

Guidelines for the Release of Public Health Data Derived from Personal Health Information

The Public Health and Safety Division (PHSD) collects a variety of Protected Health Information (PHI) through mandatory and voluntary reporting systems under Montana statutes and federal laws. These data are confidential and access to them is strictly regulated by the Montana Constitution,¹ Montana statutes,² and the federal Health Insurance Portability and Accountability Act (HIPAA).³ This document establishes Guidelines for the release of data derived from PHI for public health activities while protecting its confidentiality and integrity in compliance with state statutes and federal laws.

Data Releases Covered by These Guidelines

These Guidelines apply to all materials published by PHSD, and all replies to queries received by PHSD regardless of the source of the queries, that are based on data derived from PHI. Whenever summary data leave the immediate control of PHSD, they must be considered published. This includes, but is not limited to, providing data to local health jurisdictions within the state. In the absence of a formal confidentiality agreement, no data may be provided to a local health jurisdiction that would not be publishable under these Guidelines. Conversely, any data released under these Guidelines to any organization or individual must be made available, without discrimination, to any other organization or individual that requests the data.

Federal laws and state statutes recognize the need for public health entities to have access to data at fine levels of resolution, but also specify that the information released be *the minimum amount necessary to accomplish essential public health functions*.⁴ Many local health jurisdictions request very detailed data, but such detail may not be necessary to conduct their designated public health functions. The desire of local jurisdictions to obtain data on small numbers of events, rendering individuals potentially identifiable, must be weighed against the state's obligation to protect individual confidentiality. Depending on the specific request, the PHSD may need to enter into formal agreements with local jurisdictions to ensure adequate protection of confidentiality. Such confidentiality agreements must include the explicit agreement that local jurisdictions will comply with the provisions in these Guidelines.

Data Releases Not Covered by These Guidelines

Some programs within PHSD have established agreements to share PHI with other programs within PHSD. Those agreements are generally formalized in writing and include provisions prohibiting the disclosure of the data to third parties. These Guidelines do not affect those internal data sharing agreements.

Some programs within PHSD have legal or contractual requirements to report PHI at varied levels of detail or identification to specific third parties. These reporting requirements generally fall under HIPAA's Covered Entity provisions and Montana code 50-16-603, and may be covered by additional formal data sharing agreements. These Guidelines do not affect those reporting requirements or data sharing agreements. Program-specific reports or summaries that may require additional detail to be released may be allowed after appropriate review by the state Medical Officer and State Epidemiologist to ensure confidentiality is preserved.

Some programs within PHSD have formal processes to respond to requests for data sets containing PHI from scientific investigators. The processes include detailed confidentiality agreements and usually include IRB approvals. These Guidelines do not affect those data sharing processes.

Some programs within PHSD conduct anonymous surveys of Montana citizens. Those survey results do not collect individually identifiable PHI. These Guidelines do not apply to reporting the results of those surveys.

Confidentiality

All identifying information is confidential and may not be released. Identifying information includes name, address, birth date, social security number, *or any other information which, alone or in combination with other information, could be used to determine with reasonable accuracy the identity of an individual.*

With the wide availability of powerful computer software and the proliferation of publicly accessible databases containing individual information, the potential for probabilistic or deterministic re-identification of individuals included in apparently de-identified statistical reports has become an increasing threat to personal privacy. The risk of re-identification increases as both the numerator and the denominator of the data decrease. The numerator and denominator decrease with each subdivision of the data into strata such as time interval, sex, age group, race, or county.

In order to ensure confidentiality for Montana citizens whose PHI is collected and analyzed by the PHSD, it is necessary to suppress the reporting of small numbers of events (counts). Minimum cell sizes for reporting can be achieved by aggregating strata or several years of data.

Precision

The PHSD is responsible for providing summary statistics that are precise and therefore useful for planning, evaluation, and comparisons. The most common summary statistics in public health are rates. Rates and other summary statistics based on small numbers of events, or derived from small segments of the population, or both, are inherently imprecise. Imprecision can lead to incorrect, potentially wasteful, and potentially harmful public health decisions.

Precision is measured by the Confidence Interval (CI) around a rate.⁵ It is conventional to use the 95% CI, which contains the true rate with 95% certainty. If a CI is wide, the point estimate is not precise. As a rule of thumb, a CI with a width less than +/- 10% of the point estimate would be considered very precise and a CI +/- 20% is adequately precise for some purposes. However, the degree of precision needed for a given public health activity or decision may be more or less stringent. Users of public health data sets are encouraged to examine and evaluate the width of the CIs for their own needs.

Computed rates usually achieve adequate precision if the number of events is 20 or more.⁶ In view of the small number of many health events in Montana, an event threshold of 20 provides adequate precision without unduly restricting reporting.

In order to ensure adequate precision of summary statistics derived from PHI collected by the PHSD, it is necessary to suppress rates and other summary statistics based on small numbers of events. Precision can be improved by aggregating strata or several years of data.

