Project Tycho

Preliminary data for the state of Hawaii

Tycho database beta test version

The data presented in this report are of preliminary nature and should not be used for publication or other types of official use

June 2011

University of Pittsburgh Graduate School of Public Health

University of Pittsburgh

Table of Content

Acknowledgements

The development of this project would not have been possible without data collected every week by staff from city-, county-, state- and federal public health departments. We would like to acknowledge their invaluable contribution to the overall body of surveillance data that were digitized during this project. Particularly, we acknowledge staff from the Council of State and Territorial Epidemiologists (CSTE) and the US Centers for Disease Control for their overall efforts in disease surveillance and their particular input in this project.

We also acknowledge the help of many contributors that have contributed to the digitization and processing of weekly surveillance data included in this database. Digital Divide Data: Linda Thomas, Eric Gold, Heak Hok, Villa Kem and Hetel Patel. University of Pittsburgh Graduate School of Public Health: Steve Wisniewsky, Heather Eng, Sharon Crow, David Galloway, Jia Xu and Raaka Kumbhakar. University of Pittsburgh School of Information Sciences: Vladimir Zadorozhny, Ying-Feng Hsu, David Paul Koenig , Fatimah Ridwan and Divyasheel Sharma. Johns Hopkins Bloomberg School of Public Health: Derek Cummings. Pittsburgh Supercomputing Center: David Graham and Andrew Adams.

This project was made possible with funding from the Bill & Melinda Gates Foundation through the Vaccine Modeling Initiative (grant ID 49276).

The Tycho development team at the University of Pittsburgh Graduate School of Public Health:
Willem G. van Panhuis
John Grefenstette
Shawn T. Brown
Su Yon Jung
Nian Shong Chok
Anne Cross
Donald S. Burke

Contact information

The Tycho database is currently being tested and a login account will be required to use the database and website (www.tycho.pitt.edu). All State Epidemiologists, CDC partners, other federal agencies and our research collaborators will be provided with user accounts. User accounts can also be provided upon specific request. Note that the quality of the data during the testing phase cannot be guaranteed to be sufficient for publication or official use.

Point of contact for the Tycho project:
Dr. Willem G. van Panhuis, MD PhD
Graduate School of Public Health
University of Pittsburgh
130 DeSoto Street
704 Parran Hall
Pittsburgh PA, 15261
Email: wav10@pitt.edu
Tel: 412-624-7693

Introduction

Project Tycho started in 2009 as part of the Vaccine Modeling Initiative (VMI) at the University of Pittsburgh Graduate School of Public Health, funded by the Bill & Melinda Gates foundation. This project aims to contribute to the availability of public health data for research and policy making. The vision for this project as described in the next section is a central access point for public health data of a detailed spatial and temporal scale for all countries. Currently, the Tycho database contains 10 million records that each represent a weekly report from a location for a specific disease. These records were extracted from weekly US Nationally Notifiable Disease Surveillance Reports between 1888 and 2009 (6300 reports) using double data entry (200 million keystrokes). The current database includes a total of 90 million reported cases and 4 million reported deaths due to notifiable diseases in the United States for the last 122 years.

Project **Tycho** is named after the Danish nobleman Tycho Brahe (1546 – 1601), who meticulously collected astronomical data. After Brahe's death, his assistant Kepler used these data to develop area laws of planetary motion.

Background

"No health department, State or local, can effectively prevent or control diseases without knowledge of when, where and under what conditions cases are occurring"
(Quote stated above weekly surveillance reports published in the Public Health Reports, 1888-1951)

This quote has accompanied weekly reports on the prevalence of notifiable diseases in the United States as long as these were published in the Public Health Reports, emphasizing the importance of data for disease control policy making. In the current day and age, this statement has lost none of its significance. On the contrary, as infectious diseases continue to pose challenges to the global public health system in an increasingly interconnected global society. The public health system has responded to this challenge by evermore sophisticated disease surveillance systems. The availability of internet and information technology has facilitated collection of detailed data on infectious disease incidence, the spread of pathogens, disease determinants, health behavior, etc. New technologies such as remote sensing and mobile reporting systems have advanced measurement of these factors in both high and low resource settings. The enormous wealth of public health information that is currently being accrued also poses its own challenges for data storage, management, preservation and dissemination. Public health agencies will need to allocate resources and develop capacity to specifically address these issues.

Although the importance of public health data collection has been widely recognized, the dissemination of data has been neglected for many reasons. The above quote does not only imply data collection but also its use and dissemination as knowledge on disease occurrence can only be derived from data after processing and analysis. Not only has disease surveillance become more sophisticated, analytical tools have changed as well and now include computational models to assess disease transmission and to evaluate disease control options. These methods have introduced the opportunity for data intensive studies of fine spatial-temporal patterns using dynamic models. Disease surveillance data at such scale are often not available for analysis for multiple reasons. First, surveillance data are not collected for research purposes but for disease monitoring and planning. Secondly, public health staff often lack time or capacity to prepare these data for analysis. Third, surveillance methodology often lacks transparency and standardization, reducing options for scientific analyses. Fourth, disease surveillance data are often

kept confidential and ownership or data use requirements are often unclear. Finally, archiving and preservation of surveillance data are often not a priority and disaggregated records may get lost, losing the opportunity to ever use these data for analysis.

The Tycho project aims at overcoming some of these challenges by providing open access to large public health datasets of great spatial-temporal detail. Open access to these data will provide great opportunities for scientific analysis and better informed policy making on disease control.

The Tycho database and website

As described in the methods section, a large heterogeneity in reporting practice was observed over the past 122 years of weekly disease surveillance and substantial efforts were made to clean and standardize these data. The current online database includes all cleaned and standardized data in a format that allows comparisons across time and space. It does not include parts of the data that require additional cleaning and research such as classification of disease subcategories.

Each record in the database provides a reported number of cases or deaths due to a specific disease for a specific location, for a certain week. The database can be searched online (www.tycho.pitt.edu) by selecting a dataset, an aggregation method and an output format. Data can be viewed in tables, graphics and maps and these can be downloaded for further use.

Data availability

The data provided by the Tycho database reflect the weekly US Nationally Notifiable Disease Surveillance System between 1888 and 2009. It includes all diseases that were reported as part of this system, but it does not yet include records from other surveillance systems such as the annual system or disease specific subsystem (such as the influenza surveillance system).

The available data also reflects changes in reporting practices over time. This explains why city or state level reports are only available for a certain time period (1888-1953 for cities and 1927-2009 for states) and why morbidity and mortality reports are available for different time periods. We are currently working on inclusion of additional data as described in the section on collaboration.

Testing and release of the Tycho database

The Tycho database is currently in a beta testing phase that will start during the 2011 annual meeting of the Council of State and Territorial Epidemiologists (CSTE) in Pittsburgh June 12-16. This phase will end with a release to the general public in the fall of 2011. During the testing phase, invited users will be provided with login accounts that will enable full use of all features of the website and database. Invited users will include all state epidemiologists, partners in the CDC and other federal agencies as well as research collaborators. During this phase, data cleaning and standardization will continue. The quality of the data will not be sufficient for publication or official use during this phase.

Future developments

After optimization of the database and website during the testing phase, the first version of the Tycho database will be released to the general public. After that, efforts to clean and standardize the data will continue and more data will be released in subsequent versions. In addition, ongoing and new collaborations for inclusion of new data will allow gradual expansion of the dataset over time and increased opportunities for data driven analysis and policy making.

Vision and goal

The vision for project Tycho is based on the value of open access to detailed, disaggregated public health data for scientific analysis and policy making. Starting with 122 years of weekly US surveillance data, it is our hope that this example will convince public health authorities worldwide of the value of this resource and will stimulate similar contributions from other countries.

In the past, data sharing and archiving have not been a priority in public health or among health sciences in general. Continuous and diligent public health data collection has led to a large pool of disaggregated data scattered over the world including the internet, ranging from paper archives to basements of local health departments. The need for open access to a central repository of public health data has been recognized and both benefits and challenges of open access to data in general have been well described. Open access to data could lead to increased accountability, transparency, innovation, collaboration, cost-effectiveness, replication of results, the development of new methods and insights, and a narrower data access gap between low- and higher income countries. Ultimately, this would result in the advancement of science and technology to the benefit of all.

Challenges to open access include inconsistent formats, lack of annotation or metadata, lack of information technology (IT) capacity, a lack of incentive, and cultural norms around data ownership. Open access to public health data in particular is confronted by decentralized archiving of disease reports, lack of time and capacity for data management, degraded physical format of records and archives, and governmental level legal constraints. Multiple principles and requirements for a central open access repository for public health data have been described and include (1) capacity building for data management in low- and middle income countries, (2) feedback loops to data contributors, (3) sustainability, (4) common standards, (5) interoperability, and (6) user-friendliness.

The ultimate goal of the Tycho project is to provide a central global public health data access point. Historical as well as current public health data are of great value if archived and accessible for research and analysis. Open access will enable the use of analytical capacity from around the globe which will lead to new discoveries of disease patterns and control policies. Furthermore, a central data access point will facilitate data archiving and preservation into the future which will be an increasing need in a data rich public health environment.

Collaboration

The application of the surveillance data from the Tycho database as well as the inclusion of new data will require a wide range of collaborations with public health offices in the US and abroad. The Tycho project has currently made the first step of digitizing 122 years of weekly US surveillance reports that had already been published. The next step will be the completion of the existing dataset as well as the inclusion of new data from unpublished sources.

This report provides an overview of data available from the Tycho database for the state of Hawaii. This includes all data that has ever been published at state or city level for Hawaii in the weekly US Nationally Notifiable Disease Surveillance System. As you will find, the data availability will vary greatly over time, between city and state level reports, between diseases and between morbidity vs. mortality reports.

The Tycho development team would look forward to work with state health departments in the US to:

1. Better understand the current Tycho data for each state (eg. by collecting historic documentation)
2. Provide better usability and applications of current Tycho data for each state by continued joint development of website and database features.
3. Collect and/or digitize official, confirmed data to validate the current (preliminary) weekly data.
4. For each state, collect and/or digitize new data that has not yet been included. For example all city level reports discontinued in 1953 and completion of city level data until 2009 would greatly increase opportunities for analysis of disease patterns and trends.
5. Provide support to state health offices to manage, preserve and provide access to public health data.

Summary of methods

Detailed documentation on the methods used for the Tycho database has been provided on the website (www.tycho.pitt.edu). This section describes these methods in short.

Data collection

Weekly reports that contain tables on the occurrence of nationally notifiable diseases have been published since 1888 by public health authorities at the Federal level in various journals. Table 1 provides the list of publications and the responsible agency since 1888. All weekly nationally notifiable disease reports between 1888 and 1951 could be retrieved from the PubMed Central repository of the National Library of Medicine [1]. For 1995 to 2009, these reports could be retrieved from the MMWR digital archive on the CDC website [2]. Most weekly reports between 1952 and 1995 could be retrieved from the HathiTrust Digitial Library [3] [4], but many could not be found and had to be copied from hard copies of MMWR issues in the University of Pittsburgh library.

Table 1, publications and responsible Federal agencies for nationally notifiable disease reports

Time period	Publication title	Responsible federal agency
1888-1889	Weekly Abstract of Sanitary Reports	US Marine Hospital Service
1890-1895	Abstract of Sanitary Reports	US Marine Hospital Service
1896-1901	Public Health Reports	US Marine Hospital Service
1902-1911	Public Health Reports	US Public Health and Marine Hospital Service
1912-1951	Public Health Reports	US Public Health Service
1952-1960	Morbidity and Mortality Weekly Report	National Office of Vital Statistics, US Public Health Service
1961-1969	Morbidity and Mortality Weekly Report	Communicable Disease Center
1970-1991	Morbidity and Mortality Weekly Report	Center for Disease Control
1992-2009	Morbidity and Mortality Weekly Report	Centers for Disease Control and Prevention

Inclusion criteria

Weekly reports of each year were reviewed systematically to assess the diseases reported. We included all tables that provided disease specific information by week for US cities, townships, counties or states. Tables that provided summary or aggregated information by month, year or at the national level were not included. Similarly, tables that did not contain disease specific information (such as all cause mortality) were not included.

Data entry

Weekly reports were downloaded or scanned as PDF files and selected tables with notifiable disease reports were entered into computer spreadsheets in a highly standardized fashion using double data entry. During the second round of data entry, operators could not see what had been entered in the first round and could not continue if the system detected a discrepancy between the second and first entry

[1] http://www.pubmedcentral.nih.gov/tocrender.fcgi?journal=333&action=archive
[2] http://www.cdc.gov/mmwr
[3] http://catalog.hathitrust.org/Record/003910026
[4] http://catalog.hathitrust.org/Record/003843660

for a specific value. Such discrepancies could only be resolved by checking the PDF file and try again or discussion with the group leader and verification of a value in the source documents.

Quality control for data entry
The accuracy of data entry was checked at various levels. First, completeness of data was verified by comparing the content of entered data with PDF sources files. Secondly, accuracy of data entry was verified by multiple rounds of comparing random samples of entered files with PDF source files. Thirdly, data formatting was verified by various checks to ensure appropriate formatting for data loading.

Data loading and standardization
All data was entered in Excel spreadsheets and various components of these spreadsheets were loaded in data files. Table titles, column headers, place names and reported numbers were loaded in separate files. These files were used to extract information on each reported number, including:

1. the disease reported
2. the disease subcategory reported
3. cases or deaths reported
4. the reporting location (name, state and type of location)
5. the time period for which a number was reported
6. the date of publication of the original weekly report associated with a reported number

Integration
All reported numbers and extracted information was integrated in one database with one record per reported number and associated information.

Post-processing quality control
After integration of all data in one database, checks were performed to detect duplicate reports and data inconsistencies. Duplicate records were removed and inconsistencies resolved by verification with original PDF source files.

Data filtering
The digitized version of all historical weekly US nationally notifiable disease surveillance records is a very heterogeneous dataset. It took a substantial number of data processing protocols to standardize time and space variables as well as disease names. All standardized records have been separated from non-standard records by a filter. All remaining heterogeneity in non-standard records is inherent to the surveillance system and can only be standardized after further analysis will have been completed (eg. remaining heterogeneity in reporting periods for reports before 1953 and standardization of disease subcategories). The largest proportion of data (>4 million records) has been standardized however and has been made available in the current testing version.

Data visualization
Figures in this reports were made with the R system, version 2.9.2 and the maps were generated by the GAIA platform developed at the University of Pittsburgh Graduate School of Public Health in collaboration with the Pittsburgh Supercomputing Center (PSC). See http://midas-pitt.psc.edu/gaia for more information.

