



RHODE ISLAND SMALL NUMBERS REPORTING POLICY

Introduction

The release of public health statistics relies on the degree of confidence that this information is reliable and protects the privacy of individuals included in printed reports and online query systems. There are three accepted levels of confidence associated with the reporting of health information, from the unrestricted release and cautionary reporting of probable unstable estimates, to their suppression when statistical credibility is low and individual privacy could be compromised. Criteria for each level need to justify on statistical and confidentiality grounds the reason(s) why these estimates can be reported without reservation, reported with a warning note, or suppressed altogether.

Providing justification for the cautionary release or the suppression of public health information is as important as is the reporting of reliable statistics that conform to confidentiality safeguards. This requires uniform rules that guarantee that health-related estimates can be continuously used to support critical public health functions without undermining public trust or the credibility of information being reported.

When this information is prepared for release, public health data is aggregated as counts, rates and proportions. The latter are usually accompanied by confidence intervals (CIs), relative standard errors (RSEs) and relative confidence intervals (RCIs) to describe their level of statistical credibility.

Public policy decisions are informed by statistical information. Often, information reported consists of few cases or events and/or small denominators. Small numbers are associated with estimates of lower statistical credibility and an elevated risk of a privacy breach. Small denominators may facilitate identification of individuals and are mainly associated with confidentiality concerns, and indirectly with statistical imprecision to the extent that smaller samples contain fewer cases. Estimates with few cases tend to be unstable and prone to misinterpretation, which could mislead policy decisions. With the exception of rare conditions and physical characteristics that may help determine the identity of reported cases, small numerators do not pose a privacy threat.

Estimates with small numerators and/or denominators constitute a special challenge to the reporting of statistical information. State and federal guidelines rely on case counts, denominator size, RSEs, CIs, or on a combination of these criteria to make suppression decisions. Criteria based mainly or solely on case and denominator size may not be sufficient for determining all instances when estimates with these characteristics should be reported or suppressed.

One of the strategies used to determine the credibility of estimates (half-width confidence intervals) is based on absolute reliability, which may not be relevant to all small denominator and numerator combinations, especially when denominator/numerator differences are small. This may result in the suppression of otherwise reliable estimates or the reporting of imprecise public health information with the resultant risk of a privacy breach. Guidelines should provide uniform reporting rules and requirements that take into account all possible combinations of numerator and denominator size. A description and discussion of the statistical model used for developing this guideline can be found in *A Technical Paper Related to the Development of Guidelines for Reporting Public Health Data with Small Numbers*.¹



Purpose

This policy is intended to assist the Rhode Island Department of Health (RIDOH) and external users (general public, the media, lawmakers, researchers and scholars) of public health data with decisions concerning analysis and reporting, and specifically to address the challenges of handling small numbers.

A question often asked in public health is to what extent statistical reporting reflects population health status. The significance of a guideline that addresses small numbers is not related only to the suppression of unreliable or sensitive information. A limitation often encountered in sub-group analyses is that with smaller units statistical samples are no longer an adequate natural reference for population inquiries. A main goal of a public health information policy should then be to ensure that estimates can be generalized to their target populations, and that information can be reliably and safely translated into public policy. A subsequent section of this guideline describes strategies for enhancing public health reporting when estimates do not conform to rules and requirements for data release.

Guidelines provided in this policy contain recommendations to ensure credible and confidential reporting of rates and proportions from aggregated data. It provides also criteria for the release of counts related to vital statistics and health-related conditions. Because these recommendations represent minimum criteria users may choose to follow stricter rules in their reporting.

Recommendations do not cover rules and requirements related to the release of data sets or the disclosure of individual-level information. Guidance for handling record-level data is part of database release procedures and need to be consulted with the appropriate data administrator. Criteria to prevent disclosure of identifying health information and to protect the privacy of individuals included in publicly released data sets can be found in several publications on the topic.²⁻⁸

Scope

This policy was developed to meet eight specific objectives:

