Linkoping University
5th June, 2025

Understanding Sweden'S FHM COVID-19 Data updates.

Olayemi Morrison

Supervisor: Krzysztof Bartoszek

Examiner: Frank Miller

Agenda Overview

01 Background of the study

02 Why Nowcasting?

03 Research Questions

04 Data Description

05 Methodology

06 Implementation

07 Results

08 Conclusion

Background

- The COVID-19 pandemic posed unprecedented challenges for public health authorities, requiring innovative responses to effectively manage the crisis.
- Timely and accurate reporting of COVID-19 data—such as cases, hospitalizations, and fatalities—was essential for informed decision-making and resource allocation.
- However, delays in data collection often resulted in significant discrepancies, leading to underreporting of infections and deaths. This underreporting made it difficult for authorities to accurately assess the pandemic's severity and respond appropriately.
- Epidemiology has increasingly adopted nowcasting, a statistical technique that provides estimates of current conditions based on incomplete data. This method helps public health officials track the true incidence of cases in near real-time, addressing reporting delays.
- In Sweden, the Public Health Agency (Folkhälsomyndigheten, FHM) issues daily and weekly COVID-19 updates, including confirmed cases and hospital data. However, these reports often face delays in data collection, necessitating frequent revisions as retrospective analyses uncover additional cases and outcomes.

Why Nowcasting?

Importance of Adjusting for Delays in Case Reporting

COVID-19 case counts are often underreported in real-time due to delays in test results, administrative processing, and data collection. These delays lead to a distorted view of the actual infection trends, which can mislead policymakers and public health officials. Nowcasting is essential because:

Why Nowcasting?

Importance of Adjusting for Delays in Case Reporting

COVID-19 case counts are often underreported in real-time due to delays in test results, administrative processing, and data collection. These delays lead to a distorted view of the actual infection trends, which can mislead policymakers and public health officials. Nowcasting is essential because:

Delays can vary based on the day of the week, holidays, or changes in reporting policies (e.g., shifting from daily to weekly reports).

Why Nowcasting?

Importance of Adjusting for Delays in Case Reporting

COVID-19 case counts are often underreported in real-time due to delays in test results, administrative processing, and data collection. These delays lead to a distorted view of the actual infection trends, which can mislead policymakers and public health officials. Nowcasting is essential because:

Delays can vary based on the day of the week, holidays, or changes in reporting policies (e.g., shifting from daily to weekly reports).

Without adjustments, real-time data underestimates true infections, potentially leading to late or inadequate public health interventions.

Why Nowcasting?

Importance of Adjusting for Delays in Case Reporting

COVID-19 case counts are often underreported in real-time due to delays in test results, administrative processing, and data collection. These delays lead to a distorted view of the actual infection trends, which can mislead policymakers and public health officials. Nowcasting is essential because:

Delays can vary based on the day of the week, holidays, or changes in reporting policies (e.g., shifting from daily to weekly reports).

Without adjustments, real-time data underestimates true infections, potentially leading to late or inadequate public health interventions.

Nowcasting corrects these distortions by using statistical models to estimate the actual number of cases that occurred on a given date, even before all reports are finalized.

Example:

This table highlights the reporting delay in COVID-19 cases in Stockholm from April 11, 2020, to April 21, 2020, showing reported vs. unreported cases.

