



Open Access to Newly Digitized US Weekly Notifiable Disease Surveillance Data 1888-Present

ISDS, 23 April 2014

Wilbert van Panhuis, MD PhD

wav10@pitt.edu



University of Pittsburgh





Content:

1. Background
2. Project Tycho: digitization and content
3. Future direction



Summary

1. Access to public health data in usable format

- disaggregated, computable, long-term

2. All US weekly data since 1888

- 56 diseases, 125 years, 50 states, >3000 cities, 200 M keystrokes
- 1030 users, 15,000 new site visits, ~400 data downloads

3. Opportunities for new data use

- Easy access to historical data for analysis and decision making
- Integration with other datasets: new discoveries

Public health decision making

Public health officials make decisions that affect millions of lives



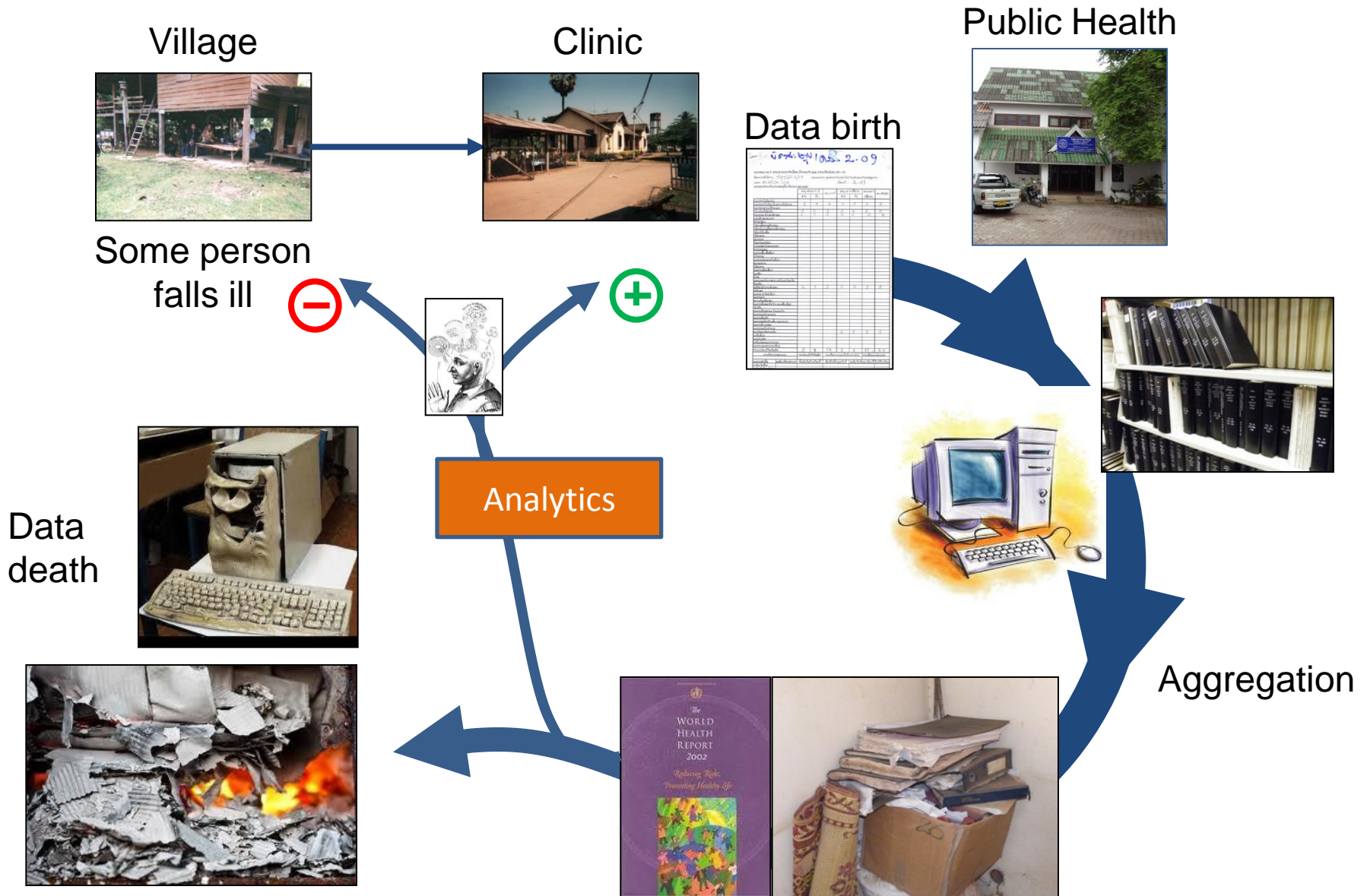
1910-1993

When major health problems arise, someone must make decisions. This is not the primary responsibility of the epidemiologist. Administrative and political as well as technical considerations must also be brought to bear. It is the epidemiologists' function to get the facts to the decision makers.

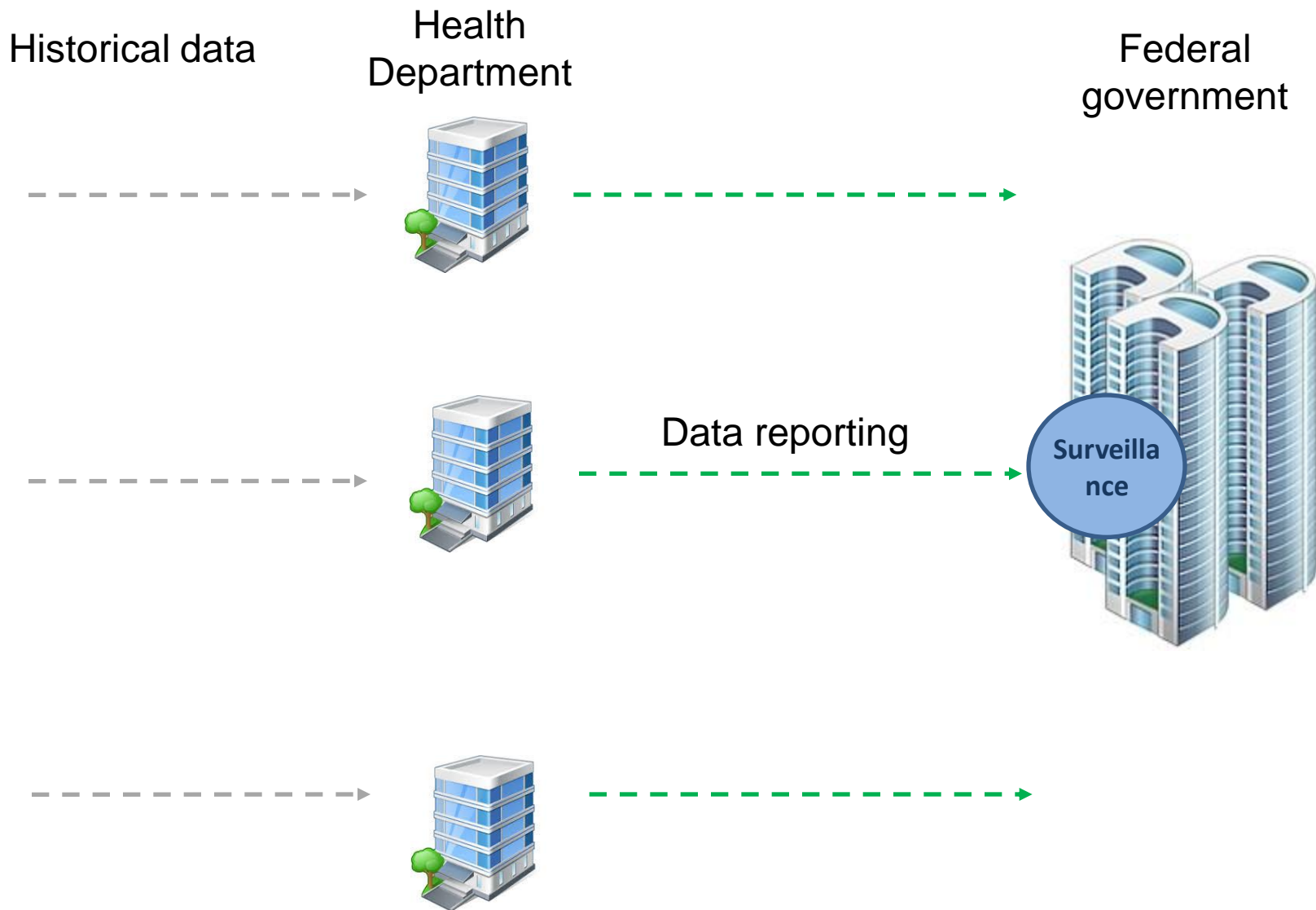
Good surveillance does not necessarily ensure the making of the right decisions, but it reduces the chances of wrong ones.

