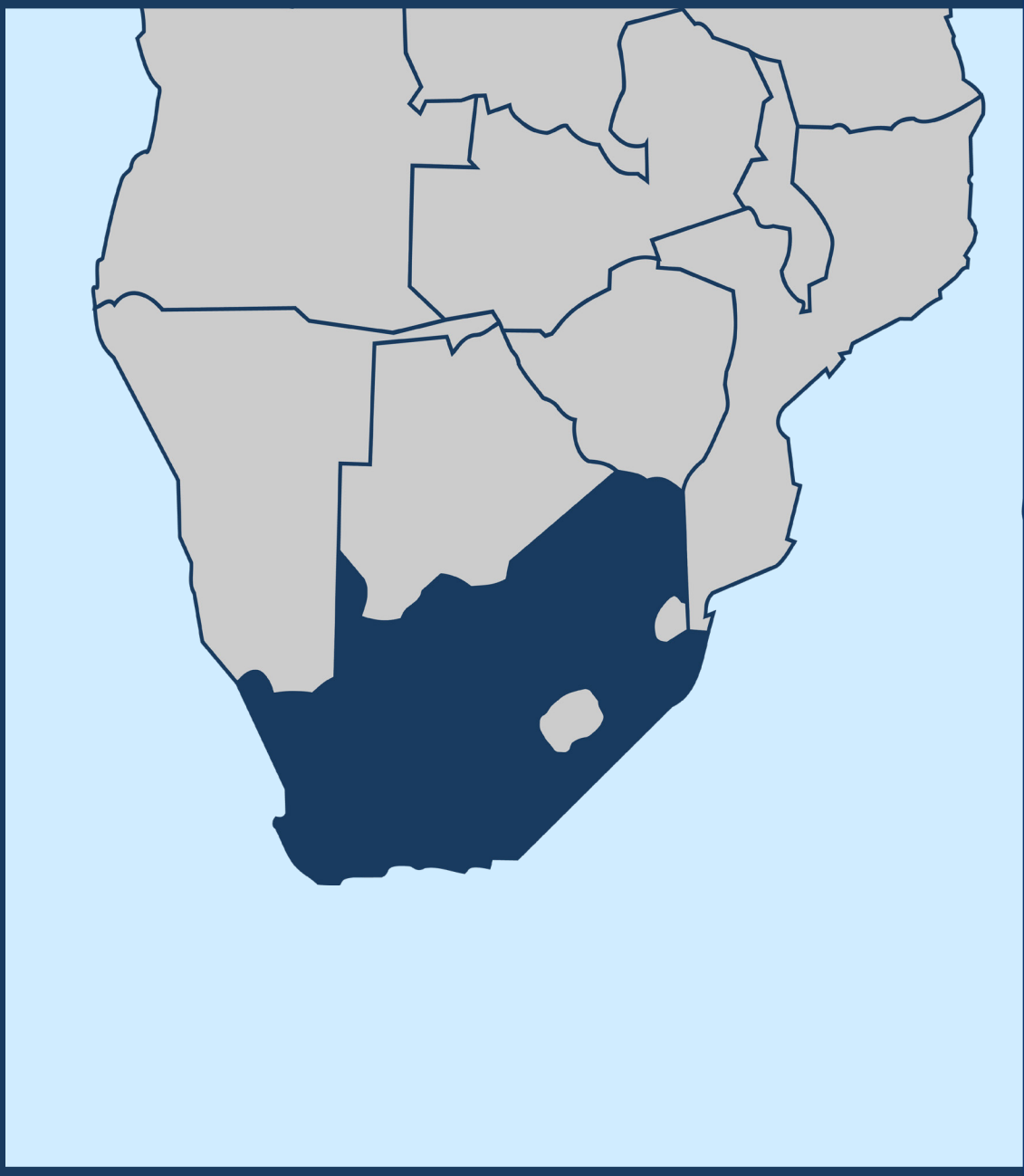
The Pan African Medical Journal

Conducting a Secondary Data Analysis to Estimate the Incidence of Congenital Syphilis in South Africa, 2010 – 2016

Participant Guide



African Case Studies in Public Health

Volume 2

March 2018
Special Issue

EMORY ROLLINS SCHOOL OF PUBLIC HEALTH

AFENET African Field Epidemiology Network
A healthier Africa

PanAfrican Medical Journal

Conducting a Secondary Data Analysis to Estimate the Incidence of Congenital Syphilis in South Africa, 2010 – 2016

Authors: Lazarus Rugare Kuonza[1,2,3], Tuyakula Nakale[4], Sinesia Lucinda Jose Sitao[5], Casey Daniel Hall[6]

