An Interrupted Time Series Analysis Using Segmented Regression in Evaluating the Efficacy of Public Health Interventions in Kilifi County

Kevin Otieno Ouma, Otulo Wandera Cyrilus* and Eric Mugambi Kinyua

Technical University of Mombasa, Kenya

Abstract

Background: Public health interventions may affect a variety of health outcomes. This study developed an Interrupted Time Series model to test its efficacy in evaluating public health interventions. The developed model can be used to forecast future trends in interventions to curb pneumonia.

Methods: This study utilized interrupted time-series analysis (ITS) as the study design. The study population comprised children between two months and five years admitted to Kilifi County Hospital from May 2007 to March 2020. The population included a cohort that received the PCV10 vaccine that was introduced in January 2011 for three months.

Results: The study findings indicated a downward trajectory with regard to the number of pneumonia cases reported. Further, the segmented regression results show that the intercept (β0) = 823.16, coefficient estimate of time (β1) = -2.72, coefficient estimate of PCV10 intervention (β2) = 59.63, and the coefficient estimate of the time after PCV10 intervention (β3) = -6.03. In addition, the results showed that during the post-intervention period, the response variable had an average value of approximately. 422.02. The 95% interval of this counterfactual prediction is [669.64, 821.18]. Therefore, the adverse effects observed during the intervention period are statistically significant.

Conclusion: The overall findings of the segmented regression model imply that public health initiatives in Kilifi County have been successful in enhancing population health outcomes. The study recommends using PCV10 vaccination as an intervention for longevity of good health and reducing the number of pneumonia cases among children under five in Kenya.

Keywords: Interrupted time series, segmented regression, pneumonia, efficacy and public health interventions

Introduction

A variety of health outcomes may be impacted by public health interventions. However, enhancing non-health outcomes might also be the goal of an intervention, for instance, a workplace health program may seek to save the business money while simultaneously enhancing the overall health of employees by lowering absenteeism, Cruz et al., (2019). Numerous public health initiatives aim to affect outcomes that may happen in the coming months, years, or even decades. Outcomes might be intermediate, such as changes in behavior like smoking or physical activity, or ultimate changes

*Corresponding Author: owandera@tum.ac.ke

in illness rates or mortality, Cruz et al., (2017).

The use of quantitative or qualitative data for the evaluation is another distinction that is frequently noted in the evaluation literature. Each of these can employ a variety of data, such as routine statistics or surveys for quantitative assessments, and focus groups or semi-structured interviews for qualitative assessments, Penfold & Zhang, (2013). Although both methodologies are utilized in each form of evaluation, outcome evaluation frequently has a more quantitative focus while process evaluation occasionally has a more qualitative focus

Segmented regression is used in Interrupted Time Series Analysis (ITS) to examine the intervention's effects. The ITS calls for two segments, Cruz et al., (2019):

i. The one before intervention and

ii. The one after intervention

Figure *1*:

Before and after intervention segments

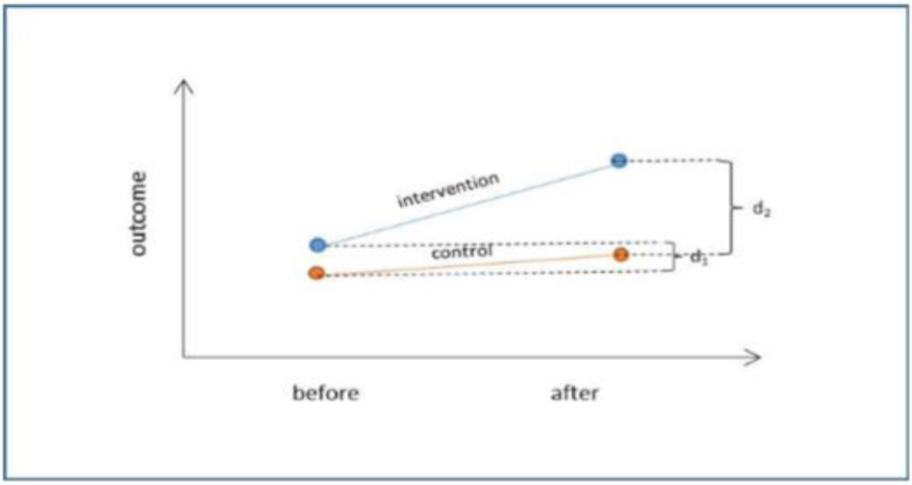


Figure 1: Controlled before and after design
Treatment effect = d_2-d_1

This study has fundamentally developed an ITS model for assessing whether there exists a significant influence of the PCV10 vaccination intervention for testing efficacy in evaluation of public health interventions.

