DRAFT:Measuring the impact of new vaccines using mortality and administrative hospitalization data

Pneumococcal conjugate vaccine as a case study

Table of Contents



WHO working group on the use of administrative data to evaluate the impact of pneumococcal conjugate vaccines:

* Adam Cohen (World Health Organization)
* Lucia Helena De Oliveira (Pan American Health Organization)
* Fernanda Lessa (Centers for Disease Control and Prevention)
* Jennifer Loo Farrar (Centers for Disease Control and Prevention)
* Tomoka Nakamura (World Health Organization)
* Cristiana Toscano (Universidade Federal de Goiás)
* Daniel Weinberger (Yale School of Public Health)

## Introduction (Cris)

Once vaccines or other public health interventions are deployed, it is often desirable to measure their impact on health. This information is often critical for policy-makers who prioritize funding and implementation of different programs. However, conducting robust and credible evaluations of the public health impact of interventions is challenging. Real-world data are complex, and decisions about how to clean, format, analyze, and interpret the data can influence the conclusions about the impact of the intervention.

In this manual, we discuss issues in performing impact analyses using routinely-collected administrative data sources. As a specific case study, we discuss challenges in evaluating the impact of the introduction of pneumococcal conjugate vaccines among children on rates of hospitalization and death due to pneumonia.

## What is vaccine ‘impact’?

Many vaccines influence disease rates in 2 ways. Direct effects reduce the risk for an individual to become ill and directly results from the immune response from the vaccine. Indirect effects are the benefit that an individual receives as a result of decreased transmission of the pathogen. Both vaccinated and unvaccinated individuals benefit from this indirect protection. When we refer to *vaccine impact* we are referring to the overall change in disease rates in the population that results from the combination of direct and indirect effects.

## Mortality data from Vital Statistics (Cris)

## Hospitalization data from administrative sources (Cris)

## Study design (DAN)

### Overview

Observational studies present a number of analytical challenges. The introduction of vaccines is often occurring concurrently with other oublic health and social interventions that can influence disease rates. Other relevant changes include changes in social welfare systems, changes in the capacity of the healthcare system, changes in the efficincy of registering hospitalizations and deaths, and other pharmacological interventions that could influence susceptibility (e.g., increasing use of anti-retroviral therapy in sub-Saharan Africa). Because vaccines are introduced at the same time as these other changes, it can be challenging to attribute specific changes to the introduction of a vaccine. The methods that we will discuss here attempt to address this issue using different approaches and different assumptions. It is important to be aware of the assumptions and limitations of the different approaches.

### Possible study designs for impact assessment

The analysis goal is to disentangle changes in disease rates that are caused by the introduction of a vaccine program with changes that are due to these other factors. There are several approaches that could be taken. First, in a *pre-licensure* study, vaccine impact could be evaluated using a cluster-randomized study design, where disease rates are compared between spatial clusters that have been randomized to receive that vaccine or not. Assuming that there is not transmission between clusters, that the assignment of clusters was random and succesful (i.e., that the vaccinated and unvaccinated clusters are comparable), this provides an unbiased estimate of the total effect of the vaccine.

In a *post-licensure setting*, the cluster-randomized design can be approximated by using a ‘step-wedge’ study design. This design can be implemented in settings where the vaccine has not yet be introduced and where it might not be feasible to introduce to the entire population at the same time. With this design, geographic areas are randomized to receive the vaccine earlier or later. All geographic units eventually receive the vaccine. With this phased rollout, the geographic units that introduce the vaccine later serve as controls for the geographics units that introduce the vaccine earlier. Because the control group is changing over time and the comparisons between the vaccinated and unvaccinated groups is taking place over some time period, estimates of vaccine impact could be cofounded by underlying temporal trends. Therefore, it is important to appropriately adjust for time-varying rates of disease when analyzing these studies.

In most settings, such controlled roll-outs of vaccine are not possible. Therefore, the most common study designs are purely observational, in which changes in disease rates are evaluated over time or between regions. Extreme caution needs to be used when performing and interpreting these studies to ensure that factors unrelated to vaccination are appropriately adjusted in the analysis. This guide will focus on the analysis situation where there is a single time series of interest from a country or region and the goal is to detect changes in incidence following vaccine introduction from this time series. When time series from multiple subnational regions are available, additional types of analyses are possible, including performing spatiotemporal analyses in which the declines in regions with higher or lower coverage are compared.

