



THE STATE
of **ALASKA**
GOVERNOR BILL WALKER

**Department of
Health and Social Services**

DIVISION OF PUBLIC HEALTH
Section of Epidemiology

3601 C Street, Suite 540
Anchorage, Alaska 99503
Main: 907.269.8000
Fax: 907.562.7802

**Alaska Section of Epidemiology
Confidentiality Policies and Procedures and
Data Release Protocols
Updated February 2018**

Summary Data Release Protocol14

<http://dhss.alaska.gov/dph/Epi/Pages/confidentiality/default.aspx>

II. Summary Data Release Protocol

Requests for Alaska Section of Epidemiology (SOE) data should be submitted on an SOE Summary Data Request and Utilization Agreement Form (see Attachment 3).

The purpose of this protocol is to protect the confidentiality of patient information when SOE releases public health data to external stakeholders. Summary data are information grouped by age, sex, geographic area, or other variables and displayed so that individual patients cannot be directly identified. Summary data may be presented as a table, figure, diagram, chart, narrative, line list, or other similar format. Examples of summary data are the annual infectious disease reports, the sexually-transmitted disease summaries, and the HIV and AIDS summaries routinely published in the Epidemiology *Bulletin*. Summary data also may be produced on an *ad hoc* basis for various agencies and entities upon request.

In general, data will be reported to requestors in counts and not as rates. Rates may be requested; however, depending on the counts involved and the purpose for the data, the Section Chief may decide that rate reporting is not appropriate for certain situations. If researchers or other stakeholders would like more specific information about incident cases of a reportable condition than the rule of ones allows, such requests will be reviewed on a case-by-case basis, in consultation with the Epidemiology Section Chief, taking into consideration the potential risks and benefits of such disclosure.

Although summary data do not include confidential data such as a patient's name, summary data may lead to *de facto* identification of a particular person if the combination of age, sex, place of residence, or other variable(s) defines only one person — this is particularly important for communities with a relatively small population. Both numerators and denominators should be considered when releasing data; it is never acceptable to release summary data that could reasonably be expected to lead to the identification of an individual patient through indirect means. Because of this potential for a breach of confidentiality, great caution must be exercised in the distribution and use of such data.

Before summary data are released, SOE program managers must carefully review and approve the data format to ensure that the release is consistent with the guidelines in this protocol. If there is a question as to whether the release is consistent with the guidelines in this protocol, program managers should consult with the Section Chief prior to granting approval. In some circumstances, the guidelines might not be sufficiently restrictive to prevent identification of individuals because of the distribution of the health condition or the population affected. In those instances, parameters more restrictive than the general guidelines should be instituted before data can be released. Program Managers are responsible for maintaining records of data requests and subsequent data releases.

In certain situations, a limited data set (LDS) may be requested and subsequently released by SOE following review and approval by both the appropriate Program Manager and the Section Chief. See Section G of the SOE Confidentiality Policies and Procedures document for more information about LDS. Record-level data involving personal identifiers will not generally be released except by formal application and approval by the Section Chief. Such data will be released to researchers or public health partners only for approved research and surveillance activities with a signed data

sharing agreement and/or a signed memorandum of understanding, per the Section Chief's discretion.

Overview of General Algorithm/Process

1. Apply the scoring schema from Table 1, Alaska "Rule of Ones" (see below).
 - Adjust variables as appropriate to obtain a score ≥ 1 .
2. Ensure that all cells with a numerator value < 5 are evaluated for possible confidentiality concerns. Suppress data in these cells or aggregate to generate larger cell sizes as appropriate.
3. Ensure that all cells with a denominator of < 500 people are evaluated for possible confidentiality concerns.
 - Avoid sub-stratification if cell denominators are < 250 people.
4. Review final data to be released to ensure that users cannot derive confidential information through a process of subtraction.
5. Ensure that standard caveats about the limitations or generalizability of the data accompany the data to the requestor, including caveats regarding stability of rates calculated with small numbers, etc.