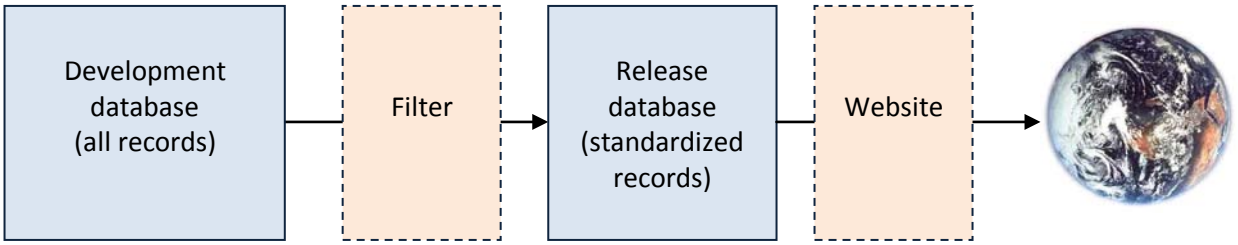

Figure 1, Schematic of the Tycho data architecture. Access to a consistent, standardized subset of the digitized US weekly surveillance reports will be provided for beta testing through a website that will allow querying and downloading of data. Data from the development database will be pushed through a filter to feed the Tycho database.

The current Tycho database will provide the reported number of cases or deaths reported by city or state health authorities to the federal health agency for all weeks between 1888 and 2009. Currently, only standardized, consistent data are being released for testing, according to the following criteria:

1. Only weekly reports are provided (reports for other periods such as 10 days, 2 weeks, 1 month, 1 year) are currently not included due to lack of comparability with the weekly reports.
2. Only reports that were published in the same year as the reporting period were included (this excludes updates or comparisons that were published more than a year after the original reporting period)
3. Only numbers were included for which information about the location, time period and disease could be extracted.
4. Only numbers or time periods for which no disease subcategories were reported were included.

Each of the steps described above have been described in detailed protocols that are posted in the documentation part of the Tycho website (www.tycho.pitt.edu).

Website

The Tycho website features simple and advanced searches of the data, visualizations and downloading of machine readable files. A dataset can be selected by specifying: 1) the disease of interest, 2) the location(s) of interest and 3) the time period. In a second step, the type of aggregation and length of increments can be specified and in a third step, the type of output (table, graphic or map) can be selected. After output has been generated, data can be downloaded.

As mentioned above, data availability varies largely by disease, time period and location. We aimed to restrict selection options by data availability to avoid "no results available.

Access to the most of the website will be password protected during the testing phase. For invited users after login, full functionality of the database and website will be available. During this phase however, the data quality will not be sufficient for publication or official use.

Overview of data available for the state of Hawaii

In this section, an overview of data available for the state of Hawaii will be provided. As described earlier, data availability depends entirely on historical reporting practices in the weekly US Nationally Notifiable Disease Surveillance System between 1888 and 2009. For this report, we only included morbidity reports (cases) to improve the format and limit the size of this report. As mentioned earlier, the analysis and standardization of has not been completed yet for all records in the database and preliminary data are provided here that may not yet be available in the online database.

This section will provide a general overview of data availability. The next sections will provide disease specific data from state and city level reports of Hawaii. Data for a maximum of 3 major cities were provided depending on availability (city data was only reported until 1953). Some summary data will be provided at the end of this report, for a subset of diseases for which data was fragmented over time.

Table 1 lists the number of weekly state or city reports that are available for Hawaii per each disease and subcategory.

Table 1, Number of weekly state or city reports per disease and subcategory

Disease	City	State
Aids	-	937
Brucellosis [undulant fever]	1	81
Chickenpox [varicella]	-	520
Chlamydia	-	663
Cryptosporidiosis	-	227
Diphtheria	-	84
Encephalitis		
Post infectious	-	142
Primary [infectious] including unspecified	-	397
Escherichia coli		
EHEC 0157	-	152
O157:H7 NETSS	-	99
O157:H7 PHLIS	-	244
STEC	-	176
Giardiasis	-	373
Gonorrhea		
Civilian	-	753
Unspecified	-	1016
Haemophilus influenzae		
Age <5 non-serotype B	-	35
Age <5 unknown serotype	-	36
All ages all serotypes	-	602

Table 1, Number of weekly state or city reports per disease and subcategory, *continued*

Disease	City	State
Hepatitis		
Acute type A	-	331
Acute type B	-	310
Acute type C	-	80
All types, <20 years	-	50
All types, >=20 years	-	59
All types, all ages	5	496
Type A [infectious]	-	1314
Type B [serum]	-	1099
Type NA NB [including C]	-	515
Type unspecified	-	359
Influenza	-	53
Legionellosis	-	588
Leprosy	-	354
Listeriosis	-	132
Lyme disease	-	91
Malaria		
Military	-	22
Unspecified	-	1613
Measles		
Imported	-	635
Indigenous	-	536
Unspecified	23	1663
Meningitis		
Aseptic	-	974
Meningococcus	-	50
Meningococcal disease		
All serogroups	-	41
Invasive all serogroups	-	161
Invasive serogroup unknown	-	40
Other serogroup	-	2
Serogroup A C Y and W-135	-	2
Serogroup B	-	34
Serogroup unspecified	1	1791
Mumps	-	1631
Pneumonia	-	18
Poliomyelitis		
Non paralytic	-	69
Paralytic	-	291
Total	12	409
Rabies in animals	1	33

Table 1, Number of weekly state or city reports per disease and subcategory, *continued*

Disease	City	State
Rocky mountain spotted fever	-	16
Rubella		
Congenital	-	32
Unspecified	-	1461
Salmonellosis		
NETSS	-	108
PHLIS	-	123
Unspecified	-	383
Scarlet fever		
Including streptococcal sore throat	8	376
Unspecified	-	84
Shigellosis		
NETSS	-	109
PHLIS	-	119
Unspecified	-	375
Streptococcal disease, invasive group a	-	374
Streptococcal sore throat	-	76
Streptococcus pneumoniae invasive disease		
Drug resistant <5 years	-	87
Drug resistant A	-	5
Drug resistant all ages	-	150
Drug resistant B	-	5
Non drug resistant <5 years	-	153
Syphilis		
Civilian primary and secondary	-	721
Congenital	-	32
Primary and secondary	-	863
Tetanus	-	111
Toxic shock syndrome	-	148
Trichiniasis	1	5
Tuberculosis [phthisis pulmonalis]		
New active	-	150
Unspecified	-	1485
Typhoid fever [enteric fever]		
Including paratyphoid fever	-	12
Unspecified	5	1370
Typhus fever	2	26
Whooping cough [pertussis]	14	1237

The diseases included in the weekly US Nationally Notifiable Disease Surveillance System varied largely over time and reflected the historical social-political priorities of each time period. Note that the diseases in the weekly system were a subset of all diseases included in the annual Notifiable Disease Surveillance System (that was not entered as part of this project).

The number of locations reporting diseases also varied over time depending on the type of reports submitted. Figure 1 on the next page provides an overview of the weeks for which a report was available per disease at the state or city level. Colors indicate the total number of reporting locations. Only state level reports for Hawaii were available.

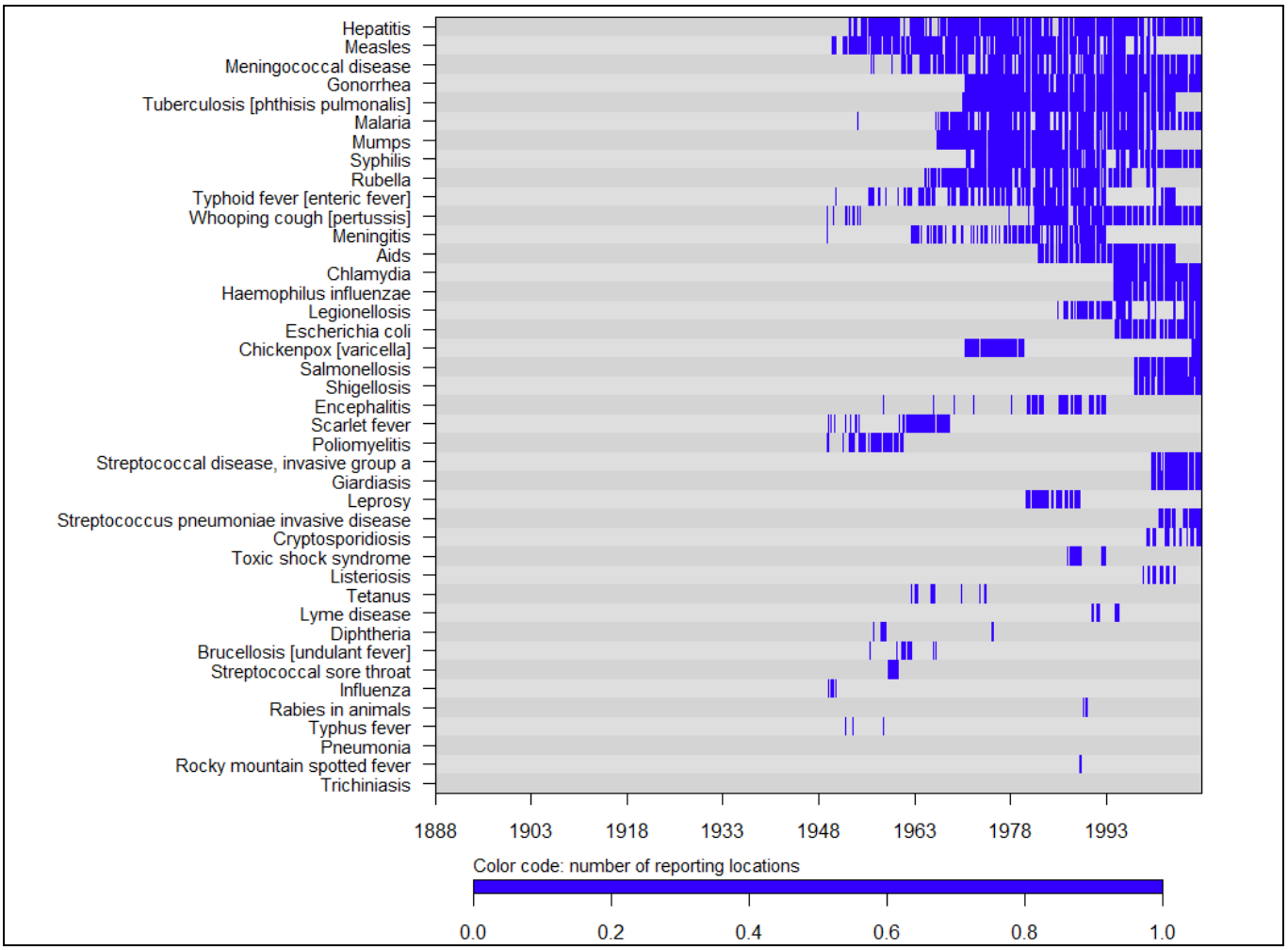

Figure 1, The number of locations (in color code) reporting on each disease per week between 1888 and 2009

The map in figure 2 (next page) illustrates the geographical distribution of the available city data. It displays the number of weekly reports for any disease available per location that was included (and for which coordinates could be derived)

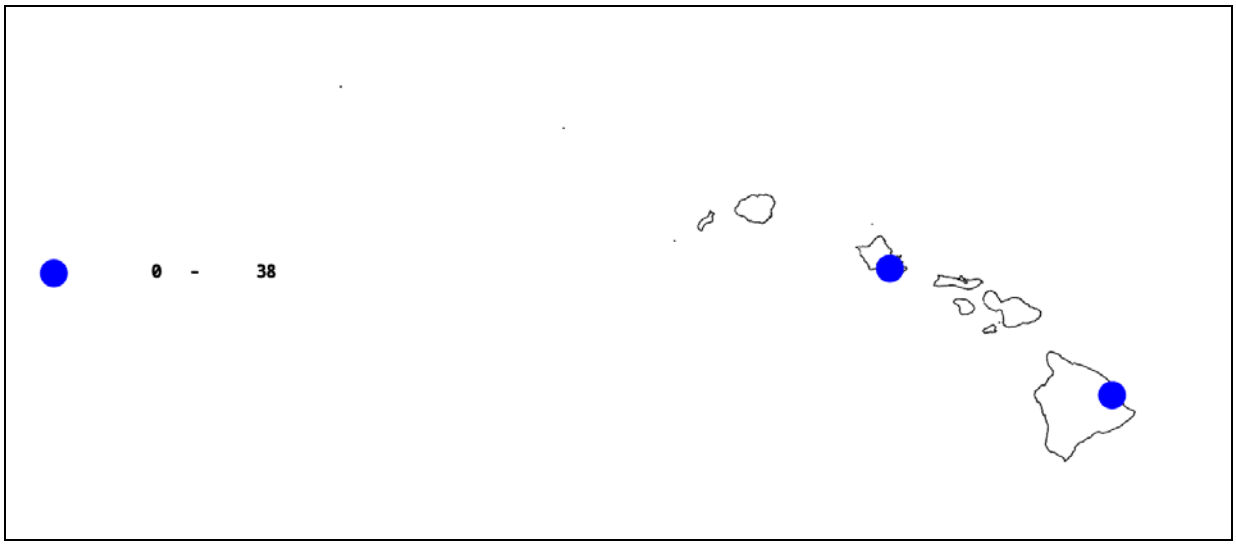

Figure 2, This map displays the number of weekly reports available for any disease per city (note that city reports were only available until 1953)

Figure 3 below provides an overview of the number of diseases for which a weekly report was included at the state state level and for individual cities. All cities for which at least 100 weekly reports were available have been listed. For most states, it shows a major drop in the number of cities after 1925 and after 1953, city reports were discontinued except for New York City, which became its own reporting jurisdiction. For Hawaii, no city reported more than 100 weekly reports.

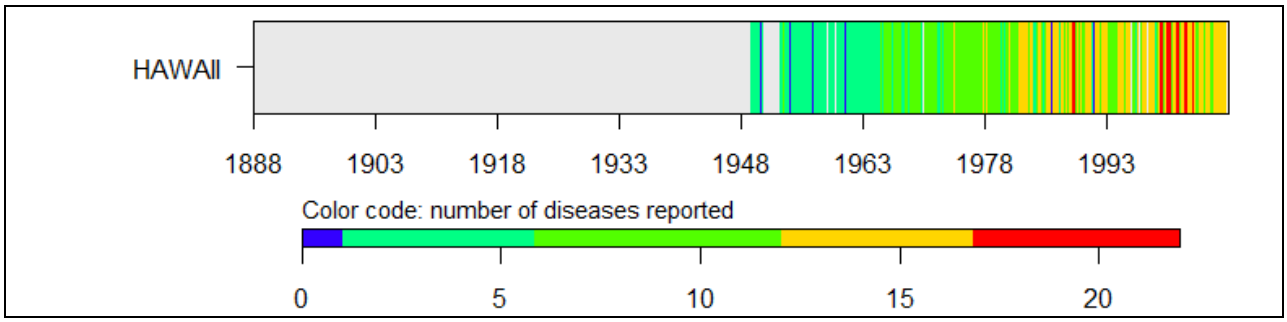

Figure 3, The number of diseases (in color code) that were reported for the state and city level (state on top row) for each week between 1888 and 2009

Disease specific data for Hawaii

AIDS

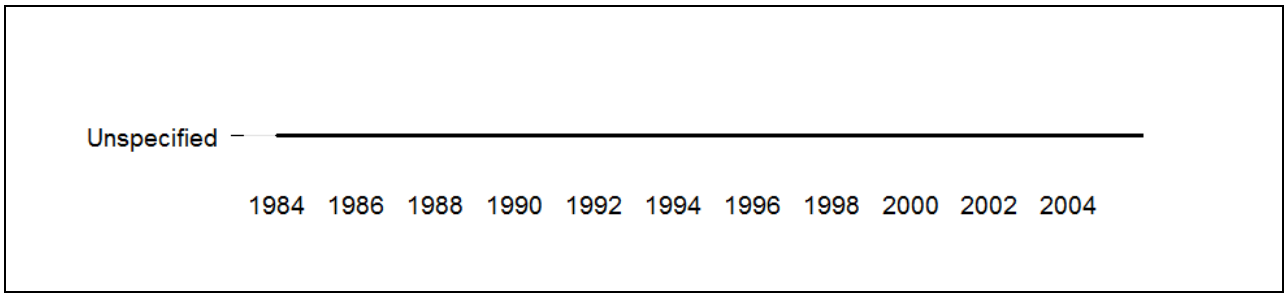

Figure D1, Weeks between 1888 and 2009 for which data on the disease and subcategories (if applicable) are available in the Tycho database. If no subcategory was reported, a subcategory of "Unspecified" was assigned.