- Provide criteria that can justify inclusion and exclusion of statistical information
- Safeguard the privacy of all individuals included in public health reports
- Apply reliability and confidentiality criteria concurrently
- Maximize utilization of statistical information that initially does not qualify for reporting (due to highly sensitive information or reliability reasons) without compromising individual privacy and estimate credibility
- Be applicable also to small numbers scenarios in sub-group analyses of large samples
- Ensure use to all users including those with low statistical expertise
- Provide alternative criteria for public health programs with specific federal or state reporting requirements as well as for programs with established reporting practices



- o Address internal (RIDOH) needs for data use other than public releases of statistical information

Policy

The set of guidelines contained in this policy defines small numbers as any numerator smaller than 20 cases and any denominator of 100 or less, and recommends the use of relative measures of reliability (RSEs and RCIs) to ensure meaningful reporting. These statistics describe the credibility of estimates with numerators and denominators of any size. Data simulations conducted for the testing of the statistical model for these guidelines indicated that half-width CIs are not useful for all small number combinations. We compared half-width CIs cut-offs of 10% with cut-offs of 20% (RSEs) and 80% (RCIs).¹ The former only provided information for 48% of all the possible small numerator/denominator combinations defined above.

These recommended guidelines define estimates as credible ($RSE < 20\%$; $RCI < 80\%$); questionable ($20 \leq RSE < 30\%$; $80 \leq RCI < 120\%$); and not credible ($RSE \geq 30\%$; $RCI \geq 120\%$). Questionable estimates can be reported provided they are accompanied with a warning note indicating their probable instability. All estimates with low credibility require suppression. These recommendations do not apply to RIDOH programs with specific RSE and RCI reporting requirements. Despite not being proposed as a reliability criteria for this policy, CIs constitute a reporting requirement for all estimates.

Estimate suppression should be considered a measure of last resort. Before suppressing estimates with low credibility or flagging estimates of questionable credibility every effort should be made to improve the reliability of overall and stratum-specific estimates as well as to protect the privacy of individuals included in public health reports.

Several strategies are available for improving reporting when estimates have low statistical precision and individual privacy could be compromised. Imputation, aggregation and collapsing methods can help improve estimate reliability. Protection of individual privacy relies mainly on sample increases, collapsing of table strata, and data suppression. These techniques protect confidentiality but may undermine information value in analyses where geographical information is essential.⁹ For example, aggregation can provide adequate protection but may not be recommendable when exact location (e.g.; city or town), physical (e.g.; race, birth defects) and other population characteristics (e.g.; age, education, marital status, insurance) are part of reporting requirements.¹⁰ Small number scenarios may occur in any of these situations. In some cases, the number of observations may not be as small as to meet NCHS's zero cell suppression guideline¹¹ but may nevertheless have a low number of non-cases.

Counts

Counts (vital events, health conditions, etc.) constitute a special case when reporting public health information. Reporting only cases or events is not restricted by guideline reliability criteria because there is no reliability associated with a count. An estimate requires a denominator to be statistically reliable or unreliable. Only when estimates are constructed using counts and presented in relation to a reference (i.e.; a statistical sample) they acquire this statistical property. Users should be aware, however, that applying a denominator to any count already released by the RIDOH in order to generate percentages or averages confers reliability properties to the estimate and makes it subject to the rules and requirements of this guideline. If in a future time, the reported count becomes the source



for rates or proportions with small numbers (≤ 20 for numerators; ≤ 100 for denominators) or small subsets of large numbers, the resulting estimates should be verified against this policy in order to ensure a meaningful interpretation. A warning footnote accompanying the initial count release should make users aware of this requirement for any future intended or unintended use of this information.

Privacy can be compromised from counts as well as from any other reported statistic. This policy recommends the suppression of any count, proportion, average or rate that may compromise the privacy of any person included in a public health report. More stringent criteria are required for safeguarding confidentiality than for ensuring statistical reliability. Numerator counts greater than 20 may constitute a potential privacy risk if reporting includes observable and other personal characteristics (e.g.; rare conditions or detailed disease profiles) that can be used to identify individuals. These situations can be avoided by omitting excessive reporting detail. The reporting of counts is not recommended if the privacy criterion is not met.