### Counterfactuals: What would have happened without a vaccine?

With any analysis of vaccine impact, the goal is to compare the observed disease rates in the post-vaccine period with an estimate of what would have happened if the vaccine had not been introduced. This value is called the **counterfactual**. There are many ways to estimate counterfactuals from very simple approaches (as is done in pre/post comparison of incidence) to more complex approaches that adjust for trends and dynamics of the disease. In each of the followinng sections, we will discuss different methods to obtain this quantity.

### Key considerations for analyses

#### Denominators

It is common when analyzing trend data to analyze changes in incidence (i.e., cases/population size) rather than the raw number of cases. The assumption here is that the number of cases should scale with population size. However, when using administrative data, using a population denominator might not effectively adjust for underlying change. Often, the administrative database does not cover the entire population, or other factors (e.g., access to healthcare) are larger drivers of trends than population size. Using alternative denominators, such as total number of hospitalizations in the system, can sometimes be a more effective approach (see Schuck-Paim et al., *Vaccine* 2017). Whatever the chosen denominator, it can either be included as an offset term (regression coefficient is held to 1), or as a covariate in a regression model.

#### Transition period

After a vaccine is formally introduced into a national immunization program, it typically takes some amount of time to implement the vaccine and vaccinate a large proportion of eligible infants. Additionally, it takes time for children who are vaccinated as infants to age and become toddlers (an age group that might be more important for transmitting pneumococcus). Due to these considerations, a transition period is typically defined to allow the program to reach maturity before trying to estimate changes in disease rates. The data are therefore subset into a pre-vaccine period, a transition period, and a post-vaccine period, and vaccine impact is defined by comparing the post-vaccine period and the pre-vaccine period, while ignoring the transition period. The length of transition period depends on how quickly the vaccine program is implemented. In most settings, 1-2 years are defined as the transition period. In many of the Latin American countries, 3-dose coverage rapidly reached a plateau within a year, so a 12 month transition period was used.

#### Choosing control variables

As discussed in the following sections, control variables can be used to strengthen analyses of vaccine impact in several ways. They can help identify data quality issues and also to adjust for time-varying factors that could lead to misleading conclusions. For instance, social welfare programs might influence the general health of the population and drive down rates of pneumonia, independent of a vaccine. Ignoring the effects of this social welfare program would lead one to incorrectly attribute declines in disease rates to the vaccine. A control disease that is also influenced by the general health improvements could allow an analyst to detect and adjust for these improvements when calculating vaccine impact

The key challenge comes in deciding which control diseases would be appropriate controls for the disease of interest. As a general principle, the control disease and the disease of interest should share relevant risk factors and etiologies. For instance, if the goal is to adjust for the effect of the social welfare program on the risk of infectious diseases, using fractures as a control when estimating the effect of a vaccine against pneumonia might be a poor choice. This is because trends in fractures have a completely different set of causes than trends in acute respiratory infections. Perhaps a different acute infectious event would be more appropriate. Conversely, if the goal is to adjust for changes in healthcare capacity, trends in fractures could capture important trends. Pathogen-specific time series, such as for influenza or RSV activity, could also be used to adjust pneumonia rates for short-term epidemics and tease out the long term trends. It can sometimes be difficult to determine *a priori* which controls would be appropriate. For instance, in mortality databases, some deaths are attributed to unspecified causes.

A key assumption when selecting control diseases is that the relationship between the outcome of interest (e.g., pneumonia hospitalizations) and the control time series is stable over time, and that the only thing that causes a change between them is the intervention of interest. Therefore, diarrhea might be a poor control variable for pneumonia when evaluating pneumococcal conjugate vaccines if rotavirus vaccine has been introduced recently as well. This is because the rotavirus vaccine changes the relationship between diarrhea and pneumonia. In this example, using diarrhea as a control would lead to a bias towards detecting no effect because diarrhea and pneumonia will both be going down due to their respective vaccines.

Another key assumption is that the control variables are themselves not influenced by the intervention of interest. Therefore, invasive pneumococcal disease (IPD) would be a poor control for evaluating the impact of pneumococcal conjugate vaccines on pneumonia, because IPD is itself influenced by the same vaccine. The inclusion of a control that is influenced by the same intervention will bias the estimates towards no effect because both will be going down in parallel due to the same cause.

Low-quality databases often have a large proportion of deaths coded using these junk codes, but in many settings, the quality of coding has changed over time. A decrease in junk codes over time would be expected to be associated with an increase in cause-specific codes. Therefore, the time series of junk codes itself could be useful for adjusting the reported rates of pneumonia deaths for changes in coding practices. Conversely, if actual pneumonia cases are incorrectly coded with a junk code, then the vaccine will cause a decline in the junk codes, and including these junk codes as control variables

**In summary** Control diseases should share some relevant etiological drivers with the disease of interest and should themselves not be influenced by the intervention or by other interventions that would change the relationship between the control and the outcome of interest. The sections that follow discuss some specific uses of control diseases in analyses.

