringent rules.

With approval from the Office of the Health Officer, case-by-case exceptions to the suppression rule can be made, so that the public may receive information when public concern is elevated and/or protective actions are warranted. (see "Exceptions to the suppression rule" below).

## Summary DOH Data Presentation for the Public – Reliability Recommendation

Include a "flag" indicating rate instability when the Relative Standard Error (RSE) of the rate or proportion is greater than or equal to 25%.

## These standards are concisely represented in this diagram.

## Purpose

The Assessment Operations Group in the Washington State Department of Health (department) develops guidelines related to data collection, analysis and use in order to promote good professional practice among staff involved in assessment activities within the department and in local health jurisdictions in Washington. While the guidelines are intended for audiences of differing levels of training, they assume a basic knowledge of epidemiology and biostatistics. They are not intended to recreate basic texts and other sources of information; rather, they focus on issues commonly encountered in public health practice and, where applicable, refer to issues unique to Washington State.

## What is new and how does this affect public health assessment?

The department has recently adopted standards for presentation of static and interactive query-based tabular data. The standards are designed to address privacy concerns and represent minimum requirements. We also discuss statistical accuracy and how that is affected by small numbers. We make recommendations about how staff can address statistical reliability, but implementation is left to programs.

## Scope of the "Standards and Guidelines for Working with Small Numbers"

The department and local health jurisdictions routinely make aggregated health and related data available to the public. Historically, these data were presented as static tables. Over the past decade, however, interactive Web-based data query systems allowing public users to build their own tables have become more common. The standards and guidelines apply to releases of aggregated population-based and survey data available to the public. These releases include both static data tables and graphics, such as charts and maps, as well as tables and graphics produced through interactive query systems.

The department and local health jurisdictions also release files containing record-level data. These standards and guidelines do not apply to release of record-level data. Release of record-level data is governed by federal and state disclosure laws, which can be specific to a dataset, and by Institutional Review Boards if the data are used for research. We will develop a guideline on generating public use data files with record-level data, but do not have an estimated date for when it will be completed.

## Why are small numbers a concern in public health assessment?

Public health policy decisions are fueled by information, which is often in the form of statistical data. Questions concerning health outcomes and related health behaviors and environmental factors often are studied within small subgroups of a population, because many activities to improve health affect relatively small populations which are at the highest risk of developing adverse health outcomes. Additionally, continuing improvements in the performance and availability of computing resources, including geographic information systems, and the need to better understand the relationships among environment, behavior and health have led to increased demand for information about small populations. These demands are often at odds with the need to protect privacy and confidentiality. Small numbers also raise statistical issues concerning the accuracy, and thus usefulness, of the data.

## Why do we have a new standard?

Our adoption of a standard requiring the suppression of cells reporting between 1 and 9 events is primarily based on the example of the federal Centers for Disease Control & Prevention (CDC) National Center for Health Statistics (NCHS), which now requires that all data originating from NCHS that is released by CDC (such as in tables produced by online query systems WISQARS <http://www.cdc.gov/injury/wisqars/fatal_injury_reports.html> and WONDER <http://wonder.cdc.gov/>) can only be released after suppression of counts which are less than 10. They adopted this standard in 2011 after finding that a previous rule of suppressing cell counts between 1 and 4 failed to prevent disclosure of an individual's information. The DOH standard allows release of tabular data where the count is zero, on the basis that a count of no events in the cell is unlikely to be a threat to confidentiality.

It is impossible to absolutely guarantee against disclosure risk in data release, because it is impossible to know how much outside information is available to the data user. Data users may have information from personal knowledge of people in the population from which the data were drawn, from searching for information on the Internet, or from other tables of the same data released by different agencies, or by the same agency at a different time.

Here we illustrate disclosure risk with an example from birth data. These are real Washington state data, but we have changed the county names and ZIP Codes to prevent disclosure of sensitive data.

> ZIP Code 47863 overlaps Bush and Clinton counties. In 2005, there were 82 births to mothers whose resident ZIP Code was 47863; 81 of those mothers lived in Clinton county, and 1 lived in Bush county. For the sake of this example, we pretend that no other ZIP Codes overlap Bush and Clinton counties. Let's say that one agency has provided, or posted on the Internet, a table that shows the number of prior pregnancies for birth mothers by resident ZIP Code, and another agency has provided or posted the same data by county of residence. By adding up the figures for all ZIP Codes in Clinton County, including 47863, a data user could ascertain that there was only 1 birth to a mother who lived in Bush County in ZIP Code 47863. If the data user happened to know this woman (say, as a neighbor), then the data user would know the number of her prior pregnancies. We can guard against this type of disclosure by suppressing some cells. In 2005, some of the ZIP Codes in Clinton County had fewer than 10 births, and a rule requiring suppression of those numbers would make it harder for the data user to figure out how many births were in the overlap area. A detailed explanation of the effects of suppressing counts of 1-4 or 1-9 is provided below.
>
> In practice, we cannot anticipate or analyze all of the data tables that will be released. We cannot guarantee either that a rule requiring only the suppression of counts between 1 and 4 will lead to disclosure of sensitive data, or that a rule requiring suppression of counts between 1 and 9 will prevent it. However, it is clear that the 1-9 rule will make disclosure substantially less likely.