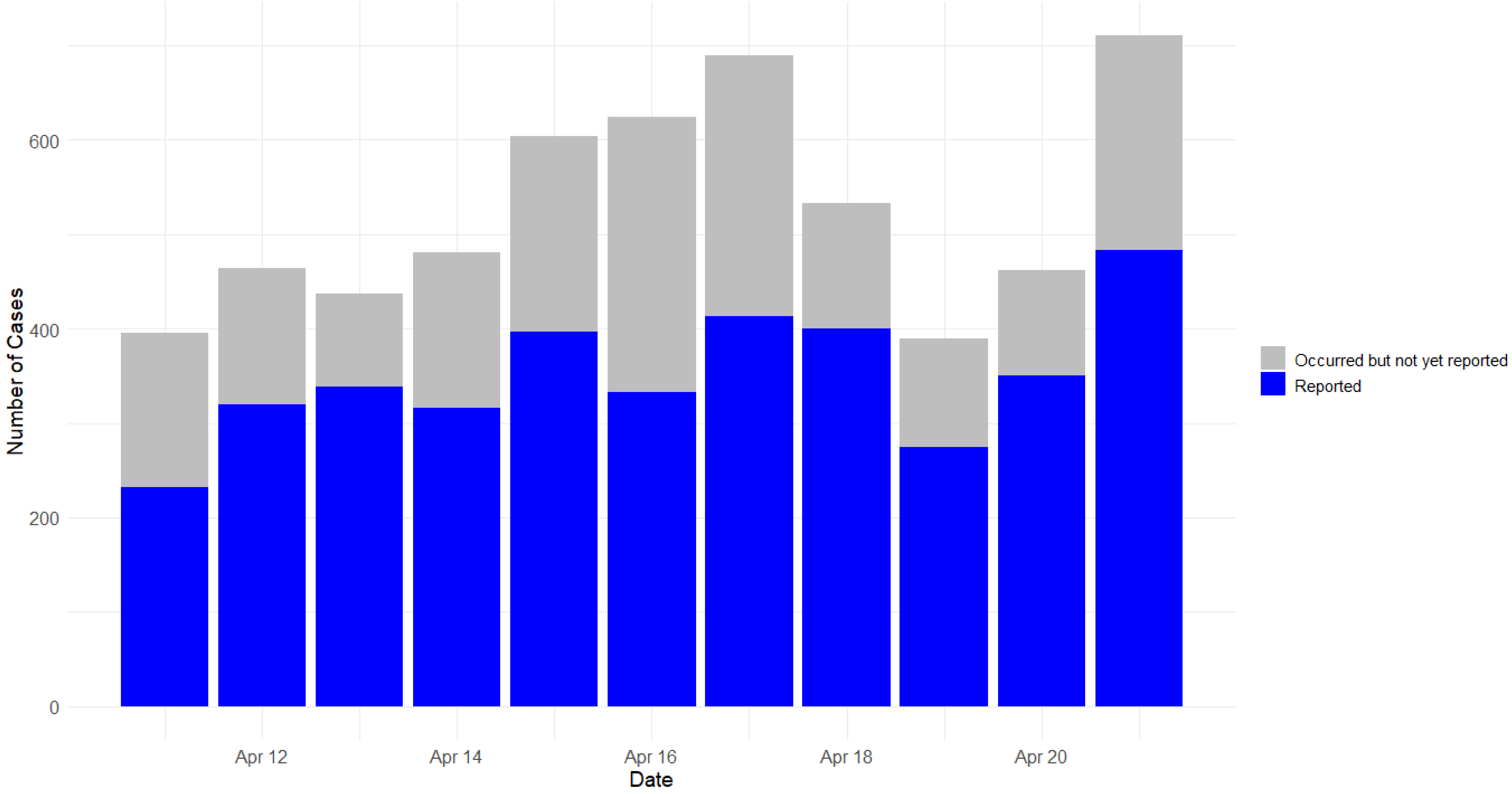
event_date	rep_date	reported_so_far	cases_occurred	cases_unreported
4/11/2020	4/12/2020	232	395	163
4/12/2020	4/13/2020	320	464	144
4/13/2020	4/14/2020	338	437	99
4/14/2020	4/15/2020	316	480	164
4/15/2020	4/16/2020	396	604	208
4/16/2020	4/17/2020	332	624	292
4/17/2020	4/18/2020	413	689	276
4/18/2020	4/19/2020	400	532	132
4/19/2020	4/20/2020	274	389	115
4/20/2020	4/21/2020	350	462	112
4/21/2020	4/22/2020	483	710	227

This table highlights the reporting delay in COVID-19 cases in Stockholm from April 11, 2020, to April 21, 2020, showing reported vs. unreported cases.

Example:

This graphs shows the reporting delay in COVID-19 cases in Stockholm from April 11, 2020, to April 21, 2020, showing reported vs. unreported cases. The blue bars represent cases reported by the Swedish Public Health Agency as of each reporting date, while the pink bars indicate additional cases that had occurred but were not yet reported at the time.