#### How many years of data required?

In many instances, vaccines are introduced in the context of unrleated changes that influence disease rates. Changes in the underlying health of the population (e.g., due to changes in social welfare programs, declines in indoor air pollution), changes in healthcare utilization, and changes in coding can all affect the reported rates of disease. Therefore, it is important to have sufficient data from the pre-vaccine period to detect and adjust for such trends. *At least 4-5 years* of pre-vaccine data are typically needed to adequately adjust for existing trends in the data. This is often context specific and depends on the particular factors driving underlying trends. Analyses using fewer years of pre-vaccine data should be interpreted with *extreme caution*.

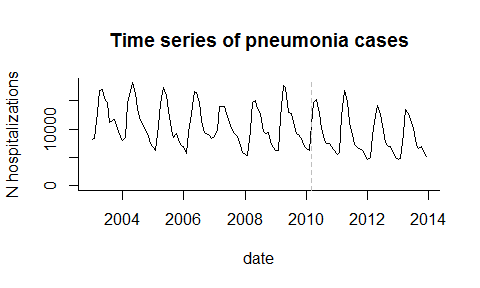
Likewise, it is ideal to accumulate several years of post-vaccine data prior to evaluating vaccine uptake. This allows the vaccine program to mature. Additionally, pneumonia rates during short periods of observation can be strongly influenced by short epidemics such as those cuased by influenza or RSV. This can potentially obscure or exagerate the true vaccine effects.

#### What to do if you have an epidemic disease

All of these methods assume that the disease patterns follow a preditable pattern and can be captured either using a straight line relationship (ITS), or that the relationship with control variables is stable. Therefore, these methods are generally only appropriate for **endemic** diseases. For diseases that are epidemic (e.g., meningococcal meningitis), other approaches might be required that account for the dynamics of the pathogen and the build-up of immunity in the population. If the epidemic is widespread but the intervention is limited to a smaller region, it might be possible to use the time series from an unvaccinated control population (e.g., another region within the country) as the control variable to generate a counterfactual in a synthetic control-type analysis. Or it might be necessary to use a dynamic transmission model that can capture non-linear dynamics (e.g., a compartmental model with Susceptiple, Infected, and Resistant classes).

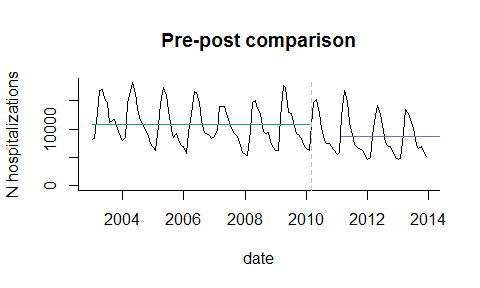
## Types of analyses

In this context, a ‘time series’ is defined as a variable in which the number of cases is tallied in each unit of time (week/month/quarter/year). The goal for the analysis is to detect changes in the average number of cases or incidence. In this example, we have the number of hospitalization coded as having a diagnosis for pneumonia (ICD10 codes J12-J18) in Brazil in infants

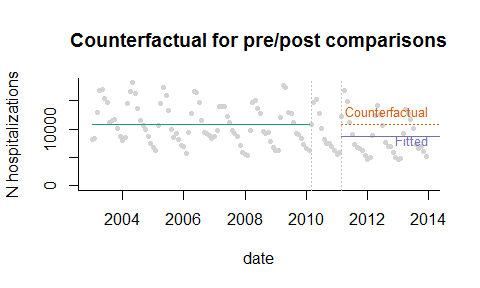


### Pre-Post comparison

The simplest possible analysis approach is to compare the average number of cases or incidence in the post-vaccine period with that in the pre-vaccine period (a ‘pre-post comparison’ study). This method is easy to implement and easy to understand. The analyst needs to define the pre-vaccine period and the post-vaccine period. Typically, the first year or two after vaccine introducion are excluded from the analysis because vaccine coverage has not yet reached full coverage levels. The decision about where to set the pre- and post-vaccine periods should be made *a priori* and should not be influenced by observed aberrations in the data (unless these are due to a known data quality issue); otherwise the estimation of the variability in disease rates will not be accurate.