### What constitutes a breach of confidentiality?

A breach of confidentiality occurs when analysts release information in a way that allows an individual to be identified and reveals confidential information about that person (that is, information which the person has provided in a relationship of trust, with the expectation that it will not be divulged in an identifiable form). In data tables, a breach of confidentiality can occur if knowing which category a person falls in on one margin (i.e. row or column) of the table allows a table reader to ascertain which category the person falls in on the other margin. The following guidelines describe situations that present high risk for a breach of confidentiality and suggestions on how to reduce this risk. In addition to these guidelines, analysts should be familiar with relevant federal and Washington State laws and regulations and department policies. (See Relevant Policies, Laws and Regulations.) **Federal and state laws and regulations and department policies supersede guidance provided in this document.**

### Why do we question the reliability of statistics based on small numbers?

Estimates based on a random sample of a population are subject to sampling variability. Rates and percentages based on full population counts are also subject to random variation. (See Guidelines for Using Confidence Intervals for Public Health Assessment for a short discussion of variability in population-based data.) The random variation may be substantial when the measure, such as a rate or

percentage, has a small number of events in the numerator or a small denominator. Typically, rates based on large numbers provide stable estimates of the true, underlying rate. Conversely, rates based on small numbers may fluctuate dramatically from year to year, or differ considerably from one small place to another, even when differences are not meaningful. Meaningful analysis of differences in rates between geographic areas or over time requires that the random variation in rates be quantified; this is especially important when rates or percentages are based on small numerators or denominators.

### Why do we have guidelines and standards?

The department has a policy governing the release of confidential information, Policy 17.0006. This policy incorporates these standards for data reporting. Protecting confidentiality starts with understanding the considerations that have gone into developing the standards, which are discussed below. There is no way to completely (100%) guarantee privacy when presenting tabular data. A user might possess external information, including information published in the news media, which may be used in such a way to result in loss of privacy. Therefore, analysts should be aware of these considerations and approaches so they can minimize the risk of a breach of confidentiality despite adhering to the minimum standards. For example, the program may adopt more stringent rules as program-specific standard practice. If the program needs to request an exception to the agency standard, the issues described below should be carefully considered and addressed in the exception request.

## Guidelines for Working with Small Numbers

### General Considerations

These standards and guidelines address both confidentiality and statistical issues in working with small numbers. In some data systems, such as the AIDS registry, the entire database is considered confidential. In other systems, such as the birth certificate system, many but not all data items are confidential. In yet other systems, none of the items are confidential, such as most records in the death certificate system. Survey data often contain confidential information and may also contain information that could be used to identify an individual (e.g., there might be small numbers of individuals with a particular visible characteristic in a small geographical area). A first step in using these guidelines is to determine if the datasets you are working with contain confidential or potentially identifiable information. If so, the following section on protecting confidentiality is relevant. Otherwise, you need only concern yourself with the statistical issues section.

### Assessing Confidentiality Issues

With population-based data, most problems with confidentiality occur when the population from which the events arise (i.e., denominator) is small, but the number of events (i.e., numerator) might also be important. For example, if there are 5,000 individuals in a specific age-race-sex group in a single county, the likelihood of identifying a single individual from data in a published table is quite small. In smaller populations, it is more likely that an individual might be identifiable. However, even in larger populations, it is conceivable that a single individual might be identifiable, if there are only one or two individuals with some special characteristic. For example, in a modest sized community, it may be commonly known that there is only one child who is frequently hospitalized, and a table showing that this community has one case of pediatric HIV-AIDS could unintentionally allow knowledgeable residents to infer the child's illness. Similarly, if a unique individual, such as one of the parents of the frequently hospitalized child described above, were drawn into a survey, knowledgeable residents

might infer the illness of the child from survey data  indicating one child with HIV-AIDS in that community. Thus, the same cautions for population data  generally apply to survey data as well.

***Know the identifiers.*** Data analysts should assess each field in the dataset to determine whether it is a “direct identifier” or an “indirect identifier”. These terms are admittedly somewhat imprecise... Direct identifiers which uniquely identify a person are never released in tabular aggregate data. The federal HIPAA Privacy Rule (section 164.514(e)) defines direct identifiers as:

- Name
- Street name or street address or post office box
- Telephone and fax numbers
- Email address
- Social security number
- Certificate/license numbers
- Vehicle identifiers and serial numbers
- URLs and IP addresses
- Full-face photos and other comparable images
- Medical record numbers, health plan beneficiary numbers, and other account numbers
- Device identifiers and serial numbers
- Biometric identifiers, including finger and voice prints

Indirect identifiers are fields which, when combined with other information, can be used to uniquely identify a person. Examples include:

- Detailed demographic information (e.g., age, gender, race, ethnicity)
- Detailed geographic information (e.g., census tract of residence, 5-digit ZIP code)
- Hospital name or location
- Detailed employment information (e.g., occupational title)
- Exact date of event (e.g., birth, death, hospital discharge)

WAC 246-455, defines direct and indirect identifiers for Comprehensive Hospital Abstract Reporting System (CHARS) data. In this case, direct identifiers includes:

- Patient first name
- Patient middle name(s)
- Patient last name
- Social security number
- Patient control number or medical record number
- Patient zip code + 4 digits
- Dates that include day, month and year
- Admission and discharge dates in combination

This new rule defines indirect identifiers as information that may identify a patient when combined with other information. Indirect identifiers includes:

- Hospital or provider identifiers
- 5-digit ZIP code
- County, state and country of residence
- Dates that include month  and year
- Admission and discharge hour

- Secondary diagnosis, procedure, present on admission, external cause of injury, and payer codes
- Age in years
- Race and ethnicity

Datasets can be linked using only indirect identifiers (Hammill and colleagues, 2009; Pasquali and colleagues, 2010; Lawson and colleagues, 2013). Although aggregate data presented in tabular format is unlikely to be used in this fashion, analysts should examine each field for the potential of use to identify a person.

***Examine numerator size for each cell.*** Data analysts should consider the number of events in each cell of a table to be released (i.e., the numerator for a calculation of a rate or proportion). There is no single national standard for determining when small numerators might lead to breaches of confidentiality. In fact, disclosing that there has been one case of a disease in a state or county might not breach confidentiality if no other detail is given. Small numerators are of increasing concern for confidentiality if there are also small numbers of individuals with the reported characteristic(s) in the population. If the characteristic is observable (e.g., distinctive physical characteristics) or the participants in the survey are known, risk for identification may be further increased. For data tables, the 2004 NCHS Staff Manual on Confidentiality requires:

- No single cells containing all observations of a row or column.
- At least five observations for a row or column total in a cross-tabulation.
- At least five observations total.