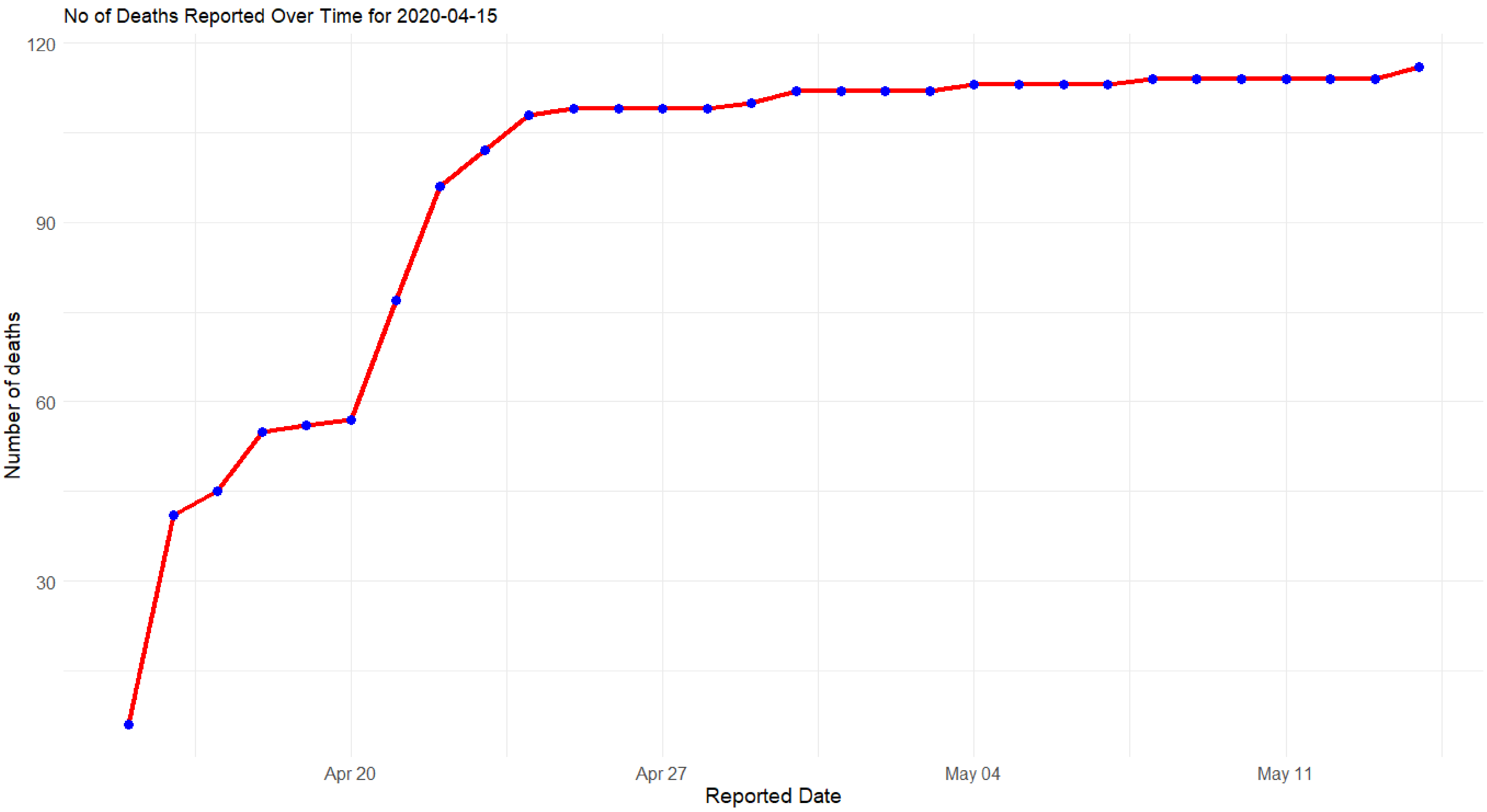
COVID-19 Reported vs. Unreported Cases



Example:

This plot visualizes how the number of reported deaths evolved over time for cases corresponding to 2020-04-15, highlighting fluctuations in delayed reporting.