Alexander Langmuir, NEJM 1963

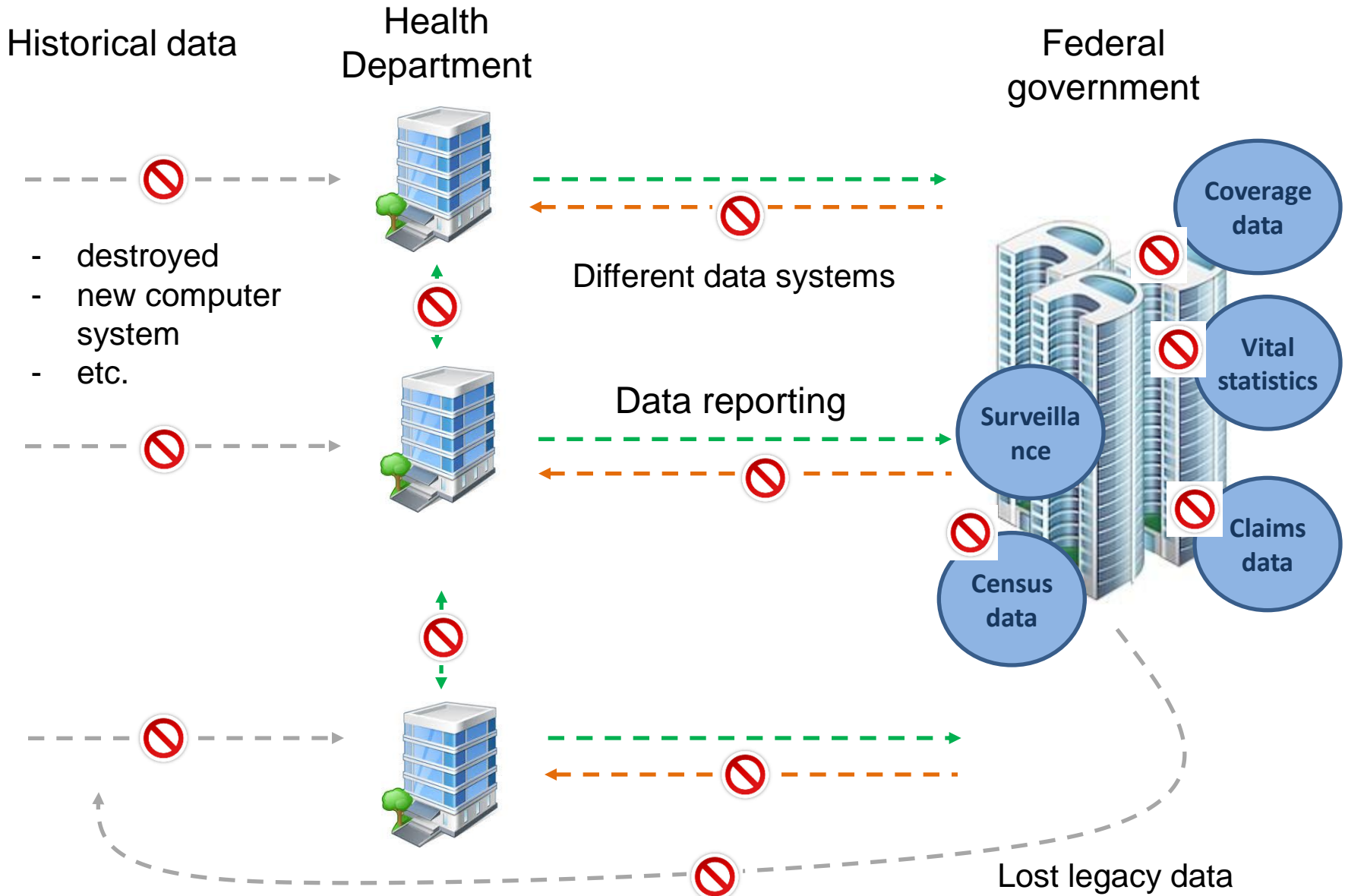
Public health data life cycle



Limited access to essential data

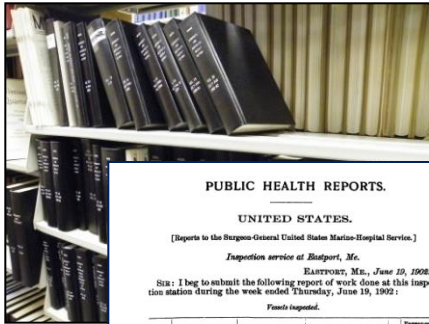


Limited access to essential data



Project Tycho™: US surveillance data

Un-usable



PUBLIC HEALTH REPORTS.
UNITED STATES.
 [Reports to the Surgeon-General United States Marine-Hospital Service.]
Inspection service at Eastport, Me.,
 EASTPORT, ME., June 19, 1902.
 Sir: I beg to submit the following report of work done at this inspection station during the week ended Thursday, June 19, 1902:

Patients hospitalized.

Date.	Name.	From—	To—	Age	Sex	Particulars.
June 13	Mr. W. P.
June 14
June 15
June 16
June 17

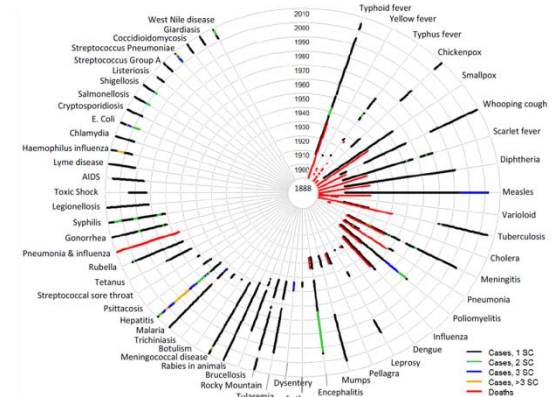
1907
PNEUMONIA (ALL FORMS).
 City Reports for Week Ended June 5, 1908.

City.	Cases.	Deaths.	City.	Cases.	Deaths.
Albany, Ohio	4	1	Lawrence, Mass	2	1
Albany, N. Y.	1	0	Lebanon, N. H.	1	0
Albany, N. Y.	1	0	Lebanon, N. H.	1	0
Albany, N. Y.	1	0	Lebanon, N. H.	1	0
Albany, N. Y.	1	0	Lebanon, N. H.	1	0

See: Through service on Delaware arrival at this station from Boston, with five officers. Just no passengers; 112

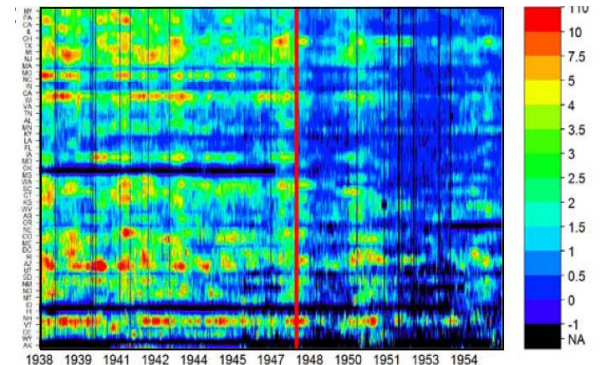
Ready for secondary use

History of US disease surveillance



- Digitization
- Extraction
- Standardization
- Metadata

Pertussis: 1938-1954



CDC MMWR
 Morbidity and Mortality Weekly Report

TABLE II. (Part 6) Provisional cases of selected notifiable diseases, United States, week ending March 9, 2013 (WEEK 10)*

Reporting Area	Legionnaires					Eyes disease					Measles				
	Cases/week	Previous 52 weeks		Cases/week	Previous 52 weeks		Cases/week	Previous 52 weeks		Cases/week	Previous 52 weeks		Cases/week	Previous 52 weeks	
		2013	2012		2013	2012		2013	2012		2013	2012		2013	2012
UNITED STATES	17	62	113	134	407	25	435	1,413	1,014	2,518	6	23	43	141	180
NEW ENGLAND	3	19	18	22	—	114	407	159	779	—	2	7	17	19	—
Conn.	1	7	2	4	—	18	102	3	100	—	0	2	3	1	—
Mass.	0	3	2	1	—	16	69	28	31	—	0	2	—	—	—
Maine	—	1	1	8	10	—	47	163	2	215	—	0	4	13	8
N.H.	0	2	1	3	—	8	53	3	92	—	0	2	—	—	—
R.I.	—	0	3	4	—	1	34	3	6	—	0	7	—	—	—
Vt.	0	2	2	—	—	4	29	10	20	—	0	3	—	—	—
MID ATLANTIC	1	22	44	102	22	197	457	635	1,294	1	6	14	29	26	—
D.C.	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
N.J.	—	2	12	5	19	—	50	166	164	442	—	0	2	4	—
N.Y. (except N.J.)	6	11	14	22	16	41	252	92	140	—	1	4	10	2	—
N.Y. City	1	2	11	17	24	—	1	12	—	—	—	1	11	20	—
Pa.	4	19	28	34	6	96	403	379	611	1	1	4	3	4	—
SE. CENTRAL	4	11	14	22	16	21	160	16	141	—	2	8	11	21	—
Fl.	2	2	3	17	—	2	13	2	8	—	—	—	—	—	—
Ind.	1	2	6	28	13	—	1	6	—	—	0	2	3	5	—
Mich.	—	2	9	—	14	—	2	9	5	8	—	0	3	2	—
Ohio	1	1	11	18	40	—	1	4	8	—	—	1	3	7	—
Tenn.	1	7	—	4	—	17	138	15	116	—	—	0	2	1	—
W. CENTRAL	—	3	10	13	24	—	2	18	2	5	—	1	3	4	7
Ark.	—	0	2	—	—	1	10	1	3	—	—	0	2	1	1
Kans.	—	0	2	1	1	—	0	2	—	—	—	0	1	2	3

