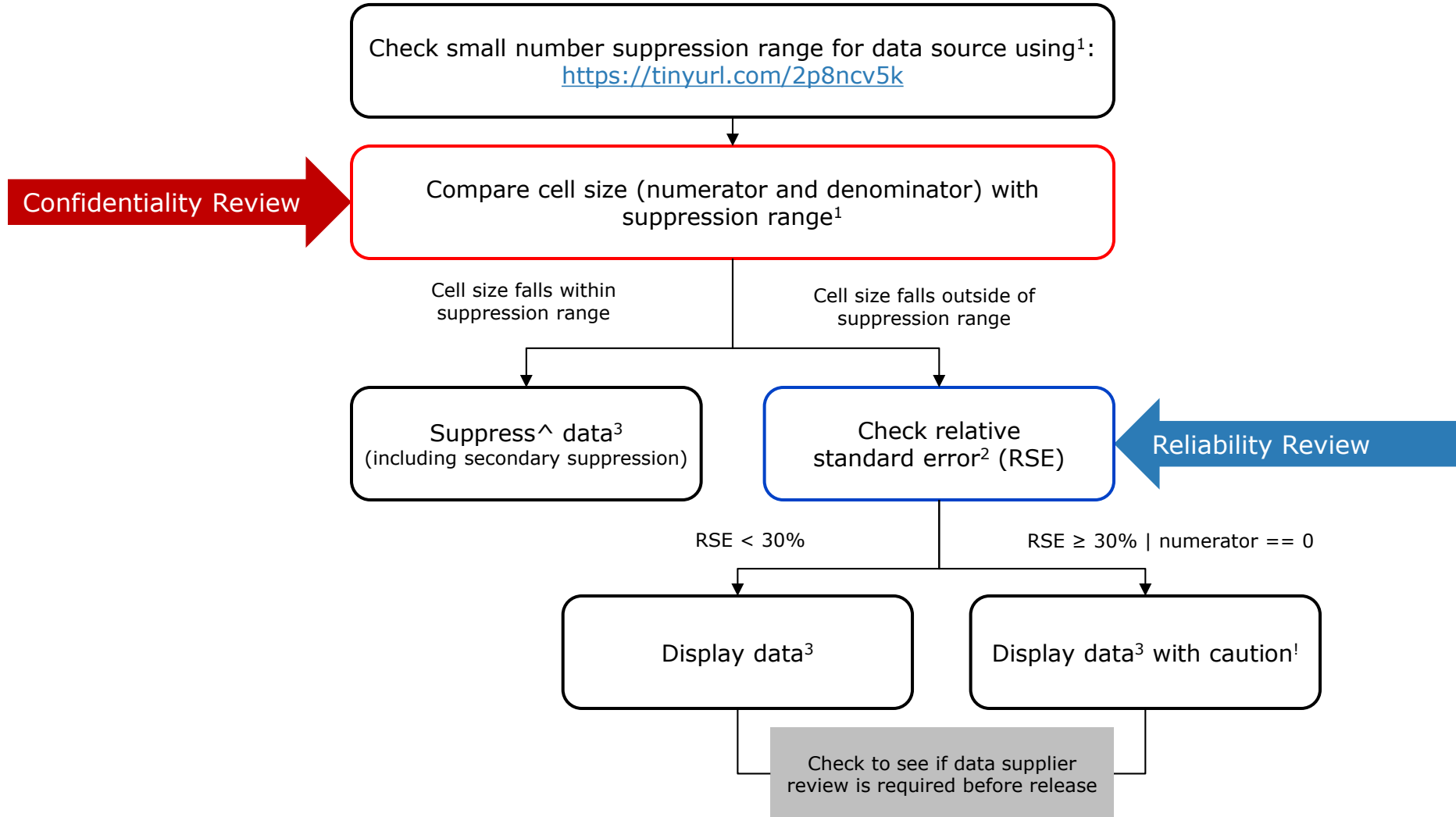


APDE Data Presentation Algorithm



[^] Data suppressed if too few people/events to protect confidentiality.
[!] Interpret with caution, sample size is small and thus estimate is imprecise.



APDE Data Presentation Algorithm

GENERAL NOTES:

- For routine data analysis, single-year data may be used for estimates for the total King County population (i.e. no cross tabulation), while multiyear averages will be used for estimates cross tabulated by demographics or sub-county place.
- This data presentation algorithm applies to estimates for the total county population and cross tabulations by a **single** demographic characteristic or sub-county geography. For example, presentation of infant mortality rate by race/ethnicity alone, but not cross tabulated by sub-county region AND race/ethnicity simultaneously.
- Please refer to each data sharing agreement for data source-specific requirements around small numbers suppression (summarized in link on page 1) and re-disclosure requirements.
- In charts/tables, a bar/table row label should be shown for all suppressed data points.
- Exceptions to small number standards/requirements can be requested – see [APDE Analysis Guide](#) for guidance.
- Analysts should always check to see if secondary suppression is needed to assure that suppressed cells cannot be recalculated through subtraction – see [APDE Analysis Guide](#) for guidance.

FOOTNOTES:

1. For case data, number of cases (i.e., numerator) is over all years included, for example, 5 cases over 5 years.
2. For binary variables, RSE should be calculated for the response that is less than or equal to 50%, and this RSE should be used for both levels. For variables with 3 or more levels, only those levels being presented need to be assessed for reliability concerns. RSE is calculated as follows:

$$\text{Counts: } \frac{1}{\sqrt{\text{numerator}}} \quad \text{Proportions: } \frac{\text{standard error}}{\text{proportion}}$$

3. “Data” represents estimated number of individuals/events and percentages for survey data, and counts and rates for all other data (e.g. vital statistics, administrative data).



APDE Data Presentation Algorithm

Why do we suppress small numbers?

- We suppress small numbers to protect the identity of the people represented in the data we use.
- We use data to protect and promote public health, and we are committed to assuring the public and all stakeholders that we can be trusted to use data responsibly.

Why do we question the reliability of statistics based on small numbers?

- Typically, statistics (e.g., a percentage) based on large numbers provide stable estimates of the truth we are seeking to understand.
- For example, imagine two different surveys aiming to identify the most popular ice cream flavor among children living in King County:
 - In survey #1 we asked half of all children in King County (~ 225,000 children) and learned that 70% of children reported chocolate as their most popular flavor.
 - In survey #2 we asked 10 children (0.002% of all children in King County) and learned that 10% of children reported chocolate as their most popular flavor.
 - Which of these two estimates is a more trustworthy or reliable estimate of the percentage of children in King County who prefer chocolate ice cream?
- Typically, statistics based on small numbers provide unstable estimates of the truth we are seeking to understand.
- It is important to note that the stability or reliability of a statistic does not indicate anything about the importance of the data. However, statistics based on small numbers may fluctuate dramatically from year to year, or from survey to survey, or from place to place, and these fluctuations may not be true differences but instead are due to the instability of the statistic.