Since May 2011, the CDC interactive query system, WONDER, has suppressed birth and death data if there are not at least 10 observations (WONDER 2012). Other groups at CDC use different criteria. For example, the Environmental Public Health Tracking Network currently suppresses rates based on non-zero counts less than six.

The department standards require suppression when the number of cases or events in a cell is less than 10, due to the likelihood of a possible breach of confidentiality. A count of no events in the cell is unlikely to be a threat to confidentiality unless it provides meaningful information about the remaining 100% of participants, but a count of one to nine events may be a threat to confidentiality. A data analyst may choose a higher threshold, if other information indicates a greater likelihood of a possible breach of confidentiality in a specific situation.

***Consider the proportion of the population sampled.*** For survey data, the potential for breaches of confidentiality decreases as the proportion of the population in the sample decreases. Surveys that include 80% or more of the eligible population should be treated in the same way as non-survey data. Surveys of facilities or surveys conducted within facilities, such as schools, sometimes fall into this category. If the survey includes less than 80% of the eligible population, and if the identity of the respondents is kept private, then the risk of disclosing identifying information is far lower than for non-survey data, particularly if weighted survey estimates are presented, instead of respondent cell sizes.

***Consider the nature of the information***. The federal Office of Management and Budget (OMB) *Checklist on Disclosure Potential of Proposed Data Releases* identifies examples of variables that are visible and, therefore, pose increased risk of disclosure. Examples include income and related variables such as property value and rent or mortgage payments; unusual occupation; unusual health condition; very old age; and race or ethnicity. Physical characteristics such as obesity are also visible and might increase risk of individual identification.

## How to Reduce the Risk of a Confidentiality Breach

***General Approach.*** The general approach to privacy protection involves what has been termed "computational disclosure control," which includes both aggregation of data values in the dataset before analysis, and cell suppression in a table after analysis (Sweeney 1997). Web-based query systems, such as that developed by the Washington Tracking Network (WTN), aggregate data using rule-based static and dynamic parameter control in order to minimize suppression. Appendix 1 outlines the aggregation rules used by the WTN to protect confidentiality.

***Aggregation.*** Aggregation of data values is appropriate for fields with large numbers of values, such as dates, diagnoses and geographic areas; it is the primary method used to create tables with no small numbers as denominators or numerators. Granularity refers to the degree of detail or precision in data, or the fineness with which data fields are subdivided. The following table shows examples.

| | | **Granularity: Aggregation** | | |
|---|---|---|---|---|
| **Field** | **Type** | *Fine* | *Medium* | *Coarse* |
| Age | Continuous | Year of birth | 5-year age group | 10-year age group |
| Date of occurrence | Continuous | Month | Year | Multiple years combined |
| Diagnosis | Nominal | Complete ICD code | Three-digit ICD | "Selected cause" Tabulation |
| Geography | Ordinal (spatial) | Zip code, census tract | County | State |

In addition to considering each field on its own, aggregation should consider each field in combination with others. When numbers are large, data are commonly disaggregated across multiple fields, resulting in release of multiple data tables. However, when numbers are small, protecting confidentiality often requires limiting the number of fields which are disaggregated simultaneously, resulting in release of fewer data tables. When numbers are tiny, tables may be limited to those where only one field is disaggregated at a time.

***Cell suppression.*** When it is not possible, or desirable, to create a table with no small numbers, then cell suppression is used. "Primary" cell suppression is used to withhold data in the cell that fails to meet the threshold, followed by secondary (also termed “complementary”) suppression of three other cells in order to avoid inadvertent disclosure through subtraction. Secondary cell suppression is a method of last resort, due to the often unavoidable side-effect of suppressing releasable data values, and due to the amount of labor necessary to implement the method. The following table shows an example of secondary suppression. In this example, even if all the cells except for the cell in the upper left (0–34 Black) meet the threshold for release, data in three additional cells need to be suppressed to prevent the ability for back-calculating the suppressed cell.

| **Age** | *Black* | *White* | *Other* | *Total* |
|---|---|---|---|---|
| 0–34 | Suppress | 30 | Suppress | 60 |
| 35–64 | Suppress | 60 | Suppress | 150 |
| 65+ | 70 | 90 | 80 | 240 |
| Total | 120 | 180 | 150 | 450 |

If the value of the information in all cells is not the same, data analysts should suppress cells that provide less useful information. In the previous table, “other” includes a diversity of racial groups and such aggregation is usually not meaningful for addressing public health problems in Washington State. In the same table, suppressing information for the two youngest age groups might be best, if

the condition is one that primarily affects older individuals. Alternatively, if the goal of the table is to provide data for targeting prevention to middle-aged people, complementary suppression of data for the youngest and oldest age groups might be preferable. The software program tau-ARGUS uses mathematical algorithms to perform secondary suppression in a way that assures that the suppressed data cannot be uncovered by back calculation. However, tau-ARGUS may be difficult to use.

***Omission of stratification variables.*** When neither of these methods (aggregation of data values to create coarser granularity or cell suppression) is satisfactory, the data analyst might want to omit certain fields from analysis entirely. For example, for a department release of asthma data, it was not possible to achieve adequately large cell denominators in annual county-level data showing both age-specific and gender-specific counts and rates. Those publishing the data opted to omit the gender-specific data, and display only tables of age-specific data, on the grounds that no intervention programs targeted groups differently on the basis of gender, but most intervention programs target age groups differently.