1888-2013

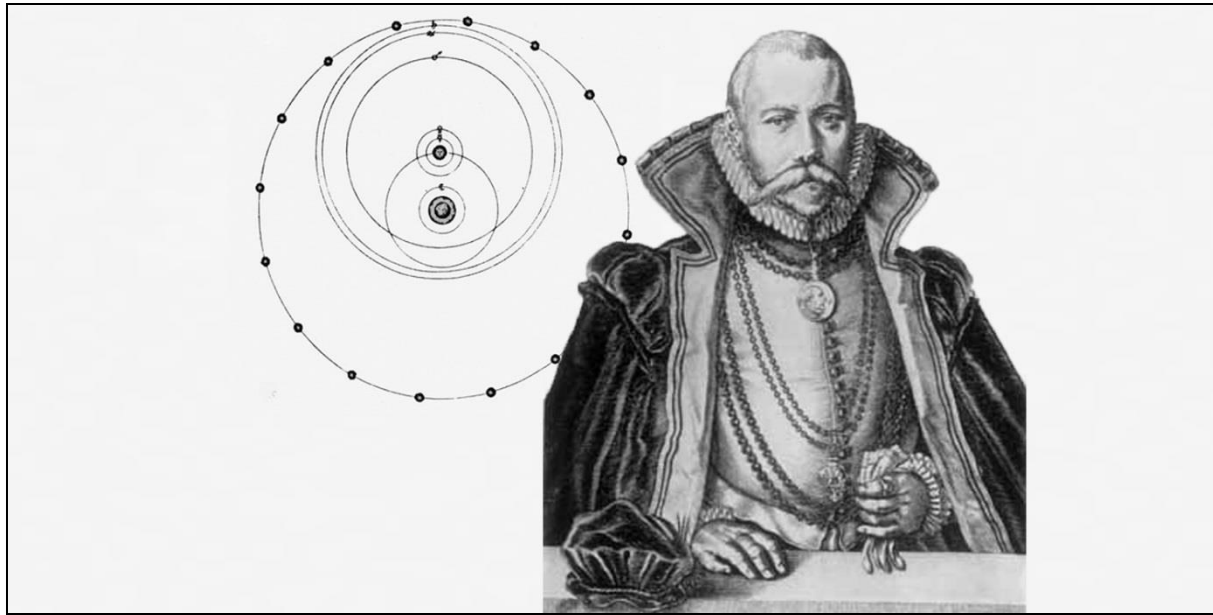
125 years

6500 weeks

35,000 tables

200 M keystrokes

Tycho Brahe



1546 – 1601

Danish nobleman who made accurate and comprehensive observations of the positions of the stars and planets. After his death, Tycho's assistant Johannes Kepler used these data to derive the laws of planetary motion.

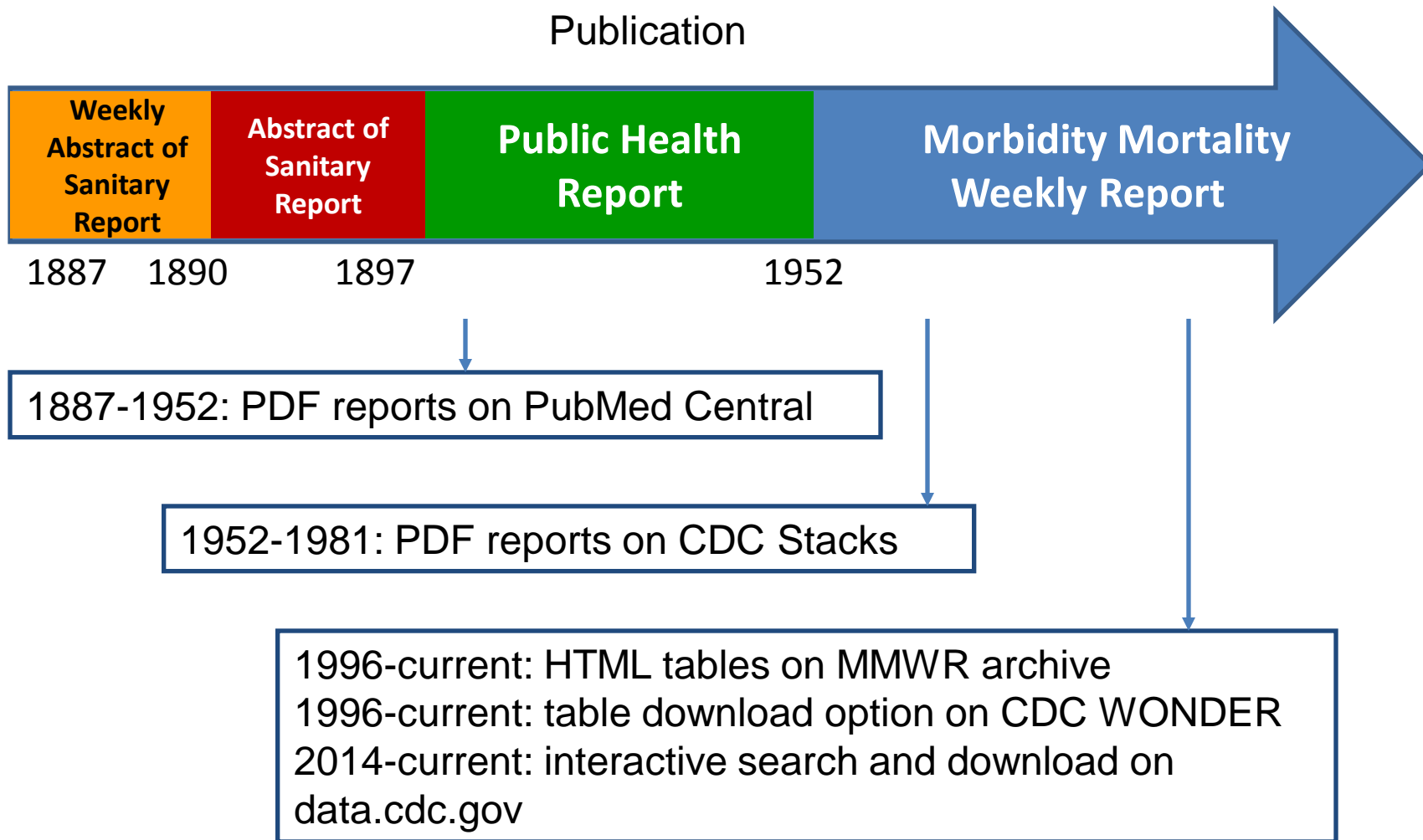


Digitization



DDD Operators in Phnom Penh, Cambodia

Was digitization/integration necessary?



Mortality reports for cities: 1887-1947

MORTALITY TABLE, CITIES OF THE UNITED STATES.

Cities.	Week ended.	Estimated population.	Total deaths from all causes.	Deaths from—						
				Cholera.	Yellow fever.	Small-pox.	Typhus fever.	Enteric fever.	Scarlet fever.	Diphtheria.
Philadelphia, Pa.....	June 16.....	1,016,758	353	3	14	4	4
Philadelphia, Pa.....	June 30.....	1,016,758	536	2	12	5	7
Baltimore, Md.....	June 30.....	431,879	248	4	1
New Orleans, La.....	June 23.....	216,090	120	1	4
Richmond, Va.....	June 30.....	63,600	61	1
Charleston, S. C.....	June 30.....	60,145	34

Weekly Abstract of Sanitary Reports: 1888 July

State morbidity reports: 1928-current

TABLE II. Cases of selected notifiable diseases, United States, weeks ending January 7, 1995, and January 8, 1994 (1st Week)

Reporting Area	AIDS*	Gonorrhea		Hepatitis (Viral), by type						Legionellosis	
				A		B		NA,NB			
				Cum. 1995	Cum. 1994	Cum. 1995	Cum. 1994	Cum. 1995	Cum. 1994		
UNITED STATES	-	2,732	6,882	144	407	33	172	7	71	5	27
NEW ENGLAND	-	76	117	2	6	1	6	-	3	-	1
Maine	-	1	-	-	-	-	-	-	-	-	-
N.H.	-	1	-	-	-	-	-	-	-	-	-
Vt.	-	-	-	-	-	-	-	-	-	-	-
Mass.	-	70	64	-	2	1	5	-	1	-	-
R.I.	-	4	10	2	4	-	-	-	2	-	1
Conn.	-	-	43	-	-	-	1	-	-	-	-
MID. ATLANTIC	-	250	637	1	35	-	33	-	5	-	-
Upstate N.Y.	-	-	-	-	2	-	1	-	1	-	-
N.Y. City	-	-	475	-	27	-	13	-	-	-	-
N.J.	-	-	18	-	1	-	10	-	2	-	-
Pa.	-	250	144	1	5	-	9	-	2	-	-
E.N. CENTRAL	-	83	1,448	40	44	5	41	-	11	2	8
Ohio	-	-	46	32	4	-	4	-	-	2	1
Ind.	-	60	133	2	8	2	4	-	-	-	1
Ill.	-	-	718	-	21	-	13	-	3	-	3
Mich.	-	-	397	6	5	3	12	-	8	-	2
Wis.	-	23	154	-	6	-	8	-	-	-	1
W.N. CENTRAL	-	54	161	1	21	1	11	1	-	1	-
Minn.	-	54	-	-	1	-	-	-	-	-	-
Iowa	-	-	24	-	1	1	1	1	-	1	-
Mo.	-	-	137	-	13	-	10	-	-	-	-
N. Dak.	-	-	-	-	-	-	-	-	-	-	-
S. Dak.	-	-	-	-	-	-	-	-	-	-	-
Nebr.	-	-	-	-	6	-	-	-	-	-	-
Kans.	-	-	-	1	-	-	-	-	-	-	-
S. ATLANTIC	-	1,637	1,843	3	12	12	10	2	5	1	-
Del.	-	18	26	-	-	1	1	-	-	-	-
Md.	-	200	225	1	5	1	3	-	-	-	-
D.C.	-	-	170	-	3	1	1	-	-	-	-
Va.	-	174	396	-	-	-	-	-	-	-	-
W. Va.	-	38	9	1	-	2	-	-	-	-	-
N.C.	-	266	567	1	1	6	4	2	2	1	-
S.C.	-	141	258	-	3	-	-	-	-	-	-
Ga.	-	487	-	-	-	-	1	-	2	-	-
Fla.	-	313	192	-	-	1	-	-	1	-	-