death_date	N	rep_date
4/15/2020	6	4/15/2020
4/15/2020	41	4/16/2020
4/15/2020	45	4/17/2020
4/15/2020	55	4/18/2020
4/15/2020	56	4/19/2020
4/15/2020	57	4/20/2020
4/15/2020	77	4/21/2020
4/15/2020	96	4/22/2020
4/15/2020	110	4/29/2020
4/15/2020	113	5/7/2020
4/15/2020	114	5/8/2020
4/15/2020	114	5/9/2020
4/15/2020	114	5/10/2020
4/15/2020	114	5/11/2020
4/15/2020	114	5/12/2020
4/15/2020	114	5/13/2020
4/15/2020	116	5/14/2020



Research Questions

The primary objective of this study is to assess the quality of Sweden's COVID-19 case reporting over time and develop a Bayesian nowcasting framework to correct for reporting delays. Specifically, the study aims to:

Research Questions

The primary objective of this study is to assess the quality of Sweden's COVID-19 case reporting over time and develop a Bayesian nowcasting framework to correct for reporting delays. Specifically, the study aims to:

Evaluate reporting accuracy trends throughout the pandemic by analyzing how data quality has changed over time. This includes assessing whether early pandemic data (e.g., 2020) were less reliable than later periods (e.g., 2022)

Research Questions

The primary objective of this study is to assess the quality of Sweden's COVID-19 case reporting over time and develop a Bayesian nowcasting framework to correct for reporting delays. Specifically, the study aims to:

- Evaluate reporting accuracy trends throughout the pandemic by analyzing how data quality has changed over time. This includes assessing whether early pandemic data (e.g., 2020) were less reliable than later periods (e.g., 2022)
- Determine the reliability of real-time data for nowcasting by identifying periods where reporting delays were minimal and assessing when data quality deteriorated, making nowcasting less reliable.

Research Questions

The primary objective of this study is to assess the quality of Sweden's COVID-19 case reporting over time and develop a Bayesian nowcasting framework to correct for reporting delays. Specifically, the study aims to:

Evaluate reporting accuracy trends throughout the pandemic by analyzing how data quality has changed over time. This includes assessing whether early pandemic data (e.g., 2020) were less reliable than later periods (e.g., 2022)

Determine the reliability of real-time data for nowcasting by identifying periods where reporting delays were minimal and assessing when data quality deteriorated, making nowcasting less reliable.

Analyze regional and demographic variations in reporting quality by identifying which regions demonstrated consistently accurate or poor reporting practices and evaluating disparities in data completeness across regions.

Data Description

Data Source:

The dataset used in this study originates from the Public Health Agency of Sweden (Folkhälsomyndigheten, FHM), which is responsible for monitoring and reporting COVID-19 statistics across Sweden. The dataset provides a comprehensive view of the pandemic's progression, covering the period from January 2020 to December 2024.
It includes records of new hospital admissions due to COVID-19, differentiating between general hospitalizations and ICU treatment, offering insights into case severity.

Mortality data is also featured, with daily and weekly death counts at national and regional levels, alongside cumulative totals for long-term trends. Epidemiological metrics include weekly incidence rates per 100,000 inhabitants and 14-day incidence rates, which reflect broader infection trends. Additionally, cumulative case counts for infections, ICU admissions, and deaths are provided.