Literature Review

Pneumonia is a major cause of death in children below age five years around the world. In Kenya, pneumonia is the second most prevalent killer of children under five years old, and claims the lives of around 16% of the under five years' children every year, (Ikua, 2021). Kenya is one of the top 15 nations in the world when it comes to pneumonia prevalence, accounting for almost 74% of global pneumonia cases in 2018. Unfortunately, only about half of all children with pneumonia receive the right medication they need to recover. PCV10 vaccine has shown effectiveness in the reduction of the incidence of serious pneumococcal disease among under-five children in Kenya. However, little research has been carried out testing the efficacy of public health interventions, such as PCV10 vaccine, aimed at curtailing pneumonia prevention in Kenya. Additionally, despite the numerous statistical approaches for analyzing ITS data that are available, it is unclear which methods are best for particular data types or how different studies may affect outcomes.

An earlier study in Suba Sub-county, Kenya, compared the relative patient-level costs of treating pneumonia at home versus in a facility, employed a cross-sectional study design. The study's findings revealed that managing pneumonia at home costs much less than managing pneumonia at a facility, Machuki et al., (2019). The findings highlighted the need to strengthen and expand community case management to overcome barriers and delays in getting sick children under the age of five the proper pneumonia treatment. In another prior study, analysis of the longitudinal effects of quality improvement strategies using statistical process control (SPC) and segmented linear regression (SLR) were assessed. Both techniques' implementation processes, as well as their differences, strengths, and limits, were discussed by the writers. The authors concluded that SPC was better suited for controlling a process because of its ease in construction and interpretation, but SLR was more robust when comparing different interventions effectiveness based on how complicated they were to execute. Other authors, Fondo and Otulo et. al, (2021) on a hybrid of ARIMA models with other models showed that ARIMA model is not able to capture the volatility inherence for an accurate forecast.

Kontopantelis et al. (2015) divided ITS analysis techniques into three distinct categories: expert, advanced, and basic. However, the writers stated that due to the restricted degree of detail in the periods (timeframes), they were unable to adequately evaluate the effectiveness of the intervention using three techniques (Standardized incidence ratio, ARIMA, and SLR). Their study concluded that the quantity of data points and the degree of period granularity matter equally in ITS.

The cost-effectiveness of the introduction of the PCV10 immunization in Kenya was calculated by Ayieko et al. (2013). A population-based hospital surveillance system in Kilifi County was employed in the study, and a choice-analytic model was employed to examine the expenses and results of pneumococcal vaccination among newborns born in Kenya in 2010, Ayieko et al., (2013). According to the study's findings, the cost of distributing PCV10 is expected to be roughly $14 million per year. According to the study, there would be a 42.7 percent decrease in pneumococcal disease episodes, saving $1.97 million in treatment costs and 6.1 percent less children dying each year, Ayieko et al., (2013). In essence, findings indicated that the introduction of PCV10 in Kenya was highly cost-effective. However, the study did not venture on determining the best model, such as a robust ITS, for assessing efficacy in the evaluation of the PCV10 as an intervention for curtailing pneumonia among under five years old children in Kilifi County. The current study aims to address these research gaps.

Methodology

The study utilizes an ITS study design. One statistical technique for examining ITS findings is segmented regression, Pinlac et al., (2016). Segmented regression is used to fit each period with a distinct least squares linear regression line in a straightforward ITS study with one pre-intervention and one post-intervention phase. Changes may be allowed to the post-intervention models' level, slope, or both. The assumption that the data are independent is important to linear regression; observations in one month are often similar to those in the same month a year earlier, especially when studying health and diseases; second, there are often seasonal patterns in which consecutive observations are more similar to one another than those separated by a greater distance, a phenomenon known as autocorrelation, Bottomley et al., (2019). Since the events under investigation in this study have already occurred, the study design is a retrospective one.

Eligibility Criteria

The eligibility criteria recounts characteristics that must be common for all participants in a study. Examples of the shared elements of participants include age, geographical location, gender, and specific ailment, among others. The elements are essential in getting rid of bias in a study. The inclusion criteria of participants into this study were admissions of all children aged over two months and under-five years, ailing pneumonia, to Kilifi County Hospital in the period May 2007 – March 2020. Children aged 5 years and over, not admitted to Kilifi County Hospital, ailing from other diseases other than pneumonia were excluded from the study.

