So, how many did die due to Maria?

Dr. Wolfgang Rolke

The Problem

- Official Government Estimate until mid 2018: 64 deaths
- Official Government Estimate since then: 2975 deaths
- Where do these come from?
- How good are they?

Why is this so hard?

- A person gets killed by a falling tree during storm - easy
- A person has a heart attack during storm - harder
- A person has a heart attack a month after the storm - ???

Six Peer-Reviewed Publications:

Estimating the death toll of Hurricane Maria

Authors: Roberto Rivera and Wolfgang Rolke

Date: February 2018

Journal: Significance

Data: Deaths September 1-19, 2017: 1582
Deaths September 20 - October 31: 4319

Source: Demographic Registry Office

Method of Analysis: Poisson counts

Point estimate: 822

95% confidence interval (605, 1039)

Mortality in Puerto Rico after Hurricane Maria

Authors: Kishore, Nishant, Domingo Marqués, Ayesha Mahmud, Mathew V. Kiang, Irmary Rodriguez, Arlan Fuller, Peggy Ebner, et al

Date: June 2018

Journal: New England Journal of Medicine

Data: results of a survey of 3299 randomly chosen household across PR, using some stratification based on geographic location. Then estimated total deaths are compared to pre-hurricane mortality rate.

Method of Analysis: Poisson counts

In the surveyd households 38 people had died, equivalent to a mortality rate of 14.3 deaths per 1000 persons for September 20 to December 31.

Point Estimate: 4645

95% confidence interval: (793, 8498)

Differential and persistent risk of excess mortality from Hurricane Maria in Puerto Rico: a time-series analysis

Authors: Carlos Santos-Burgoa, John Sandberg, Erick Suárez, Ann Goldman-Hawes, Scott Zeger, Alejandra Garcia-Meza, Cynthia M Pérez, Noel Estrada-Merly, Uriyoan Colón-Ramos, Cruz María Nazario, Elizabeth Andrade, Amira Roess, Lynn Goldman

Date: October 2018

Journal: Lancet

Data: mortality data, including deaths by age, sex, and residential municipality, for July 1, 2010, to Feb 28, 2018, from the Puerto Rico Vital Statistics System.

Method of Analysis: Time series

Two estimates, one using US Census data to estimate the population, the other accounting for the large-scale population displacement.

Point Estimate(s): 1191, 2975

95% confidence intervals: (836, 1544), (2658, 3290)

Note this last one is now the official death toll estimate of the government of Puerto Rico.

Use of Death Counts From Vital Statistics to Calculate Excess Deaths in Puerto Rico Following Hurricane Maria

Authors: Alexis R. Santos-Lozada and Jeffrey T. Howard

Date: October 2018

Journal: Journal of the American Medical Association

Data: Monthly death counts, from January 2010 through December 2017, includingp reviously unavailable death counts for January through December 2017, from the Puerto Rico vital statistics system.

Method of Analysis: Normal theory

Point Estimate: 4645

95% confidence interval: (1139, 1272)

Causes of Excess Deaths in Puerto Rico After Hurricane Maria: A Time-Series Estimation

Authors: Raul Cruz-Cano and Erin L. Mead

Date: January 2019

Journal: American Journal of Public Health

Data: monthly vital statistics data on all deaths from January 2008 through October 2017.Weconducted a time-series analysis to estimate excess mortality in September and October 2017 overall and by age, sex, and cause of death.

Method of Analysis: time series, auto-regressive integrated moving-average (ARIMA)

Point Estimate: 1205

95% confidence interval: (1069, 1568)

Modeling Excess Deaths After a Natural Disaster with Application to Hurricane Maria

Authors: Roberto Rivera and Wolfgang Rolke

Date: October 2019

Journal: Statistics in Medicine

Data: monthly vital statistics data on all deaths from January 2015 through February 2018.

Method of Analysis: Poisson counts, log-linear model

Point Estimate(s): 1453, 1293

95% confidence intervals

Poisson model: (1116, 1791) log-linear model: (1086, 1495)