Alaska "Rule of Ones"

Table 1 is adapted from the New Hampshire Division of Public Health Services¹ and involves the "Rule of Ones" for six different characteristics or variables described below. Each characteristic is given a value; the product of the six values must be ≥ 1 to allow for acceptable release of data. Details are given about what constitutes a value of one, and how to adjust that value up or down based on the granularity of the characteristic. In addition, regardless of adjustments, minimum and maximum values are given for each characteristic.

Notes:

- For *annual* state-wide reports of common diseases, sex, race, and age-group (5-year intervals) will usually be given and may be given for certain rare diseases.
 - The "Rules of Ones" applies mainly to counts of disease; however, the general guidelines also apply to the subsequent calculation of rates, i.e., avoiding calculation when the numerator is < 5 or the denominator is < 500 .
- a. **Incidence rate:** What is the disease of interest? More common diseases (e.g., chlamydia) are assigned a value of 1; less common diseases (e.g., botulism) may be discounted, per the discretion of the Section Chief. As a general rule of thumb, if the incidence of the disease is < 5 cases per 100,000 population per year, use $1/2$.
- Minimum value:* $1/2$ *Maximum value:* 1

¹New Hampshire Division of Public Health Services, Health Statistics and Data Management Section. Guidelines for the Public Release of Public Health Data. September 2008. Available at: <http://www.dhhs.nh.gov/dphs/hsdm/documents/publichealthdata.pdf>.

- b. **Population size:** In what region or population strata will the data be provided (total denominator across all age and sex groups)? A population or strata of 5,000 is assigned a value of 1; therefore, populations/strata of 10,000 have a value of 2 and populations/strata with 2,500 have a value of 1/2.
- Use the current Alaska Population Overview as the reference for the population size in a region <http://live.laborstats.alaska.gov/pop/popestpub.cfm>. Note that the size of boroughs and census areas range from <500 to ~300,000 people.
 - Community sizes should be rounded down to the nearest 500.
 - See #3 in the overview above for communities with populations <500 people.
- Minimum value: 1/10 Maximum value: 50, for any population 250,000+*
- c. **Time interval of data:** In what unit of time, months or years, will the data be provided? A year is assigned a value of 1; a single month has a value of 1/12. Data will not be provided in intervals smaller than 1 month.
- Minimum value: 1/12 Maximum value: 5, for any interval ≥ 5 years*
- d. **Age-group:** In what number of years will age-groups be provided? A 5-year age-group is assigned a value of 1; therefore, a 10-year age-group has a value of 2. Note that data with age-groupings of less than 5 years will generally not be provided. Age-grouping must be divisible by five; and map to standard intervals if possible, i.e., 0–4 years of age instead of 1–5 years of age.
- Minimum value: 1 Maximum value: 5, for ≥ 25 -year age-groups*
- e. **Sex distribution:** Will data be stratified by sex? If data are not stratified by sex, a value of 1 is assigned. If data are stratified by sex, a value of 1/2 is assigned.
- Minimum value: 1/2 Maximum value: 1*
- f. **Race distribution:** Will data be stratified by race? If data are not stratified by race, a value of 1 is assigned. If data are stratified by all four race groupings (e.g., white, American Indian/Alaska Native [AI/AN], black, Asian/Pacific Islander), value of 1/8 is assigned. If data are stratified by AI/AN and non-AI/AN only, a value of 1/5 is assigned.
- Minimum value: 1/8 Maximum value: 1*

Table 1. Scoring of Criteria Used to Evaluate Acceptable Summary Data Releases.

Characteristic	Standard Denomination	Score
a. Incidence rate	Common	1
b. Population size	5,000	1
c. Time interval	1 year	1
d. Age-group	5-year	1
e. Sex distribution	All	1
f. Race distribution	All	1
Product (a)(b)(c)(d)(e)(f)		1

Example A. Is it acceptable to release 2008 data for chlamydia in Palmer (2009 population 5,500) by 5-year age-groups?

Characteristic	Standard Denomination	Score	Example A
a. Incidence rate	Common	1	Chlamydia = 1
b. Population size	5,000	1	Palmer = 1
c. Time interval	1 year	1	2008 = 1
d. Age-group	5-year	1	5-year = 1
e. Sex distribution	All	1	All = 1
f. Race distribution	All	1	All = 1
Product (a)(b)(c)(d)(e)(f)		=	1

Answer: *Score ≥ 1 ; this is acceptable.*

Example B. Is it acceptable to release 2008 data for chlamydia in Palmer (2009 population 5,500) by 5-year age-group and by sex?

