AGGREGATE DATA SUPPRESSION POLICY
(Issued 01/14/2025, Policy 25-01)
Date Reviewed: 01/14/2025
Date Revised:
Office: Office of the Director
Division: N/A
Category: General Administration

Table of Contents

A. PURPOSE

Under privacy laws, such as the federal Health Insurance Portability and Accountability Act (HIPAA) for covered programs, the Illinois Personal Information Protection Act (PIPA), the Illinois Health Statistics Act (410 ILCS 520/), Illinois Access to Records of the Department of Public Health (Title 2, Part 1127), and some agreements, institutions reporting or releasing data are required to protect personally identifiable health information shared with the public. Thus, this policy is intended to assist data users in ensuring that data is appropriately suppressed to prevent the identification of individuals in aggregate population health datasets and tables, particularly with underlying small populations or rare conditions, to promote consistency in the release of aggregated data while balancing individual privacies and the public right to know**.**

This policy outlines the procedures and guidelines for the suppression of aggregate population health data to protect individual privacy and ensure the confidentiality of health information.

B. SCOPE

This policy must be followed by Illinois Department of Public Health (IDPH) employees, contractors, local health departments (LHDs), and other IDPH partners, etc. who collect, analyze, publish, and disseminate, or otherwise release IDPH population-based aggregate data in any form. **This does not apply to data shared within IDPH between data analysts for programmatic operation or people authorized to receive data through legislation, grant requirements, data use agreements, and authorized delegate agents.** In addition to these standards, those using IDPH data must be familiar with relevant federal and Illinois state laws, regulations, and IDPH policies. More restrictive **federal and state laws and regulations supersede suppression standards provided in this document.**

If needed, alternative data suppression parameters for ad-hoc aggregate data requests, media requests, and open records requests may also be assessed on an individual basis by the Data Governance Board or Institutional Review Board.

The release of individual health information is out of the scope of this policy. Refer to the Data Specific Statutes and Regulation Section below for additional information.

C. DATA SPECIFIC STATUTES AND REGULATIONS

In addition to the general acts and rules that control data release, most IDPH datasets have specific Illinois statutes and regulations that govern how the data must be managed. Some examples include the Illinois Health and Hazardous Substances Registry Act, Immunization Data Registry Act, Vital Records Act, and the Genetic Information Privacy Act. There are also applicable administrative rules adopted pursuant to these statutes under the Illinois Administrative Code that govern the release of these specific data sets.

A list of Illinois laws controlling the release of data is provided here: Final Assessment and Analysis Report.pdf (illinois.gov).

D. RATIONALE

Why are small numbers a concern in public health assessment?

At least three types of concerns are raised regarding the publication of aggregate statistics from population data sources when those statistics involve relatively few cases.

1. Useful and reliable quantitative information is critical for local health officials, program and policy developers and managers, health care providers, advocacy groups, and the media. Questions concerning health outcomes, related health behaviors, and environmental factors often are studied within small subgroups of a population. Statistics for rare cases are susceptible to misinterpretation and misunderstanding. Small changes in counts can result in significant changes in rates. Some calculations, such as age-adjusting rates, may compromise reliability when denominators are small.
2. Legal and ethical considerations require that the privacy and confidentiality of persons regarding their health and health care be respected. Violations may injure individuals and their families and compromise the future reliability of population data sources.
3. Data owners and decision-makers have statutory or assigned responsibilities to protect the privacy of individuals in data systems. Without clear, evidence-based rules, decisions often may rely on convention, custom, and personal intuition or comfort levels, leading to inconsistent practice.

In the context of aggregate data from whole populations, the question is whether someone would *learn something they did not already know* about a person who is included in an aggregate statistic, such as a cell count (i.e., number of individuals represented in a specific category or data field). A breach of confidentiality occurs when data users release information in a way that allows an individual to be identified and reveals confidential information about that person to unauthorized individuals. This policy ensures that any tabulated dimensions, such as geography, time period, health condition, and demographic group, must be suppressed adequately to prevent the unauthorized disclosure of information. This document provides cues to situations that present some risk to privacy and suggestions on how to reduce this risk.