Table D1, Summary information for AIDS

Indicator	Hawaii
Report period	1984-2005
Total weeks	937
Total cases	2,658
Max. cases per year	341
Year (max)	1993
Max. cases per week	83
Week (max)	1993, wk 19
Average cases per year	121
95%CI	(89-153)
Average cases per week	3
95%CI	(3-3)

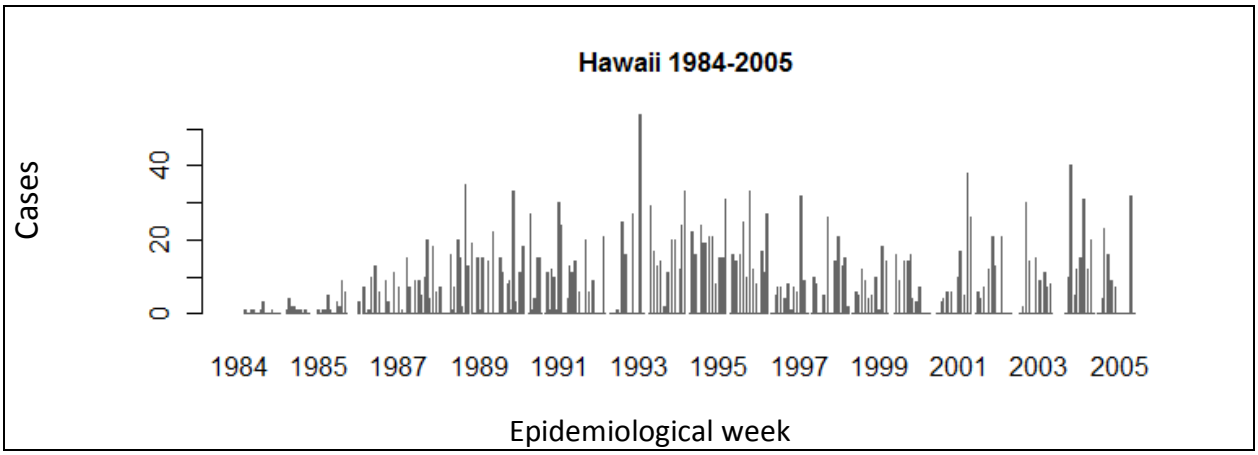

Figure D2, Number of cases reported for AIDS per epidemiological week

Chlamydia

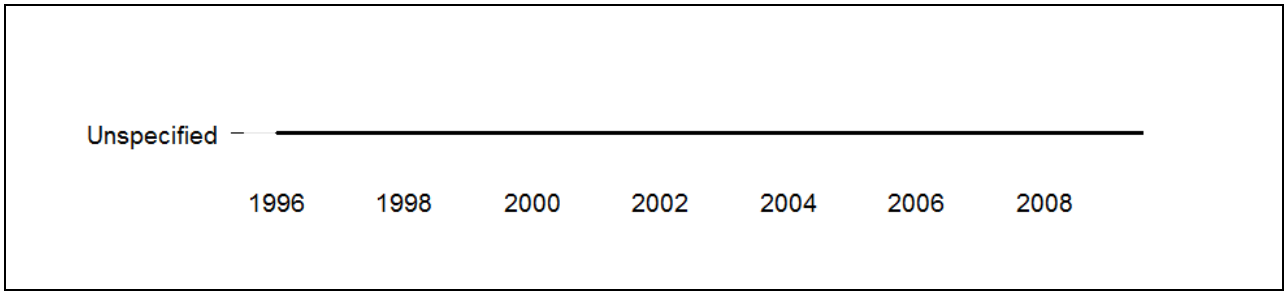

Figure D1, Weeks between 1888 and 2009 for which data on the disease and subcategories (if applicable) are available in the Tycho database. If no subcategory was reported, a subcategory of "Unspecified" was assigned.

Table D1, Summary information for Chlamydia

Indicator	Hawaii
Report period	1996-2009
Total weeks	663
Total cases	49,945
Max. cases per year	5,195
Year (max)	2005
Max. cases per week	1,902
Week (max)	2004, wk 40
Average cases per year	3,568
95%CI	(2,758-4,378)
Average cases per week	75
95%CI	(65-85)

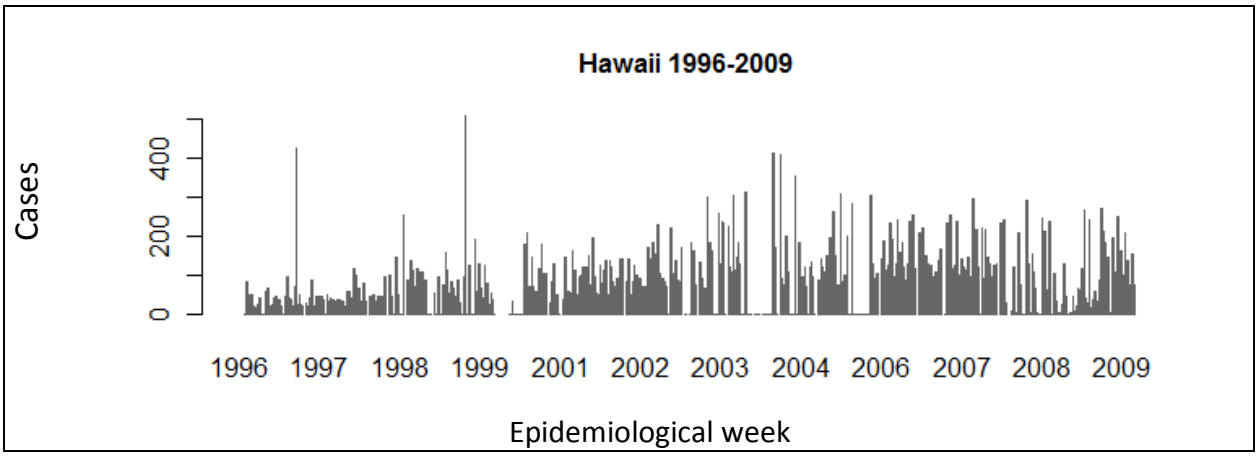

Figure D2, Number of cases reported for Chlamydia per epidemiological week

Cryptosporidiosis

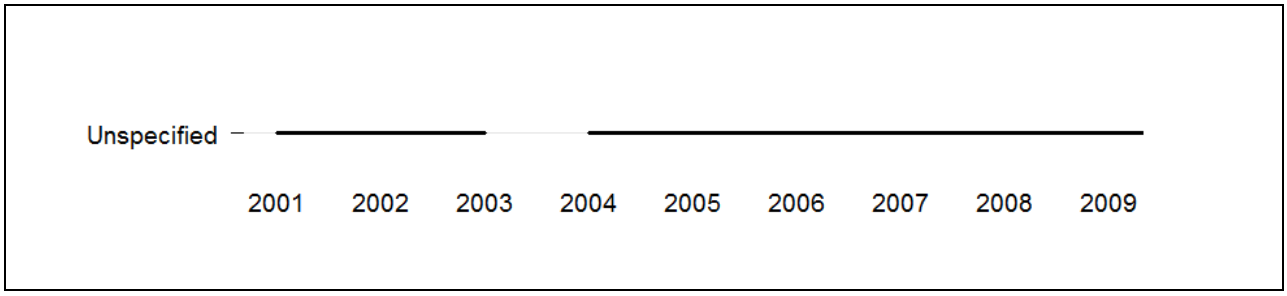

Figure D1, Weeks between 1888 and 2009 for which data on the disease and subcategories (if applicable) are available in the Tycho database. If no subcategory was reported, a subcategory of "Unspecified" was assigned.

Table D1, Summary information for Cryptosporidiosis

Indicator	Hawaii
Report period	2001-2009
Total weeks	227
Total cases	7
Max. cases per year	2
Year (max)	2001
Max. cases per week	1
Week (max)	2001, wk 22
Average cases per year	1
95%CI	(0-2)
Average cases per week	0
95%CI	(0-0)

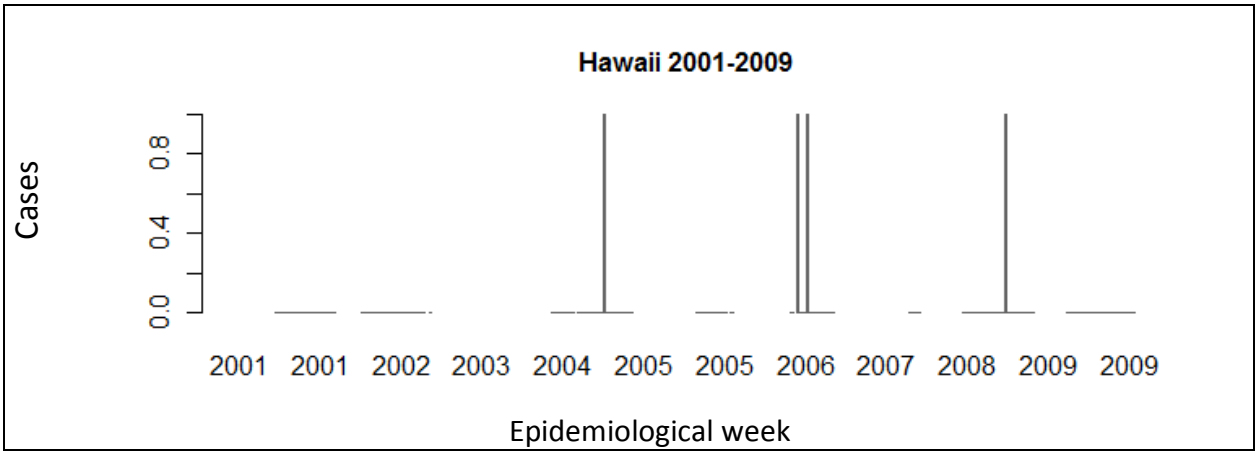

Figure D2, Number of cases reported for Cryptosporidiosis per epidemiological week

Escherichia Coli

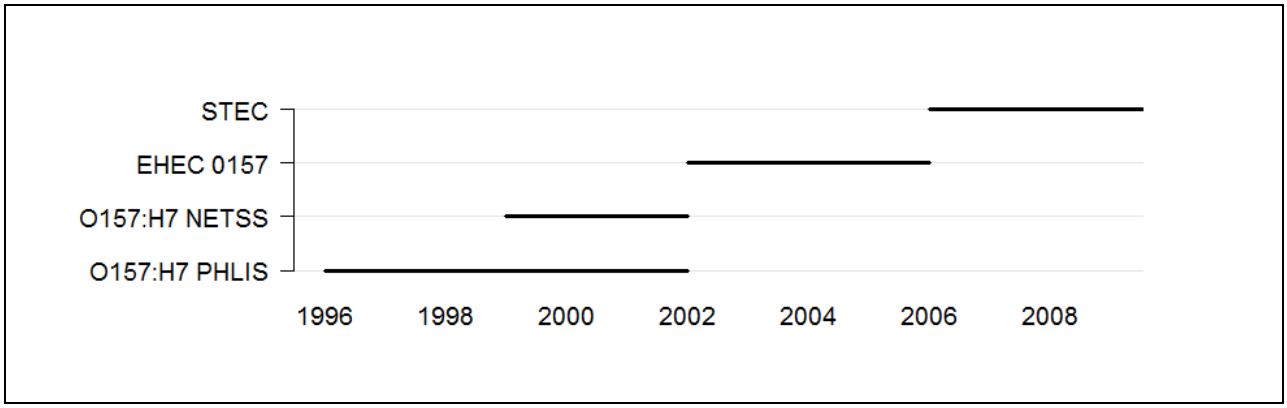

Figure D1, Weeks between 1888 and 2009 for which data on the disease and subcategories (if applicable) are available in the Tycho database. If no subcategory was reported, a subcategory of "Unspecified" was assigned.

Table D1, Summary information for Escherichia Coli (0157:H7 PHLIS, EHEC 0157, and STEC)

Indicator	Hawaii
Report period	1996-2009
Total weeks	572
Total cases	179
Max. cases per year	39
Year (max)	2002
Max. cases per week	6
Week (max)	2002, wk 33
Average cases per year	13
95%CI	(8-18)
Average cases per week	0
95%CI	(0-0)

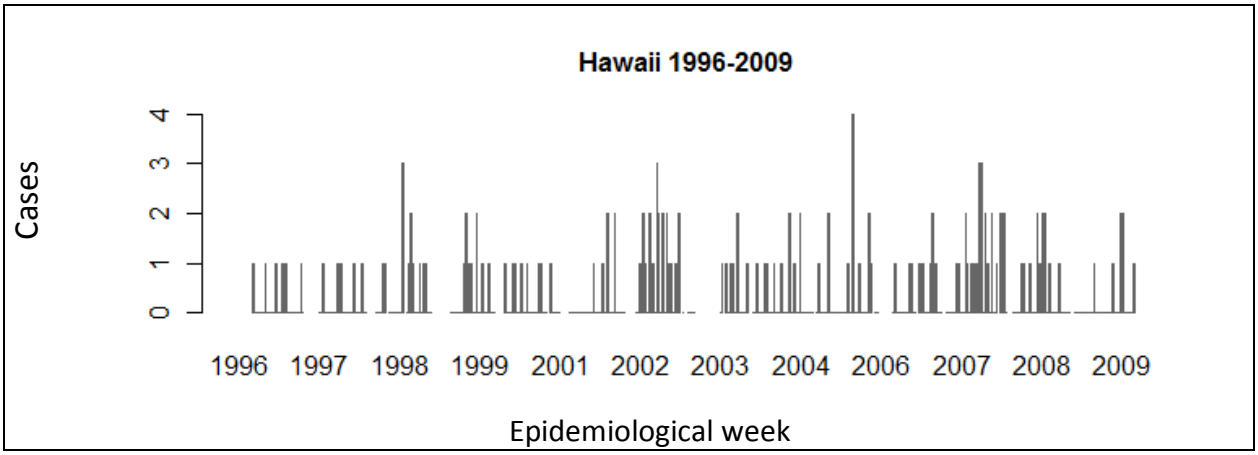

Figure D2, Number of cases reported for Escherichia Coli per epidemiological week

Giardiasis

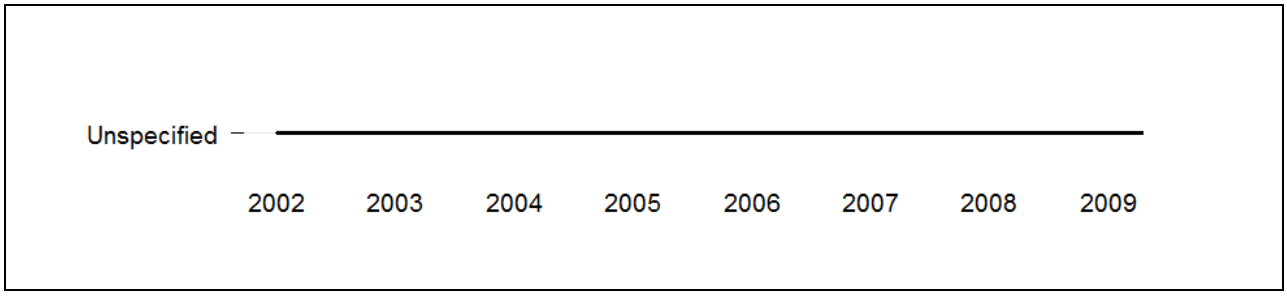

Figure D1, Weeks between 1888 and 2009 for which data on the disease and subcategories (if applicable) are available in the Tycho database. If no subcategory was reported, a subcategory of "Unspecified" was assigned.