***Group identification.*** Data in a table provides information on the probability that someone in a defined group has a given characteristic. The 2004 NCHS Staff Manual on Confidentiality describes this as "probability-based disclosure". The manual describes the problem as follows:

> Data in a table may indicate that members of a given population segment have an 80-percent chance of having a certain characteristic; this would be a probability-based disclosure as opposed to a certainty disclosure of information on given individuals. In a sense, every published table containing data or estimates of descriptors of a specific population group provides probability-based disclosures on members of that group, and only in unusual circumstances could any such disclosure be considered unacceptable. It is possible that a situation could arise in which data intended for publication would reveal that a highly specific group had an extremely high probability of having a given sensitive characteristic; in such a case the probability-based disclosure perhaps should not be published.

***Exceptions to the suppression rule.*** The agency standards allow for case-by-case exceptions, with advance approval from senior management. To request an exception, send an email to both State Epidemiologists with the subject line: Small Numbers Exception Request. In your email, please include:

- Brief description of the health data you are releasing
- Identifiers you are stratifying data by
- Rationale for exception, including why aggregation is not acceptable approach

You will receive a response within one week.

Two examples of situations when an exception would likely be approved are:

- In a cluster investigation, intense public interest often combines with very small numbers of cases. In order to be responsive to the community and allay fear, DOH may decide it is important to make an exception to the standard while still protecting privacy.

- Similarly, in a public health emergency such as a communicable disease outbreak or other all-hazards incident, case counts may be released when the numbers are very small. This should be done in the context of an imminent public health threat, such as person-to-person spread of disease, where immediate action is indicated to protect public health.

When releasing small numbers to the public in the context of the above exceptions, DOH recommends limiting the amount of information shared in order to protect the identity of the person(s) involved. In these cases, DOH recommends reporting only the person's gender, decade of age and county of residence. For minors, ages should be reported as <18

***Additional Concerns***. The following guidelines can be used to alert data analysts to situations that require particular attention to avoid breaches of confidentiality, when the suppression rules (i.e., the standards) may be seen as inadequate. These situations may occur when:

- Denominators are less than 20,000.
- Counts are less than 20.
- Reporting a specific confidential characteristic of a population if a very high proportion of the population has this characteristic.
- Producing multiple tables from the same dataset; in this case, be careful that users cannot derive confidential information through a process of subtraction.

If data are suppressed, provide an indicator (e.g., asterisk) in the suppressed cell and a legend under the table explaining the reason for suppression.

For example, the department routinely publishes data by county. In 2010, nine counties had populations less than 20,000; three of those had populations less than 20,000 person-years when combining three years of data (i.e., 2009–2011). Even though some counties do not meet a 20,000 threshold, most department programs are comfortable publishing numbers or rates by county when the population denominator is the entire county population. However, programs carefully evaluate the potential for breaches of confidentiality when considering publishing the same data by demographic characteristics, because denominators shrink when considering subpopulations within counties. Depending on the type of data and the types of demographic characteristics, programs might conclude that there is not a risk for a breach of confidentiality and they can safely publish the data. Alternatively, they might conclude there is a risk of inadvertent disclosure and decide not to publish such tables at all or not publish for selected counties.

## Assessing and Addressing Statistical Issues

***Relative standard error.*** The relative standard error (RSE) provides a measure of reliability (also termed “statistical stability”) for statistical estimates. When the RSE is large, the estimate is imprecise and we term such rates or proportions “unstable” or “Not Reliable”. In these instances, the data analyst needs to balance issues of the Right-to-Know with presenting data that might be misleading. The standards require annotation: inclusion of a “flag” when the RSE of the rate or proportion is greater than 25%; this rule simplifies to annotation of rates based on counts of 16 or less (see below for implementation issues when calculating proportions). The Not Reliable flag can be displayed by use of an asterisk or as the designation “(NR)” next to the rate/proportion in a table, and the use of diagonal hatched shading on a map; a legend is necessary for explanation. Inclusion of Confidence Intervals on a chart can satisfy the requirement for annotation with that form of data display.

There is no single national standard for deciding when the RSE is so large that one should not annotate the data. Federal agencies and even units within a single federal agency use different approaches, including use of a RSE-based rule for data suppression. For example, within the Centers for Disease Control and Prevention:

- A 2009 NCHS report suppressed data with RSEs greater than 40% and noted that data with RSEs of 30–40% were unreliable. (Fryar 2009) A 2010 NCHS publication suppressed data when RSEs were greater than 50% and noted that estimates with RSEs of 30–50% were unreliable. (NCHS 2010)
- A 2011 NCHS publication suppressed data based on sample size, but not RSEs. Estimates with RSEs of 30–59% were marked as unreliable. (Bercovitz 2011)

- CDC Environmental Public Health Tracking Network national portal displays all rates that are not suppressed for confidentiality protection. Rates with RSEs of 30% or greater are annotated as unreliable. (NEPHTN 2008)
- The National Program of Cancer Registries suppresses data due to concerns about the statistical stability when the number of events is less than 16, stating that a count of fewer than about 16 results in an RSE of about 25%. (NPCR 2008)

Different programs at the department use different practices based on RSE. Currently, some programs do not publish data when RSEs are greater than 30%. In contrast, the Washington Tracking Network follows standards for the CDC Environmental Public Health Tracking Program and marks data with RSEs greater than 30% as unreliable, but does not suppress data for statistical reasons. A middle ground is to suppress data with RSEs above a given cut point, such as RSEs of 40, 50 or 60% as in the NCHS examples given above, and mark as unreliable data with RSEs between 25% and the cut point. The approach taken by different data analysts might vary depending on the primary audience and purpose of the publication.

The threshold for annotation may be modified by data analysts. For example, when there is increased concern over statistical stability, a public health program may decide that program-specific standard practice will routinely use a RSE of >22%; for Poisson-based rates, this simplifies to annotation of rates based on counts of 20 or less.

***Calculation of the RSE.*** Depending on whether the data follow a Poisson or a Binomial statistical distribution, methods for calculation of the RSE differ.