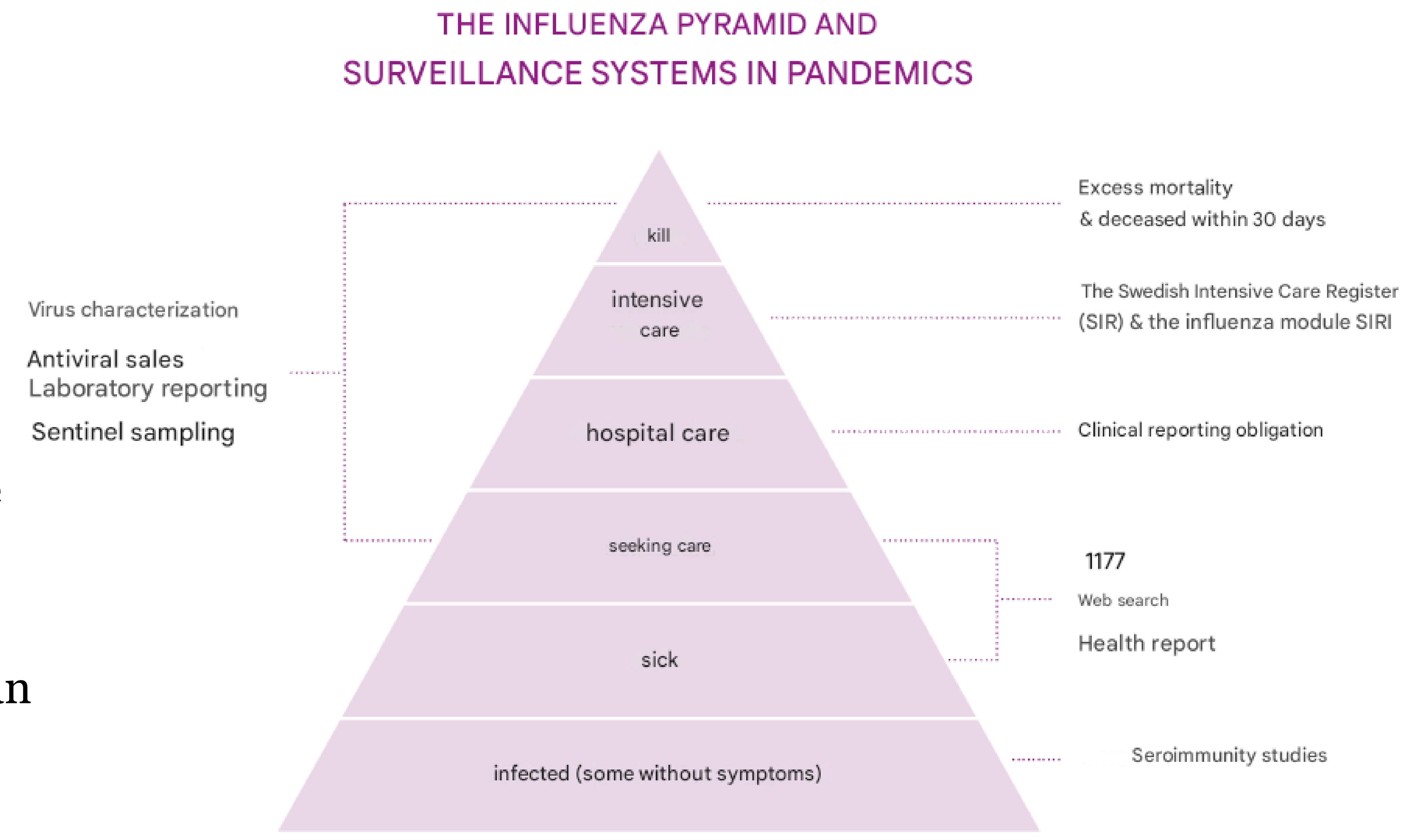
Data Description

The Public Health Agency of Sweden uses a number of different surveillance systems to monitor the spread of COVID-19 in Sweden.

The purpose of surveillance during a pandemic is to:

- Ensure that collected data provides a basis for decisions on measures that minimize mortality, morbidity and negative effects on society
- Monitor and evaluate the effects of measures taken
- Make it possible to follow and to some extent predict the development of the pandemic.

The figure comes from the pandemic preparedness document "Pandemic Preparedness – How societal actors can prepare – a knowledge support for preparedness planning", published in December 2019, and presents the systems that were intended to be used during a pandemic. Below is a description of the indicators followed and the data sources used in the surveillance of COVID-19.



Raw Data:

Variable	**Description**
år (Year)	The calendar year of the reported data.
veckonummer (Week Number)	The week of the year (1–52).
Antal_fall_vecka (Weekly Cases)	The number of new COVID-19 cases reported in a given week.
Antal_fall_100000inv_vecka (Weekly Cases per 100,000 Inhabitants)	Weekly cases normalized by population size.
Antal_fall_100000inv_14dagar (14-Day Incidence Rate per 100,000)	The cumulative 14-day incidence rate, a common metric for epidemiological trends.
Kum_antal_fall (Cumulative Cases)	The cumulative total of reported cases up to and including the given week.
Antal_nyaintensivvårdade_vecka (Weekly Intensive Care Admissions)	The number of new admissions to intensive care units.
Kum_antal_intensivvårdade (Cumulative ICU Admissions)	The cumulative total of ICU admissions.
Antal_avlidna_vecka (Weekly Deaths)	The number of COVID-19-related deaths reported during the week.
Kum_antal_avlidna (Cumulative Deaths)	The cumulative total of COVID-19-related deaths.