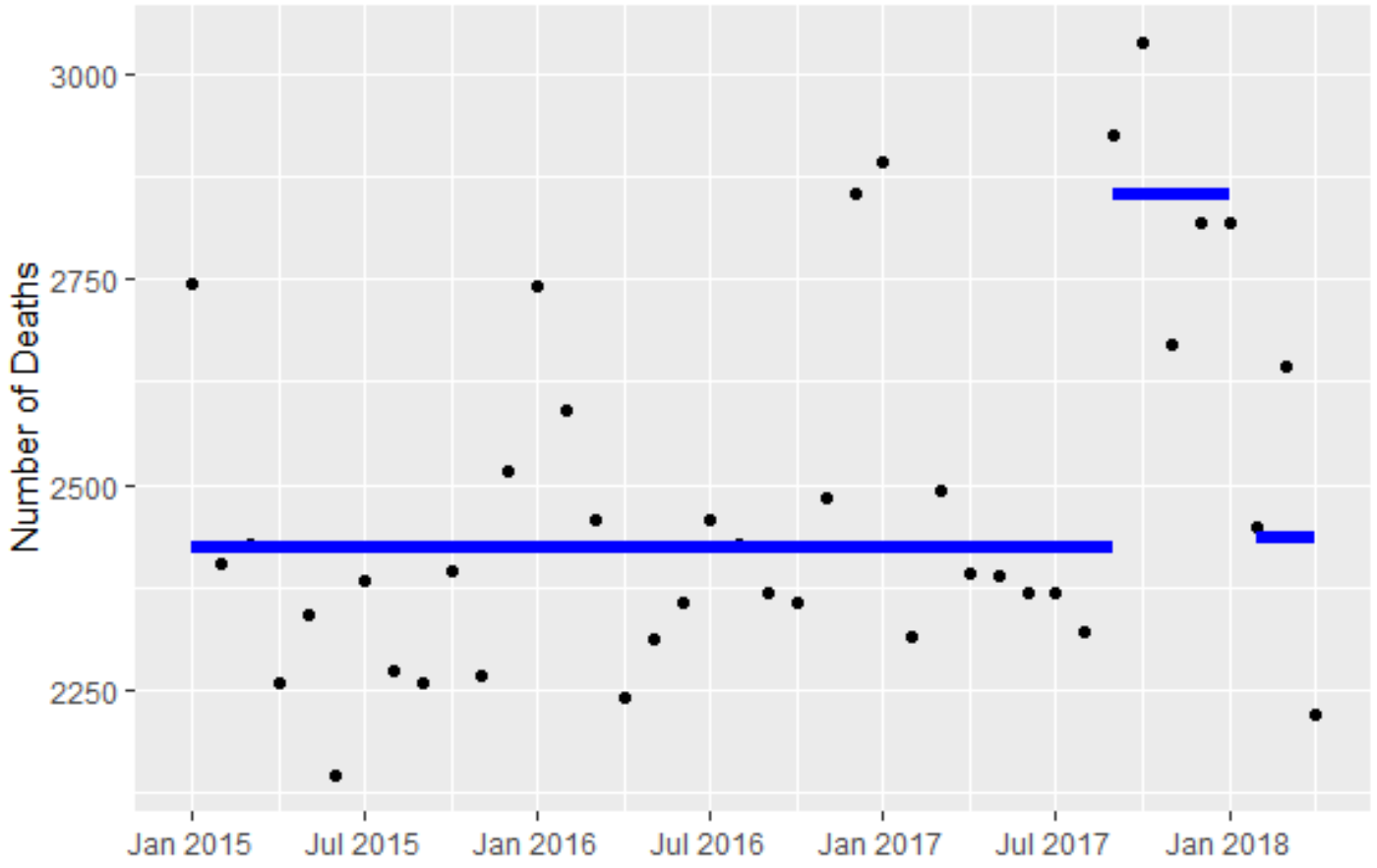
Basic Idea of Excess Counts

- an *event* (a death) happens or not
- due to one of two causes (but never both)
- we know total number of events
- we have an independent estimate of one cause
- we want to estimate number of events due to the other cause
- used in other fields: physics: particle decay, astronomy: extraneous light sources, etc.

Official Death Statistics in PR

Total de Defunciones por Mes

Mes	2015	2016	2017	2018
Jan	2744	2742	2894	2821
Feb	2403	2592	2315	2448
Mar	2427	2458	2494	2643
Apr	2259	2241	2392	2218
May	2340	2312	2390	1892
Jun	2145	2355	2369	0
Jul	2382	2456	2367	0
Aug	2272	2427	2321	0
Sep	2258	2367	**2928**	0
Oct	2393	2357	**3040**	0
Nov	2268	2484	2671	0
Dec	2516	2854	2820	0
Total	**28407**	**29645**	**31001**	**12022**


Number of Deaths
3000
2750
2500
2250
Jan 2015
Jul 2015
Jan 2016
Jul 2016
Jan 2017
Jul 2017
Jan 2018

Probability Model

A person dies or doesn't → Bernoulli trial

Number of people who die in some time period has Binomial distribution

In short time period probability of dying is small (!!)