The Poisson distribution applies whenever **rates** are calculated. Note that the Poisson-based calculation of RSE does not use the population.

- Notation:
    - A = count of events
    - B = population
    - Rate = A/B
    - SE = Standard Error = SQRT((Rate*(1-Rate))/population)
- RSE = Relative Standard Error = 1/(SQRT(A))

A simplified method can be used: any result of a rate calculation where the count of events is less than 17 can be annotated NR, as any rate calculation where the count of events is 16 or less results in a RSE >25%.

The Binomial distribution applies whenever **proportions** are calculated. This distribution is generally used when the result of a proportion calculation is >10%. If the result is <10%, a Poisson distribution is assumed to apply.

- Notation:
    - A = numerator
    - B = denominator
    - Proportion = A/B

    - SE = Standard Error = SQRT((Proportion*(1-Proportion))/B)
- RSE = Relative Standard Error = SE/Proportion

The simplified method for NR annotation of proportions, which parallels the Poisson-based method, is accurate for all binomial distributions where the denominators are >1000. When the denominator is smaller, more proportions are labeled as NR than would be if the full RSE calculation was used. Thus, the simplified method is conservative: it over-annotates some results as NR, when the numerator and denominator numbers are small.

***Increase numerator size for rare events.*** As the proportion of data suppressed or annotated as unreliable increases, the value of the data table decreases. Increasing the numerator will improve the stability of the estimate and reduce the RSE. Techniques to improve stability within a fixed sample size or population include the following aggregation methods:

- Combining multiple years of data
- Collapsing data categories
- Expanding the geographic area under consideration

***Include confidence intervals.*** We recommend including confidence intervals (CIs) when presenting rates. (See Guidelines for Using Confidence Intervals for Public Health Assessment.) CIs can be displayed on every table, or, optionally, can be displayed only when the user of an online data query system selects a "Display CI" button. When CIs are included on bar charts and line graphs, no NR annotation is needed; however, charts need some visual indication of data precision. A "hover-over pop-up" which uses a small window to separately display the rate/proportion with its CI for each data-point on an online chart is a possible method. On maps, there is no practical method for inclusion of CIs, and use of the NR shading is adequate; in addition, a hover-over pop-up which shows the rate/proportion with its CI for a specific map area can be used.

***Bias from a known undercount.*** The issue of bias differs from the issue of the precision of estimates in that bias is non-random error. When the data analyst is aware that the count of persons or events is systematically under-ascertained in the data, the data user should be informed by annotating the data. For example, agency studies have shown that hospitalization rates for some border counties are subject to an undercount if the CHARS dataset is used without inclusion of Washington residents hospitalized out-of-state. The WTN metadata explains this in its Caveats section:

> Without reciprocal agreements with abutting states, statewide measures and measures for geographic areas (e.g., counties) bordering other states may be underestimated because of health care utilization patterns. The Tracking Network rules currently call for exclusion of hospitalization data obtained from adjacent states, regardless of whether a state has reciprocal agreements with the adjacent states. In eight Washington counties, hospitalization rates are biased due to county residents traveling out-of-state for hospital care. For these counties (Asotin, Clark, Cowlitz, Garfield, Klickitat, Pacific, Skamania, and Wahkiakum), more than 15% of hospitalizations of county residents occur outside of the state. The bias from this undercount is judged to be excessive, and WTN annotates these hospitalization rates with the NR designation.

We recommend annotating the data when the data analyst is aware that the count of persons or events is systematically under-ascertained.

## Recommendations to Address Statistical Issues

- Include confidence intervals to show the extent of variation that might occur by chance.

- Consider suppression as a method of last resort when data are so unreliable and imprecise that they cannot be used effectively for planning programs or informing policy decisions. If data are suppressed, provide an indicator (e.g., asterisk) in the suppressed cells and a legend under the table explaining the reason for suppression.

## Glossary

***Confidential data/information:*** Information that an individual or establishment has provided in a relationship of trust, with the expectation that it will not be divulged in an identifiable form. The confidentiality of specific data elements or information in individual databases or record systems may be defined by federal or state laws or regulations, or policies or procedures developed for those systems.

***Confidentiality breach:*** An unauthorized release of identifiable or confidential data/information, which may result from a security failure, intentional inappropriate behavior, human error or natural disaster. A breach of confidentiality may or may not result in harm to one or more individuals.

***Individually identifiable data/information:*** Data/information that identifies, or is reasonably likely to be used to identify, an individual or an establishment protected under confidentiality laws. Identifiable data/information may include, but is not limited to, name, address, telephone number, Social Security number and medical record number. Data elements used to identify an individual or protected establishment can vary depending on the geographic location and other variables (e.g., rarity of person's health condition or patient demographics). For purposes of this guideline, "identifiable information" includes potentially identifiable information.

***Number of events:*** The number of persons or events represented in any given cell of tabulated data (e.g., numerator). (See Guidelines for Using and Developing Rates for Public Health Assessment.)

***Population or sample size:*** The total number of persons or events included in the calculation of an event rate (e.g., denominator). (See Guidelines for Selection of Population Denominators)

***Potentially identifiable information:*** Information that does not contain direct identifiers, such as name, address or specific dates, but provides information that could be used in combination with other data to identify individuals.

***Rate:*** A measure of the frequency of an event per population unit. (See Guidelines for Using and Developing Rates for Public Health Assessment.) In these guidelines the terms rate, proportion and percent are interchangeable.

***Sensitive personal information:*** Whereas confidential personal information means information collected about a person that is readily identifiable to that specific individual, sensitive personal information extends beyond that to information which may be inferred about individuals, where that information is associated with some stigma. Examples are certain diseases, health conditions or health practices. The sensitivity of certain personal information may vary between communities.

## References

Bercovitz A, Moss A, Sengupta M, et al. An overview of home health aides: United States, 2007. National health statistics reports; no 34. Hyattsville, MD: National Center for Health Statistics; 2011. http://www.cdc.gov/nchs/data/nhsr/nhsr034.pdf Accessed May 23, 2017.