Data content



Tycho analysts, University of Pittsburgh

Example source report: terminology

Vol. 44 / No. 1

MMWR

9

TABLE II. Cases of selected notifiable diseases, United States, weeks ending January 7, 1995, and January 8, 1994 (1st Week)

Title

Reporting Area	AIDS*	Gonorrhea		Hepatitis (Viral), by type						Legionellosis	
				A		B		NA, NB			
				Cum. 1995	Cum. 1994	Cum. 1995	Cum. 1994	Cum. 1995	Cum. 1994		
UNITED STATES		2,732	6,882	144	407	33	172	7	71	5	27
NEW ENGLAND		76	117	2	6	1	6	-	3	-	1
Maine		1	-	-	-	-	-	-	-	-	-
N.H.		1	-	-	-	-	-	-	-	-	-
Vt.		-	-	-	-	-	-	-	-	-	-
Mass.		70	64	-	2	1	5	-	1	-	-
R.I.		4	10	2	4	-	-	-	2	-	1
Conn.		-	43	-	-	-	1	-	-	-	-
MID. ATLANTIC		250	637	1	35	-	33	-	5	-	-
Upstate N.Y.		-	-	-	2	-	1	-	1	-	-
N.Y. City		-	475	-	27	-	13	-	-	-	-
N.J.		-	18	-	1	-	10	-	2	-	-
Pa.		250	144	1	5	-	9	-	2	-	-
E.N. CENTRAL		83	1,448	40	44	5	41	-	11	2	8
Ohio		-	46	32	4	-	4	-	-	2	1
Ind.		60	133	2	8	2	4	-	-	-	1
Ill.		-	718	-	21	-	13	-	3	-	3
Mich.		-	397	6	5	3	12	-	8	-	2
Wis.		23	154	-	6	-	8	-	-	-	1
W.N. CENTRAL		54	161	1	21	1	11	1	-	1	-
Minn.		54	-	-	1	-	-	-	-	-	-
Iowa		-	24	-	1	1	1	1	-	1	-
Mo.		-	137	-	13	-	10	-	-	-	-
N. Dak.		-	-	-	-	-	-	-	-	-	-
S. Dak.		-	-	-	-	-	-	-	-	-	-
Nebr.		-	-	-	6	-	-	-	-	-	-
Kans.		-	-	1	-	-	-	-	-	-	-
S. ATLANTIC		1,637	1,843	3	12	12	10	2	5	1	-
Del.		18	26	-	-	1	1	-	-	-	-
Md.		200	225	1	5	1	3	-	-	-	-
D.C.		-	170	-	3	1	1	-	-	-	-
Va.		174	396	-	-	-	-	-	-	-	-
W. Va.		38	9	1	-	2	-	-	-	-	-
N.C.		266	567	1	1	6	4	2	2	1	-
S.C.		141	258	-	3	-	-	-	-	-	-
Ga.		487	-	-	-	-	1	-	2	-	-
Fla.		313	192	-	-	1	-	-	1	-	-

Headers

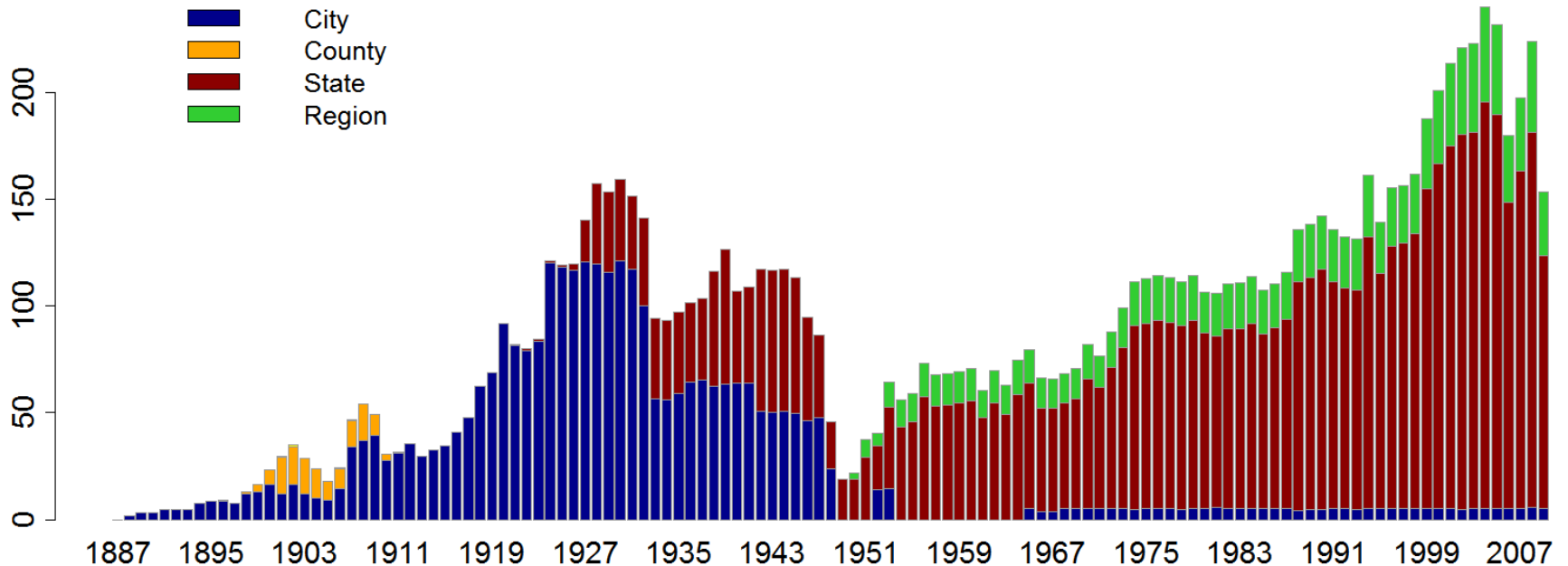
"Count"

Places

MMWR: 1995 January

Level of geography reporting

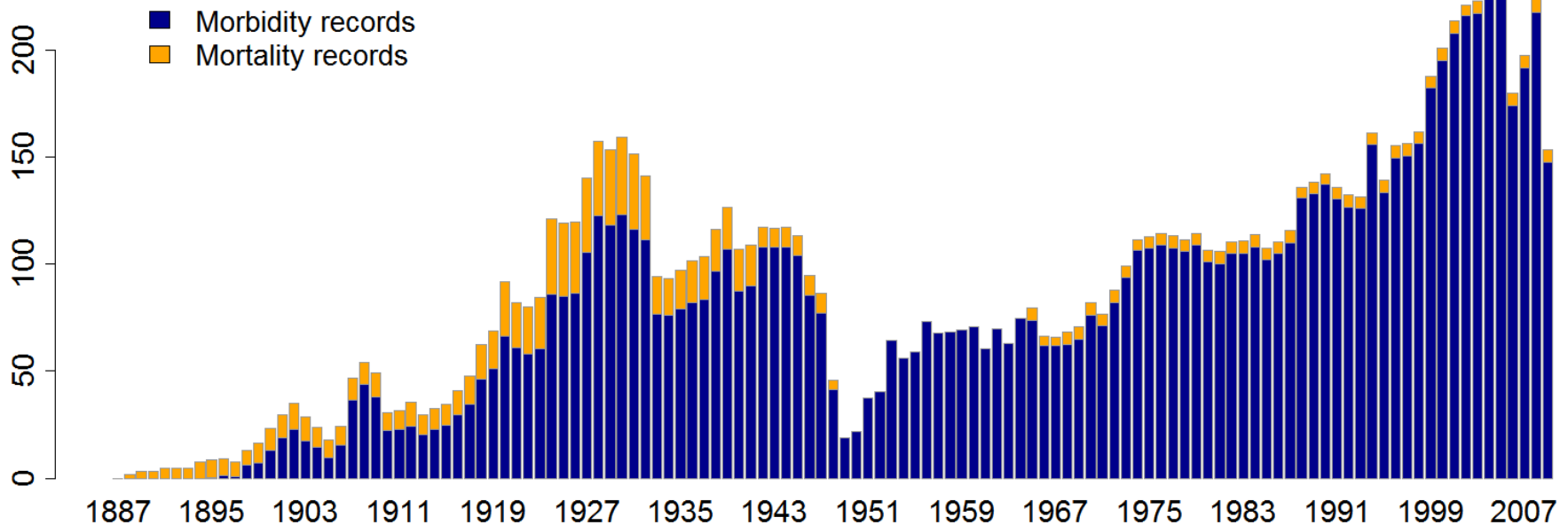
Number of counts per reporting location



- State reports since 1926
- Cities discontinued after 1953, except for pneumonia and influenza mortality
- County reports only for Smallpox

