Data Imperfections in Environmental Epidemiology: A Case Study from Ecuador

A Zoom on National Birth-Record Data Imperfections

Presented by **Mahmoud HASHOUSH**

PhD Student in Environmental Epidemiology [second year/four]

Under the supervision of Dr. Emmanuelle CADOT

April 30, 2025

UNIVERSITÉ DE MONTPELLIER

HSM

Institut exposUM UNIVERSITÉ DE MONTPELLIER

La Région Occitanie Pyrénées · Méditerranée

Introduction

What is epidemiology?

"The study of the **distribution** and **determinants** of disease frequency in human populations and the application of this study to control health problems."

→ Heavy reliance on robust **quantitative methods**

Source: Aschengrau, A., & Seage III, G. R. (2020). Essentials of Epidemiology in Public Health. Jones & Bartlett Learning.

Introduction

What is epidemiology?

"The study of the **distribution** and **determinants** of disease frequency in human populations and the application of this study to control health problems."

→ Heavy reliance on robust **quantitative methods**

The objectives of epidemiology

(1) To study the natural course of disease,
(2) To determine the **extent** of disease in a population,
(3) To identify **patterns** and **trends** in disease occurrence,
(4) To identify the causes of disease,
(5) To evaluate the effectiveness of measures that prevent and treat disease.

Source: Aschengrau, A., & Seage III, G. R. (2020). Essentials of Epidemiology in Public Health. Jones & Bartlett Learning.

So, why is it important?

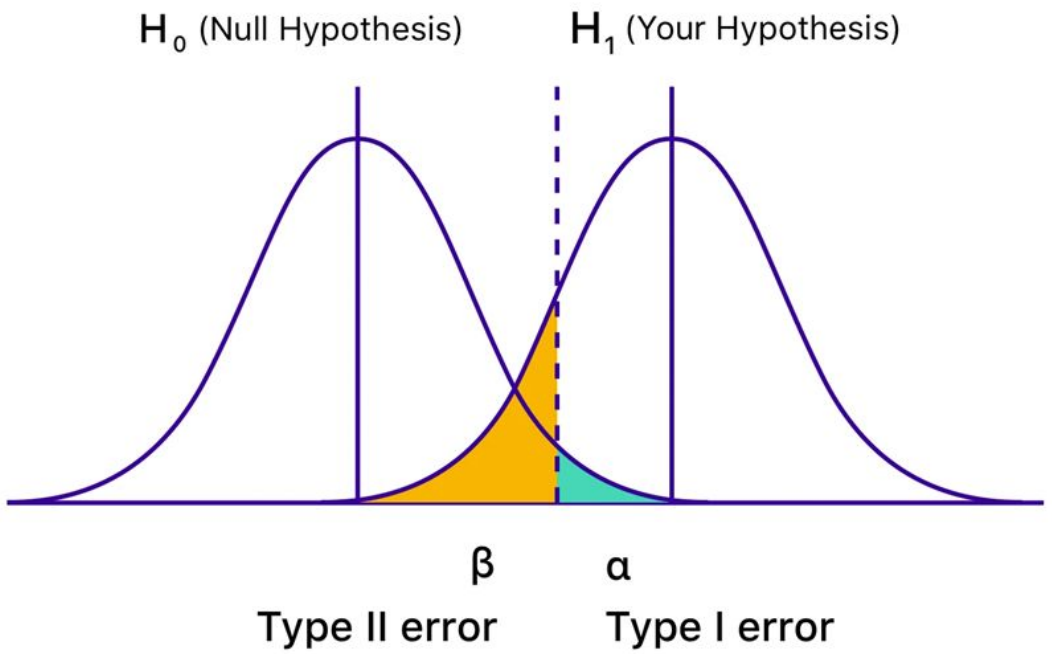
Data imperfections can **compromise the robustness** of the studies, and even **deviate** the results.