Fryar CD, Merino MC, Hirsch R, Porter KS. Smoking, alcohol use, and illicit drug use reported by adolescents aged 12-17 years: United States, 1999-2004. National health statistics reports; no 15.Hyattsville, MD: National Center for Health Statistics; 2009. http://www.cdc.gov/nchs/data/nhsr/nhsr015.pdf. Accessed May 23, 2017.

Hammill BG, Hernandez AF, Peterson ED, Fonarow GC, Schulman KA, Curtis LH. Linking inpatient clinical registry data to Medicare claims data using indirect identifiers. Am Heart J. 2009 Jun;157(6):995-1000. (http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2732025/) Accessed May 23, 2017.

Lawson EH, Ko CY, Louie R, Han L, Rapp M, Zingmond DS. Linkage of a clinical surgical registry with Medicare inpatient claims data using indirect identifiers. Surgery. 2013 Mar;153(3):423-30.

National Center for Health Statistics. *Table 1. Health insurance coverage status, coverage type, and selected characteristics, for persons of all ages, January-June 2010.* Hyattasville, MD: National Center for Health Statistics; 2010. http://www.cdc.gov/nchs/data/health_policy/Health_Insurance_Selected_Characteristics_Jan_June_2010.pdf. Accessed May 23, 2017..

National Environmental Public Health Tracking Network (NEPHTN). *Data Re-release Plan Version 2.5*. Atlanta, GA: Centers for Disease Control and Prevention, National Center for Environmental Health, Division of Environmental Hazards and Health Effects, Environmental Health Tracking Branch; 2008. (http://ephtracking.cdc.gov/docs/Tracking_Re-Release_Plan_v2.5.pdf) Accessed May 23, 2017..

National Institutes of Health. *Research Repositories, Databases, and the HIPAA Privacy Rule*. Washington, DC: U.S. Department of Health and Human Services, National Institutes of Health; Posted January 12, 2004 (revised: 7/02/04). http://privacyruleandresearch.nih.gov/research_repositories.asp. Accessed May 23, 2017..

*NCHS Staff Manual on Confidentiality*. Hyattsville, MD: Department of Health and Human Services, Public Health Service, National Center for Health Statistics; 2004. http://www.cdc.gov/nchs/data/misc/staffmanual2004.pdf. Accessed May 23, 2017..

*NPCR Technical Notes: Statistical Methods: Suppression of Rates and Counts.* Atlanta, GA: Centers for Disease Control and Prevention, National Program of Cancer Registries, Division of Cancer Prevention and Control, National Center for Chronic Disease Prevention and Health Promotion; 2008. http://www.cdc.gov/cancer/npcr/uscs/technical_notes/stat_methods/suppression.htm. Accessed May 23, 2017..

*OMB Checklist on Disclosure Potential of Proposed Data Releases.* Washington, DC: Office of Management and Budget; 1999. http://fcsm.sites.usa.gov/committees/cdac/cdac-checklist/ or https://fcsm.sites.usa.gov/files/2014/04/spwp22.pdf. Accessed May 23, 2017..

Pasquali SK, Jacobs JP, Shook GJ, et al. Linking clinical registry data with administrative data using indirect identifiers: implementation and validation in the congenital heart surgery population. Am Heart J. 2010 Dec;160(6):1099-104. (http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3011979/) Accessed May 23, 2017.

*WONDER Multiple Cause of Death 1999-2009.* Atlanta, GA: Centers for Disease Control and Prevention; 2012. http://wonder.cdc.gov/wonder/help/mcd.html#Assurance of Confidentiality.

Accessed May 23, 2017..

## Resources

Klein RJ, Proctor SE, Boudreault MA, Turczyn KM. Healthy People 2010 criteria for data suppression. Statistical Notes, no 24. Hyattsville, MD: National Center for Health Statistics; June 2002.

*NCHS Staff Manual on Confidentiality*. Hyattsville, MD: Department of Health and Human Services, Public Health Service, National Center for Health Statistics; 2004. http://www.cdc.gov/nchs/data/misc/staffmanual2004.pdf. Accessed May 23, 2017..

Federal Committee on Statistical Methodology. Confidentiality and Data Access Committee (CDAC) Resources for Confidentiality and Data Access Information. Including *OMB Checklist on Disclosure Potential of Proposed Data Releases.* Washington, DC: Office of Management and Budget; 1999. https://fcsm.sites.usa.gov/committees/cdac/cdac-resources/ Accessed May 23, 2017.

Statistics Netherlands. Statistical Disclosure Control: τ-ARGUS home page. http://neon.vb.cbs.nl/casc/..%5Ccasc%5Ctau.htm. Accessed May 23, 2017.

Sweeney L. Weaving technology and policy together to maintain confidentiality. *Journal of Law, Medicine & Ethics.* 1997;25:98-110.

## Relevant Policies, Laws and Regulations

Release of Confidential Information: Department Policy 17.006 (link accessible to department employees only)

Medical records—health care information access and disclosure: Chapter 70.02 RCW

Public records act. Chapter 42.56 RCW

Executive Order on Public Records Privacy Protections: EO 00-03. (http://www.digitalarchives.wa.gov/GovernorGregoire/execorders/recpriv/recpriv2.htm)

Vital records

- Requesting a listing or file of vital records with personal identifiers: WAC 246-490-030
  Requesting vital records information without personal identifiers: WAC 246-490-020

The following examples, provided by the department data custodians, include the major datasets used for assessment in Washington.

***Birth records:*** RCW 70.58.055 and WAC 246-491-039

***Death records:*** RCW 9.02.100 and WAC 246-490-110 (deaths related to abortion), WAC 246- 491-039 (fetal death records), RCW 70.24.105 (deaths related to HIV-AIDS).