Basis of the Guidelines

The Data Reporting Criteria outlined in Table 1 (page 3) are a compromise between the need to use public health data to the fullest extent possible while protecting confidentiality and ensuring acceptable precision. The Montana Office of Epidemiology and Scientific Support of the PHSD reviewed a variety of reference materials and source documents from the federal and other state governments to arrive at these Guidelines and Criteria.⁷

Table 1. Criteria for Reporting Public Health Data Derived from Protected Health Information

**Statewide data
(no strata)**

If number of events ≥ 20	Report counts and rates with confidence intervals †
If number of events < 20 and ≥ 5	Do not compute rates; report counts
If number of events < 5	Reporting counts is discretionary

**Stratified data
(by demographic characteristics, county, etc.)**

If denominator of cell ≥ 300

If number of events in cell ≥ 20	Report counts and rates with confidence intervals †
If number of events in cell < 20 and ≥ 5	Do not compute rates; report counts
If number of events in cell < 5	Do not compute rates; suppress count(s) ‡§ or aggregate strata or years

If denominator of cell < 300

If number of events in cell ≥ 20	Report counts and rates with confidence intervals †
If number of events in cell < 20	Do not compute rates; suppress count(s) ‡§ or aggregate strata or years

† All rates should be accompanied by confidence intervals. Please refer to Appendix A for guidance to compute exact Poisson confidence intervals for small numbers of events ($n = 20$ to 99). For 100 or more events, the Normal Approximation may be used to compute confidence intervals.

‡ May require suppression of complementary cell(s) if number in suppressed cell(s) can be computed from other cells in row or column.

§ Cell should contain symbol indicating < 5 or < 20 events. Zero (0) may be reported if applicable.

Appendix A

Using the Poisson Exact Method to Compute the 95% Confidence Interval for 20 to 99 Events

Example: Refer to Table 2 on the next page to look up the critical values corresponding to the number of events.

If there were 52 events in a population of 129,936, the crude rate is

$$(52/129,936) * 100,000 = 40.0 / 100,000$$

From the table below, the critical values for 52 events are

Lower bound critical value = 38.8

Upper bound critical value = 68.2

Computation:

$$\text{Lower bound} = (38.8/129,936) * 100,000 = 29.9$$

$$\text{Upper bound} = (68.2/129,936) * 100,000 = 52.5$$

Table 2. Critical Values for Computing 95% Confidence Intervals for Small Numbers of Events by the Poisson Exact Method.

6

Number Of Events	95% CI: Lower Critical Value	95% CI: Upper Critical Value	Number Of Events	95% CI: Lower Critical Value	95% CI: Upper Critical Value
20	12.2	30.9	60	45.8	77.2
21	13.0	32.1	61	46.7	78.4
22	13.8	33.3	62	47.5	79.5
23	14.6	34.5	63	48.4	80.6
24	15.4	35.7	64	49.3	81.7
25	16.2	36.9	65	50.2	82.8
26	17.0	38.1	66	51.0	84.0
27	17.8	39.3	67	51.9	85.1
28	18.6	40.5	68	52.8	86.2
29	19.4	41.6	69	53.7	87.3
30	20.2	42.8	70	54.6	88.4
31	21.1	44.0	71	55.5	89.6
32	21.9	45.2	72	56.3	90.7
33	22.7	46.3	73	57.2	91.8
34	23.5	47.5	74	58.1	92.9
35	24.4	48.7	75	59.0	94.0
36	25.2	49.8	76	59.9	95.1
37	26.1	51.0	77	60.8	96.2
38	26.9	52.2	78	61.7	97.3
39	27.7	53.3	79	62.5	98.5
40	28.6	54.5	80	63.4	99.6
41	29.4	55.6	81	64.3	100.7
42	30.3	56.8	82	65.2	101.8
43	31.1	57.9	83	66.1	102.9
44	32.0	59.1	84	67.0	104.0
45	32.8	60.2	85	67.9	105.1
46	33.7	61.4	86	68.8	106.2
47	34.5	62.5	87	69.7	107.3
48	35.4	63.6	88	70.6	108.4
49	36.3	64.8	89	71.5	109.5
50	37.1	65.9	90	72.4	110.6
51	38.0	67.1	91	73.3	111.7
52	38.8	68.2	92	74.2	112.8
53	39.7	69.3	93	75.1	113.9
54	40.6	70.5	94	76.0	115.0
55	41.4	71.6	95	76.9	116.1
56	42.3	72.7	96	77.8	117.2
57	43.2	73.9	97	78.7	118.3
58	44.0	75.0	98	79.6	119.4
59	44.9	76.1	99	80.5	120.5
			100	81.4	121.6

7

¹ <http://data.opi.mt.gov/bills/mca/Constitutions/II/10.htm>

² http://data.opi.mt.gov/bills/mca_toc/50.htm

³ <http://www.hhs.gov/ocr/privacy/hopaa/understanding/special/publichealth/index.html>: 45 CFR 164.502, 45 CFR 164.512, 45 CFR 164-514

⁴ <http://www.hhs.gov/ocr/privacy/hipaa/understanding/spcieal/puclichealth/index.html>: 45 CFR 164.502(b), 164.514(d), 45 CFR 164.513(b)

⁵ KJ Rothman et al. 2008. *Modern Epidemiology*, 3rd ed. Philadelphia: Lippincott Williams and Wilkins

⁶ . Snedecor GW, Cochran WG. 1989. *Statistical Methods*, 8th ed. Ames: Iowa State University Press; Casella G, Berger RL. 2002. *Statistical Inference*, 2nd ed. Pacific Grove, CA: Duxbury/Thomas Learning.

⁷ . <http://www.cdc.gov/mmwr/preview/mmwrhtml/ss5806a2.htm>; <http://www.heath.state.ny.us/diseases/chronic/ratesmall.htm>; <http://www.doh.wa.gov/Data/Guidelines/Rateguide.htm>; <http://health.utah.gov/opha/IBIShelp/DataSuppression.pdf>; www.doh.wa.gov/Data/guidelines/WordDocs/SmallNumberes.doc