Supplementary Methods

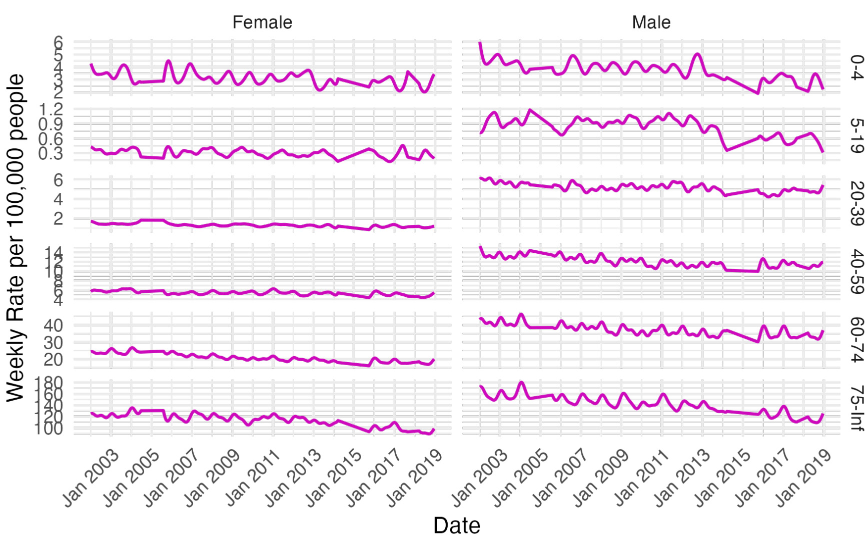
J. Christopher Polanco

Extra Results

Decision Making for Age Grouping

Age groups were collapsed into broader categories (5-19, 20-39, 40-59, 60-74, and 75+) based on their similar expected death rates per 100,000 population. This approach ensured that the analysis maintained interpretability while preserving key demographic distinctions, reducing potential overfitting and noise in the mortality trends.

Expected death rates after excluding dates concerning for excess mortality



Extra Figure 1. Expected Death Rates After Excluding Dates Concerning for Excess Mortality in Puerto Rico Over Time

Methods

Data Preparation and Visualization

Before analysis, the data underwent several cleaning steps:

1. **Aggregation by Age and Sex:** Mortality counts and population estimates were grouped by age categories and sex using the `collapse_counts_by_age` function.
2. **Weekly Aggregation:** Daily mortality counts were aggregated into weekly intervals using the `floor_date` function from the `lubridate` package, ensuring each interval represented a complete seven-day period, enabling consistent temporal comparisons. This process also allowed for a more reliable calculation of expected mortality trends while preserving essential demographic details.
3. **Exclusion of Anomalous Periods:** Periods with potential confounding effects, such as Dengue Fever or Chikungunya outbreaks or previous hurricanes (i.e., Irma, just prior to María's landfall), were excluded from baseline calculations.

Visualizations played a critical role in presenting the results, including population trends, observed versus expected mortality rates, and excess mortality stratified by age group and sex. These visualizations were created using the `ggplot2` package. Specifically, line plots were used to illustrate population trends, time-series plots compared observed and expected mortality rates, and ribbon plots highlighted periods of significant excess mortality with confidence intervals.

Functions to Estimate Expected and Excess Mortality

Our analytical framework utilized two complementary functions, `compute_expected` and `excess_model`, each serving distinct purposes. The `compute_expected` function was used to estimate baseline expected mortality rates by modeling historical data with long-term trends (splines) and seasonal patterns (harmonics). These expected rates formed the basis for detecting deviations. The `excess_model` function, on the other hand, evaluated excess mortality during the event window by comparing observed mortality rates to the computed baseline while accounting for temporal dependencies using a correlated error structure.

Further reasoning for the use of some advanced modeling techniques

The quasipoisson framework was used to account for overdispersion in the mortality data, ensuring that the variance in mortality rates did not exceed the model's assumptions. Splines were employed to capture long-term mortality trends, with knots placed at regular intervals to balance model flexibility and interpretability. The correlated error structure extended the standard Poisson model to account for temporal autocorrelation in weekly mortality data, improving the precision of excess mortality estimates. These techniques were chosen to ensure robust and reliable modeling of the data.

Estimation of Expected Mortality

Model Components

1. **Data Input:**
    - The dataset consisted of weekly mortality counts and population estimates stratified by age group and sex for the years 2002 to 2018.
    - Weekly aggregation was performed to ensure uniform temporal intervals, with complete weekly data required for inclusion.
2. **Exclusion of Event-Related Periods:**
    - Dates surrounding Hurricane Irma (September 6, 2017) and Hurricane María (September 20, 2017), as well as other significant health events (e.g., Dengue Fever outbreaks in 2004–2005 and Chikungunya outbreaks in 2014–2015), were excluded from the model to prevent bias in the expected mortality estimates.
    - These exclusions ensured the model was based on baseline mortality patterns unaffected by major events.
3. **Seasonality and Harmonics:**
    - Seasonal variations in mortality were modeled using Fourier terms (harmonics). For this analysis, two harmonics (`harmonics = 2`) were used to capture seasonal cycles effectively without overfitting.
    - These terms modeled periodic fluctuations in mortality (i.e., winter peaks due to influenza).
4. **Long-Term Trends:**
    - Splines were used to capture long-term trends in mortality rates, with knots placed at intervals determined by `trend.knots.per.year = 2`.
    - This ensured that the model could adjust for gradual changes in mortality patterns over the 16-year study period.
5. **Population Scaling:**
    - Mortality rates were standardized using population estimates for each age group and sex, allowing for meaningful comparisons across demographic subgroups.
6. **Dispersion Adjustment:**

- The model used a quasipoisson framework to account for overdispersion in the mortality data, where the variance exceeded the mean.