***HIV/AIDS and other communicable disease data:*** RCW 70.24.105 and WAC 246-101.

***Hospital discharge data:*** RCW 43.70.052 and WAC 246-455.

***Cancer registry data:*** RCW 70.54.250 and WAC 246-102-070

## Appendix 1

### Detailed example of disclosure risk

Here we illustrate disclosure risk with an example from birth data. These are real Washington state data, but we have changed the county names and ZIP Codes to prevent disclosure of sensitive data.

> ZIP Code 47863 overlaps Bush and Clinton counties. In 2005, there were 82 births to mothers whose resident ZIP Code was 47863; 81 of those mothers lived in Clinton county, and 1 lived in Bush county. For the sake of this example, we pretend that no other ZIP Codes overlap Bush and Clinton counties. Let's say that one agency has provided, or posted on the Internet, a table that shows the number of prior pregnancies for birth mothers by resident ZIP Code, and another agency has provided or posted the same data by county of residence. By adding up the figures for all ZIP Codes in Clinton County, including 47863, a data user could ascertain that there was only 1 birth to a mother who lived in Bush County in ZIP Code 47863. If the data user happened to know this woman (say, as a neighbor), then the data user would know the number of her prior pregnancies. We can guard against this type of disclosure by suppressing some cells. In 2005, some of the ZIP Codes in Clinton County had fewer than 10 births, and a rule requiring suppression of those numbers would make it harder for the data user to figure out how many births were in the overlap area. A detailed explanation of the effects of suppressing counts of 1-4 or 1-9 is provided below.
>
> In practice, we cannot anticipate or analyze all of the data tables that will be released. We cannot guarantee either that a rule requiring only the suppression of counts between 1 and 4 will lead to disclosure of sensitive data, or that a rule requiring suppression of counts between 1 and 9 will prevent it. However, it is clear that the 1-9 rule will make disclosure substantially less likely.

First, we have a list of ZIP Codes by county, which shows that one ZIP Code (47863) lies in both Bush and Clinton counties:

*Table 1: ZIP Codes by county*

| | |
|---|---|
| Bush | 47863 |
| Bush | 47864 |
| Bush | 47865 |
| Bush | 47866 |
| Bush | 47867 |
| Bush | 47868 |
| Bush | 47869 |
| Bush | 47870 |
| Bush | 47872 |
| Clinton | 47863 |
| Clinton | 47873 |
| Clinton | 47883 |
| Clinton | 47884 |
| Clinton | 47885 |
| Clinton | 47886 |
| Clinton | 47887 |
| Clinton | 47888 |
| Clinton | 47889 |
| Clinton | 47890 |

| | |
|---|---|
| Clinton | 47892 |
| Clinton | 47893 |
| Clinton | 47894 |
| Clinton | 47895 |
| Clinton | 47896 |

Let's say that we have tables showing births by county of residence, and births by resident ZIP Code, and no data is suppressed (Table 2 [column 2] and Table 3). Since the sum of births in ZIP Codes that fall wholly or at least partially in Clinton County (ZIPs 47873-47896 plus ZIP 47863) is 1,422, we can deduce that there is just one birth in those ZIP Codes that is not in Clinton County (because the total for Clinton in Table 3 is 1,421), and therefore just one birth to a Bush County resident living in ZIP Code 47863. In any set of tables lacking any suppression that showed characteristics of births (such as the number of prior pregnancies) by resident ZIP Code and by county of residence, a data user could identify the characteristics of that single birth.

*Table 2: Births by ZIP Code*

| ZIP Code | Births | Births (counts of 1-4 suppressed) | Births (counts of 1-9 suppressed) |
|---|---|---|---|
| 47863 | 82 | 82 | 82 |
| 47864 | 1 | * | * |
| 47865 | 3 | * | * |
| 47866 | 34 | 34 | 34 |
| 47867 | 1 | * | * |
| 47868 | 2 | * | * |
| 47869 | 7 | 7 | * |
| 47870 | 398 | 398 | 398 |
| 47872 | 3 | * | * |
| 47873 | 148 | 148 | 148 |
| 47883 | 14 | 14 | 14 |
| 47884 | 596 | 596 | 596 |
| 47885 | 150 | 150 | 150 |
| 47886 | 43 | 43 | 43 |
| 47887 | 1 | * | * |
| 47888 | 3 | * | * |
| 47889 | 8 | 8 | * |
| 47890 | 9 | 9 | * |
| 47892 | 11 | 11 | 11 |
| 47893 | 2 | * | * |
| 47894 | 25 | 25 | 25 |
| 47895 | 229 | 229 | 229 |
| 47896 | 101 | 101 | 101 |

*Table 3: Births by county of residence*

| County | Births |
|---|---|
| Bush | 450 |
| Clinton | 1421 |

Now let's say that we have suppressed the data in all cells having a count between 1 and 4 (see column 3 in Table 2). The sum of births in non-suppressed ZIP Codes that fall wholly or at least partially in Clinton County is 1,416. A data user can see that the counts in the three ZIP Codes which are wholly in Clinton County have been suppressed, and, knowing the suppression rule, can deduce that there were between 1,419 (i.e., the sum of 1,416 and 3, assuming that each suppressed ZIP had one birth) and 1,428 (1,416 plus 12, which assumes that each suppressed ZIP had four births) births in ZIP Codes that fall wholly or at least partially in Clinton County. Since there were 1,421 births to Clinton County residents, the data user can deduce that there were 0 to 7 births to Bush County residents living in ZIP Code 47863.

The 3 Clinton County ZIP Codes in which data were suppressed had 1, 3, and 2 births. Note that if, by happenstance, these ZIP Codes had all had 4 births, then the total number of births in Clinton County would have been 1,427, and this total would have been shown in the births by county table. Then the data user, knowing that there were between 1,419 and 1,428 births in Clinton County ZIP Codes, could deduce that there were 0 or 1 births in Bush County in Zip Code 47863. If the data user knew a Bush-resident mother who lived in ZIP 47863 and gave birth in 2005, then the data user would know that was the only such mother. Additionally, this suppression rule does not suppress counts of 0, so any combination of 0 or 4 births among those 3 ZIP Codes would have allowed the data user to reach that same conclusion.