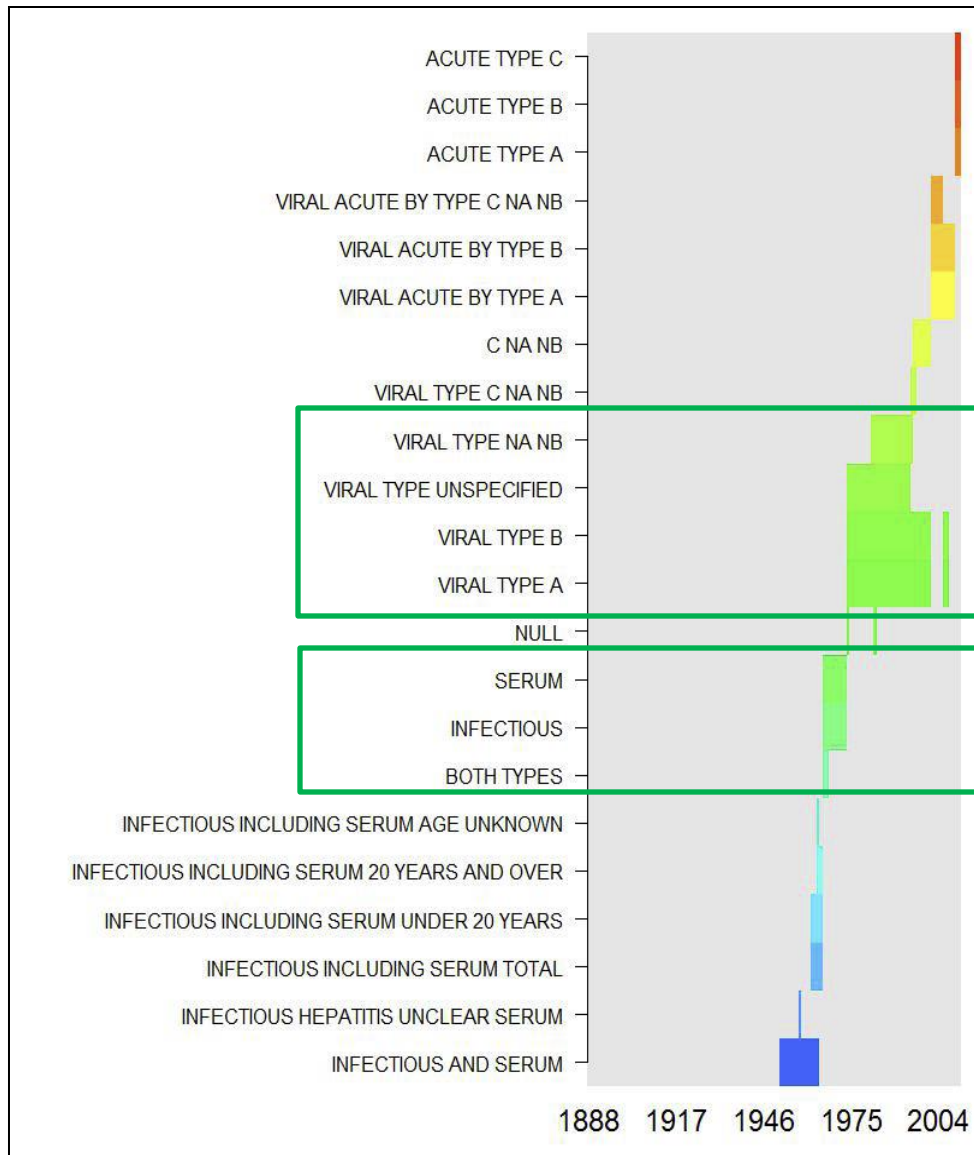
Morbidity vs. mortality reports

Number of counts (stacked) for morbidity and mortality reports



- Only mortality reports before 1895
- Mortality reports discontinued after 1948 except for pneumonia and influenza in US cities

Exploring disease subcategories: Hepatitis



MMWR 1987

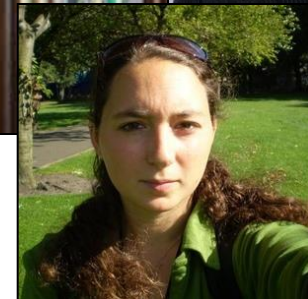
Hepatitis (Viral), by type			
A	B	NA,NB	Unspecified
1987	1987	1987	1987

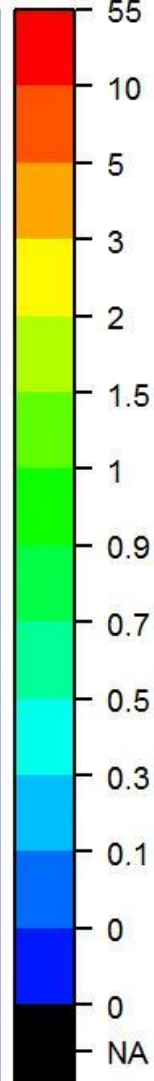
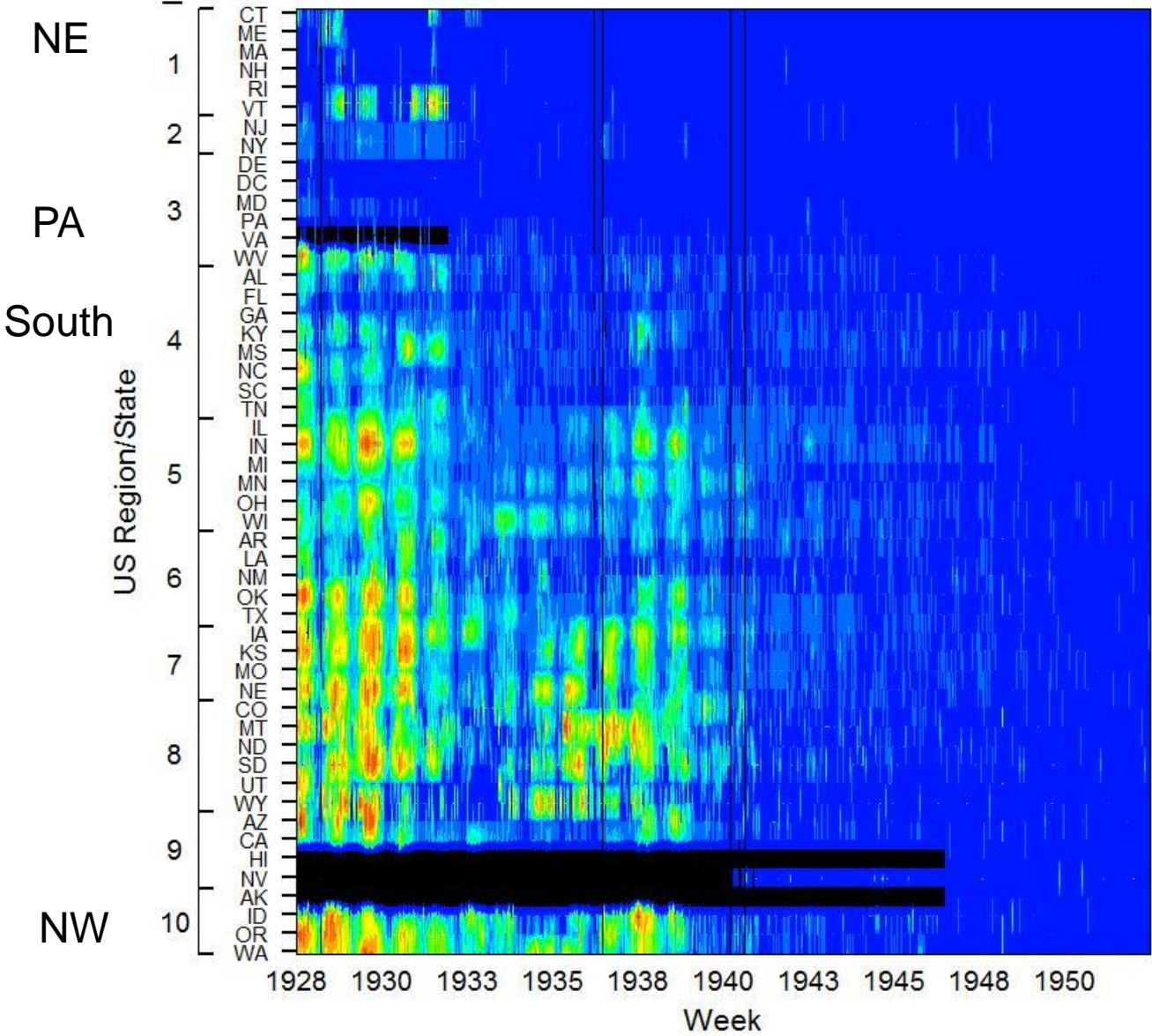
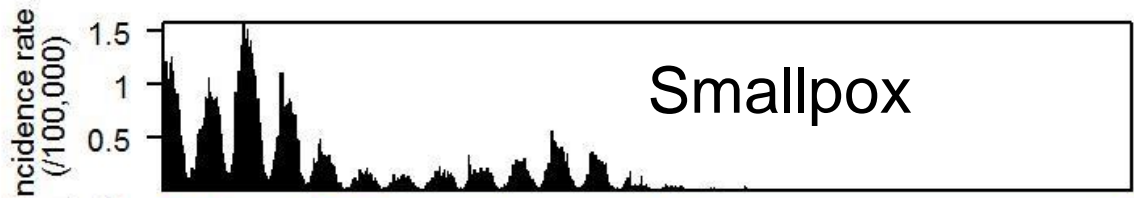
HEPATITIS		
Serum	Infectious	Both Types
1966	1966	1966