Policy for Reporting Counts

This policy does not restrict the release of counts of any size, except for the following:

1) Identifying Characteristics in Counts of Any Size

Count-only reports should not contain any level of detail that can make individual identities distinct or discernible. These situations include but are not limited to the reporting of rare conditions with observable characteristics; unusual combinations of health disorders or health-related risks; geographical (other than city or county) information about place of care, work, residence or worship; observable physical or cultural features such as weight, height, age, race, ethnicity; and other. Any identifying characteristic(s) that can be used to single out individuals must be removed prior to the release of count information, regardless of count size.

2) Sensitive Information in Counts Less than Five

Counts less than five containing information about behaviors, conditions or treatments considered to be socially-stigmatizing must be suppressed from all reports. Examples of sensitive information include behaviors (e.g.; substance abuse and chemical dependency), conditions (HIV and other sexually transmitted diseases; genetic; mental health), treatment (psychotherapy; substance abuse; HIV and sexually transmitted diseases; drug prescriptions), and other sensitive information such as sexuality and reproductive health information.

3) Counts Less than Five with No Sensitive Information

Counts less than five with non-sensitive information can be released at the discretion of the program provided they do not contain any characteristics that can make individuals identifiable.

All count-only reports need to be accompanied with a disclaimer about the expected use of this information. Any future use of counts to construct proportions, rates and other statistics is subject to reliability and privacy verifications specified under the small numerators and denominators sections of this policy, and covered under HIPPA privacy rules and privacy and confidentiality Rhode Island laws.



Rates and Proportions with Small Numerators

Estimates with small numerators may be very small fractions, of sizes difficult to grasp or convey. In these cases, health events need to be reported as rates per 1,000 (for deaths); per 10,000 (if group at risk is a sub-population); and per 100,000 if they reflect the population of cities/towns or the state.

Numerator size can be increased by reducing missing/extreme values (i.e.; imputation), reducing data detail (e.g.; variable recoding with intervals; category collapsing), and/or year aggregation. Geographical aggregation may not change --and could further reduce-- estimate reliability if cases in larger geographical subdivisions are infrequent or absent.

Aggregation or merging are not options for public health programs required to report events by city/town or year.¹¹ For these programs, decisions affecting estimate reporting rely solely on confidentiality (cell suppression) and statistical criteria (estimate suppression), and not on strategies that increase numerator size.

This policy recommends users to avoid breaking down information to a point where it becomes unreliable or identifiable. Several strategies are available^{3,7} for maintaining privacy when working with small numerators. The first is to limit data detail by recoding the contrasting variable into intervals and/or to collapse categories with small samples. After recoding and collapsing, the top and bottom variable categories can still have few observations. In these cases, privacy can be protected by leaving these categories without upper and lower boundaries (e.g.; age over 65; weight over 200 pounds; height over 6 feet; income over \$100,000; etc.). Rounding up or down each observation and adding noise to the data are strategies not discussed here as they are related to the protection of record-level information prior to the release of data sets and not for reporting aggregate data.

The risk of personal identification from estimates with a small number of rare conditions, identifiable physical characteristics, and variables with a naturally large number of strata (e.g.; race; ethnicity) can be reduced via year and/or geographical aggregation and by avoiding variables with observable characteristics.

If small numbers persist after implementing these strategies, privacy should be preserved through *primary* (exclusion of counts in sensitive cells) and *secondary* suppression (exclusion of non-sensitive cell counts to support exclusion of sensitive cell counts). To accomplish this, table row or column totals should not equal that of any single cell, and they should not be derivable via subtraction or other calculation from same table or combination of tables in a report.¹¹ The suppression recommendation includes cells with zero counts and cells with few observations accompanied by the exclusion of an entire row or table rows from reports.

Rates and Proportions with Small Denominators

Estimates with small denominators (i.e.; ≤ 100) are associated with a greater risk of individual identification and disclosure of personal health information. Remedial strategies for these situations need to stress sample size increases. Imputation methods (especially when missing data is particularly large for non-cases), limitation of data detail (recoding of variables into intervals; collapsing of categories), and year aggregation can supplement sample size increases.

Geographical aggregation is the main strategy to protect the privacy of individuals in estimates with small denominators. Year aggregation is likely to increase the statistical



reliability of estimates with small denominators. However, this alone may not reduce the risk of individual identification if one exists. Leaving the size of the referent population unchanged (i.e.; population size varies little from year to year) does not address the main issue associated with small denominators, which is the risk for privacy breaches caused by small differences between cases and non-cases. For samples <100, the smaller this difference is, the easier it may be to identify cases.