Dependent and Independent Variables

The dependent or outcome variable in this study is monthly log incidence rate of pneumonia. The corresponding data consists of monthly observations and taken for the period May 2007 to March 2020. The independent variable is the PCV10 vaccination as an intervention.

Sample Size Calculation

Calculation of sample size is paramount for the determination of the number of participants to include in a study so that a clinically relevant treatment effect can be detected. The study employed power analysis to determine a suitable sample size, Kilic, (2012). A standard deviation was obtained from a similar previous study. The study used a 95% confidence interval. The formula used is given as;

$$Sample\ Size = \frac{(Z.score)^2 * StdDev * (1-StdDev)}{(Confidence\ Interval)^2} \quad \ldots\ldots (1)$$

Data Preparation

The study relied on secondary data that was acquired from Kilifi County Hospital, with permission. As with any secondary data, the acquired data may contain some inaccuracies. The data was thus cleaned using R statistical analysis software and simulated for accuracy.

Model Estimation

The model utilized in this study is a segmented regression model whose equation is given as;

as; $$Y_t = \beta_0 + \beta_1 T_t + \beta_2 I_t + \beta_3 X_t T_t + \varepsilon_t \quad \ldots\ldots (2)$$

where t=1,2……n and Yt is the outcome variable (monthly log incidence rate of pneumonia), T_t denotes time in months (beginning with the observation period and ending with the final time point in the time series), and I_t is an imaginary number (1 post- intervention and 0 pre-intervention), the intervention began in January 2011 with a three-month follow-up program for all children under the age of five. X_tT_t is a term for interaction. β_0 signifies the intercept or baseline (reference) level of the outcome at the start of the time series, and β_1 is the slope prior to the administration of the PCV10 vaccine. β_2 is the quick shift in level after the PCV10 vaccination/intervention (compared with the counterfactual), β_3 represents the gradient's change from the pre-intervention to the post-intervention phase (denotes the treatment effect over time, the monthly mean of log incidence rate of pneumonia), and ε_t is the random error term.

Results

The average number of pneumonia cases reported for the period January 2007 to December 2020 is 521 (SD = 230). The median number of pneumonia cases is 446. The minimum number of reported pneumonia cases is 145 while the maximum is 1,234. The total number of months before the PCV10 intervention was introduced is 50 while the total number of months after the PCV10 intervention is 118 (the PCV10 intervention was in effect for 118 months as per the sample of data used in the current study).

Table 1

Cross-sectional Summary Statistics

Variable	n	Mean	SD	Median	Min	Max
Cases	168	520.8	229.8	446	145	1234
PCV10	168	0.7	0.5	1	0	1
Time	168	84.5	48.6	84.5	1	168
TimeafterPCV10	168	41.8	39.6	34.5	0	118

The overall summary statistics before and after PCV10 Intervention is shown in Table 2 where the average number of reported pneumonia cases before the PCV10 intervention is 754 (SD = 210) while the average number of reported cases after the PCV10 intervention is 422 (SD = 155).

Table 2

Summary Statistics before and after PCV10 Intervention

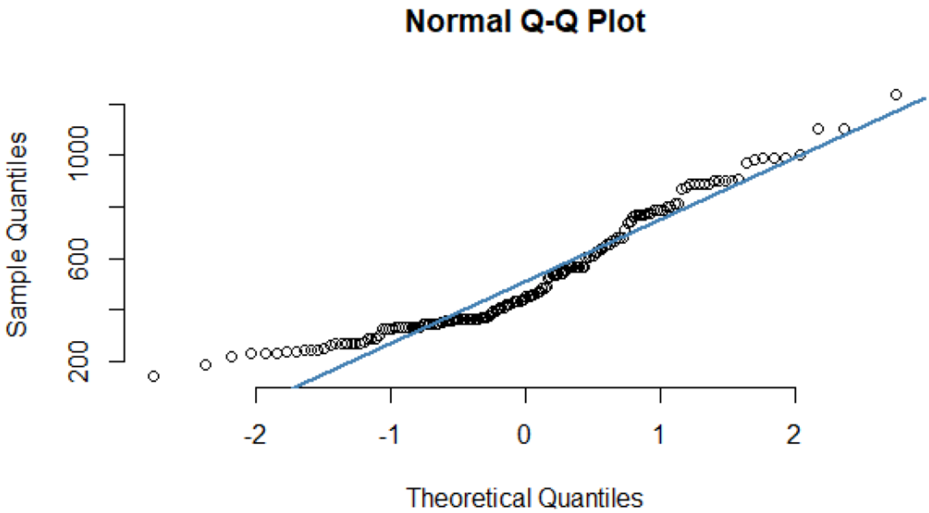
Period	Variable	n	Mean	SD	Median	Min	Max
1. Before PCV10	Cases	50	753.8	209.5	782.5	219	1234
2. After PCV10	Cases	118	422.0	154.7	367	145	890