**Counterfactual** In this study design, it is assumed that if the vaccine had not been introduced, the incidence rate after vaccine introduction would be the same as the incidence of disease before vaccine introduction. Therefore the counterfactual is simply the average incidence in the pre-vaccine period, and the comparison is with incidence in the post vaccine period. In the plot below, the counterfactual is shown with a dotted line. Comparing the counterfactual with the observed mean for the post-period gives an estimate of the vaccine effect.

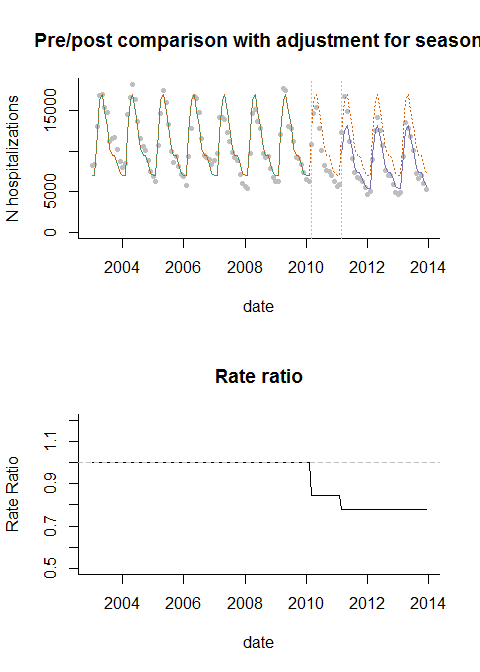


**Calculation of vaccine impact** The most common statistic reported from a pre/post comparison study is a Rate Ratio, which is simply calculated as (Average Incidence Post-Vaccination)/(Average Incidence Pre-Vaccination). Values <1 are considered evidence that the disease rates have declined. It is also possible to calculate a rate difference to obtain the number of cases prevented.

**Assumptions** This analysis assumes that the only change in disease rates that is occurring over time is due to the vaccine. This is rarely a realistic assumption. **Therefore, this is a weak study design, and the results should be interpreted with caution.**

**Extensions** It is sometimes desirable to adjust for seasonality in the analysis. This is particularly important when partial years are included in the pre and/or post period, and the pre- or post-periods are imbalanced in terms of which parts of the year are included in the analyses. This is accomplished in a regression model by adjusting for seasonality using dummy variables for month or harmonic variables. More details can be found in the hands-on exercises.

In this example, the counterfactual is seasonally adjusted, so the counterfactual for January is equal to the mean number of cases in January in the pre-vaccine period, and likewise for all of the months.



### Interrupted time series analysis (ITS)

In many instances, there is an underlying trend in the time series that is unrelated to vaccination. Such a trend can be caused by a number of factors, including changes in healthcare access, changes in susceptibility of the population, and changes in the sensitivity of surveillance. Such trends can bias the estimates of vaccine impact if they are not properly controlled. The simplest approach to adjust for trends is to fit a straight line through the data and then test whether the slope of the line or the level of the line changes after vaccine introduction. This can be accomplished by fitting a Poisson or negative binomial regression model. A number of quantities can be estimated sing these models, including the change in the slope of the trend line or in the average number of cases. However, it is often most useful to use the model to calculate the decline in incidence (rate ratio or rate difference) compared to what would be expected if the trend/level had remained constant. Seasonal variations and known trends can be adjusted for in this model using regression covariates.

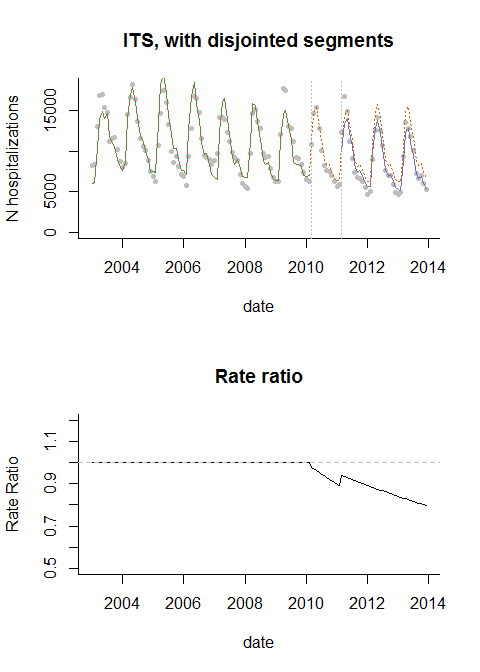
**Assumptions of ITS analysis** A key assumption is that the linear trend in the pre-vaccine period captures the important underlying trends and that these trends would have continued at the same rate if the vaccine had not been introduced. Assumptions also need to be made about which time periods to include in the ramp-up period and whether the slope should level out after a certain amount of time (and how much time should be allowed). Some of these assumptions about the timing and shape of the trajectory can be relaxed by using a flexible spline to capture post-vaccine changes (For example, see van Deursen, **Vaccine**, 2017)