Key Datasets and variables:

All files have been converted to .csv format for uniformity. Each dataset captures a different aspect of case reporting dynamics, contributing to a comprehensive understanding of how delays and underreporting affect real-time surveillance. The processed folder contains the following:

File Name	**Description**	**Key Variables**
acov19DAG.csv	Daily COVID-19 cases.	Region, Day, Cases per day
bcov19Kom.csv	Weekly data for each municipality.	Municipality, Indicator, Year and week, Cases by municipality and week
ccov19Reg.csv	Weekly confirmed cases per region.	Region, Indicator, Year and week, Confirmed cases
ccov19kon.csv	Weekly cases by gender and region.	Region, Indicator, Gender, Year and week, Cases by gender and region
dcov19ald.csv	Weekly cases by age group.	Indicator, Age group, Year and week, Cases by age group
xcov19ivavDAG.csv	Daily deaths and ICU admissions.	Indicator, Day, Intensive care and deceased per day
ycov19ivavald.csv	ICU and deaths by age group (weekly).	Indicator, Age group, Year and week, ICU cases and deaths
ycov19ivavkon.csv	ICU and deaths by gender (weekly).	Indicator, Gender, Year and week, ICU cases and deaths
ecov19sabo.csv	Cases among individuals 65+ receiving social services (weekly).	Region, Category, Year and week, Cases among 65+

Methodology

The initial stage of nowcasting involves the creation of a reporting triangle, which visually illustrates the connection between event dates—when incidents or cases actually occur—and reporting dates—when these cases are documented in the database. This relationship is crucial for understanding the flow of information over time. To quantify the gap between these two dates, we define the reporting delay (d), which captures the time it takes for a case to be recorded after it has occurred. This delay is an important factor in accurately assessing and forecasting trends in case data.

Day	$d = 0$	$d = 1$	$d = 2$	$d = 3$	$d = 4$
$t - d_{\max}$	$n_{t-d_{\max},0}$	$n_{t-d_{\max},1}$	$n_{t-d_{\max},2}$	$n_{t-d_{\max},4}$	$n_{t-d_{\max},5}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$t - 5$	$n_{t-5,0}$	$n_{t-5,1}$	$n_{t-5,2}$	$n_{t-5,3}$	$n_{t-5,4}$
$t - 4$	$n_{t-4,0}$	$n_{t-4,1}$	$n_{t-4,2}$	$n_{t-4,3}$	$n_{t-4,4}$
$t - 3$	$n_{t-3,0}$	$n_{t-3,1}$	$n_{t-3,2}$	$n_{t-3,3}$	
$t - 2$	$n_{t-2,0}$	$n_{t-2,1}$	$n_{t-2,2}$		
$t - 1$	$n_{t-1,0}$	$n_{t-1,1}$			
t	$n_{t,0}$				

Methodology

In the context of daily COVID-19 reporting, a maximum reporting delay, denoted as D, is established to define the upper limit on how long it may take for cases to be reported. The dataset is organized in the form of a matrix, referred to as n, where each row corresponds to the number of confirmed cases for a specific day. The columns of this matrix represent the various reporting delays, showcasing how case counts may vary based on when they are officially recorded.

Nt,d=cases reported on day t with delay d

This method aims to infer the total number of events on a given day t based on the Nt information available at a later day . Since case reporting is delayed, the reported $T>t$ count on a given day underestimates the true number of infections. The observed and unobserved case counts can be formulated as:

$$N_t = \sum_{d=0}^{T-t} n_{t,d} + \sum_{d=T-t+1}^{D} n_{t,d}$$

Methodology

where:

- $n_{t,d}$ is the number of events occurring on day t but reported with a delay of d days.
- The first sum represents observed deaths, while the second sum accounts for unreported deaths.

Bayesian Model Components

Epidemic Curve (Latent Process Model)

The epidemic curve (epi curve) is a graph showing the number of new COVID-19 cases over time. It helps us see the rise, peak, and decline of an outbreak. It helps us predict missing or delayed data by looking at trends.