[1]South African Field Epidemiology Training Programme, National Institute for Communicable Diseases, National Health Laboratory Services, Johannesburg, South Africa, [2]School of Public Health, University of Witwatersrand, Johannesburg, South Africa [3]School of Health Systems and Public Health, University of Pretoria, South Africa, [4]Namibia Institute of Pathology, Windhoek, Namibia, [5]Mozambican Field Epidemiology and Laboratory Training Programme, [6]Rollins School of Public Health, Emory University, Atlanta, USA

Corresponding author: Lazarus Rugare Kuonza

Email: Lkuonza@gmail.com OR Lazarusk@nicd.ac.za

Abstract

Analysis of existing health data is often used as a cost-effective and time-efficient means to provide evidence to inform important public health decisions. However, the accuracy of the resultant decisions largely depends on the quality of the accessible data, and how the data are processed, analyzed, interpreted and reported. This case study, based on an actual secondary data analysis that was conducted by a trainee of the South African Field Epidemiology Training Programme during April 2017, was designed to provide a classroom simulation of practical considerations that should be taken into account when planning an analysis of a secondary dataset. The case study is ideally suited to reinforce principles already covered in lectures or in background reading assignments.

How to Use the Case Study

General instructions: The case study is suited for a class of 15-20 trainees per session. Trainees should preferably be seated in a U-shaped setup to encourage participation and interaction. The instructor facilitating the session should direct participants to read a paragraph aloud, going around the room to give each participant a chance to read. When a participant reads a question, the instructor may choose to engage the class in large group discussion of the answer, randomly identify a participant to respond to the question, or divide the class into smaller groups for exercises, depending on the type of question. The role of the instructor is largely to coordinate the session such that participants learn from each other, and not just from the instructor.

Audience: Public health practitioners involved in the analysis and interpretation of surveillance data, and others who are interested in the topic

Prerequisites: Before using this case study, case study participants should have received lectures or other instruction on data quality, basic descriptive epidemiology, and analysis of secondary datasets.

Materials needed: Flip chart or white board with markers, graph paper, and calculator

Level of training and associated public health activity: Data management and analysis

Time required: 3 hours

Language: English

Participant Guide

Goal of Case Study: To simulate an analysis of a secondary dataset to inform public health decisions

Learning Objectives: After completion of this case study, the participants should be able to:

1. Identify possible sources of data for a secondary analysis
2. Discuss advantages and disadvantages of using secondary data to inform public health decisions
3. Formulate research questions or objectives to answer public health questions using a secondary data analysis
4. Identify components of an analysis plan prepared to analyze secondary data
5. Explain the significance and outline the process of cleaning data before analysis
6. Calculate and interpret incidence rates
7. Draw and interpret a line graph
8. Calculate and interpret incidence rate ratios
9. Make conclusions based on the findings from a data analysis
10. Discuss how analysis results are disseminated for public health decision making

Introduction

A field epidemiologist in-training has recently been seconded to the Maternal and Child Health Directorate of the South African National Department of Health, as part of her Field Epidemiology Training Program (FETP) residency. Her supervisor informs her that the department is in the process of reviewing the country's progress towards achieving targets that were set by the World Health Organization in 2007, for member states to reduce the incidence of congenital syphilis to below 50 cases per 100,000 live births. (1,2) The supervisor requests the trainee to join a team that has been tasked to compile a report on the most recent statistics of congenital syphilis in the country and make recommendations to the Directorate regarding the country's progress towards achieving the WHO target.

Question 1. What general investigative approaches could be used to obtain estimates of the magnitude of congenital syphilis in the country?

Question 2. What are the advantages and disadvantages of using existing data sources to estimate the magnitude of congenital syphilis in the country?

Question 3. What existing data sources could the trainee epidemiologist consider for analysis to address the assigned task?

Part 1

After discussions with members of the team, the trainee learns that congenital syphilis is a reportable (or notifiable) disease, [6] with health care providers expected to report all such cases to the Department of Health, through the Notifiable Medical Condition (NMC) surveillance system. However, the reporting by health care providers has been inconsistent and incomplete. This rules out the analysis of the available NMC surveillance data as a feasible option to address the task at hand, i.e. to estimate the magnitude of congenital syphilis in the country. The supervisor suggests that the trainee should submit a request to the National Institute of Communicable Diseases (NICD) to access electronic laboratory data collected through the National Health Laboratory Services (NHLS) on specimens collected from infants within the 1st year of age that were submitted for syphilis testing over the past few years. The supervisor also suggests that the trainee develops an analysis plan in preparation for the anticipated data analysis.

Question 4. What are possible advantages and disadvantages of using routine laboratory data to estimate the magnitude of disease (such as congenital syphilis) in the population?

Question 5. Why is it important to prepare an analysis plan before analyzing secondary data?