Characteristic	Standard Denomination	Standard Score	Example B
a. Incidence rate	Common	1	Chlamydia = 1
b. Population size	5,000	1	Palmer = 1
c. Time interval	1 year	1	2008 = 1
d. Age-group	5-year	1	5-year = 1
e. Sex distribution	All	1	M vs. F = 1/2
f. Race distribution	All	1	All = 1
Product (a)(b)(c)(d)(e)(f)		=	1/2

Answer: *Score < 1 ; this is not acceptable.*

Example C. But, what if age-group was aggregated to be 10-year increments?

Characteristic	Standard Denomination	Standard Score	Example C
a. Incidence rate	Common	1	Chlamydia = 1
b. Population size	5,000	1	Palmer = 1
c. Time interval	1 year	1	2008 = 1
d. Age-group	5-year	1	10-year = 2
e. Sex distribution	All	1	M vs F = 1/2
f. Race distribution	All	1	All = 1
Product (a)(b)(c)(d)(e)(f)		=	1

Answer: *Score ≥ 1 ; this is acceptable.*

Example D. Is it acceptable to release 2008 data for botulism by sex, race, and 5-year age-group for the entire state?

Characteristic	Standard Denomination	Standard Score	Example D
a. Incidence rate	Common	1	Botulism = 1/2
b. Population size	5,000	1	AK = 50
c. Time interval	1 year	1	2008 = 1
d. Age-group	5-year	1	5-year = 1
e. Sex distribution	All	1	M vs F = 1/2
f. Race distribution	All	1	4 groups = 1/8
Product (a)(b)(c)(d)(e)(f)			50/32 (1.6)

Answer: Score ≥ 1 ; this is acceptable.

Example E. Is it acceptable to release 2008 data for gonorrhea by sex, race, and 5-year age-group for Palmer (2009 population 5,500)?

Characteristic	Standard Denomination	Standard Score	Example D
a. Incidence rate	Common	1	Gonorrhea = 1
b. Population size	5,000	1	Palmer = 1
c. Time interval	1 year	1	2008 = 1
d. Age-group	5-year	1	5-year = 1
e. Sex distribution	All	1	M vs. F = 1/2
f. Race distribution	All	1	4 races = 1/8
Product (a)(b)(c)(d)(e)(f)			1/16

Answer: Score < 1 ; this is not acceptable.

Example F. Is it acceptable to release data for gonorrhea by sex, AI/AN race for Palmer by increasing age-group to 10-year increments and looking for 5 years of data?

Characteristic	Standard Denomination	Standard Score	Example D
a. Incidence rate	Common	1	Gonorrhea = 1
b. Population size	5,000	1	Palmer = 1
c. Time interval	1 year	1	2006-10 = 5
d. Age-group	5-year	1	10-year = 2
e. Sex distribution	All	1	M vs F = 1/2
f. Race distribution	All	1	AI/AN vs non = 1/5
Product (a)(b)(c)(d)(e)(f)			1

Answer: Score ≥ 1 ; this is acceptable.

Example G. How many variables could you release for gonorrhea cases in a small community, i.e., Fort Yukon (2009 population 585)?

Characteristic	Standard Denomination	Standard Score	Example D
a. Incidence rate	Common	1	Gonorrhea = 1
b. Population size	5,000	1	Fort Yukon = 1/10
c. Time interval	1 year	1	2001-5 = 5
d. Age-group	5-year	1	10-year = 2
e. Sex distribution	All	1	All = 1
f. Race distribution	All	1	All = 1
Product (a)(b)(c)(d)(e)(f)			
		=	1

Answer: To have an acceptable release of data for a common disease in Fort Yukon, you would need to use a 5-year period and release age data in 10 year intervals. Stratification by sex would only be possible if the age-group interval was expanded to ALL (product score =25/20). Stratification by race would not be possible.