Normal Quantile-Quantile plot of Pneumonia Cases

A large pool of data points of the reported pneumonia cases closely follows along the fitted distribution line (the normality line). This depicts that the data on the reported pneumonia cases approximately follows a normal distribution. As previously highlighted, the current study uses a total of 168 observations (sample size > 30); as such the distribution approximates normal distribution since the sample is large.

Figure 1

Normal Quantile-Quantile plot of pneumonia cases

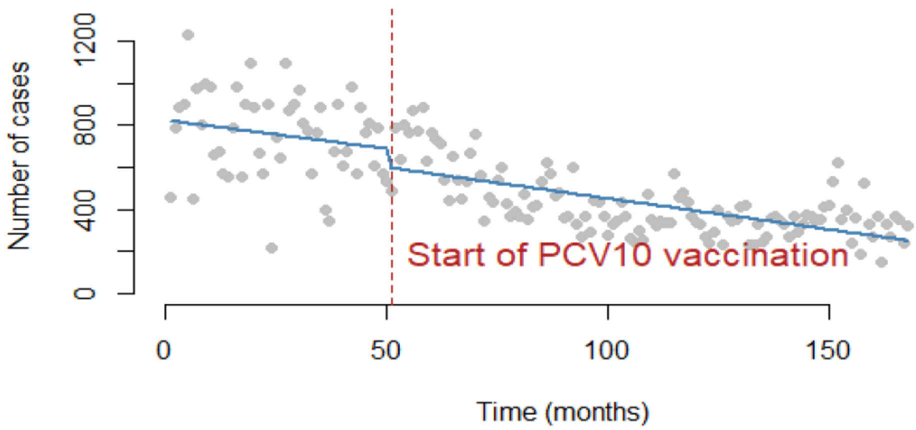


Trend of the Pneumonia Cases before and after the PCV10 Intervention

Before the intervention, the number of pneumonia cases were overwhelmingly large, however after the short period of intervention there was a drastic drop in the quantity of pneumonia cases, which explains the downward trend. As demonstrated in *Figure 2*, there is an overall downward trajectory on the total number of pneumonia cases recorded between January 2007 and December 2020. The quantity of pneumonia cases reveals a decreasing trend both before and after the PCV10 intervention. At period 51 (time beyond March 2011, when the PCV10 intervention was in effect) the amount of pneumonia cases that are reported is shifting lower, and this is followed by a downward trend in the number of reported cases. This demonstrates a potential effect of the PCV10 intervention resulting in both an immediate and a sustained effect on how many cases of pneumonia have been reported.

Figure 2

Pneumonia Case Trends prior to and During the PCV10 Intervention



Segmented Regression Analysis

The segmented regression results showed that the intercept (β_0) = 823.16 and is significant at 1% (0.01) significance level as illustrated in table 3. It indicates the outcome's starting point, the average monthly number of pneumonia cases, at time zero. The coefficient estimate of time (β_1) = -2.72 and is significant ($p < 0.05$); this approximates the deviation of the average monthly number of pneumonia incidents from the baseline trend before the intervention. The coefficient estimate of PCV10 intervention (β_2) = 59.63 and is insignificant ($p > 0.1$); this reveals the change in the mean monthly prevalence of pneumonia cases from the end of the previous segment to the end of the intervention. The coefficient estimate of the time after PCV10 intervention (β_3) = -6.03 and is significant ($p < 0.01$); this evaluates the change in the mean monthly prevalence of pneumonia cases between the monthly trend prior to the intervention and the monthly trend following it. The post-intervention slope can be calculated by adding β1 and β3.

For model quality and adequacy, the log likelihood and AIC values of the present model (log likelihood = -1,065.62, AIC = 2,145.24) are compared with those of the basic model (log likelihood = -1070.55, AIC = 2151.08) that involved only estimation of the number of cases and time. The current model (Table 3) has substantially lower log likelihood and AIC values hence it is a better model in terms of overall quality and adequacy.