We model the expected number of new cases using a log-Gaussian process:

$$\log(\lambda_t) \sim N(\log(\lambda_{t-1}), \sigma^2)$$

where:

- $\lambda_t = E[N_t]$ is the expected number of cases at time t.
- This follows a Random Walk assumption, ensuring smooth epidemic progression.

Methodology

Bayesian Model Components

Delay Distribution (Reporting Model)

The delay distribution describes how long it takes for COVID-19 cases to be reported after they actually happened.

The delay process is modeled as a hazard function:

$$Pt, d = P(delay = d | infection.\, at.\, t)$$

Following Günther et al. (2020), we assume a discrete-time hazard function:

$$ht, d = Lt, d + Wt, d$$

where:

- ht,d is the reporting hazard at delay d.
- Lt,d captures structured delay effects (i.e, weekday effects).
- Wt,d represents residual variability.

Cases reported with delay d follow a negative binomial distribution:

$$n_{t,d} | \lambda_t, P_{t,d} \sim NB(\lambda_t P_{t,d}, \phi)$$

where ϕ is an overdispersion parameter accounting for variability in reporting delays.

Methodology

Hierarchical Bayesian Nowcasting Model

<u>Latent fatalities (or true death counts)</u>

Let $\lambda_{r,t}$ be the expected number of true new cases in region r on day t.
We model this as:

$$\log \lambda_{r,t} = \alpha_r + \beta_r X_{r,t} + \epsilon_{r,t}$$

Where:

- α_r: region-specific baseline incidence (intercept)
- β_r: region-specific effects (e.g. weekday/holiday effects)
- $X_{r,t}$: covariates (like day-of-week, public holiday, testing intensity, etc.)
- $\epsilon_{r,t} \sim N(0, \sigma_\epsilon^2)$: Random Noise.

Methodology

Hierarchical Bayesian Nowcasting Model

Observed reports (with delay)

Let D be the reporting delay (e.g. 0 to 30 days), and Pr,d,t be the probability of a delay d for a case in region r at time t.

$$y_{r,t,d} \sim NB\left(\lambda_{r,t-d} \cdot p_{r,d,t}, \phi\right)$$

Where:

- Yr,t,d is the number of cases reported on day t that actually occurred on day t−d in region r
- Pr,d,t can be modeled using a delay distribution (e.g., a Log-normal structure across delays)
- ϕ is the overdispersion parameter

Methodology

Evaluation Metrics:

Metric		Role	Sensitivity	Interpretation
Absolute Error	Absolute difference between the median prediction and the retrospectively reported true value.	Point estimate accuracy	High to large errors	Straightforward but ignores uncertainty
Log Score	Log-likelihood of the true value against the predicted distribution, with lower values indicating better predictive accuracy.	Distribution quality	Very high	Penalizes when true values are in low-density regions
CRPS	evaluates the accuracy of the entire predicted distribution by accounting for both calibration and sharpness.	Distribution sharpness + calibration	Moderate	Good all-around distributional metric
RMSE	Evaluates the deviation of the predicted medians from the true values.	Point estimate deviation	High to outliers	Sensitive to large deviations

Implementation & Inference

A custom function is implemented in R to construct the reporting triangle, which is essential for modeling the relationship between the event date (when a case occurs) and the reporting date (when it is officially reported).

This function begins by extracting the relevant event dates and reporting dates from the structured dataset. Using these, it computes the reporting delay, denoted as d, for each observation. These delays are then used to populate a two-dimensional matrix y, where rows represent event days and columns represent reporting delays up to a maximum delay DT. This matrix forms the foundation of the likelihood component in the Bayesian model.

Additionally, the function includes adjustments for systematic temporal effects, such as weekday-specific reporting patterns and public holidays, which can introduce bias in the observed delay structure if left unaccounted for.

Algorithm 1: Data Preparation For Reporting Triangle And Design Matrices

01 Initialization

Define the nowcasting period from start date to current date t_0:2
T ← length of t_0:2

Initialize matrix n ∈ ℕ^(T × T) with NA values to store case counts by event date and delay.