**Calculation of vaccine impact from ITS analysis** Vaccine impact is measured in a variety of ways with ITS models. Some authors report the change in the slope parameter or change in the level. However, these values are somewhat abstract and difficult to interpret from a public health standpoint. It is more useful to calculate a relative change in incidence (rate ratio) or an absolute change (rate difference) by comparing the fitted value from the regression with the counterfactual value.

**Variations of ITS models** There are a number of ways to structure these models, but they typically include an index for time to capture the slope during the pre-vaccine period as well as terms that allow the slope or intercept to change in the post-vaccine period. The structure of the ITS model depends on how quickly you expect the vaccine effect to take hold. For instance, the change can be immediate (disjointed ITS) or more gradual. Some of these variations are detailed below. An excellent review of this topic is provided by Bernal, Cummins and Gasparrini (Intl Journ of Epidem., 2017 (46)1: 348-355).

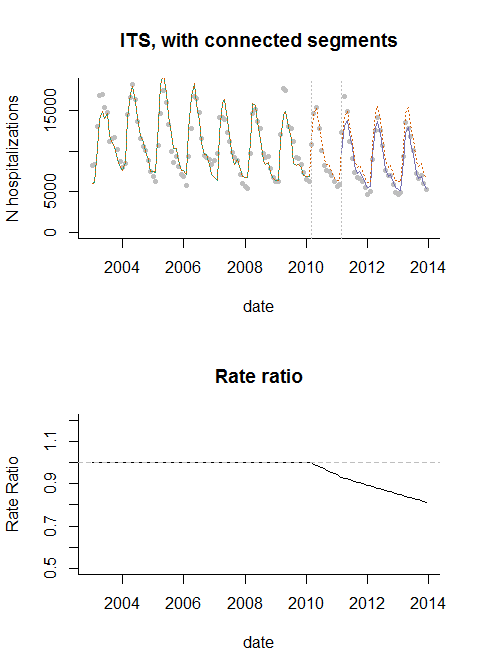
#### Disjointed ITS

In a disjointed ITS analysis, dummy variables (encoded 0 before vaccine introduction and 1 after vaccine introduction) are included to allow the level to change, and an interaction term between the dummy variable and the index for time allows the slope to change after vaccine introduction. In this model, the line segments fitted through the data do not connect. In practice, this can lead to strange and implausible shifts in the fitted values as shown in the plots of the rate ratio below.



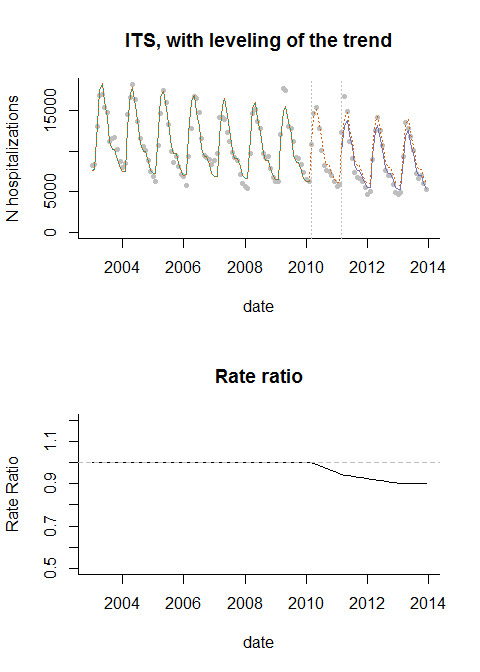
#### ITS with connected segments

A better alternative is to use a linear spline, which forces the fitted line segments to connect. Since most vaccines roll out gradually, and there is unlikely to be an immediate drop, this is a more realistic way to model the data.



#### ITS with leveling of the slope

This could be further modified to allow the slope to level out after a certain time period (in this example, the post-vaccine trend levels out 36 months after vaccine introduction)



#### The use of controls in ITS analysis

With this type of ITS analysis, there is a good chance that there are changes in the time series that are not well-captured by the linear trend assumption. There are 2 approaches that can be used to detect and control for such patterns: control outcomes and control covariates.