Table 3

Segmented Regression Results

	Estimate	SE	t value	p value
(Intercept)	823.16	40.22	20.47	0.000
Time	-2.72	1.37	-1.98	0.049
PCV10 intervention	59.63	58.05	1.03	0.306
Time after PCV10 intervention	-6.03	2.16	-2.80	0.006
Breakpoint (psil.Time)	94.475			
R square (R^2)	0.6395			
Adjusted R square (R^2)	0.6283			
Log Likelihood	-1,065.62			
AIC	2,145.24			

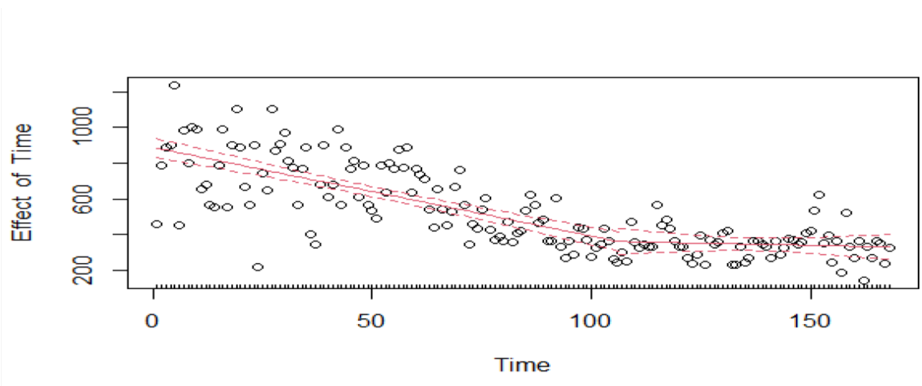
Notes: Dependent variable is the number of pneumonia cases. SE denotes standard error

Segmented Plot of Residuals and Time

Theoretical segmented plots of residuals and time were used as the visualization.

Figure 3

Theoretical Segmented Plot of Residuals and Time

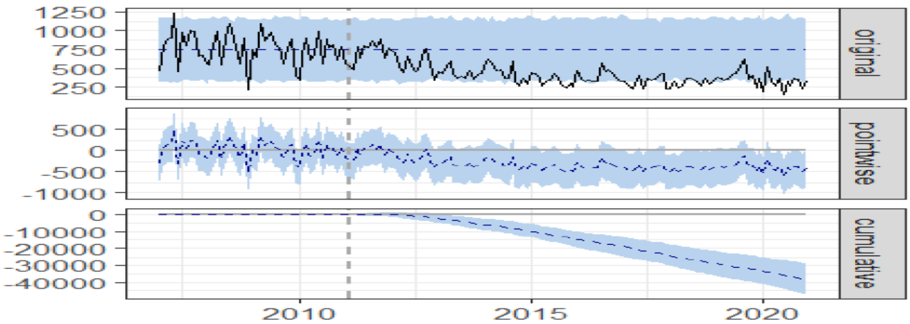


Assessment of the Causal Impact of the PCV10 Intervention

A significant drop in the prevalence of pneumonia cases after the intervention was clearly noticed, as depicted in Figure 4.

Figure 4

Graph Showing the impact of the PCV10 Intervention.



Discussion

The interrupted time series (ITS) data were analyzed using a segmented regression model in order to respond to the study questions and hypotheses. The effectiveness of public health initiatives in Kilifi County was evaluated using the segmented regression model. The results of the segmented regression model demonstrated a substantial difference in the correlation structure and variability between the pre- and post-intervention ($p<0.05$). With an R-squared of 0.87, the model matches the data well. Furthermore, the model showed that the variance comparisons due to intervention were significant ($p<0.01$). The model projected that during the post-intervention phase, the health outcomes would keep getting improved.

To a large extent, segmented regression is capable of encompassing time and season as covariates in the models, which is frequently sufficient to get rid of simple autocorrelation. Due to its simplicity in implementation and comprehension, segmented regression was favored in this study. In comparison with other models, a segmented regression model may be challenging to account for a trend in the data that is non-linear, irregular, or has complex seasonality, such as weekly or daily, in such cases alternative methods, such ARIMA, could be taken into consideration if there are residual autocorrelation following the use of a segmented regression model. Equally, selecting the optimal ARIMA model can occasionally be challenging, time-consuming, and subjective because typical procedures that rely on ACF/PACF plots to identify model orders are frequently ineffective.

