IHI COVID-19 Data Tracker

Using Shewhart Charts to Understand Variation in COVID-19 Reported Deaths

The Institute for Healthcare Improvement (IHI) and our colleagues have developed a COVID-19 Data Tracker application showing the reported deaths from COVID-19 globally using Shewhart charts. This introduction provides a brief description of how the COVID-19 Data Tracker is organized, key terms used to display the data, and guidance on how to navigate the various screens.

Because the Shewhart charts, a cornerstone of improvement methods, are designed to distinguish between common cause (random) and special cause (non-random) variation, they are ideally suited to help decision makers understand if the reported number of daily deaths is predictable within limits or unpredictable (i.e., special cause variation).

The IHI COVID-19 Data Tracker has two main sections: 1) individual countries and 2) US states and territories. Shewhart charts are updated daily for each country and each US state and territory. The model tracking the trajectory of COVID-19 reported deaths detects when the number of deaths by geographic location is entering rapid exponential growth (Epoch 2, defined below) and when exponential growth has ended. We believe that the Shewhart chart approach will prove to be useful to subject matter experts, decision makers, care providers, and the public as they address COVID-19 challenges.

Additional resources:

- Read a more detailed explanation of the methodology behind the Shewhart charts used in this application on the New York Times GitHub site.
- Read an article that provides a description of initial development of the Shewhart chart application: Understanding variation in reported COVID-19 deaths with a novel Shewhart chart application
- Watch the Science of Improvement Whiteboard videos on the IHI website, which describe Shewhart charts and their use.
- Those interested in the details behind Shewhart chart construction and use should consult:
  o *The Health Data Guide* (Provost L, Murray S. Jossey-Bass Publishers; 2011)
  o *Quality Health Care: A Guide to Developing and Using Indicators, 2nd Edition* (Lloyd R. Jones & Bartlett Learning; 2019)

Key Concepts and Terms

The IHI COVID-19 Data Tracker has two main sections: 1) individual countries and 2) US states and territories. When exploring these two sections it is important to understand the key terms used to create and interpret the Shewhart charts. By understanding these terms, you will be able to find and explore the data for the geographic unit of interest easily and quickly. The key concepts described below include:

- Epochs
- Extensions
- Adjusted versus Unadjusted Data
- Ghosting unusually large data values

Epochs

This term is used to categorize the expected growth and decline in the number of deaths plotted on the Shewhart charts during an epidemic. Four discrete Epochs have been identified; we allow more than one phase within each Epoch. Within Epochs, phases are defined by different Shewhart charts.

- **Epoch 1: Pre-Exponential Growth**
  Epoch 1 begins with the first reported death. Daily reported deaths are then plotted on a c-chart. Epoch 1 ends when a special cause signal indicates exponential growth in the number of deaths. The special cause is determined by: 1) **one day** with reported deaths above the upper limit, or 2) **eight consecutive days** with reported deaths greater than the c-chart centerline (i.e., the average). Note that Epoch 1 may have multiple phases described by c-charts with differing center lines and control limits (either increasing or decreasing).

- **Epoch 2: Exponential Growth**
  The transition to Epoch 2 begins after at least five days of plotting reported deaths that produce a significantly positive slope for the regression of log10 of the deaths on each day. Epoch 2 ends when a special cause signals the start of a new phase and the new phase no longer has positive slope on the log10 scale. In Epoch 2, the special cause is signaled by: 1) **two consecutive days** with reported deaths above the upper control limit or below the lower control limit, or 2) **eight consecutive days** with reported deaths above or below the center line. Within Epoch 2 there may be more than one Shewhart chart phase if each phase continues to show exponential growth.