Table D1, Summary information for Giardiasis

Indicator	Hawaii
Report period	2002-2009
Total weeks	373
Total cases	3,777
Max. cases per year	3,352
Year (max)	2006
Max. cases per week	3,304
Week (max)	2006, wk 35
Average cases per year	472
95%CI	(-501-1,445)
Average cases per week	10
95%CI	(-7-27)

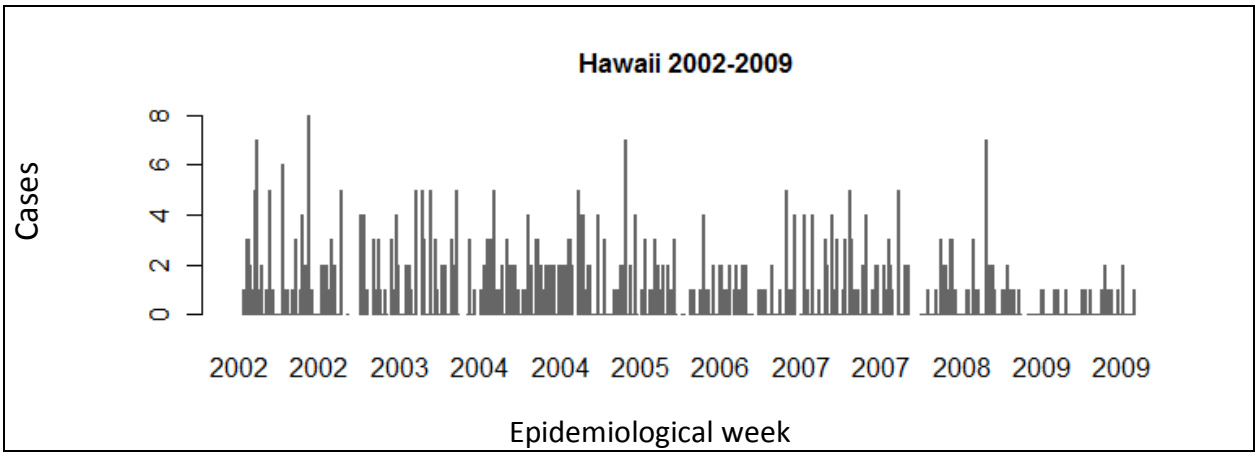

Figure D2, Number of cases reported for Giardiasis per epidemiological week

Gonorrhea

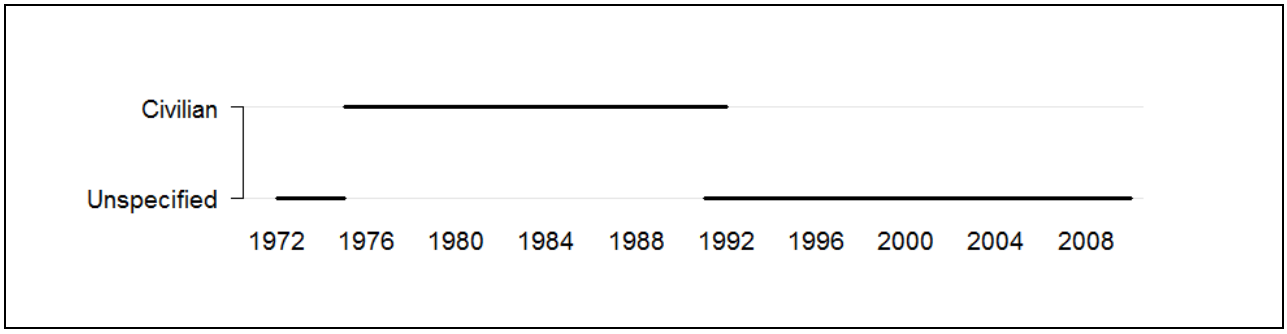

Figure D1, Weeks between 1888 and 2009 for which data on the disease and subcategories (if applicable) are available in the Tycho database. If no subcategory was reported, a subcategory of "Unspecified" was assigned.

Table D1, Summary information for Gonorrhea (Unspecified and Civilian)

Indicator	Hawaii
Report period	1972-2009
Total weeks	1,769
Total cases	48,632
Max. cases per year	3,274
Year (max)	1976
Max. cases per week	551
Week (max)	1976, wk 34
Average cases per year	1,280
95%CI	(965-1,595)
Average cases per week	27
95%CI	(25-29)

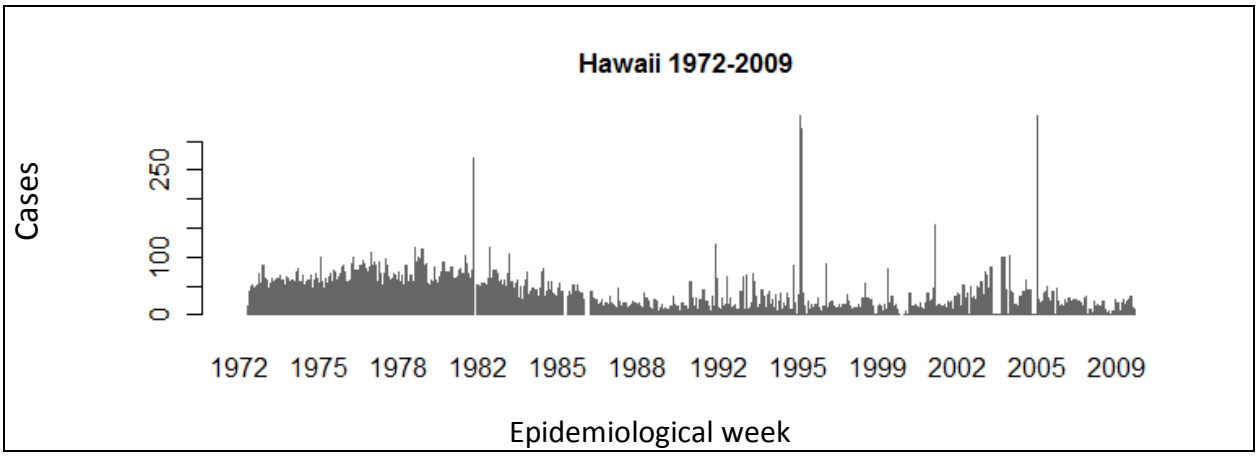

Figure D2, Number of cases reported for Gonorrhea per epidemiological week

Haemophilus Influenzae

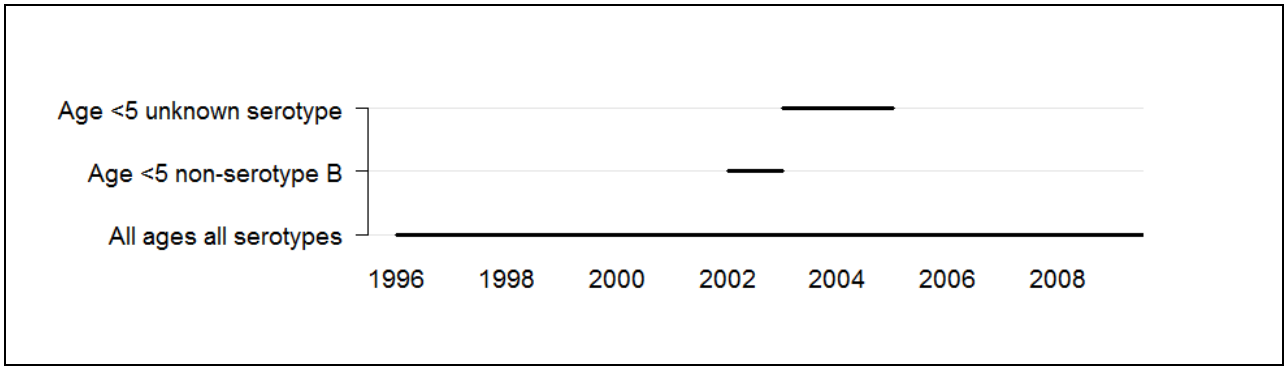

Figure D1, Weeks between 1888 and 2009 for which data on the disease and subcategories (if applicable) are available in the Tycho database. If no subcategory was reported, a subcategory of "Unspecified" was assigned.

Table D1, Summary information for Haemophilus Influenzae (All ages all serotypes)

Indicator	Hawaii
Report period	1996-2009
Total weeks	602
Total cases	231
Max. cases per year	32
Year (max)	2002
Max. cases per week	17
Week (max)	2001, wk 28
Average cases per year	16
95%CI	(10-22)
Average cases per week	0
95%CI	(0-0)

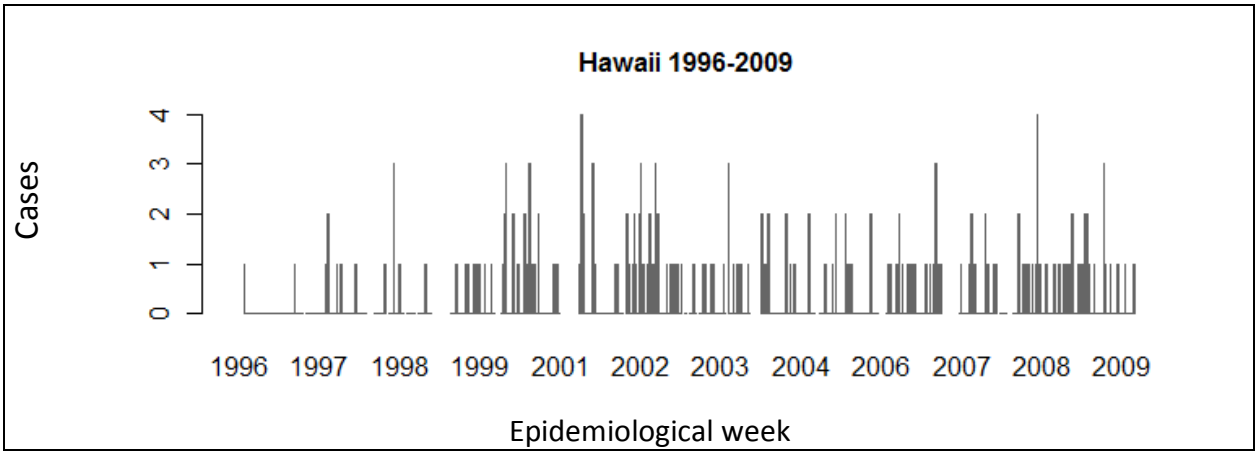

Figure D2, Number of cases reported for Haemophilus Influenzae per epidemiological week

Hepatitis

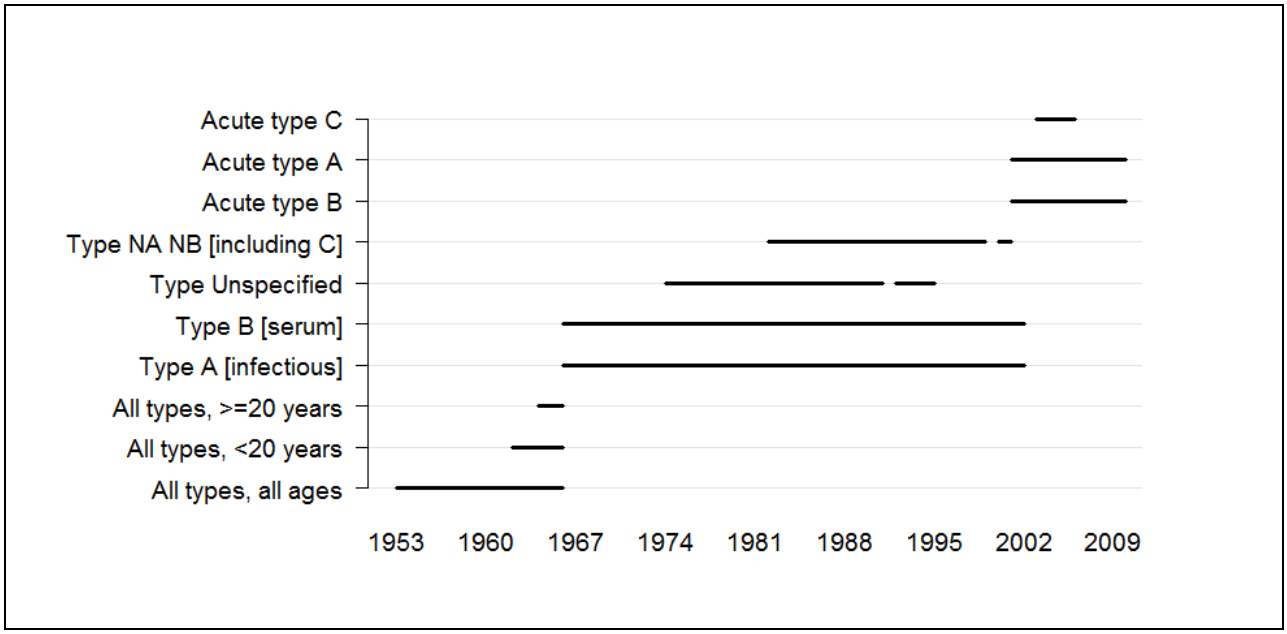

Figure D1, Weeks between 1888 and 2009 for which data on the disease and subcategories (if applicable) are available in the Tycho database. If no subcategory was reported, a subcategory of "Unspecified" was assigned.