MMWR 1966

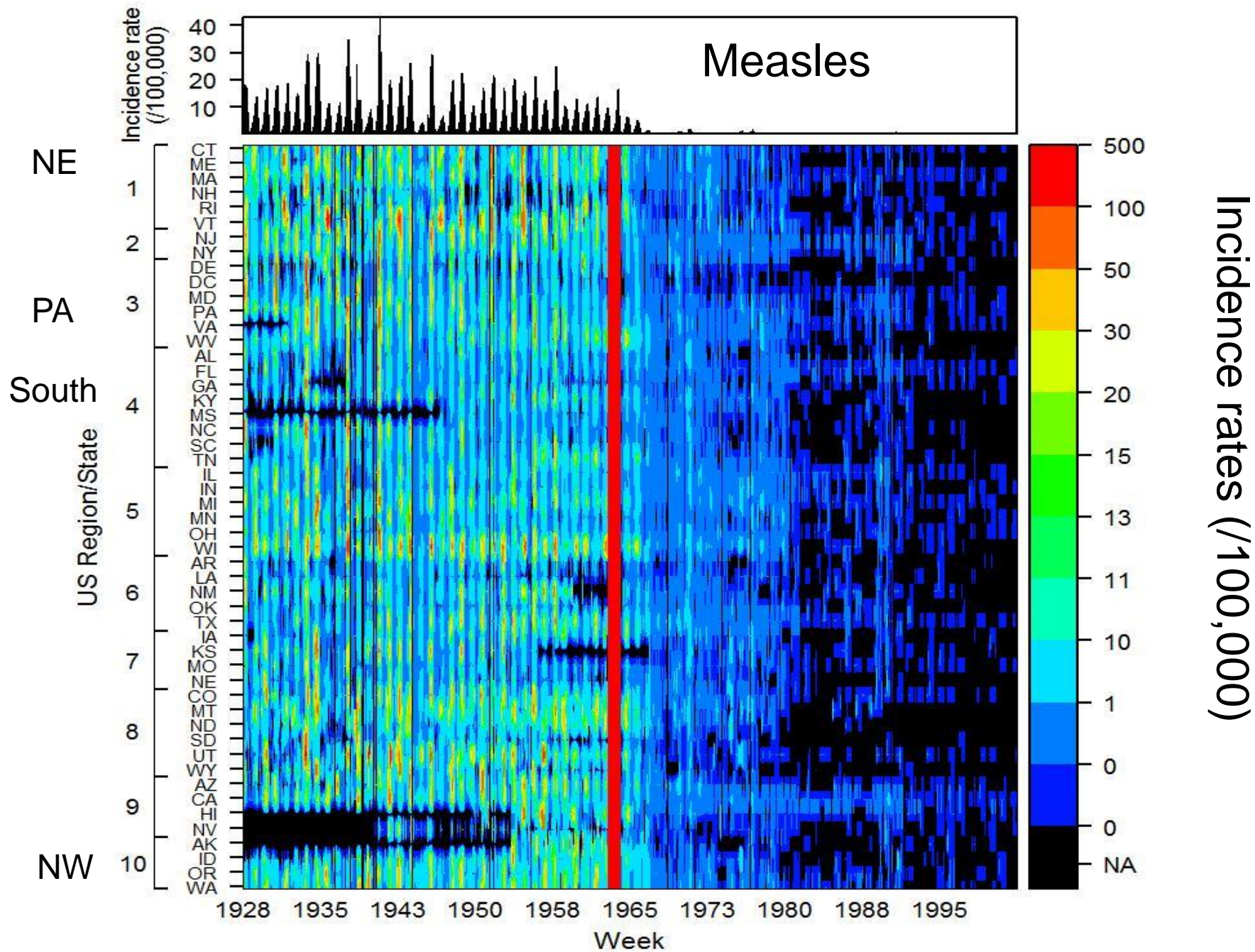


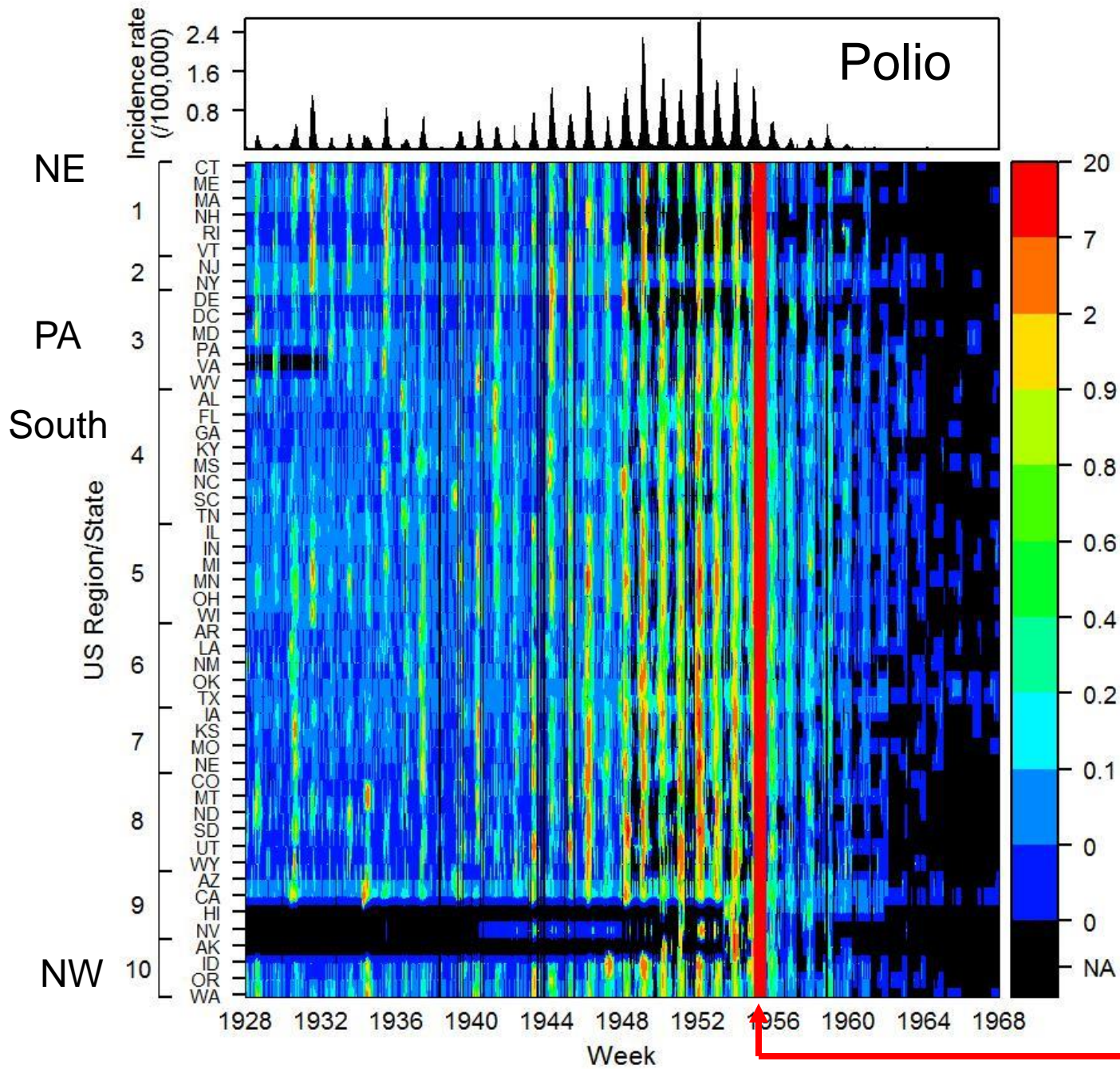
Application





Incidence rates (/100,000)