Population is large

Poisson approximation to Binomial

$$X \sim \text{Pois}(n(\pi + \mu)), Y \sim \text{Pois}(m\tau\pi)$$

- X = number of deaths after Maria
- Y = number of deaths before Maria
- π = probability of death due to normal reasons
- μ = probability of death due to Maria
- τ = relative length of time periods
- n, m = sizes of population before and after

Interval Estimation

Invert hypothesis test $H_0: \mu = \mu_0$ vs $H_0: \mu \neq \mu_0$ using likelihood ratio test and Wilk's theorem (profile likelihood method).

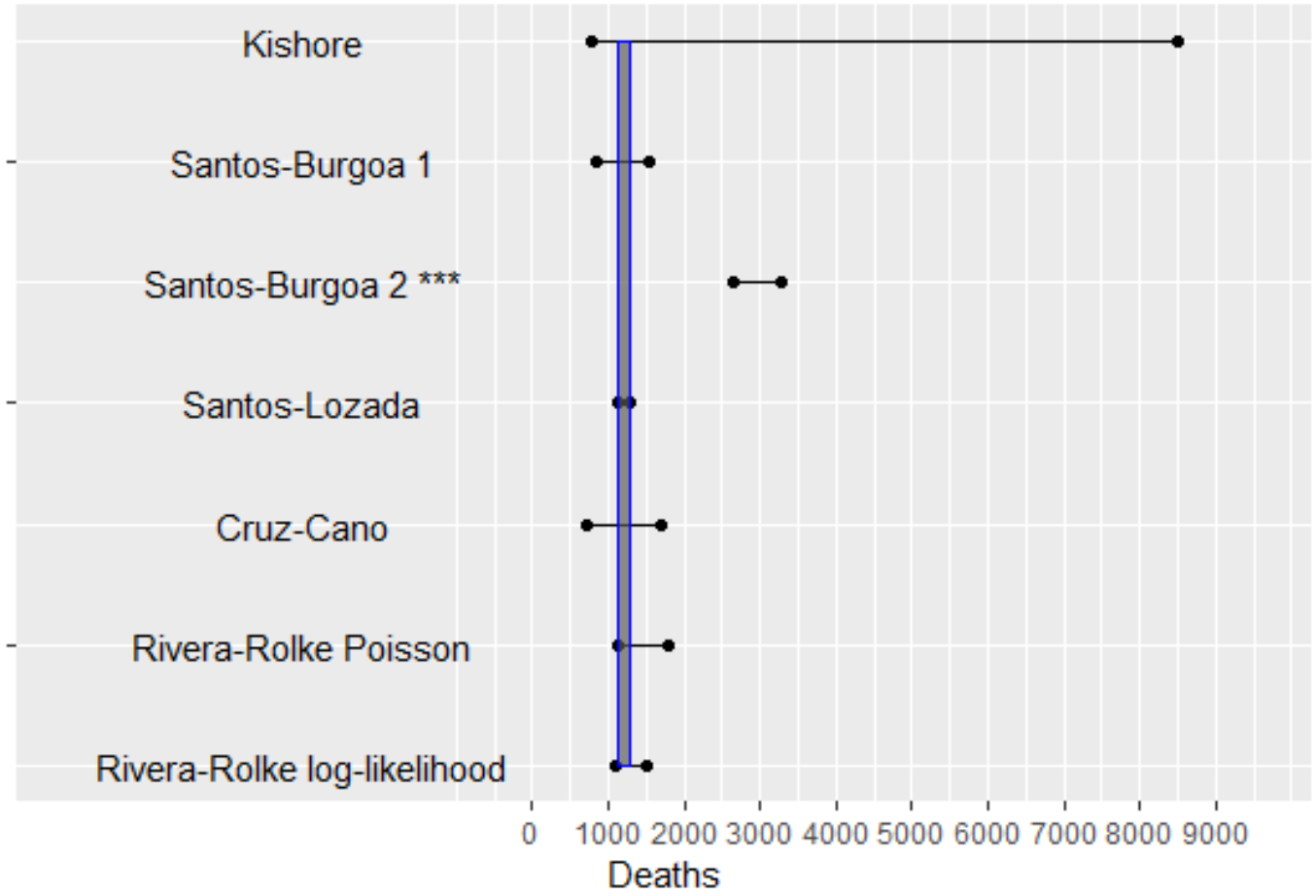
or

Bayesian credible intervals using Jeffrey's prior.

$\rightarrow$

almost same answer

Comparison of Results



Issues with the Official Death Toll Estimate

Why is the estimate of Santos-Burgoa so high? There seems to be two reasons:

- the data they used to estimate the change in population (air line passenger data) included a wrong value, which increased the estimate by about 300.
- Their model estimates the number of deaths from regular causes for December 2017 to be 2318. However, historical data shows much higher number of deaths in December in previous years, for example 2854 in 2016. For this to make sense, the population of PR had to have shrunk by 20% after the storm!
- This estimate is not really believable!