Example consequences:

- **hindering the identification of cases**
- **misallocating exposure/controls**

→ Erroneous conclusions

such as type I and type II errors

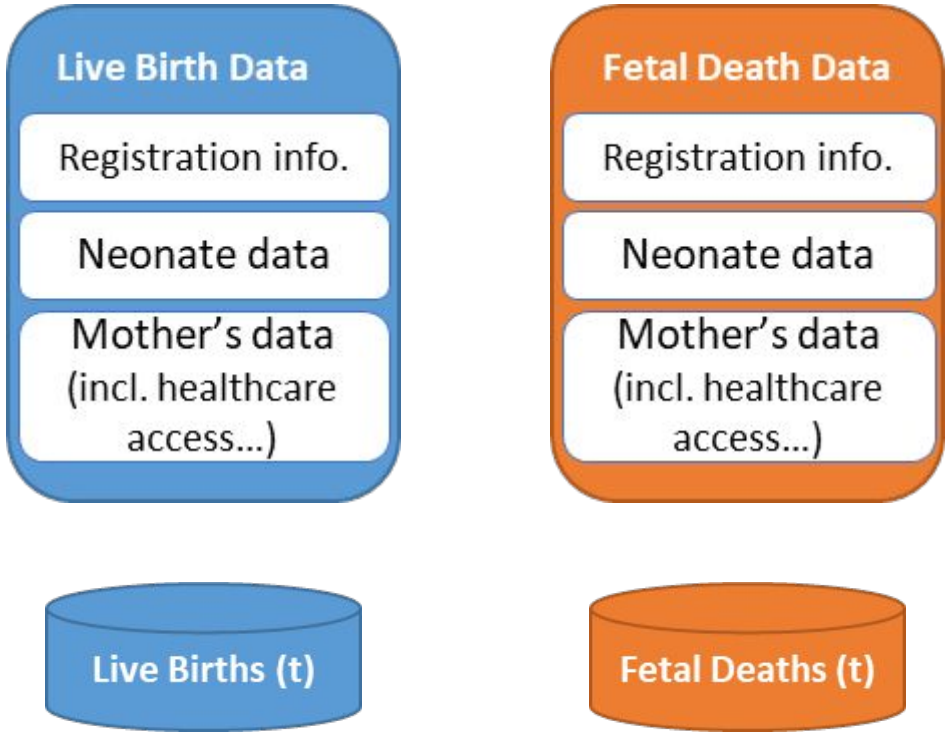


Source: Ahluwalia S. Type 1 and Type 2 errors in A/B testing and how to avoid them [Internet]. Blog. 2025. Available from: https://vwo.com/blog/errors-in-ab-testing/

A **type I error** is a false positive, a rejection of a null hypothesis that is actually true, whereas a **type II error** is a false negative, a failure to reject a null hypothesis that is false.

Birth Record Data

A dataset of more than 3,000,000 cases



Data remarks:

- ∃ various data types.
- The value ranges are massive.
- ∃ variable definitions that vary across the years.
- Spatial and temporal frames are promising.

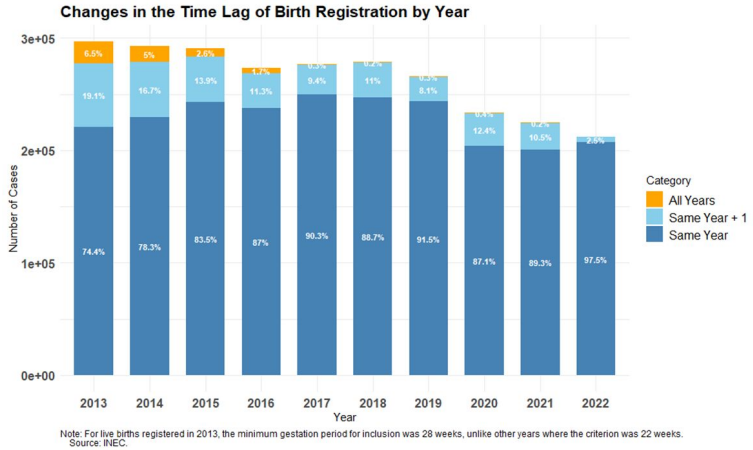
Results

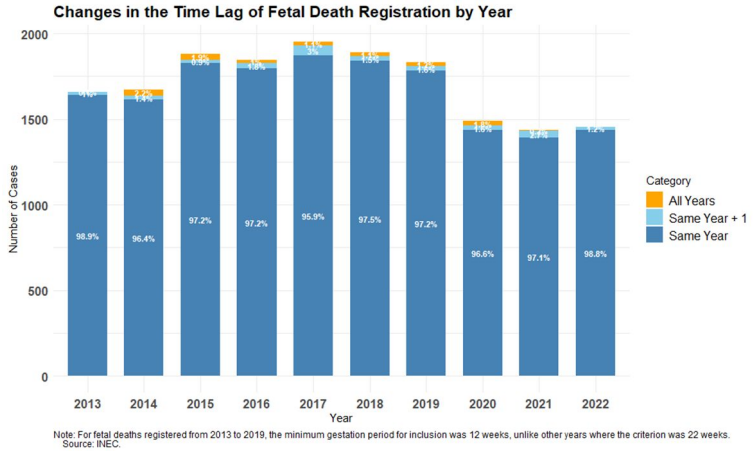
The data that will be discussed:

