

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 peer-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 SOLELY on syndromic data.