- **Epoch 3: Post-Exponential Growth with flat trajectory or exponential decline**
  Epoch 3 begins after at least five days of plotting reported deaths that produce a non-positive slope for the regression of log10 of the deaths on each day. The slope is either not significantly different from zero or significantly negative. Epoch 3 ends with the same special cause signals used to end Epoch 2: 1) **two consecutive days** with reported deaths above the upper control limit or below the lower control limit, or 2) **eight consecutive days** with reported deaths above or below the center line. Epoch 3 may also have more than one Shewhart chart phase as long as the conditions to end Epoch 3 are not met.

- **Epoch 4: Stability after descent**
  Epoch 4 must be preceded by special cause signals that mark the end of a phase in Epoch 3: **two consecutive days** below the lower control limit or **eight consecutive days** below the center-line average AND the **lower control limit must be less than 2**. Otherwise, the new phase will either continue in Epoch 3 or mark the beginning of a new Epoch 2. Our choice of lower control limit less than 2 is an empirical decision related to tracking deaths. We want to mark the transition back to relatively rare events that can be roughly described by a c-chart. A lower limit less than 2 indicates that we are anticipating 1's and 0's for the number of deaths, which is consistent with tracking "rare" events. If you analyze other measures, you could argue that the value for the lower control limit parameter might be larger. Epoch 4 may also have multiple Shewhart chart phases described by c-charts with differing averages and control limits. A country, state, or territory may return to Epoch 2 from Epoch 4 if the logic detects the beginning of exponential growth.

Extensions

When a phase is "active" the upper and lower control limits and center line are extended as reference markers for 10 days. If a new phase has begun but we only have a couple days of death records, we do not have any limits or center line to extend. The example chart for Argentina (shown below in Figure 1) demonstrates this "Extension" procedure on the right side of the chart.

Adjusted versus Unadjusted Data

The raw data (unadjusted) obtained for countries of the world and the US reflect the number of deaths submitted by the numerous sources tracking COVID-19 deaths in each geopolitical unit. As the IHI team worked with these raw data sources, it became apparent that weekday and weekend reporting patterns frequently displayed different patterns. The pattern observed most often was a weekend versus weekday difference in the number of reported deaths. For example, the number of deaths frequently decreased on weekends and then showed rather unexpected increases on Mondays or Tuesdays when the data from the previous weekend was posted. This practice also created "spikes" or extreme data points that did not typify the general historical patterns in the reported data. In order to deal with this recurring pattern, the unadjusted (raw) data have been adjusted to create a more stable stream of data from the various data sources. Read more about the methodology used to adjust the raw data on the New York Times GitHub site.

Ghosting Unusually Large Data Values

During Spring 2020, the IHI team followed news reports of "data dumps" and flagged these events manually. Data dumps are a simple and clear example of a special cause of variation in the data series that can affect Shewhart chart calculations. We developed a **detect outliers function** to identify records flagged as being unusually large. The detect outliers function examines each daily record to assess if the record is unusually large compared to days preceding and following the record. On the Shewhart charts **these unusually large data points are referred to as being "ghosted"** because the program used to create the charts, PowerBI, plotted such values as a **pale blue dot** (●) that is not connected by lines to the other dots on the sequence. While the ghosted data points are not included in the calculations of the limits, we leave them on the charts to remind us that they represent extreme values related to unusually large "data dumps."

Navigating the IHI COVID-19 Data Tracker

The IHI COVID-19 Data Tracker site has two major sections:

1. Shewhart charts for individual countries (Afghanistan through Zimbabwe) showing the daily number of COVID-19 reported deaths by day of the week and date
2. The number of COVID-19 reported deaths by day of the week and date in the United States for each state and five territories (i.e., District of Columbia, Guam, Northern Marianna Islands, Puerto Rico, US Virgin Islands)

Navigating the Individual Countries Charts

Figure 2 below is a screenshot of the Individual Countries home page. The key areas of the chart described below are outlined with red boxes and identified with numbers.