Ecuadorian birth-record data

Data quality assessment results

Late registration assessment results

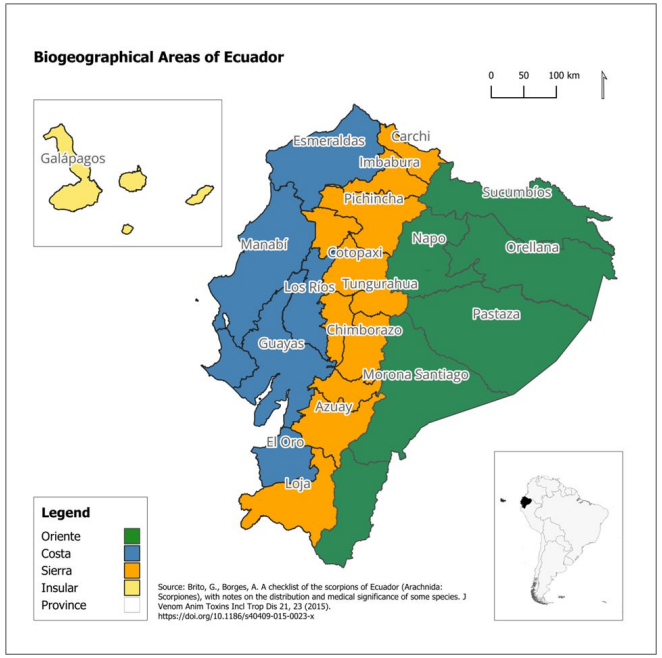


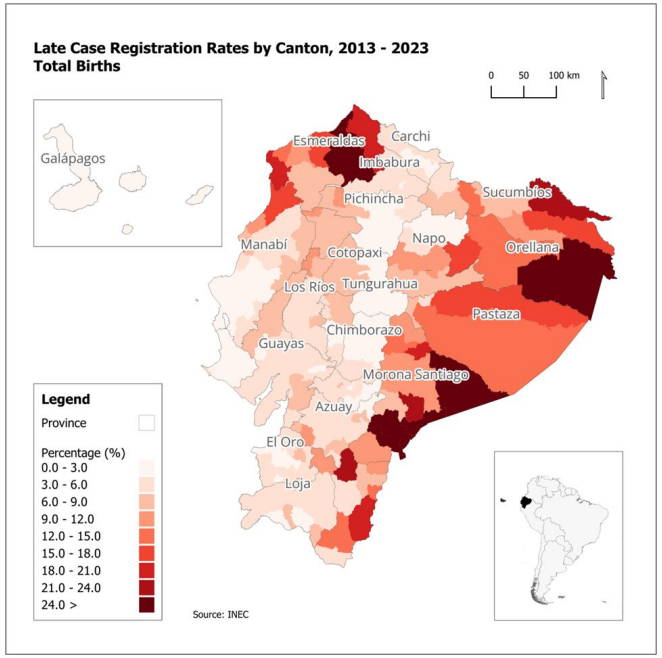


Results

The data that will be discussed:

Ecuadorian birth-record data





Spatial distribution assessment results

Conclusions

•Data imperfections are ubiquitous. They **can** also **occur indirectly**, during the process of data quality improvement.
•Imperfections **can** sometimes **be** tracked and **traced**, suggesting locations of potential improvement.
•The use of **standardized methods** is highly recommended to avoid data inconsistencies.
•**Statistical methods** can be used to mitigate the impacts of data limitations.
•Promoting timely data registration through **policies and/or programs** is of an utmost importance for the completeness of records.

Data Imperfections in Environmental Epidemiology: A Case Study from Ecuador

A Zoom on National Birth-Record Data Imperfections

Presented by **Mahmoud HASHOUSH**

PhD Student in Environmental Epidemiology [second year/four]

Under the supervision of Dr. Emmanuelle CADOT

April 30, 2025

UNIVERSITÉ DE MONTPELLIER

HSM

Institut exposUM UNIVERSITÉ DE MONTPELLIER

La Région Occitanie Pyrénées · Méditerranée