**Control Outcomes** With a control outcome, a different disease that is not affected by the vaccine is chosen, and the same model that was fit to the main outcome of interest is fit to this control time series. The ‘vaccine effect’ is estimated for both the outcome of interest and the control. The inference is typically made by comparing these vaccine effects. If the control series shows no vaccine effect, and the outcome of interest does, this would be taken as evidence that the estimate is not biased by unrelated trends. This requires that a control is chosen that would be influenced by relevant confounding factors. See the ‘special considerations’ section above for a discussion about selecting control variables.

**Control covariates** It is also possible to include other time series as control variables in the ITS regression model. The goal with this approach is to adjust for time-varying confounders. As an example, if the outcome is pneumonia, and smoking rates are changing over time, the rate of smoking in the population at each time point could be included as a covariate.

**Counterfactual** With an ITS model, the assumption is that the trend from the pre-vaccine period would have continued with the same slope into the post-vaccine period. To estimate the value of the counterfactual at each time point, we use the regression model but hold the terms representing post-vaccine changes in trend or level to 0 (no change from the pre-vaccine trends and levels).

#### Extensions

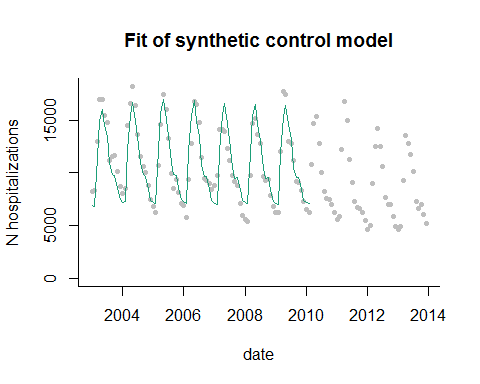
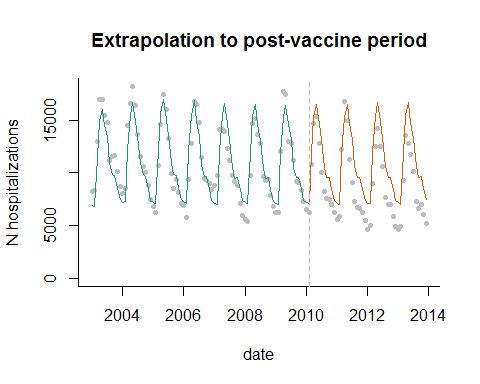
There are a number of variants on these approaches that can, in some circumstances, strengthen the analyses.

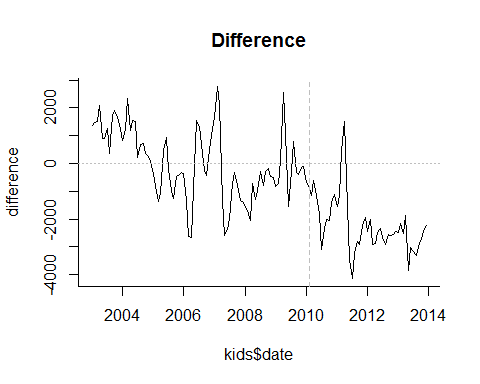
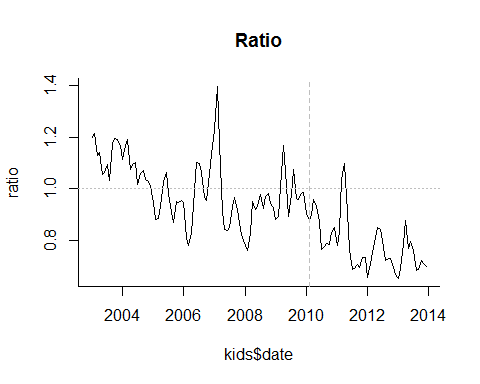
**Change point models** With the ITS model described, the user needs to specify the time at which the vaccine-associated change begins. In some circumstances, even if the change occurs before the introduction of a vaccine, the decline is incorrectly attributed to the vaccine. Change point analyses are effectively ITS models, where the timing of the change in slope or change in mean is estimated from the data. This provides an additional piece of information: how much disease rates change and **when** disease rates begin to change. The model are also more felxible and could rpovide a more accurate estimate of the decline, if the guess about when rates will decline is incorrect in the ITS model. This approach can provide a helpful check to ensure that the change is actually occurring after the data of intervention at a plausible time point. Kurum et al, *Epidemiology* 2017 provides an example of how these models can be used to estimate effects of pneumococcal vaccines against pneumonia.