Numerators (cases or events) of appropriate size do not suggest the need for suppression. However, the risk of disclosing personal information increases with decreasing denominator size, especially when the ratio of non-cases (i.e.; statistical sample minus cases) to cases approaches 1. Joint categories of contingency tables with two or more variables can produce unique or nearly unique combinations and facilitate individual identification. Small statistical samples may facilitate individual identification if they have a low non-case to case ratio and one or more identifying variables are present. Expanding sample size in order to increase the number of non-cases becomes then the main strategy to protect confidentiality and/or avoid the need for cell suppression.

If these strategies fail to increase the ratio, the recommendation is to suppress all statistics associated with the identifying variable. An alternative is to drop the identifying variable(s) from reports.

This policy recommends two criteria to ensure confidentiality protection in tables and estimates with small denominators: 1) cell suppression (1^{ry} and 2^{ry}) when cell numbers equal row or column totals, and 2) suppression of table strata when numerator counts are small and the number of non-cases (those without the condition being reported represented in the denominator) is 2.5 or less (i.e.; a non-case to case ratio ≤ 2.5) the number of cases included in the numerator. The latter criteria only applies to estimates reported with one or more identifying variables that can facilitate individual identification.

It is possible that aggregation or strata merging may not be feasible for public health programs required to report estimates by city/town or year. Strategies for increasing the statistical sample then should rely on top/bottom coding, reduction of data detail and imputation, and not on data aggregation or collapsing.

Rates and Proportions with Small Numerators and Denominators

Estimates with small denominators and small numerators represent a combined risk for individual identification and estimate inaccuracy. Recommendations provided above for small numerators and denominators address these situations separately.

Strategies for increasing reliability and improving reporting of estimates with small numerators and denominators are similar. There are differences, however, in strategies used for protecting individual privacy. This policy recommends aggregating data first by year if reporting focuses on rare conditions, requires descriptions of observable characteristics, or includes variables with a naturally large number of strata. Geographical aggregation may be also required if reporting includes identifying variables with ratios ≤ 2.5 .

Small and medium-size Rhode Island cities and towns usually have estimates with small numerators and denominators. Estimate reliability and privacy protection can be preserved if reporting includes geographical subdivisions with populations of 20,000 or more. HIPPA's safe harbor method requires that statistics for cities or towns use this as the minimum denominator for the reporting or rates.¹² We found a good correlation between HIPPA's recommendation and the RSE $\geq 30\%$ suppression recommendation in the statistical model



used for these guidelines. For instances when the population reference approaches 20,000, RSE is the preferred criteria for making suppression decisions.

Confidentiality protection is a main goal in the reporting of estimates with small denominators and numerators. This policy recommends to apply privacy protection measures in conjunction with statistical reliability measures. Estimates should be reported only if both reliability and privacy reporting criteria are met.

Attachment 1 summarizes reporting criteria for counts and estimates with small numerators and/or denominators as well as strategies for increasing reliability and protecting the privacy of individuals included in these reports.

Policy for Reporting Rates and Proportions

Use the small numbers framework (*Attachment 2*) to determine if information intended for public health reporting meets RIDOH's reliability and confidentiality criteria.

Implement remedial strategies (*Attachment 1*) if information does not meet the minimum reporting requirements contained in this policy.

After the reliability verification, suppress estimates if RSEs >30% or RCIs >120%; report estimates with RSEs 20 to <30% or RCIs 80 to <120% with warning note indicating they need to be interpreted with caution.

After the privacy verification, conduct primary and secondary cell suppression if cell counts equal row or column totals, or can be derived by any calculation from one or more tables. Suppress table strata only if an identifying variable is present and the number of non-cases (individuals without the condition being reported) is 2.5 times or less the number of those with the condition.

Report only cells, strata, tables and statistical estimates that meet reliability and privacy criteria. Suppress these estimates if either reliability or privacy criteria are not met. Provide explanatory note with reason(s) for suppression and any remedial strategy(ies) conducted to resolve reliability or privacy concerns.