- The countries of the world are listed in alphabetical order on the left side of the chart (Box 1).
- Note that the screen opens with the United States as a default (Box 2). All countries are listed in alphabetical order (Box 1).
- Click the circle to the left of a country's name to view its chart. This step will be easier if you use the *Full Screen Mode* to make the view larger (Box 3). Click *Esc* to return to the standard view.
- The data shown on the chart are *Adjusted Data* following the methodology described below (Box 4). To view the chart with *Unadjusted Data*, click on the box in the lower left corner labeled "Click for Unadjusted Daily Deaths" (Box 5). You can also view the Unadjusted number of reported deaths by clicking on the "2" in the label at the bottom center of the chart that looks like this: <1 of 2> (Box 6).
- To view a specific date range on a chart, use the *"slider"* (located along the bottom of the chart) to select the date range you wish to review (Box 7).

Figure 2.

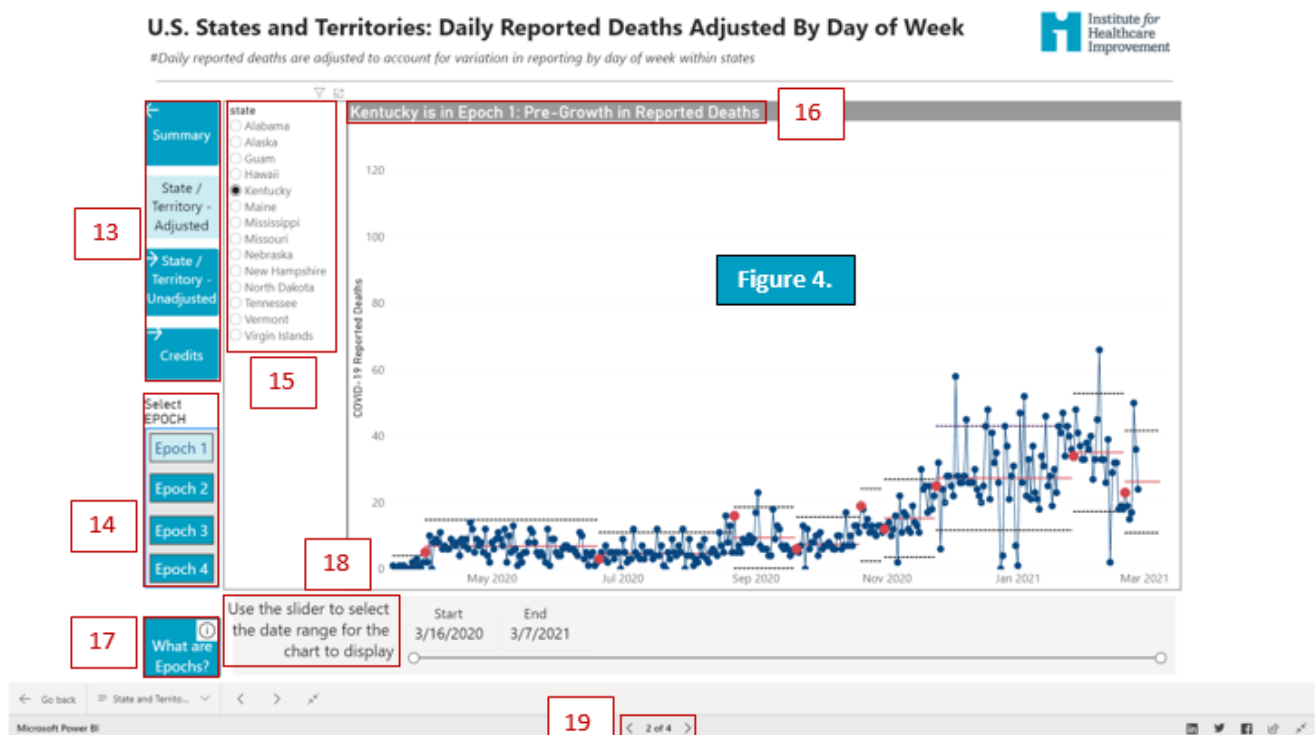## Navigating the US States and Territories Charts