**Holt-Winters models and ARIMA models** There are countless variants of ITS-type models that incorporate information about the time series as well. With the Holt-Winters approach, time points closer to the date of vaccine introduction are given more weight when estimating the slope and mean cases in the pre-vaccine period and can thus more effectively adjust for changes in the population that occur close to the time of vaccine introduction. See Andrade et al, *PLOS One* 2017 for an example of this model. ARIMA models also explicitly account for the time series characteristics of the data and decompose the series into different components. This approach allows the confidence intervals to expand appropriately as we move further away from the date of vaccine introduction.

### Synthetic controls

With the synthetic controls approach, the goal is to use time series that are not influenced by the vaccine to adjust for underlying secular trends in the data. With the “control covariates” approach described above, the control disease is included as an adjustment in the regression model, and the model is set up to detect changes in slope or level of disease after introduction of the vaccine. Synthetic controls analysis is an extenstion of this idea. The main differences are (1) that the models are only fit to data from the pre-vaccine period and used to extrapolate to the post-vaccine period and (2) typically many control time series are included, and variable selection is used to weight these controls.

**Simple example with a single time series** Using the same dataset from Brazil that was used in the ITS example, we will evaluate changes in J12-J18 following introduction of the vaccine. We will fit the model just to data form the pre-vaccine period, adjusting for seaosnality and the control variable. As an example, we will use log(all-cause)  Now that we have fit the model, we will plug in the observed values of the control variable into the model to get an estimate of pneumonia during the post-vaccine months 

The vaccine effect is calculated by taking the difference or ratio between the observed and extrapolated value during the post-vaccine period. The plots shown below give the ratio or difference at each time point. We will typically instead aggregate the observed cases and the extrapolated cases over a certain time period–by year, or during the evaluation period (e.g., excluding the first 12 months of post-vaccine data) 

**Use of multiple controls** In the example above, we have a single control time series. However, it is often difficult to choose a single control *a priori*, and there are often a number of possible candidates. With a typical synthetic control analysis, a large number of possible control variables are included in the regression, and Bayesian variable selection is used to give more or less weight to different controls. Controls that do not capture key variation in the outcome will receive a weight of 0. Controls that are more important will receive weights closer to 1. Controls are not ‘excluded’ from the model as they would be using classicla forward or backward variable selection. Rather they just receive more or less weight. Thus the final model is a composite that includes all of candidate controls.

**Counterfactual** The counterfactual is generated by plugging the observed values for the control time series during the post-vaccine period into the model. If the relationship between the outcome disease and the controls is stable over time, this represents an estimate of number of pneumonia cases that would be expected if the vaccine had not been introduced.

**Calculation of vaccine impact** The vaccine effect is calculated by taking the ratio or difference between the observed cases and the extrapolated number of cases estimated from the regression model (the counterfactual). This can be done for each time point (as in the plots above). We will typically instead aggregate the observed cases and the extrapolated cases over a certain time period–by year, or during the evaluation period (e.g., excluding the first 12 months of post-vaccine data).

**Assumptions** The synthetic controls model relaxes the assumption used in the ITS model that the trends following a linear pattern. However, synthetic controls has its ow asusmptions that need to be met for the results to be valid. 1) The control variables are not influenced by the vaccine. If the control variable is influenced by the target vaccine, then rates of disease for both the outcome and control will decline, and the estimate of the difference between the observed data and the counterfactual will be biased towards seeing no effect. Controls should be screened in advance to identify those that might be influenced by the vaccine

1. The association between the outcome and the controls is stable over time. Or to put it another way, the only factor that changes the relationship between the control disease and the outcome over time is the target vaccine. This is a key assumption. If, we are evaluating the effect of PCvs against pneumonia, and diarrhea is a control variable, this could be a problem if rotavirus vaccine was introduced around the same time. That is because we would have 2 factors that change the relationship between pneumonai and diarrhea: PCV introduction and rotavirus vaccine introduction. In this instance, both diseases would be driven down by their respective vaccines, so the counterfactual would be biased down, and the vaccine effect would be under-estimated. Control variable should be screened in advance to identify those that might be influenced by targeted interventions.

Broad interventions (e.g., social wlefare programs) will affect both the outcome and control disease. In this instance, we *want* to include the control variables that are influenced by the intervention because it will help to adjust for trends in the outcome variable that are unrelated to vaccination.