Question 6. List the logical steps that the trainee should include in the analysis plan.

Question 7. What considerations should the trainee contemplate before submitting a request for the data to the National Health Laboratory Services?

Question 8. What research questions or objectives would you propose to the trainee for the anticipated data analysis, considering the task assigned to the team?

The trainee submitted a request for data to the National Health Laboratory Services, requesting data on all specimens from children less than a year old who were tested for syphilis during the period between 1st January 2010 and 31st December 2016. The laboratory information manager informed the trainee that a total of 89,639 children were tested for syphilis through the National Health Laboratory Services laboratory network across South Africa during the seven years, and 7,381 (8.2%) of them tested positive. The information manager also informed the trainee that the available data had a limited number of variables that could be used for analysis (name, gender, date of birth, date of specimen collection, district and province).

After further consultations with her teammates, the trainee decided to focus on analyzing trends in the incidence rate of early congenital syphilis per 100,000 livebirths over a seven-year period, 2010 to 2016.

Question 9. What is an incidence rate? Are the data available from the National Health Laboratory Services suitable to determine the incidence of congenital syphilis in the South African population?

After receiving the data from the National Health Laboratory Services, the trainee began the process of cleaning the data in preparation for the analysis.

Question 10. What is data cleaning? Why is it important to clean the data before analysis?

Question 11. Discuss the steps or process that the trainee should consider when cleaning the data in preparation for analysis.

Part 2

During the data cleaning process, the trainee deleted 1,024 (13.9%) duplicate records from the initial count of 7,381 syphilis positive cases that were in the dataset obtained from the NHLS, leaving her with a total of 6,357 records for analysis. The first part of her analysis is focused on characterizing the occurrence of congenital syphilis in South Africa over the seven years (2010 to 2016). The trainee has also obtained data on the total number of livebirths that were registered in South Africa during each of the years from 2010 to 2016, from Statistics South Africa. [12] The data are shown in Table 1.

Table 1: Specimens submitted for syphilis testing and early congenital syphilis incidence for South Africa, January 2010 to December 2016

Year	Number of specimens submitted to the laboratory	Number of specimens that tested positive for syphilis	Total number of live births during the year*	Incidence of early congenital syphilis
2010	17,285	1,025	1,022,216	
2011	14,749	1,067	1,028,451	
2012	14,832	1,024	1,026,175	
2013	13,521	982	1,010,116	
2014	12,009	1,132	1,006,495	
2015	12,076	1,037	943,802	
2016	12,282	1,114	943,802	

*Livebirths figures obtained from StatsSA

Question 12. In general, what are the categories of information that are included in a descriptive epidemiological analysis of data?

Question 13. Complete Table 1 by calculating the congenital syphilis incidence rates for each of the seven years per 100,000 live births.

Question 14. Use an appropriate graph to display the congenital syphilis incidence rates for South Africa over the seven-year period. Interpret the graph.

As the trainee continued with the data analysis, she decided to characterize the congenital syphilis incidence rates by province. The analysis showed that the incidence rates of early congenital syphilis have generally been stable in most of the provinces, except for KwaZulu Natal, Gauteng, Free State and Northern Cape which showed marginal increases in incidence during the 7-year period. A colleague suggested that she calculate rate ratios (RR) as a way of comparing the congenital syphilis incidence rates for different provinces. The province level data for the latest year (2016) are presented in Table 2.

Table 2: Number of specimens submitted to the National Health Laboratory Services laboratories for syphilis testing, South Africa 2016.

Name of Province	Total number of live births during the year	Number of specimens that tested positive for syphilis	Incidence of early congenital syphilis	Rate Ratio
WHO target	-	-	50.0	1.0
Limpopo	121,973	40	32.8	
North West	66,254	37	55.8	
Mpumalanga	73,686	46	62.4	
Free State	47,473	40	84.3	
Gauteng	192,439	188	97.7	
Northern Cape	24,310	24	98.7	
Eastern Cape	109,210	149	136.4	
KwaZulu Natal	184,225	279	151.4	
Western Cape	96,626	311	321.9	
South Africa	943,802	1135	118.0	

Question 15. Complete Table 2 by calculating the rate ratios for each of the provinces. [*Hint: Use the WHO elimination target as a reference*]

Question 16. Interpret the rate ratios calculated in Table 2.

Part 3

After completing the data analysis, the trainee started preparing to give feedback to the Maternal and Child Health Directorate of the National Department of Health. She was informed that the directorate had invited internal and external stakeholders to the feedback session, including representatives from other directorates within the Department of Health, as well as representatives from the National Institute for Communicable Diseases, the World Health Organization Country Office, the Centers for Disease Control and Prevention Country Office, and others.