Implementation

This policy recommends combined application of reliability and privacy criteria followed by steps leading to their resolution, if applicable. Both criteria are operationalized through a framework (*Attachment 2*) that helps determine if estimates can be reported as initially presented and provides remedial strategies to resolve reliability and confidentiality issues.

Algorithm

These guidelines consider aggregate public health data as inputs to the entry (problem) state of the guideline's algorithm. The algorithm defines paths and processes for successive phases concluding in an ending state with a specific set of results. There are only two possible outcomes in the end stage of the algorithm: reporting or suppression of statistical information. The algorithm helps users defer these decisions until available resolution strategies to address reliability and privacy issues encountered in the initial stage are implemented. Since reliability and privacy are inter-related, the purpose of this framework



is to ensure that both issues are considered and addressed concurrently before a final reporting decision is reached.

As they move through the algorithm, users need to re-calculate statistical tables used as initial inputs in order to determine the results of the particular strategy being used. Then, re-evaluate results until small numbers strategies to increase credibility and protect confidentiality are completed.

User Level

The algorithm recognizes data users with different levels of statistical expertise and reporting needs. Reporting may range from simple counts and one-variable tables to tables with multiple variables, strata and rates. The algorithm has specific requirements for each user level. The upper half of the algorithm identifies the three small number scenarios and data elements needed for reporting single variables where small numbers are not present. This requires a minimal level of statistical expertise. The bottom half of the algorithm applies to intermediate and more advanced data users.

Simple reports require only counts and percentages provided there are no small numbers in a statistical table. Reporting of counts and estimates with small numbers is discouraged for users with low statistical expertise, as they lack the skills necessary to determine information credibility or to protect confidentiality.

Knowledge about statistical methods to increase information credibility and protect privacy is required for working with small numbers. Data users without knowledge of CIs, RSEs and other statistical concepts should seek statistical advice before analyzing and reporting this information. In the absence of this knowledge and appropriate statistical oversight, it is preferable to avoid using or reporting this information. Estimates with RSEs $\geq 30\%$ (or alternative health program policy) require suppression. Confidence intervals need to be provided for all estimates being reported regardless of RSE level. Users should also state reason for suppressing percentages, rates, indexes, averages, and any cell or table row with zero or low counts. Secondary suppression criteria also apply to this and higher reporting levels. Suppression information should be provided as a footnote in the corresponding table and in the text of the report, if appropriate.

Policy Exceptions

This policy applies to all RIDOH's external reporting and releases of statistical information. There are several exceptions related to the use of new or unverified information in internal RIDOH operations. Use of statistical information conducted in support of grant-writing, planning and other internal activities is not covered by this set of guidelines unless information later becomes public domain.

These guidelines recommend that data tables and any other statistical information used in support of these activities specify purpose of use, acknowledge any statistical limitation and potential privacy risk, and label this information as not in the public domain. Any future release of this information will make it subject to the requirements of this policy.

Public health programs with specific federal or state reporting requirements are exempt from the RSE $\geq 30\%$ suppression criteria. Recommendations related to privacy protection and statistical reporting included in this policy apply to these and all other public health programs.



Policy for Reporting Exceptions

This policy does not exclude the use of stricter reliability and confidentiality criteria for the reporting of public health information. It leaves at the discretion of users and programs the application of severer criteria for counts, rates or proportions than those specified in the policy. Users are encouraged to describe any exception(s) to the application of this policy in their reports.