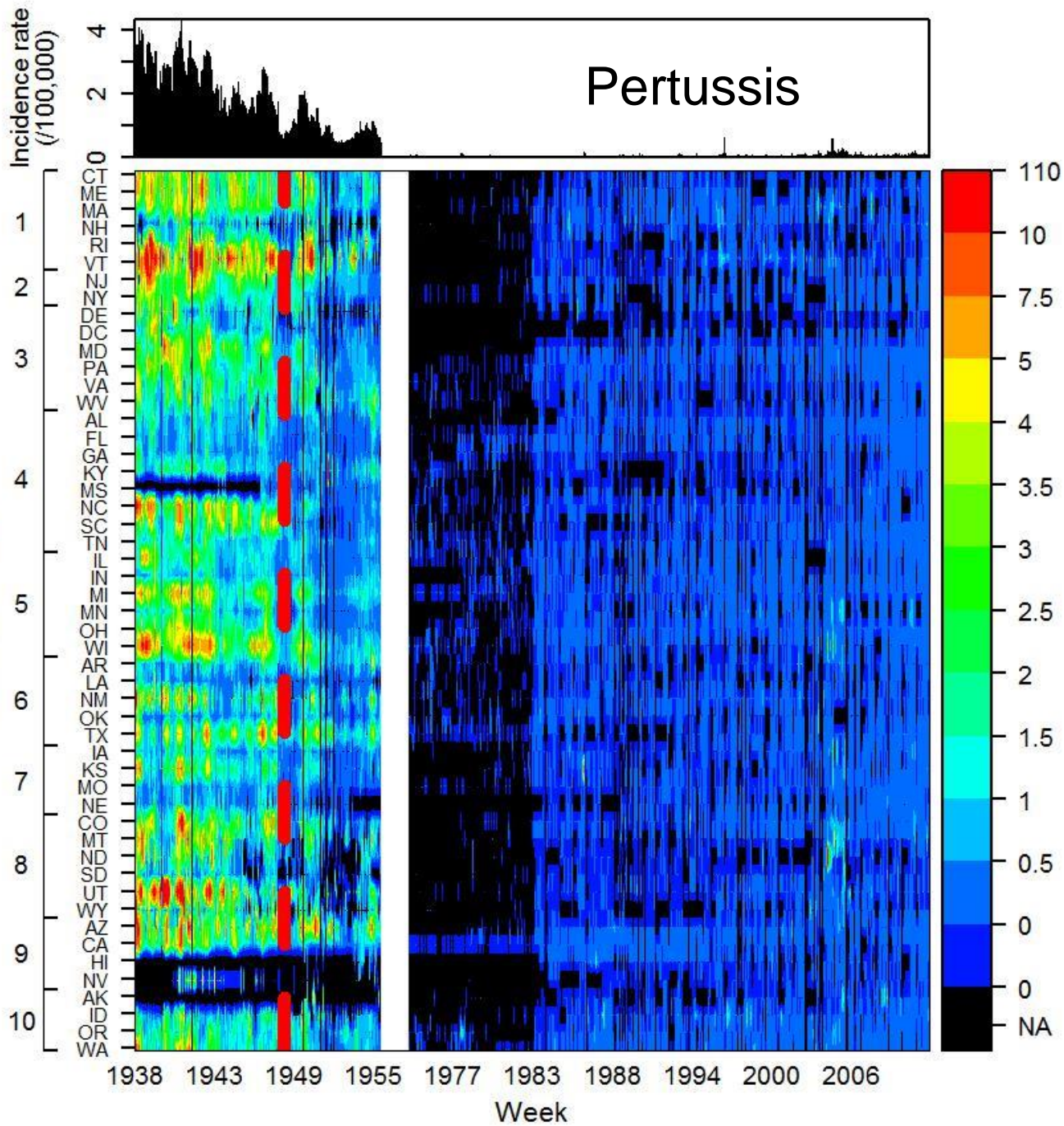


Incidence rates (/100,000)

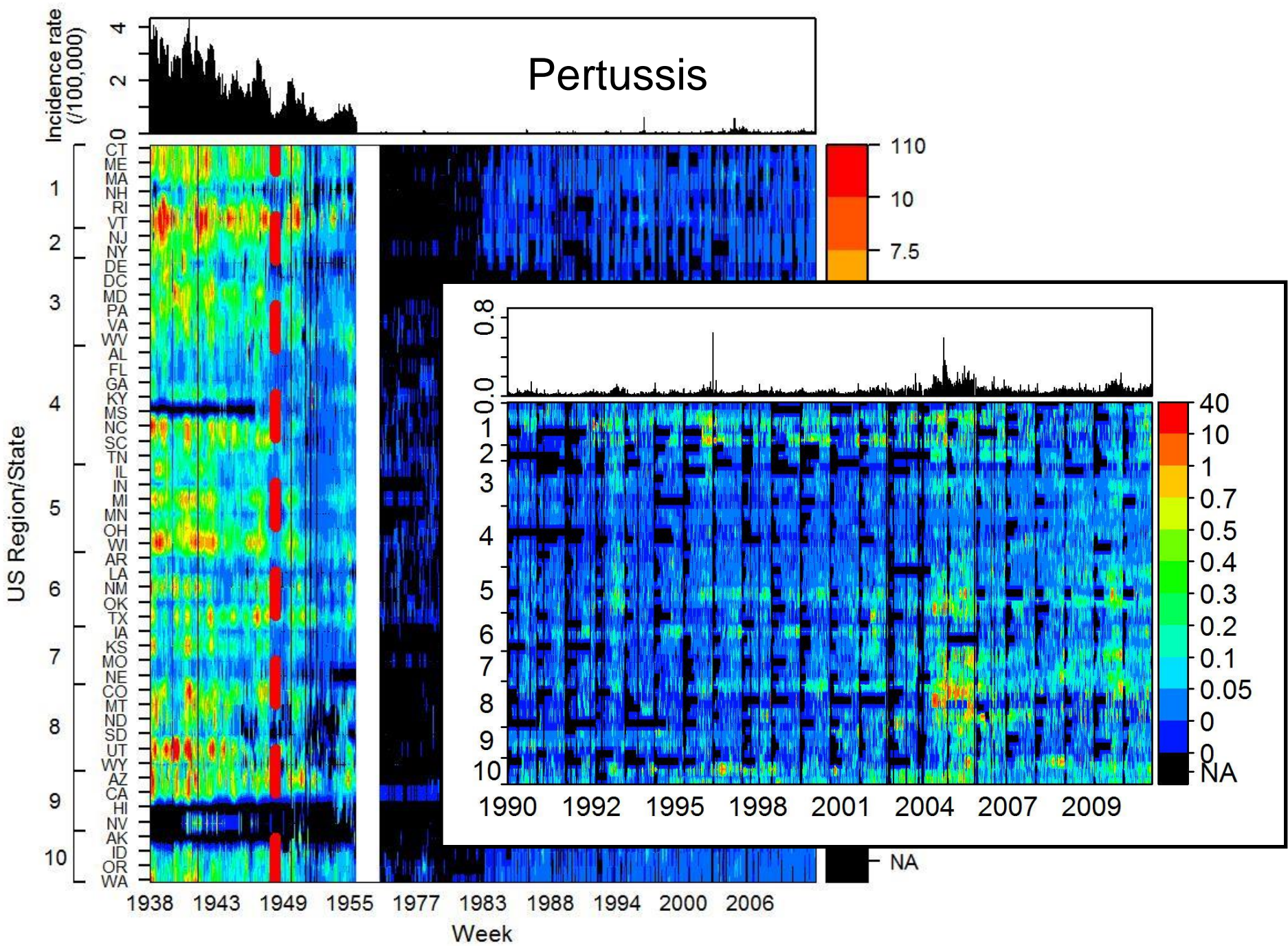


Pitt Salk Vaccine

Pertussis



Pertussis



Project Tycho release paper

The NEW ENGLAND JOURNAL *of* MEDICINE

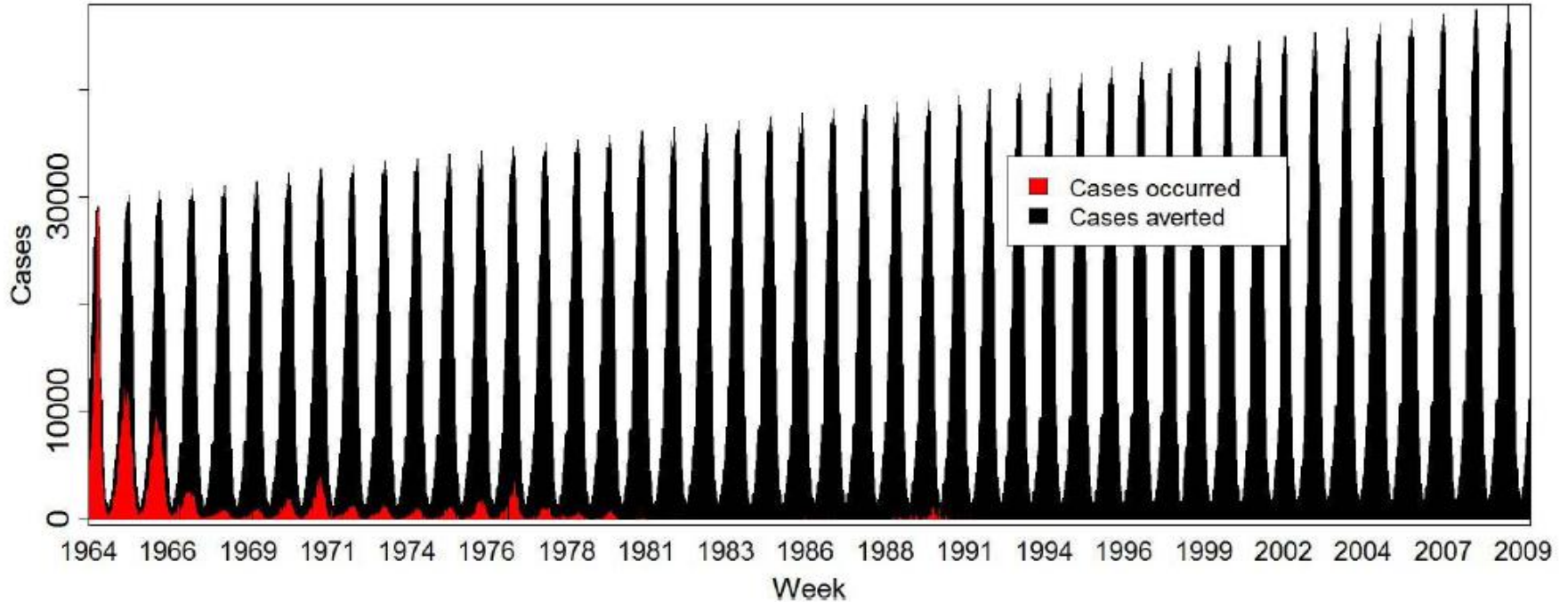
MEDICINE AND SOCIETY

Contagious Diseases in the United States from 1888 to the Present

Willem G. van Panhuis, M.D., Ph.D., John Grefenstette, Ph.D., Su Yon Jung, Ph.D.,
Nian Shong Chok, M.Sc., Anne Cross, M.L.I.S., Heather Eng, B.A., Bruce Y. Lee, M.D.,
Vladimir Zadorozhny, Ph.D., Shawn Brown, Ph.D., Derek Cummings, Ph.D., M.P.H.,
and Donald S. Burke, M.D.