The policy mandates that:

A. Any data element in a cell (e.g., admissions, discharges, patients, services, etc.) with a value between 1 and 9 cannot be reported directly.
B. Cells with a zero value are exempt from this minimum cell size restriction.
C. Furthermore, a reasonable effort should be made so that no data is reported in a manner that would allow a value between 1 and 9 to be calculated from other reported cells or information (e.g., linked data). For instance, using percentages or other mathematical formulas that, when combined with other reported data, reveal a cell value between 1 and 9 is prohibited.
D. No hospital or health care facility level data can be released that would reveal proprietary business data about the hospital or identify an individual supplying data to IDPH unless it is mandated that IDPH release that data (e.g., Hospital Report Card data).
E. When datasets are combined, the suppression rules should follow the most stringent requirements from the contributing datasets.

These guidelines should accommodate most IDPH programs' data suppression needs and requirements. Some programs have established national standards as grant requirements (e.g., the North American Association of Central Cancer Registries standards used by the national cancer surveillance system). Data owners and stewards will be responsible for ensuring that additional national standards and grant-required suppression measures are applied. There are specific populations for which additional data protection in aggregate needs to be applied, beyond cell size suppression only, which will be determined on a case-by-case basis by IDPH data owners and data stewards.

Most HIPAA identifiers[1] can never be published in aggregate data. Personally identifiable information (PII) and protected health information (PHI) free text fields should not be released; this includes any unstructured data that could inadvertently reveal individual identities or confidential details.

Data aggregation at intersections of multiple characteristics (e.g., ethnicity and gender) should be done cautiously to ensure individual identities are protected. For example, listing a female member of a rare indigenous tribe with a health condition increases the risk of identification and is considered a privacy breach.

Data users have several strategies to comply with the minimum cell size requirements specified in the IDPH cell size suppression policy. Common strategies to avoid displaying cell values between 1 and 9 are listed in Section E below.

E. SUPPRESSION STRATEGIES

1. **Combining Categories:** The primary method used to lower the risk of reidentification is to reduce the granularity of the data and thus lower the risk of re-identification. This can be done in several ways:
    a. **Limiting Geographic Detail:** To prevent the identification of individuals in small geographic areas, report data at broader geographic levels (e.g., state or regional levels) rather than more granular ones (e.g., ZIP code or census tract.) *For example, i*n the case of a rare disease, where there are five cases in each of three counties, the data could be aggregated and reported at a broader geographic level, such as by region or state.
    b. **Expanding Time Parameters:** Time parameters can be expanded to month, year, five-year, or 10-year intervals. *For example, f*or a rare disease with five annual cases in a county, the data could be reported over a five-year period as 25 cases rather than annually.
    c. **Grouping Sex and Gender or Race and Ethnicity Data**: Use an “other” category for, or do not report on, rare sex, gender, race, or ethnicity categories that do not have sufficient members to remain deidentified. *For example, r*ather than publishing data including Black, White, Asian, and **Native American**, you might consider publishing Black, White, and Other rather than suppressing many cells.

---

[1] Names of individuals, specific geographic identifiers, facility specific information, birth dates, admission dates, discharge dates, death dates, specific ages over 89, telephone numbers, vehicle identifiers, fax numbers, device identifiers, email addresses, URLs, Social Security numbers, internet protocol (IP) addresses, medical record numbers, biometric identifiers, health plan beneficiary numbers, full-face images, account numbers, any other unique identifying number, characteristic, or code, certificate/license numbers.

d. **Top or Bottom Coding Age**: Highest and lowest values in a dataset will be limited to predefined ranges, particularly extreme values, to protect the privacy of individuals. *For example*, use a < 18 or >= 90 age group to group rare age groups. This can be done even when using 5- or 10-year age group intervals.

e. **Group Outcomes**: Aggregate outcome data within rare populations to prevent identification. *For example*, combine data for all types of sensory impairments when reporting on narrower age or race groups.