REFERENCES

1. Tinajero AM. A Technical Paper Related to the Development of Guidelines for Reporting Public Health Data with Small Numbers. Rhode Island Department of Health. August 22, 2015
2. Fernández-Alemán JL, Carrión Señor I, Oliver Lozoya PA, Toval A. Security and privacy in electronic health records: A systematic literature review. *Journal of Biomedical Informatics*. 2013. 46, 541–562
3. Matthews GJ and Harel O. Data confidentiality: A review of methods for statistical disclosure limitation and methods for Assessing Privacy. *Statistics Surveys Vol. 5* (2011), 1–29
4. Zandbergen, PA. Ensuring Confidentiality of Geocoded Health Data: Assessing Geographic Masking Strategies for Individual-Level Data. *Advances in Medicine*. Hindawi Publishing Corporation. Volume 2014, 1-14
<http://dx.doi.org/10.1155/2014/567049> (accessed 6-1-15)
5. Sarathy, R and Muralidhar, K. The security of confidential numerical data in databases. *Information Systems Research*. 2002. 13 (4), 389–403
6. Aized Amin Soofi, M Irfan Khan and Fazal-e-Amin. Encryption Techniques for Cloud Data Confidentiality. *International Journal of Grid Distribution Computing*. Vol.7, No.4 (2014), 11-20 <http://dx.doi.org/10.14257/ijgdc.2014.7.4.02> (accessed 6-1-15)
7. Skinner, C. Statistical disclosure control for survey data. In: Pfeermann D and Rao, CR Eds. *Handbook of Statistics 2009*. Vol. 29A: Sample Surveys: Design, Methods and Applications. pp. 381–396
8. Shah NN and Scholar ME. A Review of Data Confidentiality in Removable Media and Symmetric Key Encryption Algorithms. *International Journal of Engineering Research & Technology*. November 2014. Vol. 3 Issue 11, 1224-1228
9. National Research Council. Putting People on the Map. Protecting Confidentiality with Linked Social-Spatial Data. Panel on Confidentiality Issues Arising from the Integration of Remotely Sensed and Self-Identifying Data. Gutmann MP and Stern PC, Eds. Committee on the Human Dimensions of Global Health, Division of Behavioral and Social Sciences and Education. Washington DC. The National Academies Press. 2007
https://books.google.com/books?id=_MvNlyhvhR4C&pg=PA61&lpg=PA61&dq=what+table+cell+suppression+does&source=bl&ots=KDK8F-xGUz&sig=imvHqbUeE8nxtOuH9OEyPObIPFM&hl=en&sa=X&ei=GJ5tVYzPEaG_sQTAlI OgCQ&ved=0CDAQ6AEwAzgU#v=onepage&q=what%20table%20cell%20suppression%20does&f=false (accessed 6-1-15)
10. Congenital Birth Defects Program. Birth Defects Data Book 2012. Rhode Island Department of Health
<http://www.health.ri.gov/publications/databooks/2012BirthDefects.pdf> (accessed 5-15-15)



11. Statistical Policy Working Paper 22 (Second Version 2005). Report on Statistical Disclosure Limitation Methodology. Federal Committee on Statistical Methodology. <https://fcsm.sites.usa.gov/files/2014/04/spwp22.pdf> (accessed 5-15-15)
12. Guidance on Satisfying the Safe Harbor Method. Guidance Regarding Methods of De-Identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. November 26, 2012
<http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/De-identification/guidance.html> (accessed 5-15-15)
13. Brown LD, Cai TT and DasGupta A. Interval estimation for a binomial proportion. *Statistical Science* 2011. Vol 16, No. 2, 101-133
14. <https://www.easycalculation.com/statistics/standard-error-sample-proportion.php> (accessed 6-1-15)
15. <http://epitools.ausvet.com.au/content.php?page=CIProportion> (accessed 6-1-15)