Now let's say that we have suppressed the data in all cells having a count between 1 and 9 (see fourth column in Table 2). The sum of births in non-suppressed ZIP Codes that fall wholly or at least partially in Clinton County is 1,399. A data user can see that the counts in 5 ZIP Codes in Clinton County have been suppressed, and, knowing the suppression rule, can deduce that there were between 1,404 (i.e., the sum of 1,399 and 5, assuming that each suppressed ZIP had one birth) and 1,444 (1,399 plus 45, which assumes that each suppressed ZIP had nine births) births in ZIP Codes that fall wholly or at least partially in Clinton County. Since there were 1,421 births to Clinton County residents, the data user can deduce that there were 0 to 23 births to Bush County residents living in ZIP Code 47863. An alternative realization of these data that would allow a data user to identify an individual mother as the only mother in Bush County in ZIP Code 47863 would require each of these 5 ZIP Codes to have either 0 or 9 births. This would be far less likely to happen than the scenario above, which only required 3 ZIP Codes to have 0 or 4 births.

## Appendix 2

***Washington Tracking Network rule-based use of aggregation***

The Washington Tracking Network (WTN) has an online data query system which displays data in tables, charts and maps, accessible by the public. In order to avoid automated production of tables where most rows are suppressed due to small numbers, WTN supplements its suppression rules with aggregation rules. The purpose is to aggregate data using static and dynamic parameter control in order to minimize suppression. [*Note*: dynamic parameter control methods have been implemented in the Washington State Cancer Registry website (https://fortress.wa.gov/doh/wscr/WSCR/Query.mvc/Query), and on the national Tracking Network portal (http://ephtracking.cdc.gov/), but are still in the planning stage for WTN at this time.]

When a health event is relatively rare, application of suppression rules can result in tables with many rows of suppressed data. Users find these tables to be extremely frustrating. Small subpopulations invariably lead to small numbers. Aggregation yields larger numbers, although stratification is needed to focus analysis, so a balance is desirable.

Fields in a dataset are commonly termed "parameters" in the context of data query systems. Parameter control can be achieved through use of static methods (within a parameter) or dynamic methods (between parameters). Dynamic parameter control is also termed "adaptive stratification." Optimal parameter control includes protocol-driven use of both static and dynamic methods.

With static parameter control, some strata can be blocked by design, limiting tables to those based on greater aggregation. Examples are: displaying only multi-year data, not annual data (temporal aggregation); or, displaying only multi-county data, not county-level data (spatial aggregation). Parameters can also be excluded entirely, as when a dataset field is not relevant to program planning or evaluation. The static parameter control design rules should be reviewed with data stewards and program partners, who may want to make refinements. The key basis for the application of static parameter control design rules is program/planning utility.

> The story of the asthma data online query system developed jointly by American Lung Association of Washington (ALAW) and the Washington State Department of Health (department) in the early 2000s is illustrative. The data shared by the department with ALAW for the query system potentially could have contained very tiny numbers, if stratified by age and gender simultaneously. The department proposed to share only one of these fields, but not both. ALAW members and department asthma program staff decided that, because intervention and prevention programs differ by age (there are programs for children and separate programs for adults), but not by sex, they wanted to see age strata in the data tables. The department excluded the gender parameter.

WTN rules for static parameter control start with count-based thresholds for Stratum Exclusion:

<u>Spatial</u>
- if <200 cases/year, then only multi-county regions available (no single county display)
- if <100 cases/year, then only state-level available (no multi-county regions or single county display)

<u>Temporal</u>
- if <400 cases/year, then only 5-year rollup available (no single year or 3-year rollup)
- if <800 cases/year but 400+ cases/year, then only 3-year rollup available (no single year)

| From | To | Temporal | Spatial |
|---|---|---|---|
| 800 | above | Single year | County |
| 400 | <800 | 3-year rollup | County |
| 200 | <400 | 5-year rollup | County |
| 100 | <200 | 5-year rollup | MCR |
| | <100 | 5-year rollup | statewide |

Consultation with data stewards and program partners has often modified these rules. For example, in order to display annual data, greater spatial aggregation can be used. Once these rules are decided upon, they become static.

With dynamic parameter control, disaggregation is dependent on interactive query choices. In other words, adaptive stratification is interdependent, conditional on whether other parameters are aggregated. With small numbers, we want more aggregation; with larger numbers, we want less aggregation. WTN separates various topic areas in differing levels for adaptive stratification, termed AS Levels.

- With an AS1 (very small numbers), only one stratification parameter is available at a time; for example, if a user selects disaggregation by geography, then the remainder of parameters are fully aggregated.
- With an AS3 (mid-range numbers), three stratification parameters are available at a time; for example, if a user selects disaggregation by geography, time and gender (e.g., annual county-level by gender), then the remainder of parameters are fully aggregated.
- With an AS5 (large numbers), five stratification parameters are available at a time; for example, if a user selects disaggregation by geography, time, age group, gender and race (e.g., annual county-level by age, race and gender), then the remainder of parameters are fully aggregated.

The WTN thresholds for Adaptive Stratification are:

- AS1 = < 100 cases per year statewide
- AS2 = 100-499 cases per year statewide
- AS3 = 500-999 cases per year statewide
- AS4 = 1000-4999 cases per year statewide
- AS5 = 5000-99,999 cases per year statewide
- AS6 = 100,000+ cases per year statewide

This WTN practice is a rule-based protocol. Thresholds between adjacent levels of Adaptive Stratification are independent of topic area (i.e., standardized across all topic areas).