7. **No Weekday Effects:**

   - Given the weekly aggregation of data, weekday effects were not included (`weekday.effect = FALSE`) to avoid overcomplicating the model without adding meaningful information.

Statistical Approach

1. **Smoothing of Mortality Trends:**

   - The model smoothed mortality rates over time using a combination of splines for long-term trends and Fourier harmonics for seasonal variations.
   - This approach ensured that expected mortality estimates reflected realistic patterns while filtering out noise.

2. **Expected Mortality Estimates:**

   - For each week, the expected mortality rate was calculated by applying the smoothed model to the population estimates.
   - The expected counts were derived as:

   $$\text{Expected Mortality} = \exp(X\beta) \times \text{Population}$$

   where X represents the covariates (trend and seasonality terms) and β are the model coefficients.

3. **Uncertainty Estimation:**

   - The standard error of the log-transformed expected mortality was computed for each week, enabling the calculation of confidence intervals.

4. **Extrapolation and Components:**

   - When necessary, extrapolation was applied for periods outside the included data range. Seasonal, trend, and baseline mortality components were preserved for interpretability (`keep.components = TRUE`).

Estimation of Excess Mortality for Hurricane María

Model Components

1. **Data Input:**

- The dataset contained weekly mortality counts and population estimates, stratified by age group and sex.
- The analysis focused on the period spanning one year before and after Hurricane María (September 20, 2017) to capture the full impact of the hurricane.

2. **Expected Mortality:**
    - Expected mortality rates were computed using smooth time functions that incorporate:
        - **Seasonality:** Captured with harmonic terms to reflect periodic mortality patterns.
        - **Long-Term Trends:** Modeled using splines, with knots placed at intervals proportional to the `knots.per.year` parameter.
    - For this analysis, `knots.per.year = 4` ensured flexibility in capturing mortality trends without overfitting.

3. **Model Specification:**
    - A **correlated error model** was chosen (`model = "correlated"`) to account for potential autocorrelation in weekly mortality data.
    - Discontinuities were included at the date of Hurricane María's landfall (`discontinuity = TRUE`) to allow for abrupt changes/increases in mortality patterns associated with the hurricane's impact.

4. **Control Periods:**
    - Control dates (2002–2013) were specified to calibrate the correlated error structure and avoid biases from overlapping events.
    - Periods with potential confounding effects, such as Hurricane Irma (September 6, 2017) and historical outbreaks (e.g., Dengue Fever and Chikungunya), were excluded to ensure the integrity of the expected mortality calculations.

5. **Event Window:**
    - The event window spanned from September 20, 2016 (one year before Hurricane María) to September 20, 2018 (one year after).
    - Weekly mortality data within this window were analyzed to identify deviations from expected rates.

6. **No Minimum Mortality Rate Specified:**
    - By not enforcing a minimum rate, the expected mortality counts purely reflect historical trends and smoothing parameters (i.e., Fourier harmonics and splines). This approach avoids artificially inflating expected counts in periods with extremely low historical mortality rates, ensuring that deviations truly represent excesses relative to realistic expectations.

7. **Population Denominator:**
    - Mortality rates were adjusted by population estimates for each age group and sex, ensuring comparability across demographic subgroups.
8. **No Weekday Effects:**
    - Again, given the weekly aggregation of data, weekday effects were not included (`weekday.effect = FALSE`) to avoid overcomplicating the model without adding meaningful information.

Statistical Approach

1. **Correlated Errors:**
    - An autoregressive (AR) structure was fitted to account for autocorrelation in mortality rates, ensuring accurate standard error estimates for excess mortality.
2. **Spline-Based Modeling:**
    - Splines were used to model seasonal and long-term trends in mortality, with knots placed at regular intervals to balance flexibility and interpretability.
3. **Excess Mortality Estimation:**
    - Excess mortality was calculated as the difference between observed and expected mortality rates, adjusted for population size.
    - Confidence intervals for excess mortality were derived using standard errors estimated from the fitted model.
4. **Detection of Significant Deviations:**
    - Periods with statistically significant deviations (above a threshold defined by `alpha = 0.05`) were flagged as excess mortality intervals.

Decision Making for Excluding Hurricane Irma

The main event window spans one year before and one year after Hurricane María's landfall on September 20, 2017, to capture both short- and long-term impacts on mortality. To accurately attribute excess mortality to Hurricane María, overlapping events such as Hurricane Irma (landfall on September 6, 2017) were explicitly excluded. Exclusion periods for Hurricane Irma's impact (September 6, 2017, to March 6, 2018) were defined using the `exclude` parameter in both the `compute_expected` and `excess_model` functions. Given Hurricane Irma's proximity to María's landfall and María's greater magnitude and overall negative impact, a distinct effect from Irma was not anticipated. Excluding overlapping periods ensured that mortality trends attributed to Hurricane María were not confounded by other significant events, allowing for a more precise analysis of its effects.