ATTACHMENT 1

STRATEGIES AND REPORTING CRITERIA FOR RATES AND PROPORTIONS

Issue	Small Numerator	Small Denominator	Small Numerator and Denominator
Strategies to Increase Reliability	<ol style="list-style-type: none"> 1. reduce missing/verify extreme values 2. reduce data detail <ol style="list-style-type: none"> a. recode variable into intervals b. collapse variable strata 3. aggregate by year and/or location 	<ol style="list-style-type: none"> 1. reduce missing/verify extreme values 2. reduce data detail <ol style="list-style-type: none"> a. recode variable into intervals b. collapse variable strata 3. aggregate by year 	<ol style="list-style-type: none"> 1. same strategies as for increasing small numerators and denominators 2. aggregate first by year
Suppression and Reporting Criteria for Reliability	<ol style="list-style-type: none"> 1. report estimates with RSEs <20% 2. report estimates with RSEs 20 to <30% with warning note 3. suppress estimates with RSEs ≥30% 4. verify if confidentiality criteria are met for RSEs <30% 	<ol style="list-style-type: none"> 1. report estimates with RSEs <20% 2. report estimates with RSEs 20 to <30% with warning note 3. suppress estimates with RSEs ≥30% 4. verify if confidentiality criteria are met for RSEs <30% 	<ol style="list-style-type: none"> 1. same criteria as for small numerators and small denominators
Strategies to Ensure Confidentiality	<ol style="list-style-type: none"> 1. reduce data detail <ol style="list-style-type: none"> a. recode continuous into interval variables b. collapse strata of interval variables c. leave upper/lower strata without boundaries (top/bottom coding) d. recode interval into categorical variables 2. aggregate by year and/or location for rare conditions, identifiable physical characteristics, and overstratification 	<ol style="list-style-type: none"> 1. reduce data detail <ol style="list-style-type: none"> a. recode continuous into interval variables b. collapse strata of interval variables c. leave upper/lower strata without boundaries (top/bottom coding) d. recode interval into categorical variables 2. aggregate by location if there is (are) estimate(s) with identifying variables and/or non-case to case ratios ≤2.5 	<ol style="list-style-type: none"> 1. reduce data detail (see small numerators and denominators) 2. aggregate by year if reporting focuses on rare conditions, requires description of observable characteristics, or variable(s) has (have) a naturally large number of strata 3. aggregate by geographical location if reporting includes identifying variables and ratio still ≤2.5 after year aggregation
Suppression and Reporting Criteria for Confidentiality	<ol style="list-style-type: none"> 1. conduct primary and secondary suppression if cell size equals row or column totals, or this can be derived by any calculation from one or more tables 2. suppress estimate(s) if reliability criteria are not met 3. suppress estimate(s) if reliability criteria are met but confidentiality criteria are not after implementing strategies for increasing numerator size 4. report estimate(s) only if reliability <u>and</u> confidentiality criteria are met 5. make above determinations for cells, strata, tables and all their associated statistics 	<ol style="list-style-type: none"> 1. suppress primary and secondary cells if counts equal row or column totals, or can be derived by any calculation from one or more tables 2. reassess the release of a table if >50% of cells are suppressed 3. suppress strata if non-case to case ratios ≤2.5 persist in cross-tabulations with an identifying variable after implementing strategies for increasing denominator size 4. suppress estimate(s) if confidentiality criteria are not met 5. suppress estimate(s) if confidentiality criteria <u>or</u> reliability criteria are not met after implementing strategies for small denominators 6. report estimate(s) only if reliability <u>and</u> confidentiality criteria are met 7. make same determinations for all cells, strata, tables and associated statistics 	<ol style="list-style-type: none"> 1. suppress primary and secondary cells if counts equal row or column totals, or can be derived by any calculation from one or more tables 2. reassess the release of a table if >50% of cells are suppressed 3. suppress strata if non-case to case ratios ≤2.5 persist in cross-tabulations with an identifying variable after implementing strategies for increasing denominator size 4. suppress estimate(s) if confidentiality <u>or</u> reliability criteria are not met after implementing strategies for small numerators and denominators 5. report estimate(s) only if reliability <u>and</u> confidentiality criteria are met 6. make same determinations for all cells, strata, tables and associated statistics

ATTACHMENT 1 (cont.)

STRATEGIES AND SUPPRESSION CRITERIA FOR REPORTING COUNTS

Issue

Counts

**Strategies for
Protecting
Confidentiality**

1. Identifying information (any count size)
 - a. observable rare conditions
 - b. geographical designation (other than city or county) about place of care, work, residence or worship
 - c. observable physical or cultural features (e.g.; weight, height, age, race, ethnicity)
 - d. unusual combinations of health disorders or health-related risks