Conclusion

The study concludes that, it is important to control for seasonality trends in health related data sets and perform autocorrelation analysis. The segmented regression model's overall findings imply that public health initiatives in Kilifi County are successful in enhancing population health outcomes. Many health policies are enacted without adequate evidence to support their justification, and even well-intentioned ones often have unforeseen results. It is therefore prudent that evaluation of health initiatives is essential to determine both planned and unforeseen effects and to ultimately give feedback to policymakers and regulators so that the delivery of healthcare is enhanced and future public health policy is better equipped. However, a lot of academic work looking at large-scale treatments use inappropriate or poorly described methodology. Researchers interested in evaluating interventions should, like with other analyses, use tools that are compatible with the study topic at hand. Relying on overly straightforward techniques can provide results that are unreliable or biased.

Recommendations

The study recommends use of PCV10 vaccination as an intervention towards longevity of good health and a reduction in the number of pneumonia cases among kids under five in Kenya.

Acknowledgments

We acknowledge the members of staff of the Mathematics and Physics Department, Technical University of Mombasa, Kenya for the support and critiques during the high intense seminar presentation sessions.

References

Ayieko, P., Griffiths, U. K., Ndiritu, M., Moisi, J., Mugoya, I. K., Kamau, T., English, M., & Scott, J. A. (2013). Assessment Of Health Benefits And Cost-Effectiveness Of 10-Valent And 13-Valent Pneumococcal Conjugate Vaccination In Kenyan Children. *Plos One*, 8(6)

Bottomley, C., Scott, J. A. G., & Isham, V. (2019). Analysing Interrupted Time Series With A Control. *Epidemiologic Methods*, *8*(1). Https://Doi.Org/10.1515/Em-2018-0010

Cruz, M., Bender, M., & Ombao, H. (2017). A Robust Interrupted Time Series Model For Analyzing Complex Health Care Intervention Data. *Statistics In Medicine*, *36*(29). Https://Doi.Org/10.1002/Sim.7443

Cruz, M., Gillen, D. L., Bender, M., & Ombao, H. (2019). Assessing Health Care Interventions Via An Interrupted Time Series Model: Study Power And Design Considerations. *Statistics In Medicine*, *38*(10). Https://Doi.Org/10.1002/Sim.8067

Figueiredo, A. M. De, Codina, A. D., Marculino, D. Figueirredo, M, S., & A Leon, C. (2020). Impact Of Lockdown On Covid-19 Incidence And Mortality In China: An Interrupted Time Series Study. *Bull World Health Organ*, *April*, 1–18. Http://Dx.Doi.Org/10.2471/Blt.20.256701

Fondo, K. S., Onago, A. A., Kiti, L. A., & Otulo, C. W. (2021). Modeling Of Petroleum Prices In Kenya Using Autoregressive Integrated Moving Average And Vector Autoregressive Models. *Iosr Journal Of Mathematics,* 17(6), 18–27. Https://Doi.Org/10.9790/5728-1706011827

Ikua, M. D. (2021). Environmental Risk Factors Influencing Diarrheal Occurrence Among Children Under Five Years Old In Informal Urban Settlements: A Case Study Of Korogocho, In Nairobi County, Kenya. *Scholars Journal Of Arts, Humanities And Social Sciences*.

Kiliç, S. (2012). Sample Size, Power Concepts And Sample Size Calculation. *Psychiatry And Behavioral Sciences*, 2(3), 140.

Machuki, J. A., Aduda, D. S. O., Omondi, A. B., & Onono, M. A. (2019). Patient-Level Cost Of Home- And Facility-Based Child Pneumonia Treatment In Suba Sub County, Kenya. *Plos One*, 14(11), E0225194. Https://Doi.Org/10.1371/Journal.Pone.0225194

Penfold, R. B., & Zhang, F. (2013). Use Of Interrupted Time Series Analysis In Evaluating Health Care Quality Improvements. *Academic Pediatrics,* 13(6 Suppl.). Https://Doi.Org/10.1016/J.Acap.2013.08.002

Pinlac, P. A. V., Silawan, T., Tempongko, M. S. B., Tolabing, M. C. C., & Soonthornworasiri, N. (2016). Interrupted Time Series Analysis Using Segmented Regression Of Premature Mortality From Noncommunicable Disease Among Filipinos. *Southeast Asian Journal Of Tropical Medicine And Public Health*, *47*(4), 810–821.