Table D1, Summary information for Hepatitis (Type A [Infectious], Acute type A, Type B [Serum], and Acute type B)

Indicator	**Hawaii**
Report period	1966-2009
Total weeks	1,853
Total cases	4,971
Max. cases per year	1,503
Year (max)	1970
Max. cases per week	1,271
Week (max)	1970, wk 44
Average cases per year	
before 1990	177
95%CI	(60-294)
after 1990	29
95%CI	(16-42)
Average cases per week	
before 1990	4
95%CI	(2-6)
after 1990	1
95%CI	(1-1)

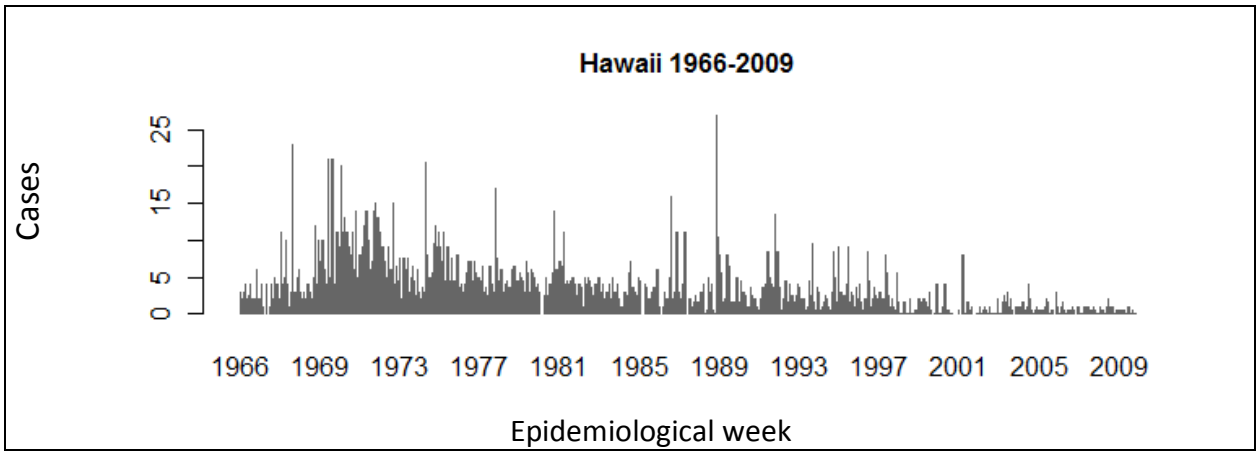

Figure D2, Number of cases reported for Hepatitis per epidemiological week

Influenza

Figure D1, Weeks between 1888 and 2009 for which data on the disease and subcategories (if applicable) are available in the Tycho database. If no subcategory was reported, a subcategory of "Unspecified" was assigned.

Table D1, Summary information for Influenza

Indicator	Hawaii
Report period	1950-1951
Total weeks	53
Total cases	2,175
Max. cases per year	1,454
Year (max)	1950
Max. cases per week	544
Week (max)	1950, wk 46
Average cases per year	1,088
95%CI	(-3,569-5,745)
Average cases per week	41
95%CI	(18-64)

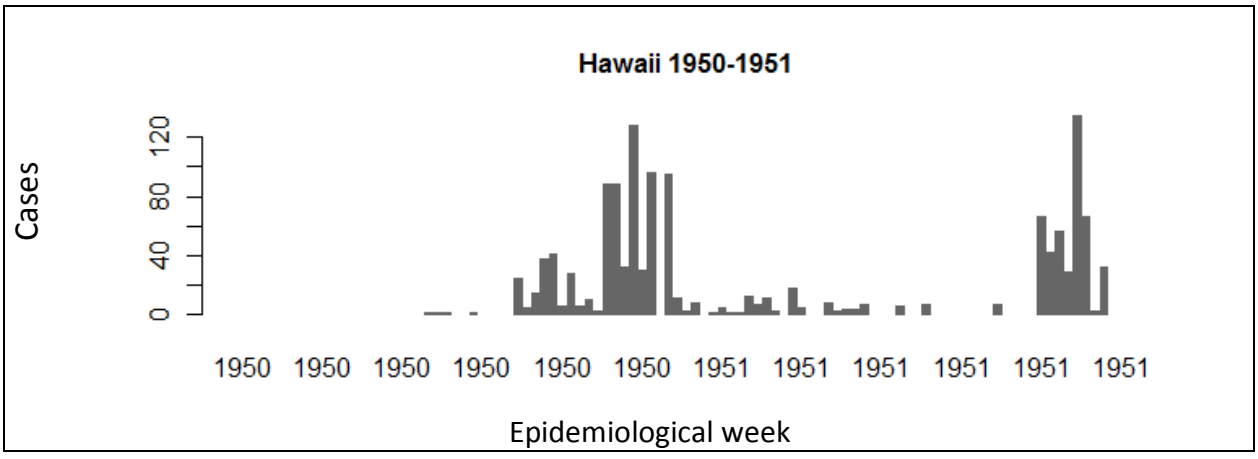

Figure D2, Number of cases reported for Influenza per epidemiological week

Leprosy

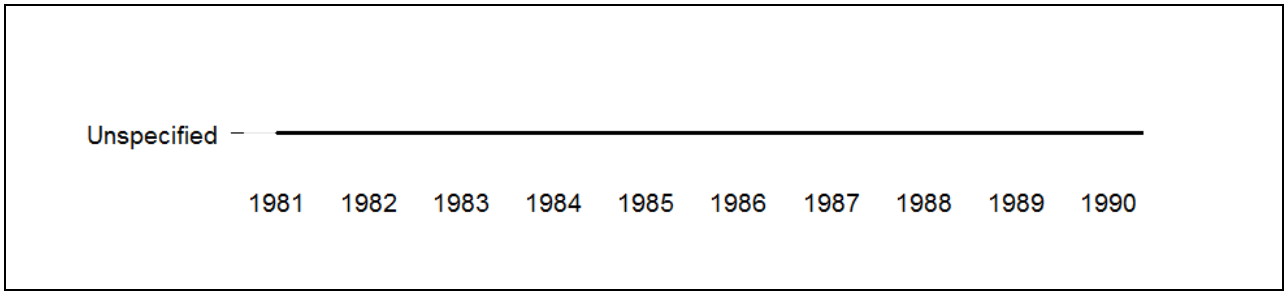

Figure D1, Weeks between 1888 and 2009 for which data on the disease and subcategories (if applicable) are available in the Tycho database. If no subcategory was reported, a subcategory of "Unspecified" was assigned.

Table D1, Summary information for Leprosy

Indicator	Hawaii
Report period	1981-1990
Total weeks	354
Total cases	211
Max. cases per year	40
Year (max)	1990
Max. cases per week	18
Week (max)	1990, wk 52
Average cases per year	21
95%CI	(13-29)
Average cases per week	1
95%CI	(1-1)

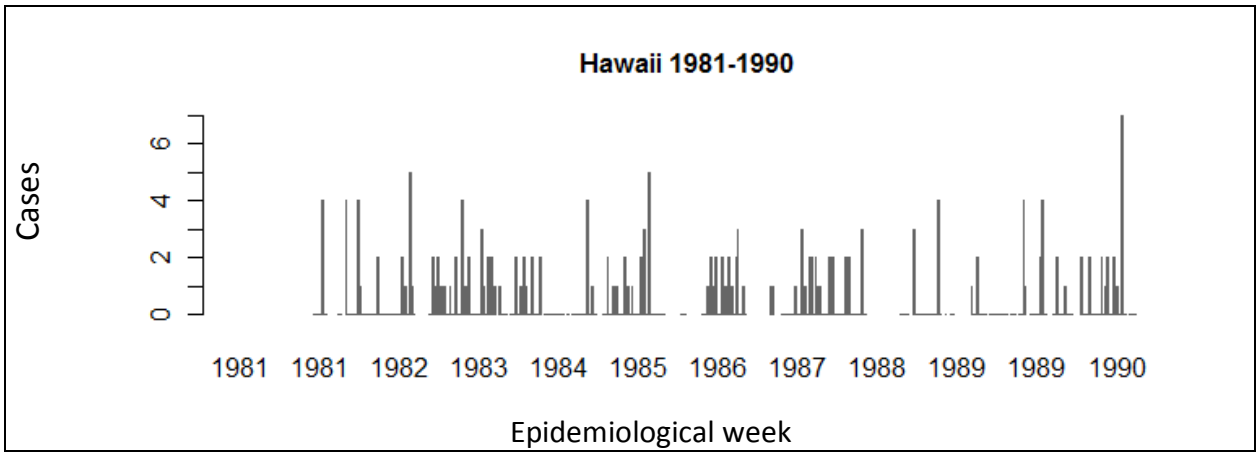

Figure D2, Number of cases reported for Leprosy per epidemiological week

Malaria

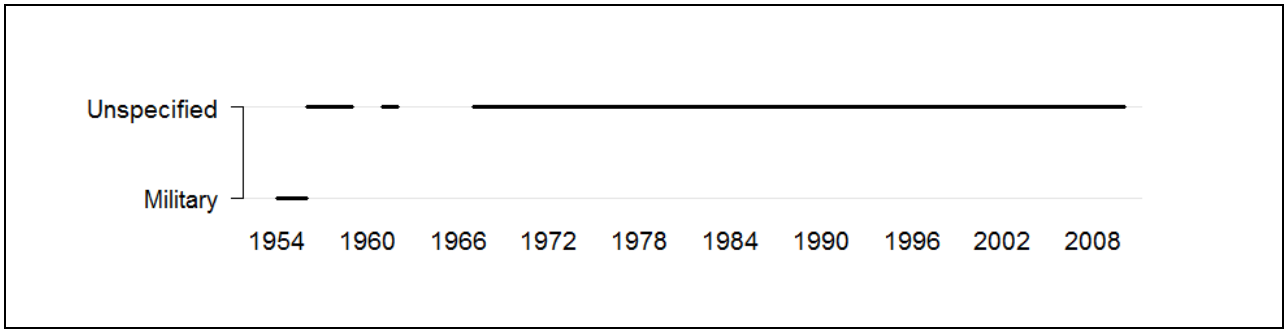

Figure D1, Weeks between 1888 and 2009 for which data on the disease and subcategories (if applicable) are available in the Tycho database. If no subcategory was reported, a subcategory of "Unspecified" was assigned.

Table D1, Summary information for Malaria (Unspecified)

Indicator	Hawaii
Report period	1956-2009
Total weeks	1,613
Total cases	839
Max. cases per year	142
Year (max)	1970
Max. cases per week	31
Week (max)	1968, wk 38
Average cases per year	18
95%CI	(8-28)
Average cases per week	1
95%CI	(1-1)

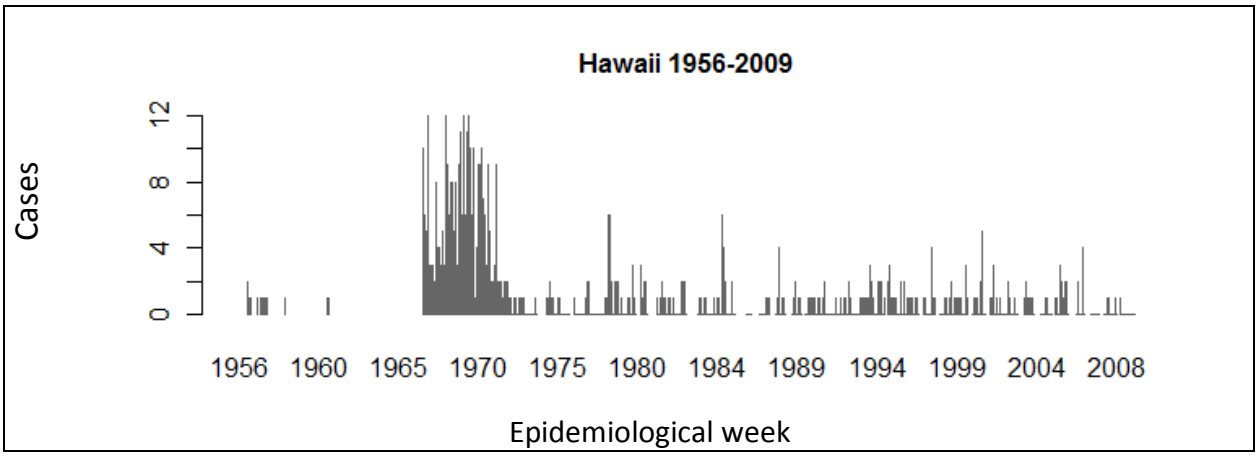

Figure D2, Number of cases reported for Malaria per epidemiological week

Measles

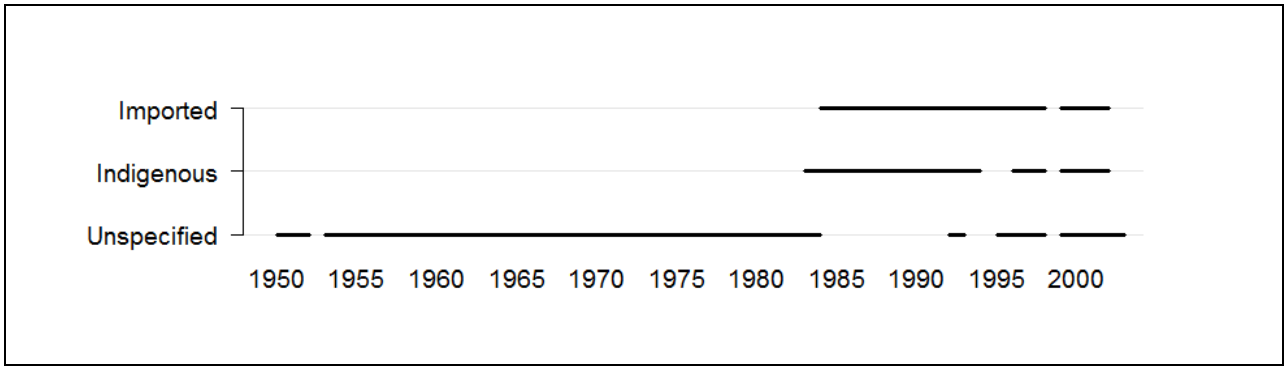

Figure D1, Weeks between 1888 and 2009 for which data on the disease and subcategories (if applicable) are available in the Tycho database. If no subcategory was reported, a subcategory of "Unspecified" was assigned.