1. Remove any level of detail that can make the identity of individuals distinct or discernible

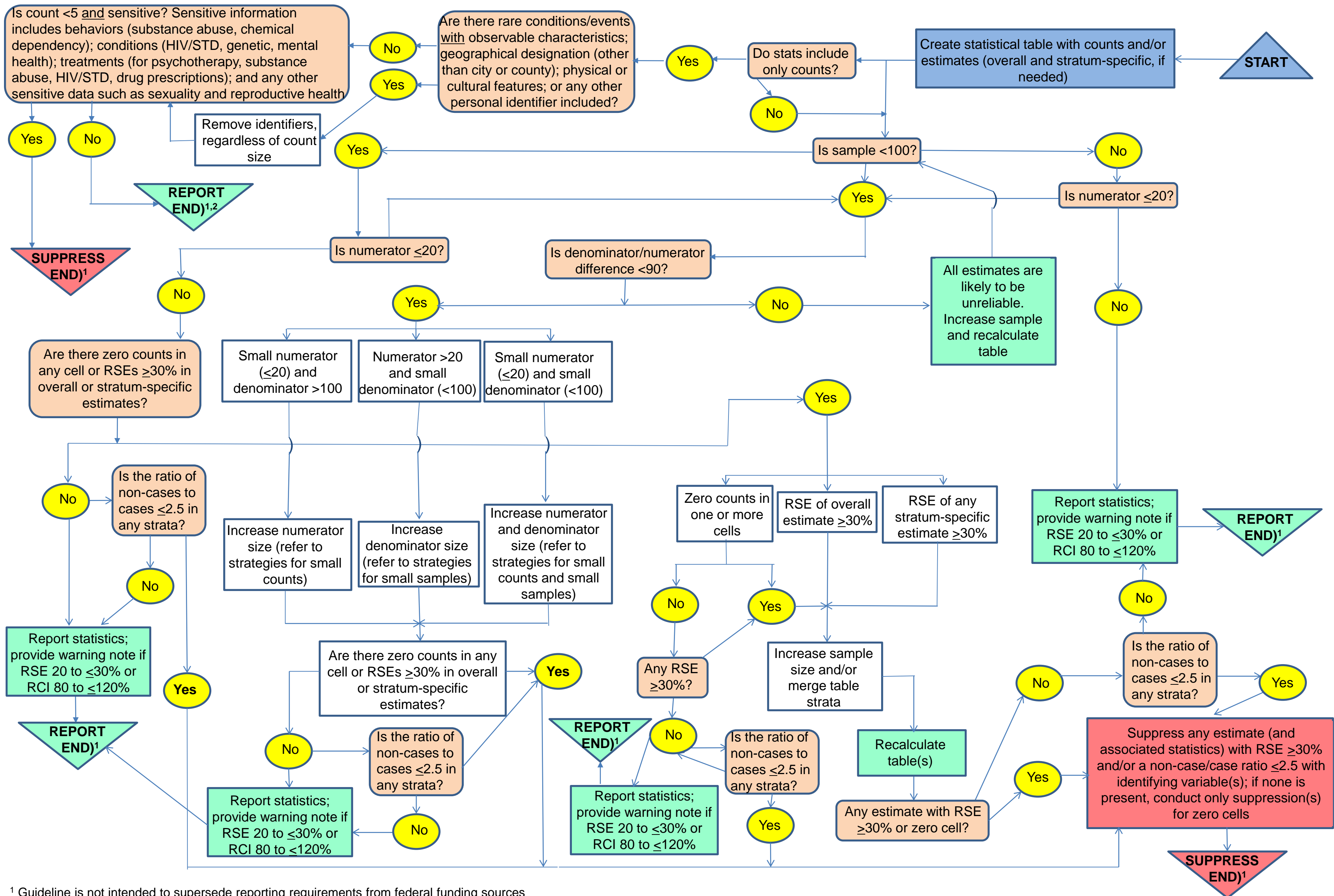
**Reporting and
Suppression Criteria**

2. Very small counts with sensitive data
 - a. behaviors (substance abuse, chemical dependency)
 - b. health conditions (HIV/STDs, genetic, mental health disorders)
 - c. medical treatments (psychotherapy, substance abuse, HIV/STDs, drug prescriptions)
 - d. Sexuality and reproductive health information

2. Suppress counts <5 that contain sensitive information

ATTACHMENT 2

ALGORITHM FOR THE REPORTING OF PUBLIC HEALTH DATA WITH SMALL NUMBERS RHODE ISLAND DEPARTMENT OF HEALTH



¹ Guideline is not intended to supersede reporting requirements from federal funding sources

² Add footnote "Any future use of counts to construct proportions, rates and other statistics is subject to reliability and privacy verifications specified under the small numerators and denominators sections of the State of Rhode Island Public Health Information reporting policy"