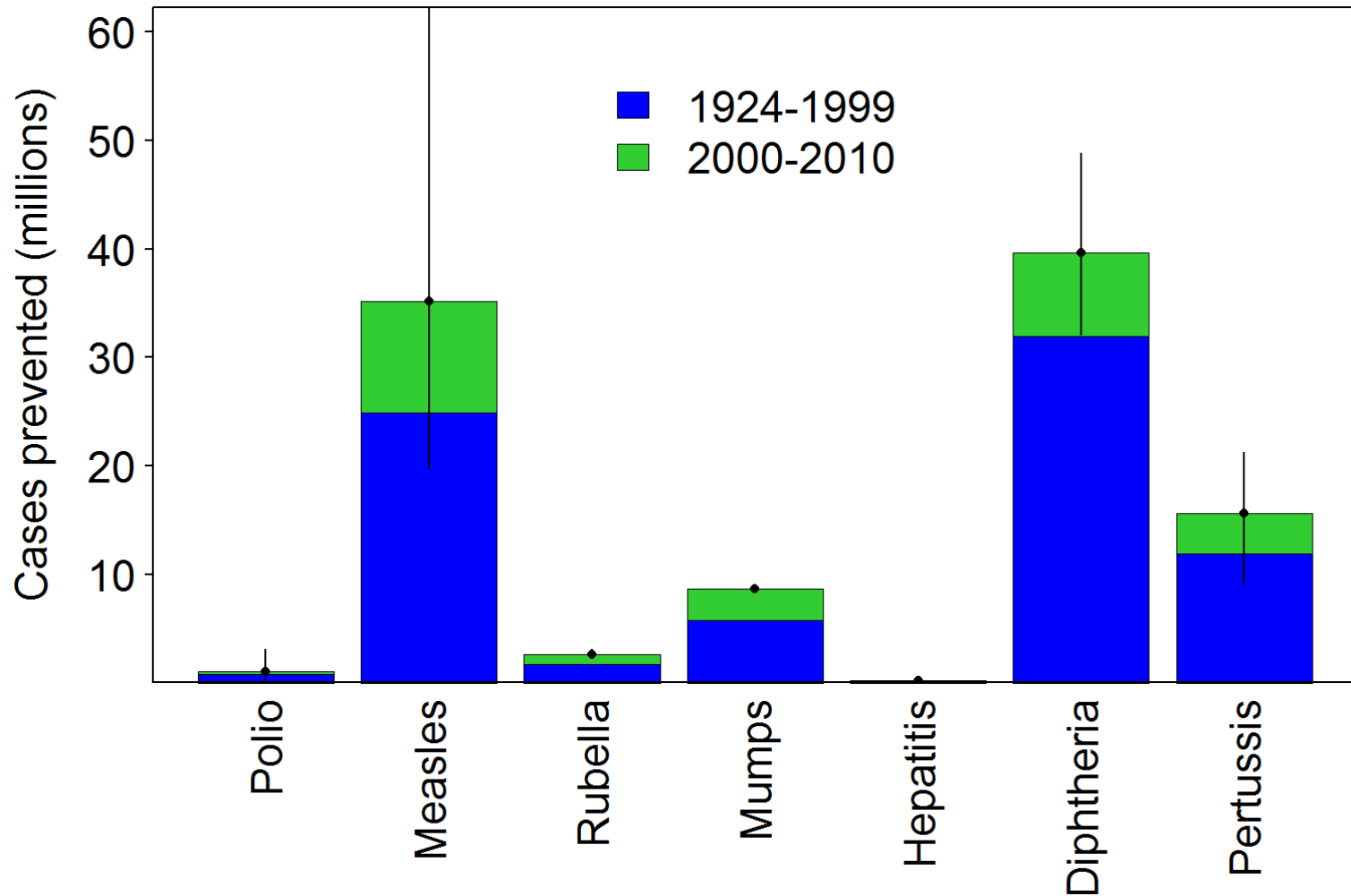
Cases of measles prevented by vaccination

Counterfactual IR x annual population -> Counterfactual cases



~ 36 million cases prevented since 1964

Disease cases prevented by vaccination



Total entire period: 103 million (P10-90: 72-148)

Total 2000-2010: 26 million (P10-90: 18-37)

Data redistribution



University of Pittsburgh

Home About Data Resources FAQ Contact Us



PROJECT TYCHO
DATA FOR HEALTH



Project Tycho™ makes data available for use by others

News

[Project Tycho™ provides open access to newly digitized US weekly surveillance data for contagious diseases from 1888 to the present for use in research, education, and policy making](#)

[Project Tycho™ data have been used to estimate that 100 million cases of contagious diseases have been prevented by vaccination programs in the United States](#)

<p>Level 1 data</p> <p>These data have been custom tailored for specific projects. These are the most complete data but for a limited number of diseases, locations, and years.</p> <p>Current version: 1.0.0</p> <p>Indicators: cases, incidence rates Diseases: smallpox, polio, measles, mumps, rubella, hepatitis A, diphtheria, and whooping cough Locations: all 50 states and 122 cities (diphtheria only) Years: varying between 1916 and 2009</p> <p>Go to level 1 data</p>	<p>Level 2 data</p> <p>These data have been cleaned for immediate use, including a large number of diseases, locations, and years. Data for some diseases, locations, and years have not been cleaned yet, resulting in gaps.</p> <p>Current version: 1.0.0</p> <p>Indicators: cases and deaths Diseases: 47 contagious diseases Locations: all 50 states, 6 territories, and 1287 cities Years: varying between 1888 to 2013</p> <p>Go to level 2 data</p>	<p>Level 3 data</p> <p>These data have not been cleaned and cannot yet be used for analysis. We are continuing data cleaning and would welcome help and support for this. Raw data are provided upon request.</p> <p>Current version: 1.0.0</p> <p>Indicators: cases, deaths, and summary statistics Diseases: 56 contagious diseases and 72 subcategories Locations: 10 regions, 50 states, 6 territories, 3165 counties, and 2906 cities, towns, and parishes Years: varying between 1887 to 2013</p> <p>Level 3 data request form</p>
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

[Log in or Register](#)
[License Agreement](#)
[Privacy Policy](#)

[News](#)
[About](#)
[Resources](#)

[Data Level 1](#)
[Data Level 2](#)
[Data Level 3](#)

Project Tycho™
University of Pittsburgh
Graduate School of Public Health
130 De Soto St.
Pittsburgh, PA 15261

Project Tycho™ is funded by the [Bill & Melinda Gates Foundation](#) and the [National Institutes of Health](#)

© 2013 [Public Health Dynamics Laboratory](#), University of Pittsburgh

www.tycho.pitt.edu (open access)

Demo

www.tycho.pitt.edu

Big Data: From “Lazy” to “Active”



“A University of Pittsburgh initiative called Project Tycho, for example, unlocks CDC data on contagious diseases which goes back all the way to 1888. Among other things, they've identified more than 100 million cases of contagious illness that were prevented by immunizations.” (Secretary Sebelius)

Reimagining Health Care Delivery: Remarks to the Aspen Institute Care Innovation Summit (Washington DC, Feb 27 2014)

Future Direction

- Data validation
 - internal consistency, final vs. weekly, additional data
 - case definitions, disease classification
- Expansion of Project Tycho
 - county level US data
 - other US and foreign disease surveillance systems
 - related data: census, demographic, climate, etc.
- Tools for specific user groups
 - national, state, and local health departments
 - academia
 - industry
 - general public

In collaboration with CDC, CSTE and health departments

Acknowledgements

Tycho development team

Don Burke, Wilbert van Panhuis, John Grefenstette, Shawn Brown, Ernesto Marques, Bruce Lee, Derek Cummings, Vladimir Zadorozhny, Steve Wisniewski, Dan Bain, Patrick Manning, Su Yon Jung, Nian Shong Chok, Heather Eng, Anne Cross, David Galloway, Suzanne Cake, Raaka Kumbhakar, Xi Huang, Yongxu Zang, Erin Jenkins



- University of Pittsburgh Graduate School of Public Health, Geology department, History department, School of Information Sciences
- Johns Hopkins Bloomberg School of Public Health
- Bill & Melinda Gates Foundation, NIH MIDAS
- CDC OSELS Office, CSTE and State Health Offices
- Open Government Initiative, Health.Data.Gov, NCBI
- MOH of Thailand, Laos, Cambodia, Vietnam, Brazil, Taiwan
- World Health Organization Geneva, PAHO