Table D1, Summary information for Measles (Unspecified)

Indicator	Hawaii
Report period	1950-2002
Total weeks	1,663
Total cases	44,225
Max. cases per year	6,369
Year (max)	1955
Max. cases per week	784
Week (max)	1951, wk 49
Average cases per year	
before 1970	2,276
95%CI	(1,172-3,380)
after 1970	47
95%CI	(-1-95)
Average cases per week	
before 1970	49
95%CI	(43-55)
after 1970	1
95%CI	(1-1)

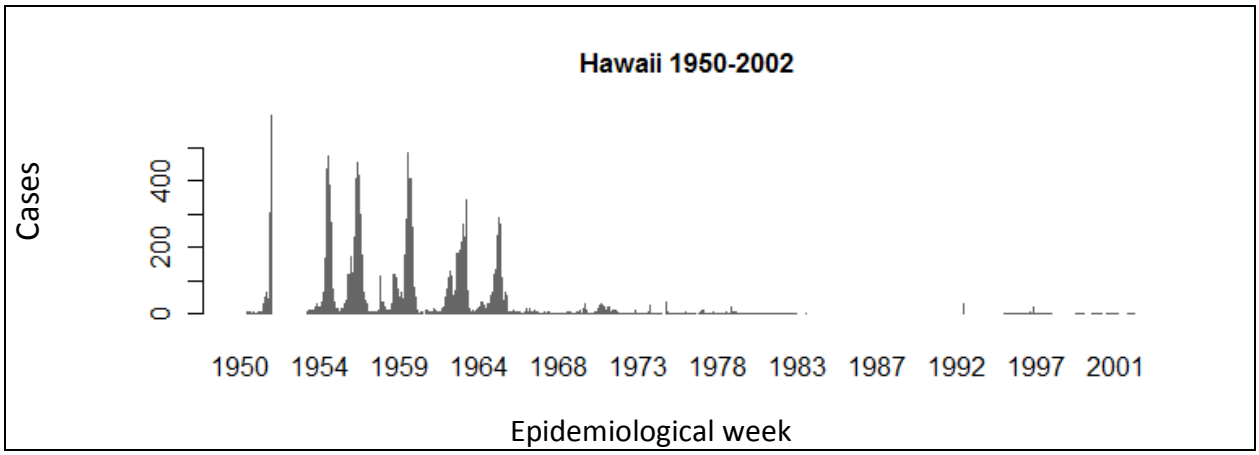

Figure D2, Number of cases reported Measles per epidemiological week

Meningococcal Disease

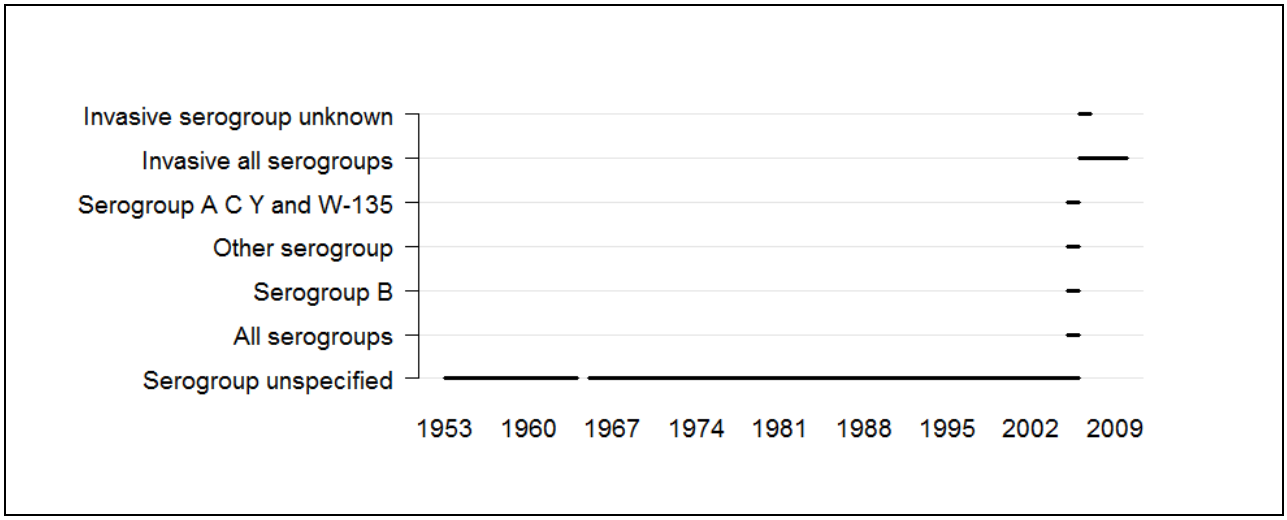

Figure D1, Weeks between 1888 and 2009 for which data on the disease and subcategories (if applicable) are available in the Tycho database. If no subcategory was reported, a subcategory of "Unspecified" was assigned.

Table D1, Summary information for Meningococcal Disease (Serogroup unspecified and Invasive all serogroups)

Indicator	**Hawaii**
Report period	1954-2009
Total weeks	1,952
Total cases	354
Max. cases per year	16
Year (max)	1988
Max. cases per week	8
Week (max)	1985, wk 46
Average cases per year	
before 1980	6
95%CI	(4-8)
after 1980	7
95%CI	(6-8)
Average cases per week	
before 1980	0
95%CI	(0-0)
after 1980	0
95%CI	(0-0)

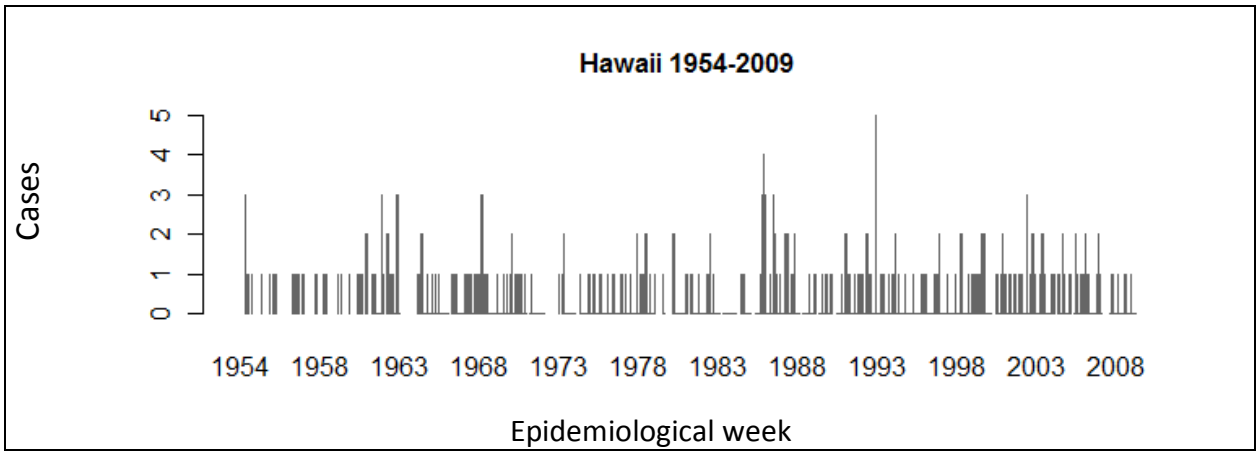

Figure D2, Number of cases reported for Meningococcal Disease per epidemiological week

Mumps

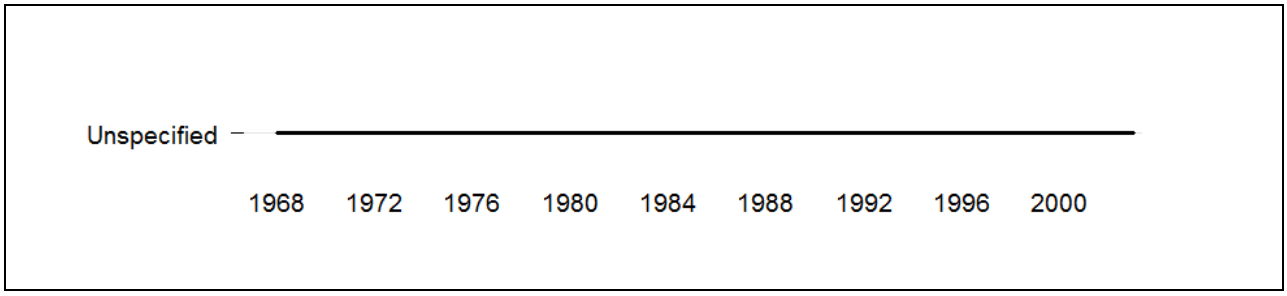

Figure D1, Weeks between 1888 and 2009 for which data on the disease and subcategories (if applicable) are available in the Tycho database. If no subcategory was reported, a subcategory of "Unspecified" was assigned.

Table D1, Summary information for Mumps

Indicator	Hawaii
Report period	1968-2002
Total weeks	1,631
Total cases	5,387
Max. cases per year	1,493
Year (max)	1970
Max. cases per week	83
Week (max)	1970, wk 19
Average cases per year	
before 1980	380
95%CI	(95-665)
after 1980	20
95%CI	(17-23)
Average cases per week	
before 1980	8
95%CI	(7-9)
after 1980	0
95%CI	(0-0)

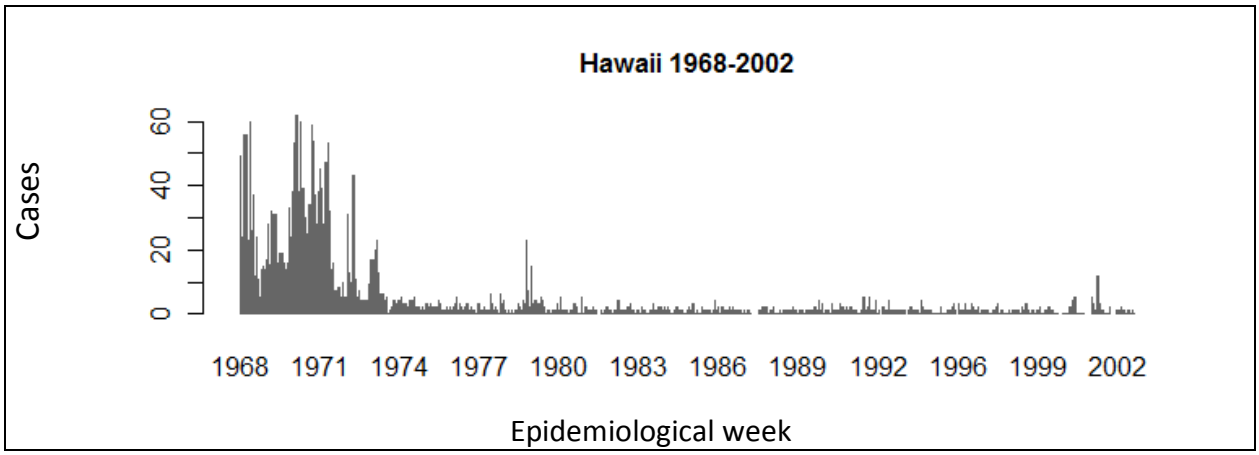

Figure D2, Number of cases reported for Mumps per epidemiological week

Rabies in Animals

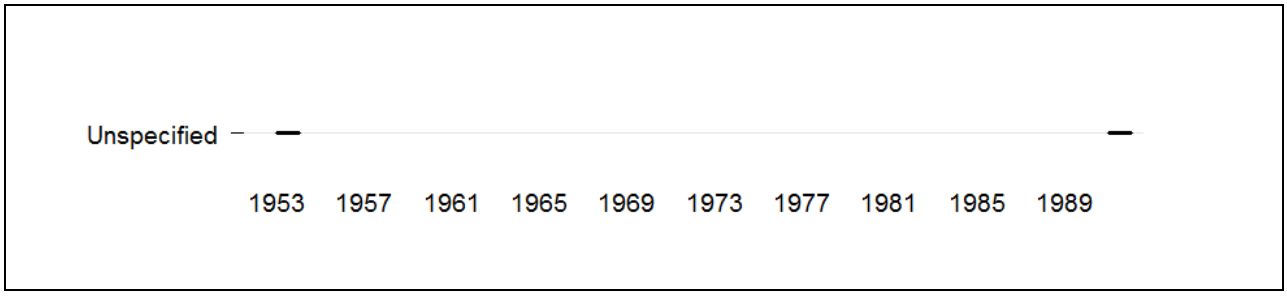

Figure D1, Weeks between 1888 and 2009 for which data on the disease and subcategories (if applicable) are available in the Tycho database. If no subcategory was reported, a subcategory of "Unspecified" was assigned.

Table D1, Summary information for Rabies in Animals

Indicator	Hawaii
Report period	1991-1991
Total weeks	33
Total cases	0

Rocky Mountain Spotted Fever

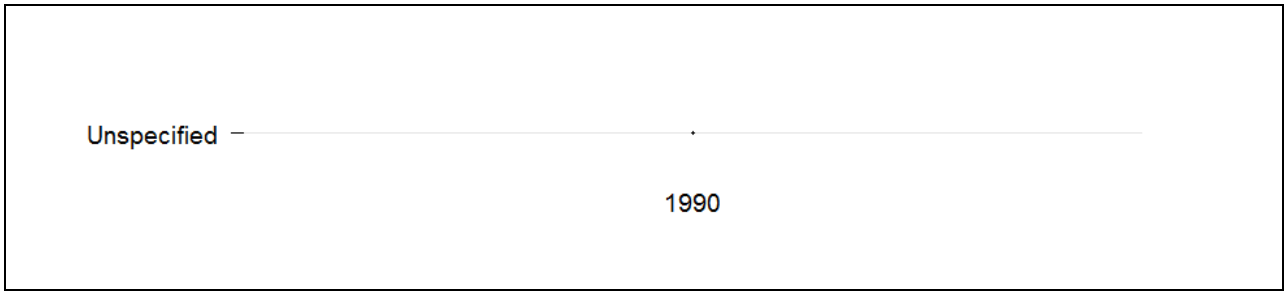

Figure D1, Weeks between 1888 and 2009 for which data on the disease and subcategories (if applicable) are available in the Tycho database. If no subcategory was reported, a subcategory of "Unspecified" was assigned.

Table D1, Summary information for Rocky Mountain Spotted Fever

Indicator	Hawaii
Report period	1990-1990
Total weeks	16
Total cases	0

Rubella

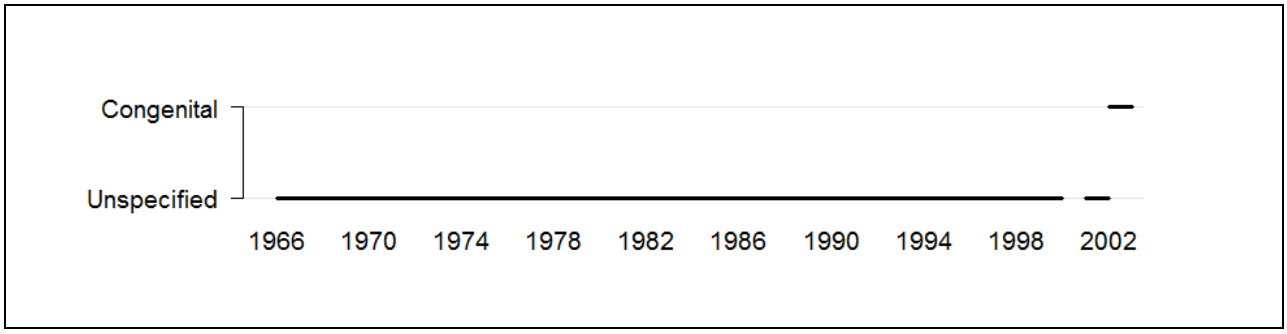

Figure D1, Weeks between 1888 and 2009 for which data on the disease and subcategories (if applicable) are available in the Tycho database. If no subcategory was reported, a subcategory of "Unspecified" was assigned.