Question 17. As the trainee begins preparing a feedback presentation to the directorate, what type of visual would you recommend she use to display the province-specific incidence rates that are presented in Table 2?

Question 18. What conclusions can the trainee draw from the analysis? Explain your answer.

Question 19. What methods would you suggest to the trainee to disseminate the findings of the analysis to stakeholders?

Question 20. What additional investigations can the trainee recommend to the directorate to further understand the burden of congenital syphilis in the country?

Conclusion

Congenital syphilis is largely a preventable disease targeted for elimination by the World Health Organization. In 2017 a secondary analysis of data routinely collected by the National Health Laboratory Services of South Africa was conducted to estimate trends in the incidence of congenital syphilis in the population during the period 2010-2016.

A report from the analysis was compiled and shared with the National Department of Health. Among the numerous recommendations that came out of the analysis, the report recommended the triangulation of congenital syphilis data with maternal syphilis data for the same period, and the strengthening of syphilis surveillance among pregnant women and their infants.

Although the data had numerous inherent limitations, the data analysis provided the first estimates of the burden of congenital syphilis in the country, and served as a baseline for future evaluations of the country's progress towards achieving international targets to eliminate mother-to-child transmission of syphilis. Furthermore, the analysis illustrated that the utilization of existing data sources provides a viable option in situations where time and resources to conduct primary data collection may be limited, and showed that secondary data analysis has the potential of having an impact on public health policies and decision making.

Background Reading

World Health Organization. The global elimination of congenital syphilis: rationale and strategy for action. Available from: http://www.who.int/reproductivehealth/publications/rtis/9789241595858/en/

Competing Interests

The authors declare no competing interest.

Author's Contributions

The case study was conceptualized by Lazarus Rugare Kuonza. All authors contributed to the design, drafting and critical revision of the content of this case study.

Acknowledgements

We wish to thank the African Field Epidemiology Network and Emory University for organizing a case study development workshop through which this case study was conceptualized and developed. We also acknowledge Ms. Rudzani Mathebula, a resident of the South African Field Epidemiology Training Programme during 2017, for conducting the secondary data analysis on which this case study is based.

References

1. WHO | The global elimination of congenital syphilis: rationale and strategy for action [Internet]. WHO. [cited 2017 May 17]. Available from: http://www.who.int/reproductivehealth/publications/rtis/9789241595858/en/

2. WHO | Eliminating congenital syphilis. A global health priority [Internet]. WHO. [cited 2017 May 21]. Available from: http://www.who.int/reproductivehealth/publications/rtis/cs_advocacy/en/

3. Johnson M. Secondary Data Analysis: A Method of Which the Time has Come. Qual Quant Methods Libr QQML. 2014; 3:619 – 626.

4. Smith AK, Ayanian JZ, Covinsky KE, Landon BE, McCarthy EP, Wee CC, et al. Conducting High-Value Secondary Dataset Analysis: An Introductory Guide and Resources. J Gen Intern Med. 2011 Aug;26(8):920–9.

5. Smith AK, Ayanian JZ, Covinsky KE, Landon BE, McCarthy EP, Wee CC, et al. Conducting High-Value Secondary Dataset Analysis: An Introductory Guide and Resources. J Gen Intern Med. 2011 Aug 26(8):920–9.

6. Notifiable disease. In: Wikipedia [Internet]. 2017. Available from: https://en.wikipedia.org/w/index.php?title=Notifiable_disease&oldid=778640998

7. WHO | Role of the laboratory in surveillance [Internet]. WHO. [cited 2017 May 21]. Available from: http://www.who.int/ihr/lyon/surveillance/lab_surveillance/en/

8. Simpson SH. Creating a Data Analysis Plan: What to Consider When Choosing Statistics for a Study. Can J Hosp Pharm. 2015;68(4):311–7.

9. CDC. Creating an Analysis Plan [Internet]. Centers for Disease Control and Prevention. [cited 2017 May 21]. Available from: https://www.cdc.gov/globalhealth/healthprotection/fetp/training_modules/9/creating-analysis-plan_pw_final_09242013.pdf

10. Data Cleaning | Better Evaluation [Internet]. [cited 2017 May 21]. Available from: http://www.betterevaluation.org/en/evaluation-options/data_cleaning

11. Statistics South Africa. Publication | Statistics South Africa [Internet]. Statistics South Africa - Recorded Live births. [cited 2017 May 22]. Available from: http://www.statssa.gov.za/?page_id=1854

12. CDC. Principles of Epidemiology | Lesson 1 - Section 6 [Internet]. [cited 2017 May 21]. Available from: https://www.cdc.gov/ophss/csels/dsepd/ss1978/lesson1/section6.html