2. Sufficient Suppression

Applying a threshold-based suppression rule to any cell that contains between 1 and 9 (called primary suppression) increases the risk of deriving the value of a suppressed cell by mathematical calculation. Therefore, complementary or secondary cell suppression may be needed. For example, in the dataset below, if the age category 85+ cell size is five, then adequate suppression for the 85+ age category requires complementary suppression of another cell. Usually, it will make sense to suppress the next smallest cell. If you have more than one variable in your dataset or table, it will be necessary to ensure that at least two cells are suppressed in each row and column in which a cell is suppressed.

Age groups	Data points
60-69	1,000
70-74	1,900
75-79	500
80-84	100
85+	5
Total	3,505

Age groups	Data points
60-69	1,000
70-74	1,900
75-79	500
80-84	**
85+	**
Total	3,505

The suppression strategies outlined above serve as an initial screening tool to assess potential re-identification risks in a data set. If these conditions are not met, and the data user believes that it is important to publish the data in the proposed form, then the data user should apply an alternative risk assessment method to evaluate the risk.

F. RISK ASSESSMENT SCORING METHOD

Before releasing data, it is essential to analyze the specific characteristics of the data set. This includes understanding how intersections between variables, such as race, age, or geographic location, may result in small numbers that could increase re-identification risks. The Risk Assessment Scoring Criteria combines several conditions that increase the risk of a given data table and allows IDPH to evaluate those risks in combination with each other. The variables included in the Scoring Criteria are routinely used to publish data but are not all-inclusive. This scoring method helps identify small values considered sensitive and combines various conditions that increase the risk of a given data table. The scoring is based on two main factors:

- **Size of Potential Population:** Smaller populations receive higher scores.
- **Variable Specificity:** Variables with more specific or unique categories receive higher scores.

This method requires a score less than or equal to 12 for the data table to be released without additional data masking. When identifying the score for each variable, use the highest scoring criteria. For example, if a table had age groups of 0-11 years, 12-14 years, and 15-18 years, then the score for the age range variable would be +5 because the smallest age range is 12-14, which is an age range of three years. If a variable has a greater granularity than the score listed, use the highest score listed. For example, if the data were to be published weekly, then the score would be +5, which is the maximum score associated with the most granular level monthly.

Any score over 12 will require using suppression strategies outlined above in Section E or a decision from the Data Governance Board.

Variable	Characteristics	Score
Events (Numerator)	1,000+ events in a specified population	+2
	100-999 events	+3
	11-99 events	+5
	<11 events	+7
Sex	Male or Female	+1
Age Range	>10-year age range	+2
	6–10-year age range	+3
	3–5-year age range	+5
	1–2-year age range	+7
Race Group	White, Black or African American	+1
	White, Asian, Black or African American	+2
	White, Asian, Black or African American, American Indian, Alaska Native, Native Hawaiian, Other Pacific Islander, or Mixed	+3
	Detailed race	+4
Ethnicity	Hispanic or Latino - yes or no	+2
	Detailed ethnicity	+4
Race/Ethnicity Combined	This applies when race and ethnicity are collected in a single data field	
	White, Black or African American, Hispanic or Latino	+1

	White, Asian, Black or African American, Hispanic or Latino	+2
	White, Asian, Black or African American, Hispanic or Latino, American Indian, Alaska Native, Native Hawaiian or Other Pacific Islander, Mixed	+3
	Detailed race/Ethnicity	+4
Time – Reporting Period	5 years aggregated	-5
	2-4 years aggregated	-3
	1 year (e.g., 2001)	0
	Bi-annual	+3
	Quarterly	+4
	Monthly	+5
Residence Geography*	State or geography with a population >2,000,000	-5
	Population 1,000,001 - 2,000,000	-3
	Population 560,001 - 1,000,000	-1
	Population 250,000 - 560,000	0
	Population 100,000 - 250,000	+1
	Population 50,001 - 100,000	+3
	Population 20,001 - 50,000	+4
	Population ≤ 20,000	+5
Service Geography*	State or geography with a population >2,000,000	-5
	Population 1,000,001 - 2,000,000	-4
	Population 560,001 - 1,000,000	-3
	Population 250,000 - 560,000	-1
	Population of reporting region 20,001 - 250,000	0
	Population of reporting region ≤20,000	+1
	Address (street and ZIP code)	+3
Variable Interactions	Only events (minimum of 5), time, and geography (residence or service)	-5
	Only events (minimum of 3), time, and geography (residence or service)	-3
	Only events (no minimum), time, and geography (residence or service)	0
	Events, time, and geography (residence or service) + 1 variables	+1
	Events, time, and geography (residence or service) + 2 variables	+2
	Events, time, and geography (residence or service) + 3 variables	+4

