Project Hg User Guide

What is Project Hg?

Project Hg (Project Mercury) is a collaborative of multiple state and municipal health departments (HD) compiling estimated background rates of food and animal exposures from two different data sources:

1) Food and animal exposure history data collected from interviews of human cases of *Salmonella* and Shiga-toxigenic *E. coli* (STEC) infection.
2) Locally implemented statewide population surveys (Hawaii and Arizona)

This collaborative built a framework to compare exposures across different HD questionnaires using a standardized exposure code (the "Keene code"). HDs aggregate their case exposure data locally to assess local exposure background rates, which are compiled for Project Hg to create a set of comprehensive background rates and allow comparisons of these rates between sites.

The name "Project Hg" has two popular meanings in our group:

1) Mercury is a chemical element with the symbol "Hg" (for us, hypothesis generation)
2) In Roman mythology, Mercury is the winged messenger of the gods.

This project was originally conceived by the late, prolific epidemiologist Dr. William "Bill" Keene of the Oregon Public Health Division.

Background

Lab-confirmed human infections of *Salmonella* and STEC are nationally reportable in the United States, meaning that each state requires the report of these cases, and these states turn and report these data to the Centers for Disease Control and Prevention (CDC). States develop internally standardized questionnaires to investigate these reportable diseases locally; clusters of disease are detected through various mechanisms (e.g., molecular subtyping, geo-temporal clusters). States and CDC work in tandem to detect multi-state clusters and partner on these investigations as they cross state lines.

Up until the 1990s, clusters were largely investigated using case-control studies as the gold standard method. In approximately 2002, Dr. Bill Keene of the Oregon Public Health Division implemented a companion to the case-control study: binomial probability calculations.

When used with standardized hypothesis-generating questionnaires (or HGQs, sometimes called trawling or "shotgun" questionnaires for the way they ask many questions all together in hopes to get a few target hits), the binomial model can develop hypotheses that epidemiologists could potentially test with traceback of commercial food product or commercially-distributed animals. If this traceback triangulates on a single product, this obviates the need for a further analytic study. Laboratory testing of suspected food products or commercial animals has also emerged as a field practice to follow up on hypotheses. Binomial probability calculations have identified food vehicles in pulsed-field gel electrophoresis (PFGE) clusters of infections by *Salmonella*, Shiga toxin-producing *E. coli* (STEC), and *Listeria.*

One required data element in the binomial model is a reasonable estimate of background rates of the exposures included on the HGQ or other outbreak questionnaire. Project Hg data will serve as a range of different Salmonella and STEC case background exposure rates across the United States; once all twenty (as of June 2019) Project Hg sites are reporting data there will be a decent geographic spread of data available, in additional to demographic granularity and exposure seasonality.

To narrow the focus of an investigation and most effectively use time and resources, investigators should begin to generate hypotheses about potential sources of the outbreak during the earliest stages of the investigation and refine them as they receive information. Part of being prepared for these investigations means having standardized questionnaires available for immediate case interviews. Many states have developed their own internal standardized questionnaires; questionnaires from Project Hg sites can be found on the Oregon CoE website. In addition to many HDs developing internal HGQs, CDC has also developed a National HGQ that can be rapidly deployed during multi-state outbreak investigations.

Binomial probabilities calculations during cluster investigations

$$\sum_{y=x}^{n} \left(\frac{n!}{y!(n-y)!}\right) p^y (1-p)^{n-y}$$

$n = $ *number of cases*
$x = $ *number of cases who ate product*
$p = $ *background consumption rate*

The cumulative binomial probability model

Binomial probability calculations are used to generate hypotheses which can help focus an investigation to particular food vehicles that should be further explored. These

hypotheses could be tested via high-powered studies including environmental trace back, sample testing, and case-control studies. The binomial probability indicates how likely we are to find x of n people (or more) eating a given product by chance alone. Vehicles that have probabilities of 0.05 or lower are generally determined as significant and should be followed-up. It is important to note that binomial probability calculations require a reasonable estimate of background food consumption rate and that it is not be used to *test* hypothesis but rather to *generate* hypothesis and then pivot to more definitive methods.

Tools to Calculate Binomial Probabilities

Several tools have been developed to assist you in calculating binomial probabilities

- Project Hg Tableau: Interactive online data visualization tool
  - As described below in this guide.

- The Keene classic: Binomial probability worksheet (.xls)
  - Just like it sounds; programmed for fast analysis! Just enter number of cases interviewed, number who said yes, and expected background to get instant calculations.

- Key Points: Using Binomial Probability Calculations During Cluster Investigations
  Information on using binomial probability calculations for investigations, including guide for using the Keene binomial probability worksheet (above).

- FoodNet Population Study Atlas of Exposures (Pop Survey): nation-wide population-based telephone surveys to assess common exposures associated as risk factors for foodborne illness. The original source of background rate estimates for binomial probability calculations.

- Epi Info StatCalc tab
  - CDC has also included options for calculating binomial probabilities in Epi Info, including the mobile app.

Project Hg Tableau online data visualization tool

Tableau is used to interactively visualize Project Hg data and to allow for easy background data extraction. The visualization (viz) can be found here: https://public.tableau.com/profile/fomes#!/vizhome/ProjectHg/Main. This link will direct you to the main dashboard.

"Main" tab

On the main dashboard is a short description of our project, an interactive map that will bring you to a data table for the selected state, and links to several program websites that support this project.

"Data table" tab

The data table dashboard displays data similarly to an excel workbook and is designed for easy data extraction. On the right of this dashboard (and all subsequent dashboard) is a filter pane that is used to filter down on a specific pathogen, state, and exposure (Keene ID and description). Once the filters are selected and the intended data is displayed, the data can be extracted by simply pressing on the download button (on the bottom right) and selecting crosstab.

"Graphs" tab

The graphs dashboard graphically displays background data for a single pathogen, single state, and single exposure for all variables (percentage of yes by sex, age, race, ethnicity, and season).

"Maps" tab

The maps dashboard helps visually compare the background rates of selected exposures across multiple states by displaying these rates as a dot-density. States with higher background rates will have larger dots.

"Keene Mapping" tab

The Keene mapping dashboard illustrates how standardized exposures (Keene Codes) are mapped to state specific exposures.

"Site Questionnaires" tab

The last dashboard, site questionnaires, is a simple repository for state-specific questionnaires that are used to capture background data.

Description of Site data sets

								Description of Project Hg Site Data Sets							
State	**Salmonella years of data**	**STEC years of data**	**Questionnaire type**	**Type of cases interviewed**	**Exposure period**	**Date used to define case year**	**Outbreak cases**	**Multi-racial practice**
CO	2014-2017	*N/A*	Standard case report form	Cluster & sporadic	7 days	Event date	Included	Multi-racial cases are only counted once in "Multiple" category
MN	2013-2018	2016	Standard case report form (MN "long form")	Cluster & sporadic	7 days	Specimen collection date	Excluded	Multi-racial cases are only counted once in "Multiple" category
NH	2016	*N/A*	Standard case report form	Cluster & sporadic	7 days	Specimen collection date	Excluded	Multi-racial cases are only counted once in "Multiple" category
OR	2013-2018	2013-2018	"Shotgun" HGQ	Cluster & sporadic	7 days	Specimen collection date	Included	Multi-racial cases are only counted once in "Multiple" category

Please direct correspondence to:

Hillary Booth, MPH

Oregon Public Health Division

hillary.booth@dhsoha.state.or.us

971-673-1072

July 30, 2019