Can we combine these estimates - meta-analysis?

Unfortunalty this is in principle not possible on the level of confidence intervals. Some of the problems are:

- data is not independent, in fact most of these intervals are based on (almost) the same data.
- the methods of analysis are very different
- reputation of investigator
- …

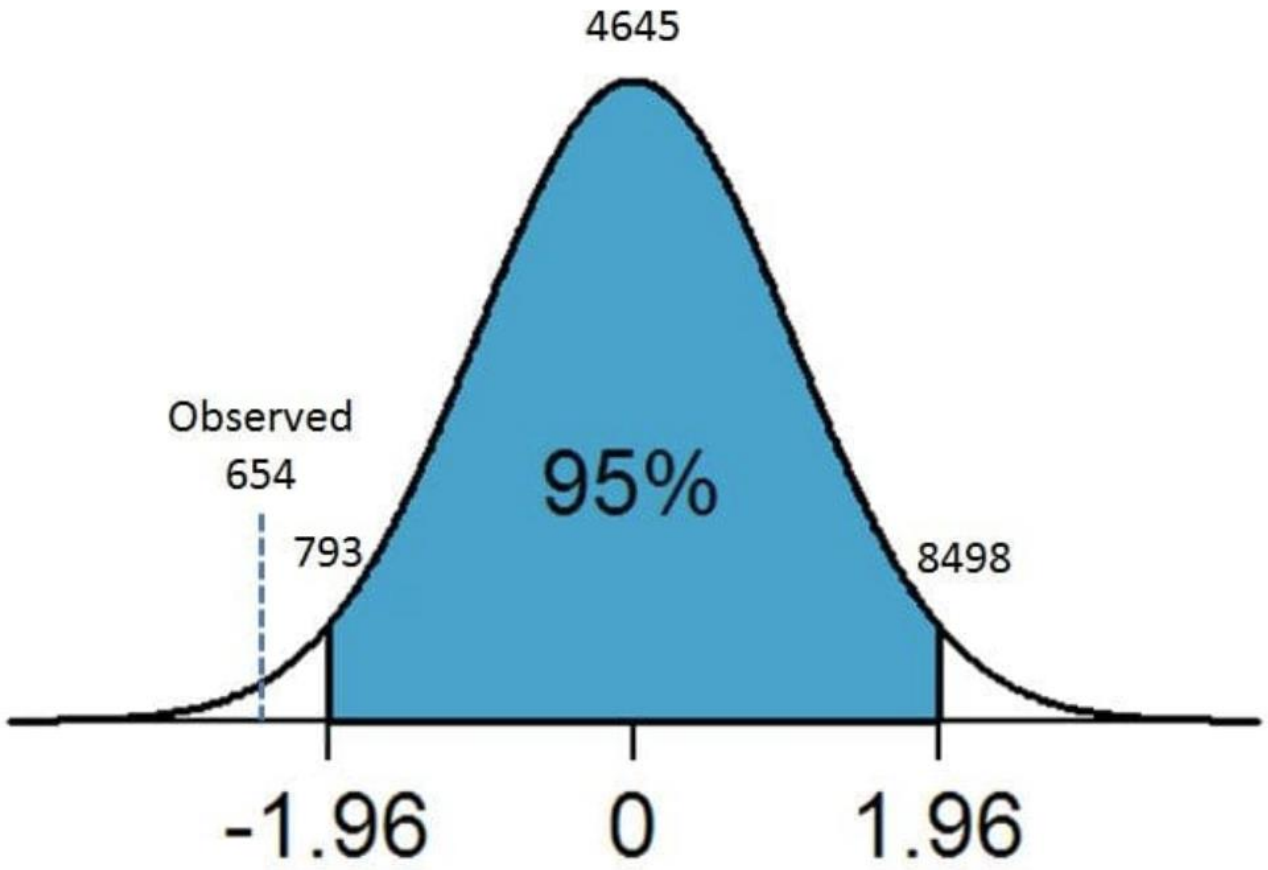
Statistics and the Public

- statistical concepts like confidence intervals are not trivial
- often misinterpreted as follows: the probability that the true number of deaths was between 2658 and 3290 is 95%

How hard are these ideas? In May 2018 I got a call from a reporter from Washington Post who wanted my expert opinion on Harvard study. Among other things I pointed out that it is not justified to pick a point in an interval as more “likely” than any other. In the articel he he writes

The widely reported number of 4,645 is simply the midpoint and is no more or less valid than any other number in the range.

Great! But a little later he has this graph:


4645
Observed
654
95%
793
8498
-1.96
0
1.96