Table D1, Summary information for Rubella

Indicator	Hawaii
Report period	1966-2001
Total weeks	1,461
Total cases	3,235
Max. cases per year	975
Year (max)	1969
Max. cases per week	234
Week (max)	1977, wk 29
Average cases per year	
before 1970	433
95%CI	(7-859)
after 1970	36
95%CI	(7-65)
Average cases per week	
before 1970	10
95%CI	(7-13)
after 1970	1
95%CI	(1-1)

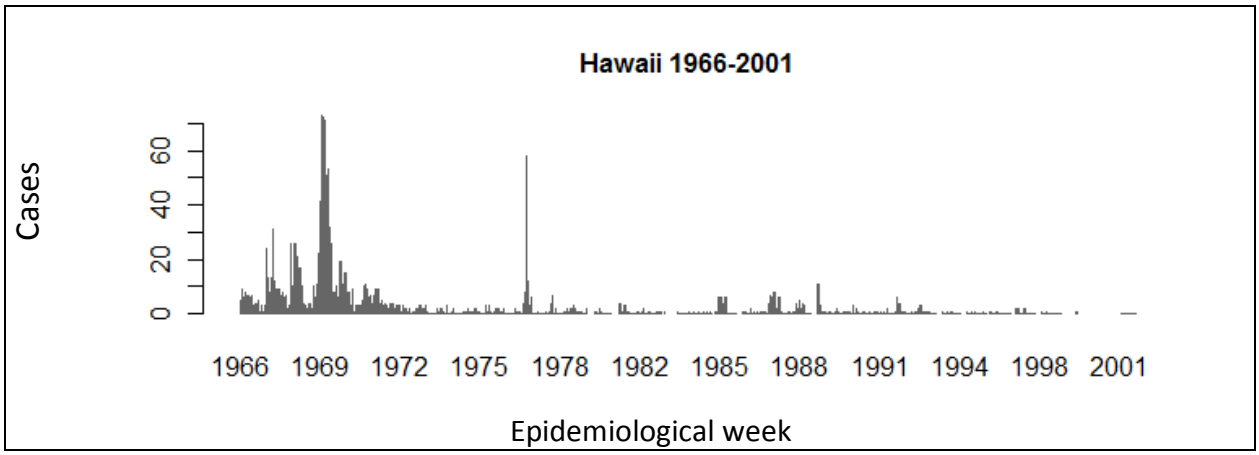

Figure D2, Number of cases reported for Rubella per epidemiological week

Salmonellosis

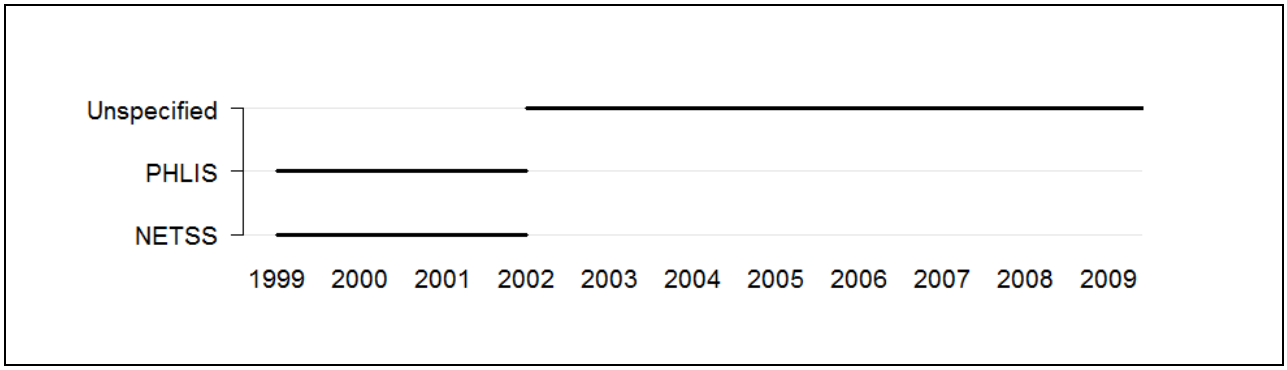

Figure D1, Weeks between 1888 and 2009 for which data on the disease and subcategories (if applicable) are available in the Tycho database. If no subcategory was reported, a subcategory of "Unspecified" was assigned.

Table D1, Summary information for Salmonellosis (PHLIS and Unspecified)

Indicator	Hawaii
Report period	1999-2009
Total weeks	506
Total cases	2,992
Max. cases per year	564
Year (max)	2005
Max. cases per week	281
Week (max)	2005, wk 50
Average cases per year	272
95%CI	(190-354)
Average cases per week	6
95%CI	(5-7)

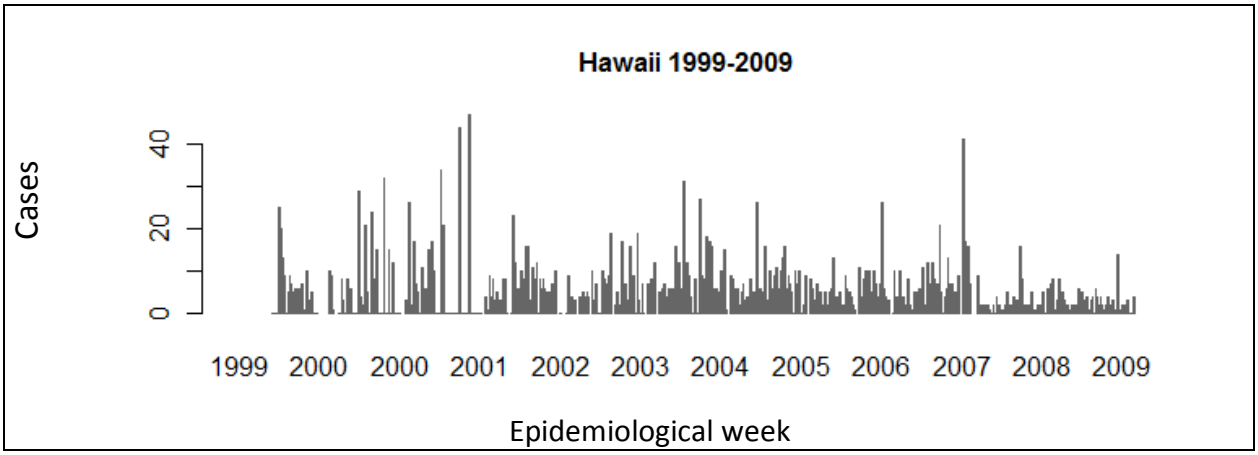

Figure D2, Number of cases reported for Salmonellosis per epidemiological week

Shigellosis

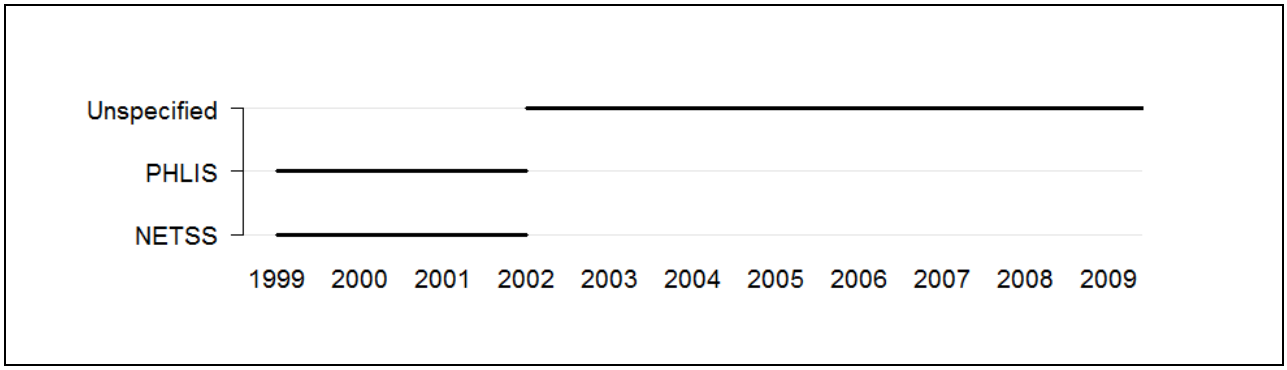

Figure D1, Weeks between 1888 and 2009 for which data on the disease and subcategories (if applicable) are available in the Tycho database. If no subcategory was reported, a subcategory of "Unspecified" was assigned.

Table D1, Summary information for Shigellosis (Unspecified and PHLIS)

Indicator	Hawaii
Report period	1999-2009
Total weeks	494
Total cases	616
Max. cases per year	181
Year (max)	1999
Max. cases per week	173
Week (max)	1999, wk 33
Average cases per year	56
95%CI	(27-85)
Average cases per week	1
95%CI	(0-2)

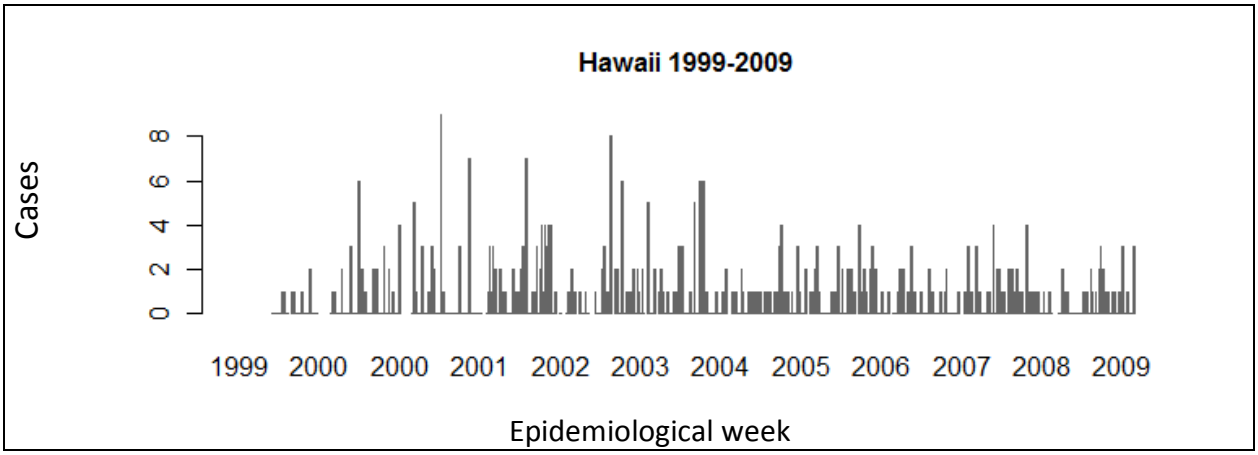

Figure D2, Number of cases reported for Shigellosis per epidemiological week

Streptococcal Disease, Invasive Group A

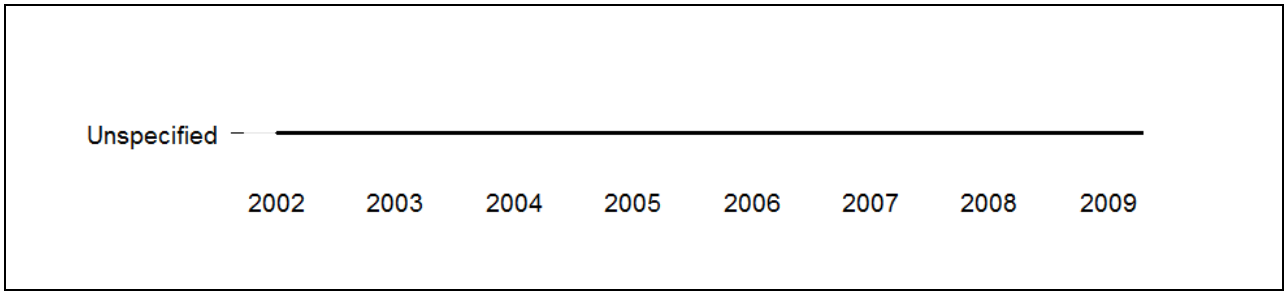

Figure D1, Weeks between 1888 and 2009 for which data on the disease and subcategories (if applicable) are available in the Tycho database. If no subcategory was reported, a subcategory of "Unspecified" was assigned.

Table D1, Summary information for Streptococcal Disease, Invasive Group A

Indicator	Hawaii
Report period	2002-2009
Total weeks	374
Total cases	964
Max. cases per year	194
Year (max)	2007
Max. cases per week	91
Week (max)	2007, wk 49
Average cases per year	120
95%CI	(91-149)
Average cases per week	3
95%CI	(2-4)

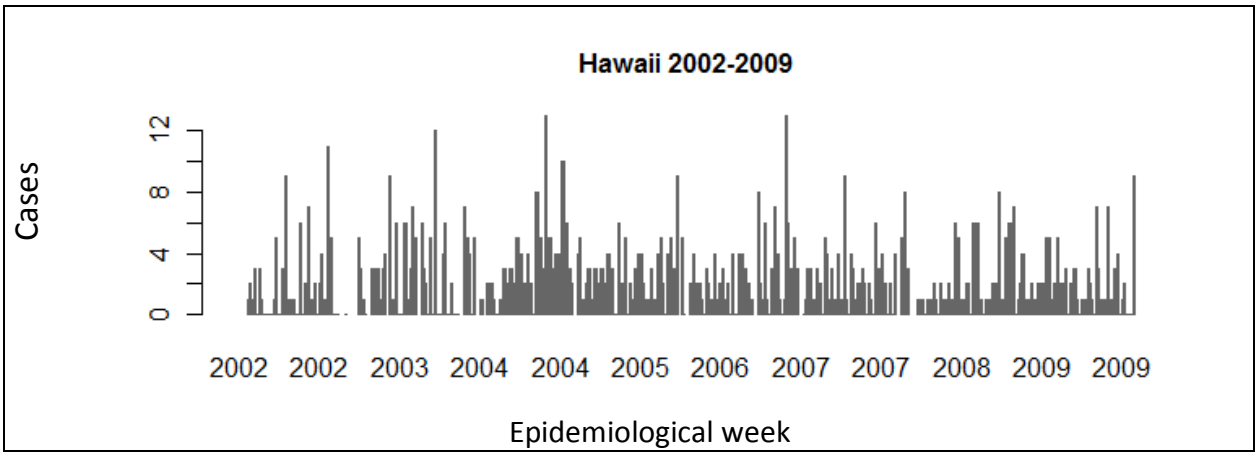

Figure D2, Number of cases reported for Group A Streptococcal Disease per epidemiological week

Streptococcal Sore Throat

Figure D1, Weeks between 1888 and 2009 for which data on the disease and subcategories (if applicable) are available in the Tycho database. If no subcategory was reported, a subcategory of "Unspecified" was assigned.