US Summary Chart

- The starting point for the US States and Territories is the Summary chart shown in Figure 3 (Box 8).
- The charts for the US States and Territories follow a similar layout as outlined above for the Individual Countries, but there are a few additional options. Again, numbers (Boxes 8-12) have been placed on Figure 3 to direct you to the different aspects described below.
- The US Summary chart provides a "system view" of the states and territories that are in each Epoch over time. Figure 3 reinforces the previous point that on any given day we are dealing with different dynamics across the US as well as within individual states and territories over time.
- Notice in Figure 3 that when we started tracking the number of deaths in March 2020, all 55 states and territories were in Epoch 1 (Pre-Exponential Growth), the light blue line on the chart (Box 9). This started to change dramatically in April and May as observed by the sharp drop in the number of states and territories in Epoch 1.
- In early April 2020, 22 states and territories moved rapidly into Epoch 2, the Exponential Growth Epoch (purple line). Box 10 highlights this point. You will then observe a sharp decline in Epoch 2 and the related increase in the number of states and territories entering Epoch 3, Post-Exponential Growth (the orange line).
- At the far right of the chart (Box 11), you can see that on February 26, 2021, 35 states and territories were in Epoch 3 (orange line), 15 in Epoch 1 (light blue line), 4 in Epoch 4 (purple line), and 1 in Epoch 2 (dark blue line).
- To review the definitions of the four Epochs, click on "*What Are Epochs?*" in the lower left corner (Box 12).
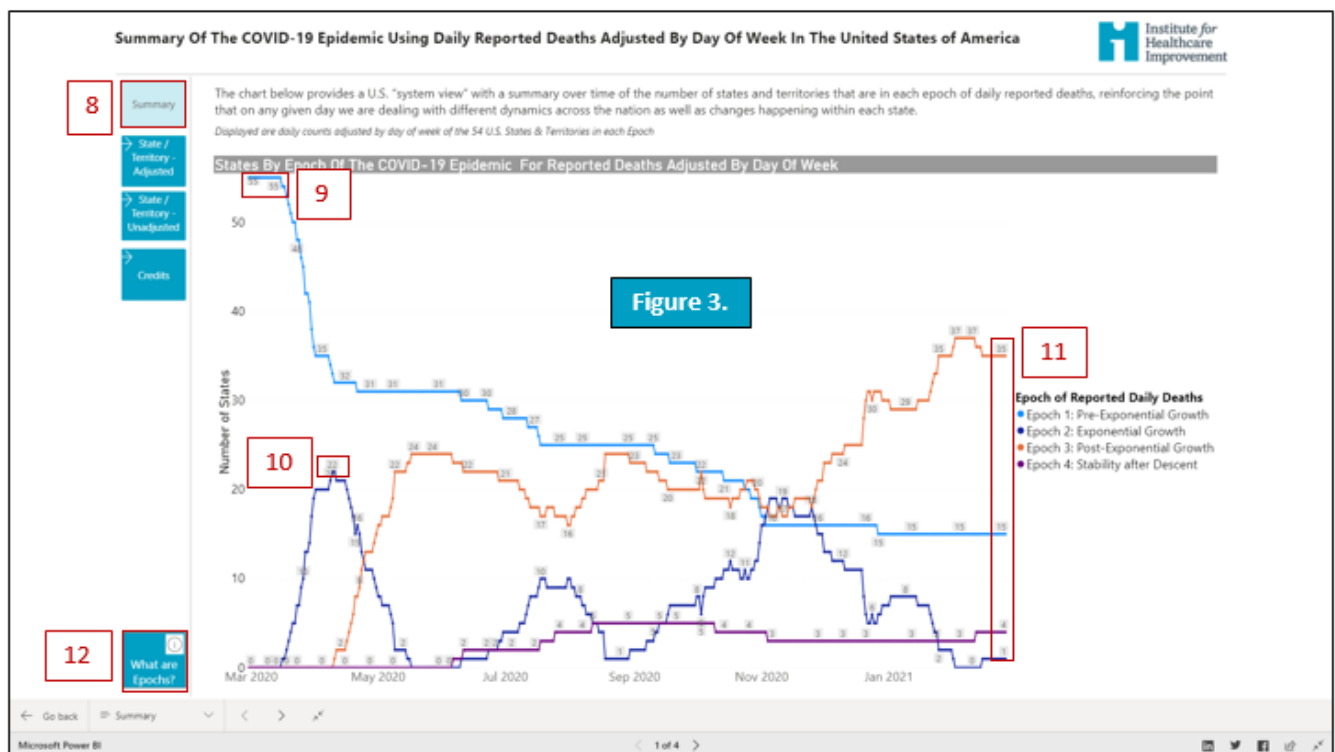
Figure 3.

