

Exploring Missouri's New, Innovative, and Multilayered Health Data Suppression Rules

Whitney B. Coffey

Bureau of Health Care Analysis and Data Dissemination, Missouri Department of Health and Senior Services, Jefferson City, Missouri, United States

Objective

By the end of this session, users will be able to describe the innovative and multilayered suppression rules that are applied to Missouri's homegrown health data web query system. They will also be able to use the lessons learned and user feedback described in the session to facilitate discussions surrounding the application of suppression to their specific data systems.

Introduction

In Spring 2017, the Missouri Department of Health and Senior Services (MODHSS) launched the Missouri Public Health Information Management System (MOPHIMS) web-based health data platform. Missouri has supported a similar data system since the 1990s, allowing the public, local public health departments, and other stakeholders access to community level birth, death, and hospitalization data (among other datasets). The MOPHIMS system is composed of two separate pieces. Community Data Profiles are topic-, disease-, or demographic-specific reports that contain 15-10 indicators relevant to the report. Because these static reports are developed in-house a multilayered suppression rule is not required. The second piece of MOPHIMS, the Data MICAs, or Missouri Information for Community Assessment, can be used to create customized datasets that slice and dice up to a dozen demographic and system-specific variables to answer complex research questions.

The MOPHIMS interface features, among other things, a new and innovative method for addressing confidentiality concerns through the suppression of health data. This pioneering approach integrates multi-level logic that uses inner and outer cell analytics, the use of exempt and conditionally exempt variables, and multiple levels of user access. Moving beyond a simple model of suppressing any values below a certain threshold, MOPHIMS takes a bold step in providing users exceptionally granular data while still protecting citizen privacy.

Methods

In order to implement this new suppression methodology, MODHSS worked with both internal information technology resources (OA-ITSD) and outside contractors to develop the suppression rules utilized in the Data MICAs. Before these meetings began, MODHSS analysts met weekly to determine the overall goals and frames for the rule, knowing that writing the code to implement the complicated and comprehensive vision would be a collaborative and iterative process.

Because the MOPHIMS system is homegrown and this specific confidentiality process is not currently utilized (to our knowledge) elsewhere, all of those at the discussion table were required to be innovative, open to criticism, and willing to engage in extremely detailed explanations. A team of users from Missouri's local public health departments provided feedback throughout this process.

A basic description of the process flow that occurs before suppression is applied in MOPHIMS follows. To begin, de-identified record-level data are loaded into online analytical processing (OLAP) cubes and relational databases. No suppression is applied to these back end databases. The information is then aggregated for display on the front end screens of the Data MICAs based on customized user selections. Depending upon which level of access a user has logged in, suppression is then applied to the data output generated using these customized selections. Not only are the rules applied to data tables but also to the MOPHIMS data visualization tools, which include multiple types of charts and maps.

Results

Table cells are suppressed in the Data MICAs if a frequency less than 5 (but greater than 0) is the result of a custom query using three or more conditionally exempt variables, which are described below. Additional rows/columns with cell frequencies above this threshold may be suppressed to ensure a total of at least 3 rows/columns are suppressed. Suppression is applied to either rows



ISDS Annual Conference Proceedings 2019. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 Unported License (<http://creativecommons.org/licenses/by-nc/3.0/>), permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

or columns, depending on which is more numerous in a given table. The suppression rules are applied separately to both the inner and outer rows/columns of a table. This layering of suppression across individual pieces of customized datasets negates the possibility of users algebraically determining the hidden values.

A small set of variables were classified as exempt (unlikely to identify individuals or break MODHSS privacy standards) and alone, or in combination, would not trigger suppression. These include state totals, multi-year ranges, gender, and basic age groups. Some variables, however, could be potentially identifying, either individually or used in tandem. These are considered conditionally exempt and include things like: county/ZIP geographies, single years of data, race, ethnicity, expanded age groups, and specific causes/diagnoses. If three or more conditionally exempt variables are used in a query then suppression is invoked.

In addition to the rules themselves, MOPHIMS contains a mechanism that allows users to log in at different levels of access. Public and Registered user levels are free and available to all operators with a valid e-mail address. Partner level access is reserved for epidemiologists at the state and local level who are using the Data MICAs for program planning, evaluation, and grant writing. Because these individuals are required to adhere to the same data dissemination policies as those who create the MOPHIMS system, Partner level access turns off suppression in the MOPHIMS system. Values that would be suppressed at the Public or Registered user levels are shown in italicized, red font. A multi-level approval process is required for individuals to obtain Partner level access to MOPHIMS.

Conclusions

MODHSS created an innovative suppression system that allows public health planners to access granular data through customizable queries without risking a confidentiality breach. Users have indicated this is highly preferable to a blanket suppression rule that hides any value under a certain threshold.

Additionally, approved MOPHIMS users can view specially formatted values that would otherwise have been suppressed. The flexibility associated with creating a homegrown web query system has allowed the formation and implementation of this multilayered rule, which likely would not have been possible if using an off-the-shelf product. Data disseminators are encouraged to review current confidentiality and suppression rules to determine whether they might be modified to provide more granular data users while still protecting the privacy of citizens.



ISDS Annual Conference Proceedings 2019. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 Unported License (<http://creativecommons.org/licenses/by-nc/3.0/>), permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.