1. The controls share key causal drivers with the controls and capture relevant variation that is unaffected by the vaccine. If there are trend sin the outcome that are not captured by the controls, the estimates will be biased. This bias could go in either direction. For instance, in many populations, reported pneumonia rates in the elderly have been increasing for unexplained reasons. If the controls cannot capture this increase, one might (incorrectly) conclude that the vaccine causes an increase in pneumonia rates. Likewise, if the outcome disease exhibits wild fluctuations due to unexplained factors, and the controls don’t adjust for this, there will remain a lot of unexplained variability, and it will be difficult to obtain an estimate of vaccine impact

## Defining the study question and objectives (CRISTIANA)

## Data Sources (international and national for both mortality and hospitalization) (Cristiana)

## Feasibility of conducting a time series analysis of secondary data to assess vaccine impact (Concepts) (CRISTIANA)

## Data management (MOSTLY CRISTIANA)

## Data analysis (text + web based tutorial) (DAN)

### Data formatting

To evaluate changes in disease rates associated with an intervention, the first step is to format the data into a **time series**. Time series enumerate the number of cases in a time period (e.g., week, month, quarter, or year). We typically do this by starting with a spreadsheet that has individual-level data, creating a variable that has the date rounded down to the date of the beginning of the nearest week/month/quarter/year, and then adding up the number of cases that occurred during that time period.

Ultimately we want to get a dataset that looks something like this, with a column for the date, and a column for each of the diseases that we want to enumerate.

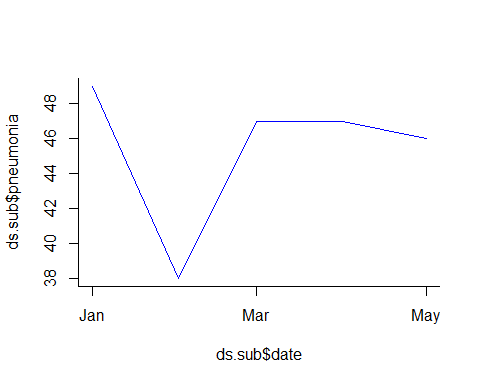
## date pneumonia control1  
## 1 2010-01-01 49 202  
## 2 2010-02-01 55 215  
## 3 2010-03-01 33 210  
## 4 2010-04-01 41 181  
## 5 2010-05-01 39 224  
## 6 2010-06-01 42 167  
## 7 2010-07-01 47 183  
## 8 2010-08-01 46 180  
## 9 2010-09-01 47 211  
## 10 2010-10-01 56 200  
## 11 2010-11-01 54 184  
## 12 2010-12-01 51 221  
## 13 2011-01-01 32 193  
## 14 2011-02-01 58 192  
## 15 2011-03-01 59 163

If you have multiple strata, such as different age groups or regions, the dataset will need to reflect this. In the table below, we have 3 age groups, each with 5 observations. Each date/age group combination should have only 1 row in this dataset, *and all date/age combinations should be represented, even if no cases were observed during a particular time period*.

## age\_group date pneumonia control1  
## 1 1 2010-01-01 49 194  
## 2 1 2010-02-01 38 235  
## 3 1 2010-03-01 47 219  
## 4 1 2010-04-01 47 206  
## 5 1 2010-05-01 46 174  
## 6 2 2010-01-01 50 208  
## 7 2 2010-02-01 50 207  
## 8 2 2010-03-01 50 219  
## 9 2 2010-04-01 45 194  
## 10 2 2010-05-01 43 240  
## 11 3 2010-01-01 56 189  
## 12 3 2010-02-01 47 189  
## 13 3 2010-03-01 47 210  
## 14 3 2010-04-01 46 179  
## 15 3 2010-05-01 50 203

In the *exercise*, we will learn how to take a individual-level dataset that has 1 observation for each hospitalization/death and convert it to a time series dataset using the statistical software R.

TODO: USE EXAMPLE FROM PH SURVEILLANCE CLASS WITH CHILE DATA, FORMATTING DATA< EXPLORING ICD CODES

* Descriptive analysis The most useful thing to do to ensure the data are approriately formatted is to plot the data as a time series line plot, where we have time on the x-axis and number of cases on the y-axis. 
* Time series models
* Synthetic Control analysis
* Incorporating co-variates into the analyses
* Exercises

#### Converting individual-level data to time series

## Interpreting results (DAN)

 ‘dummy’ results that highlight different scenarios (ie wide Cis with a point estimate <<1, wide Cis with a point estimate near 1, tight Cis with a point estimate near 1, tight Cis with a point estimate far from 1) o Communicating and Presenting results effectively (including suggested templates) (DAN AND CRIS)