Table D1, Summary information for Streptococcal Sore Throat

Indicator	Hawaii
Report period	1960-1961
Total weeks	76
Total cases	834
Max. cases per year	638
Year (max)	1961
Max. cases per week	85
Week (max)	1961, wk 47
Average cases per year	417
95%CI	(-2,391-3,225)
Average cases per week	11
95%CI	(7-15)

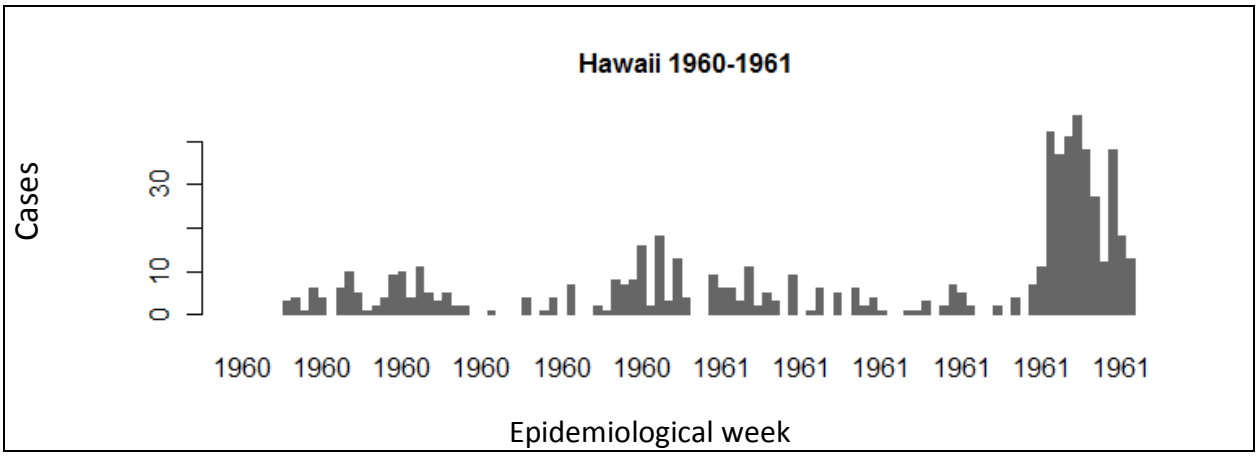

Figure D2, Number of cases reported for Streptococcal Sore Throat per epidemiological week

Streptococcus Pneumoniae Invasive Disease

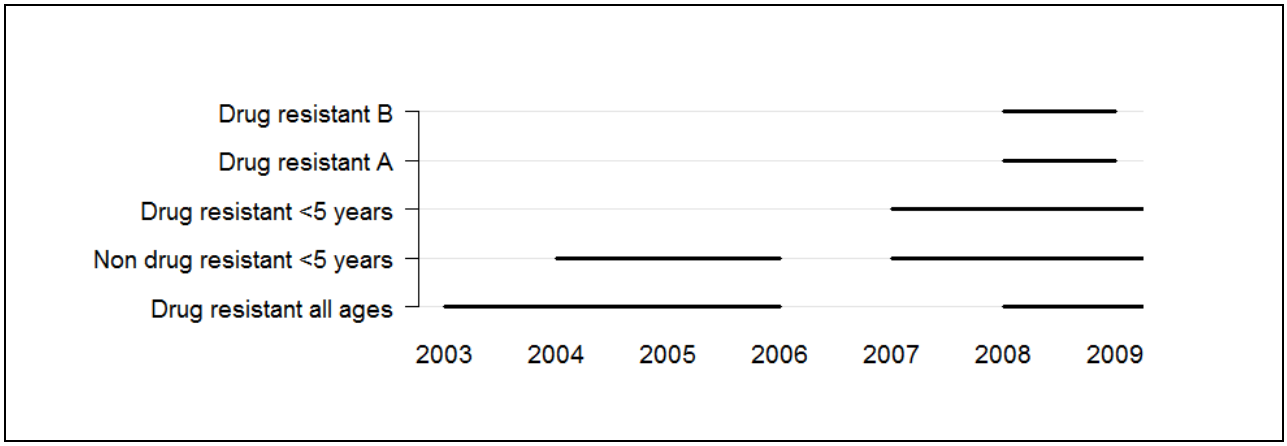

Figure D1, Weeks between 1888 and 2009 for which data on the disease and subcategories (if applicable) are available in the Tycho database. If no subcategory was reported, a subcategory of "Unspecified" was assigned.

Table D1, Summary information for Streptococcus Pneumoniae Invasive Disease (Drug resistant <5 years)

Indicator	Hawaii
Report period	2007-2009
Total weeks	87
Total cases	4
Max. cases per year	2
Year (max)	2007

Syphilis

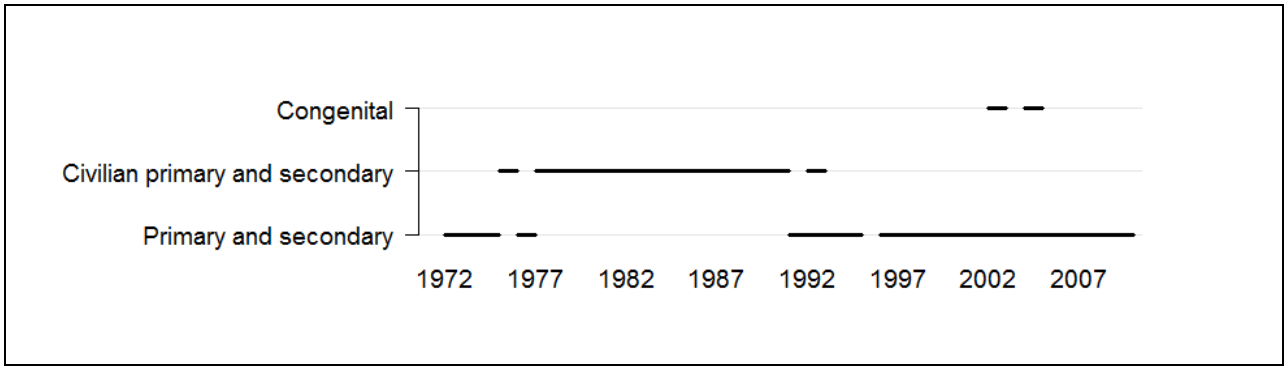

Figure D1, Weeks between 1888 and 2009 for which data on the disease and subcategories (if applicable) are available in the Tycho database. If no subcategory was reported, a subcategory of "Unspecified" was assigned.

Table D1, Summary information for Syphilis (Primary and secondary and Civilian primary and secondary)

Indicator	Hawaii
Report period	1972-2009
Total weeks	1,584
Total cases	1,221
Max. cases per year	140
Year (max)	1980
Max. cases per week	21
Week (max)	1973, wk 28
Average cases per year	33
95%CI	(21-45)
Average cases per week	1
95%CI	(1-1)

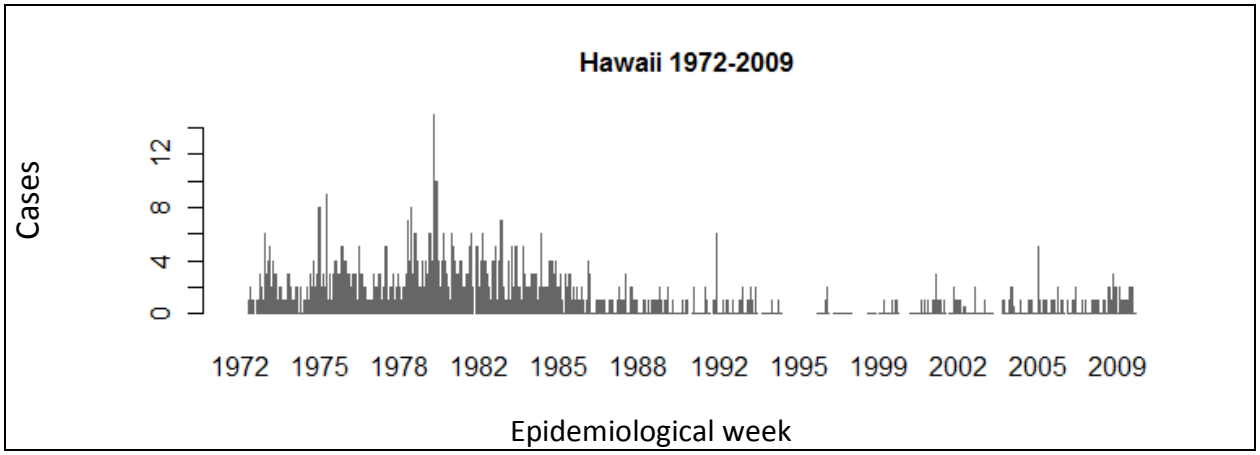

Figure D2, Number of cases reported for Syphilis per epidemiological week

Tuberculosis

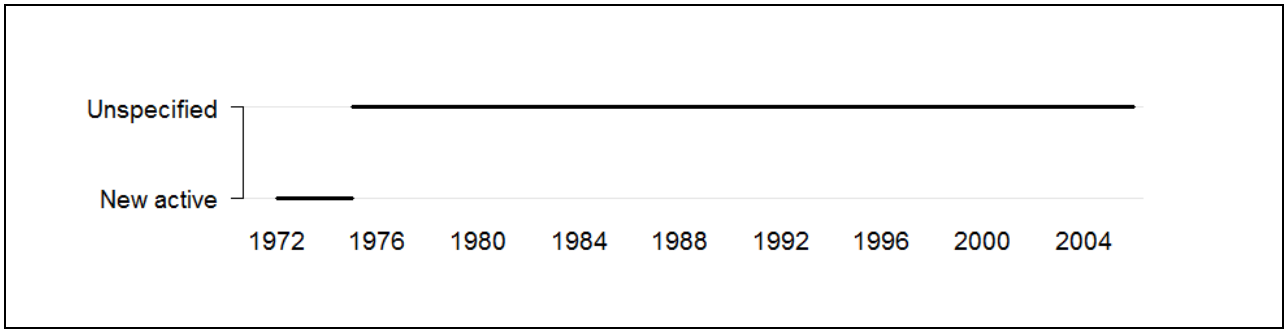

Figure D1, Weeks between 1888 and 2009 for which data on the disease and subcategories (if applicable) are available in the Tycho database. If no subcategory was reported, a subcategory of "Unspecified" was assigned.

Table D1, Summary information for Tuberculosis (Unspecified)

Indicator	Hawaii
Report period	1975-2005
Total weeks	1,485
Total cases	6,994
Max. cases per year	653
Year (max)	1976
Max. cases per week	61
Week (max)	2001, wk 28
Average cases per year	226
95%CI	(173-279)
Average cases per week	5
95%CI	(5-5)

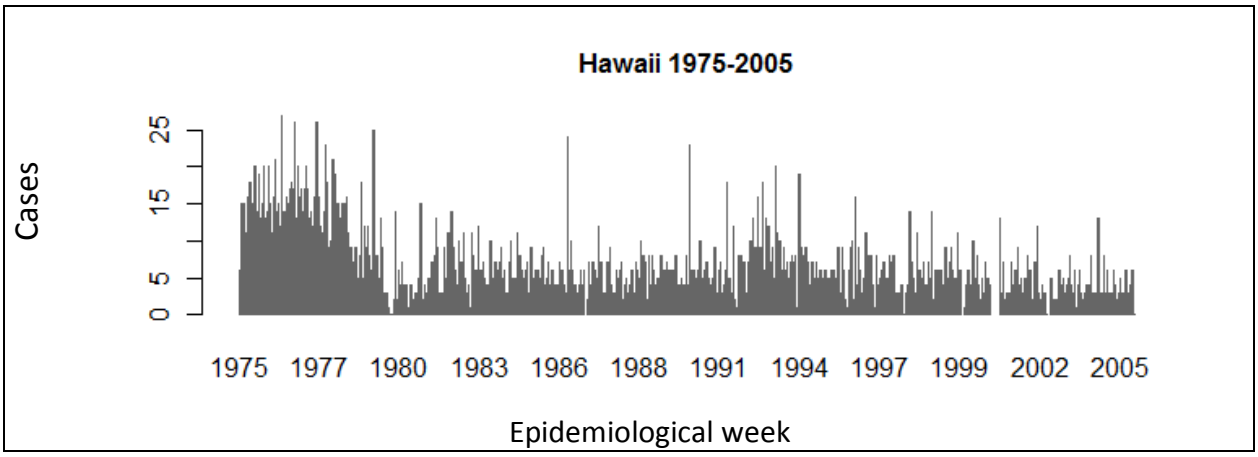

Figure D2, Number of cases reported for Tuberculosis per epidemiological week

Diseases with fragmented data over time

For a number of diseases, only fragmented data was available for Hawaii. This was due to the inclusion of these diseases in the weekly surveillance system for short periods or widely dispersed periods of time. Due to these wide gaps or limited data, data for these diseases were not presented in separate sections but in the table below. Work will continue to include more data in the Tycho database and complete data sets for each disease where possible. Collaborations with local, state and federal public health agencies will be made to explore data availability and opportunities to include these in the Tycho database.

Table 2, Summary information on the occurrence of diseases with fragmented data in Hawaii

Disease	Report type	Report period	Number of reports	Total cases
Brucellosis [undulant fever]	City	1953-1953	1	1
Brucellosis [undulant fever]	State	1954-1975	81	215
Chickenpox [varicella]	State	1972-2009	520	9510
Diphtheria	State	1950-1976	84	64
Encephalitis	State	1951-1993	397	110
Legionellosis	State	1983-2009	588	53
Listeriosis	State	2000-2005	132	18
Lyme disease	State	1992-1996	91	1
Meningitis	State	1950-1964	50	19
Pneumonia	State	1950-1951	18	29
Poliomyelitis	City	1953-1953	12	17
Poliomyelitis	State	1950-1970	409	638
Scarlet fever	City	1953-1953	8	12
Scarlet fever	State	1950-1969	460	23907
Tetanus	State	1962-1975	111	11
Toxic shock syndrome	State	1985-1994	148	5
Trichiniasis	City	1953-1953	1	1
Trichiniasis	State	1954-1955	5	10
Typhoid fever [enteric fever]	City	1914-1915	5	6
Typhoid fever [enteric fever]	State	1951-2005	1382	190
Typhus fever	City	1953-1953	2	2
Typhus fever	State	1954-1961	26	33
Whooping cough [pertussis]	City	1953-1953	14	18
Whooping cough [pertussis]	State	1950-2009	1237	1655

Project Tycho

This report provides preliminary data for the state of Hawaii available in the Tycho database. This database is currently being beta tested and these data cannot be used for publication or other official use at this time. An open access release to the general public is planned for later in 2011.

Please visit the Tycho website for more information and to query the database at: www.tycho.pitt.edu. For further information regarding the Tycho project, contact Dr. Wilbert van Panhuis at the University of Pittsburgh Graduate School of Public Health.

Dr. Wilbert van Panhuis, MD PdD
Graduate School of Public Health
University of Pittsburgh
130 DeSoto Street
704 Parran Hall
Pittsburgh, 15261 PA
Tel: 412-624-7693
Email: wav10@pitt.edu