Charts for Individual US States and Territories

- These charts can be accessed by becoming familiar with the tabs in Box 13 and Box 14 in Figure 4.
- The first tab in Box 13, *Summary*, was described above.
- The key tabs in Box 13 are the *State/Territory Adjusted* and *State/Territory Unadjusted* tabs. When one of these tabs is selected its color changes from dark to light blue. In Box 13 the *State/Territory Adjusted* tab is light blue indicating that the data shown on the chart in Figure 4 have been adjusted using the logic described above. If you click on the *State/Territory Unadjusted* tab it will turn light blue and the chart will display the unadjusted data for the selected state or territory.
- Note that when you select either tab — *State/Territory Adjusted* or *State/Territory Unadjusted* — you will automatically activate a chart showing the states and territories in Epoch 1 as shown in Figure 4.
- The last tab in Box 13 provides the "*Credits,*" which identifies the team members that have developed this site and links to read more about the Shewhart chart methodology used in the IHI COVID-19 Data Tracker.
- Box 14 lists the four Epochs. As mentioned above, Epoch 1 is the default starting place as shown in Figure 4. In Figure 4, Kentucky has been selected as an example. The complete title for this chart is then shown in Box 16. This title will change depending on the Epoch and the state or territory selected.
- When you click on a tab for any of the three remaining Epochs, it will turn light blue and you will see in Box 15 the states and territories in this Epoch. Click the circle to the left of a state or territory to view its chart.
- Again, if you need a refresher on the definition of each Epoch, click on Box 17.
- Use the "*slider*" (Box 18) to revise the dates along the horizontal axis and zoom in on a particular time period.
- Use the forward (>) or back (<) arrows (Box 19) to switch between the Summary chart (Page 1), Adjusted Deaths charts (Page 2), Unadjusted Deaths charts (Page 3), and Credits (Page 4).

Figure 4.

IHI COVID-19 Data Tracker Data Sources

The IHI Data Tracker uses COVID-19 data from the New York Times GitHub site for the US states and territories charts and the Our World in Data site for the country-specific data.

IHI COVID-19 Data Tracker Team

- Lloyd Provost, Improvement Advisor, Associates in Process Improvement
- Shannon Provost, Clinical Assistant Professor, The University of Texas at Austin
- Rocco Perla, Co-Founder, The Health Initiative
- Kevin Little, Improvement Advisor, Informing Ecological Design, LLC
- Neel Vora, Vice President, Information Technology, Institute for Healthcare Improvement
- Robert Lloyd, Vice President, Improvement Science and Senior Improvement Advisor, Institute for Healthcare Improvement

For inquiries about navigating and using the IHI COVID-19 Data Tracker, email Robert Lloyd (rlloyd@ihi.org).

How to Cite This Content: *COVID-19 Data Tracker: Using Shewhart Charts to Understand Variation in COVID-19 Reported Deaths*. Boston: Institute for Healthcare Improvement; 2021. (Available on ihi.org)