G. APPROVALS and REVISIONS

This policy and any major revisions must be reviewed and approved by the Data Governance Board. The draft policy will be circulated to relevant stakeholders for feedback prior to finalization. Following the IDPH policy review and approval process, the final version of this policy must be approved by the IDPH director before implementation.

This policy will be reviewed and updated at least every five years or as required to ensure that it remains current and effective. Any proposed changes to the policy must be documented and submitted to the Data Governance Board for consideration. The Data Governance Board will evaluate the proposed changes and will determine further review or an approval process.

Major changes: Major changes are substantial modifications that significantly impact the policy’s scope, application, or procedures. The Data Governance Board, key stakeholders, and IDPH director must review and approve major changes.

Examples of major changes:

- Alterations that change the core objectives or goals of the policy.
- Introduction of new sections or elimination of existing ones.
- Changes that affect compliance with laws, regulations, or organizational standards.
- Modifications that require significant change or impact on data life cycle management.
- Adjustments that impact large groups of stakeholders or IDPH programs.

Minor changes: Minor changes are small adjustments that do not affect the policy's overall intent or fundamental scope and core objectives. The Data Governance Board may approve minor changes without further escalation, though documentation is still required.

Examples of minor changes:

- Clarifications or corrections of typographical, grammatical, or formatting errors.
- Updates to reflect changes in terminology or titles.
- Adjustments to improve readability or understanding without altering the policy’s meaning.
- Minor procedural tweaks that do not significantly impact stakeholders or data life cycle management.

Version	Date	Reviewer(s)	Description of Change
Draft Version	7/31/2024	IDPH data and legal experts • General Counsel • Deputy General Counsel and HIPAA Privacy Officer • Assistant General Counsel • Informatics Section Chief • Communicable Disease Section Chief • Communicable Disease Assistant Section Chief • State Epidemiologist • Chief Medical Officer • Chief, Division of Emerging Health Issues • Special Projects Coordinator / Epidemiologist	Development of Policy

		• Chief, Division of Epidemiologic Studies • Cancer Registry Director • HIV Surveillance Program Administrator • Data Modernization Director	
Draft Version	*9/30/2024*	Data Governance Board • Chief Operating Officer • IDPH Privacy Officer/Legal • Chief Information Officer • IDPH Chief Medical Officer • IRB chair • Data Modernization Director • State Epidemiologist • Informatics Section Chief • Division Chief of Health Data and Policy • Communicable Disease Section Chief • Deputy State Registrar	Data Governance Board Review
Draft Version	10/25/2024	IDPH Data Users Group	Review and comments
Final Version	10/30/2024	Data Governance Board	Final review and approval
	11/20/2024	IDPH Senior Leadership • IDPH Director • Deputy Chief of Staff • Acting Chief Operating Officer	Review and approval
1 Major Change			
1.1 Minor change			

H. REFERENCES

California Department of Health Care Services, Data De-Identification Guidelines (DDG) December 6, 2022, Version 2.2 Retrieved from Public Reporting Guidelines (ca.gov)

Illinois Department of Public Health, Division of Infectious Diseases Policy on Release of Communicable Disease Surveillance Data, February 9, 2011

Centers for Medicare & Medicaid Services (CMS), CMS Cell Suppression Policy, January 01, 2020, Version date 05/08/17 Retrieved from CMS Cell Suppression Policy | Guidance Portal (hhs.gov)

Washington State Department of Health Department, Standards for Reporting Data with Small Numbers May 2018, Retrieved from Guidelines for Working With Small Numbers (wa.gov)

Wisconsin Department of Health Services, Division of Public Health, DPH Guidelines for Releasing Aggregate Population Data Retrieved from Guidelines for Releasing Aggregate Population Data | Data Modernization Community (circle.so)