02 Compute delays and populate the reporting triangle

Add column delay to D, where:
delay = report_date - event_date

For each t in {1, ..., T}:
Subset D_t ← {x ∈ D : x.event_date = t_0:2[t]}
For d = 0 to T - t:
n[t, d+1] ← count of observations in D_t with delay d

03 Truncate and correct long delays

n_long ← row sums of n[:, D_max+2:T]
If any n_long > 0:
Add n_long to column D_max+1 of n
Truncate n to columns 1 through D_max+1
Replace NA entries in n with 0

04 Build weekday design matrix W_wd

Define changepoints every 2 weeks back from now to start
Let wdays ← {4, 5, 6} (Wed, Thu, Fri)
Initialize W_wd ∈ ℕ^(T × (D_max+1) × |wdays|)
For each t = 1, ..., T:
For each w ∈ wdays:
For d = 0 to D_max:
W_wd[t, d+1, w] ← I[wday(t_0:2[t] + d) = w]

Algorithm 1: Data Preparation For Reporting Triangle And Design Matrices

05 Build changepoint matrix W_cp

Define changepoints every 2 weeks back from now to start
Initialize W_cp ∈ ℕ^(T × (D_max+1) × C)
For each t = 1, ..., T:
 For each changepoint c_i:
  For d = 0 to D_max:
   W_cp[t, d+1, i] ← I[t_0:2[t] + d > c_i ∧ ≤ c_i + 14]

06 Combine design matrices

W_wd+cp ← abind(W_wd, W_cp, along = 3)

07 Build holiday/weekend indicator Z

Initialize Z ∈ ℕ^(T × (D_max+1))
For each t = 1, ..., T:
 For d = 0 to D_max:
  date ← t_0:2[t] + d
  Z[t, d+1] ← I[wday(date) ∈ {1, 2, 7} ∨ date ∈ holidays]
Return list containing:
{T, D_max, |wdays|, n, W_wd, W_wd+cp, t_0:2, Z}

Algorithm 2: Evaluate Nowcasting

Once the reporting triangle has been constructed, the Bayesian model is compiled and executed using Stan via the RStan interface. This model execution phase involves sampling from the posterior distributions of the key parameters of interest.

Specifically, the model estimates the total number of incident cases Nt for each day t, taking into account the delay structure and incomplete reporting.

It also infers the delay-adjusted reporting probabilities Pd for each delay day d, which describe the likelihood that a case is reported d days after its occurrence. Furthermore, the model incorporates the effects of leading indicators—such as ICU admissions or other correlated data streams—to improve the precision and reliability of the nowcast estimates.

Algorithm 2: Evaluate Nowcasting

Require: *Dataset dat, Stan model model, Current date now, maximum delay Dmax (default 35) Samples from the posterior and saved results*

```
Start timer: start time ← Sys.time()
Parse date: now date ← formatted string from now
Load data: ICU data and reported case data from Excel files using now date
dat_mod ← dat filtered to include reports ≤ now, expanded by count
Set estimation window: start ← now − 8 weeks + 1
Preprocess time series:
    Merge ICU and case data on date
    Create 7-day rolling means and lagged indicators
    Replace NAs with 0
    Filter dates between start and now
Prepare Stan input: prep dat list ← prepare data list() with inputs
dat mod, now, start, and Dmax
Load Stan model from file based on model name using:
    T , D, r, Z from prep dat list
    lead_ind from preprocessed time series
    W_wd as reshaped weekday-holiday hazard matrix
    Replace NA entries in n with 0
    α as vector of ones
    seed = 1142, chains = 4, iter = 2000
    control: adapt delta = 0.98, max treedepth = 15
    save res ← {N, p, logLambda}
```

Evaluation & Results

Algorithm 3: Evaluate Model Accuracy and Interval Coverage

Require: *List of posterior samples N list, maximum delay m delay, reporting dates rep date List of RMSE, log score, CRPS, and PI coverage matrices*

Initialize matrices res_rmse, res_logs, res_crps, pi_75, pi_90, pi_95

For each reporting date i in N_list:

For each delay j from 1 to m_delay + 1:

Extract the forecast vector v from column (56 + 1 - j) in N_list[[i]]

Get the true observed count truth from n_true_retro on date rep_date[i] - j + 1

Compute the following evaluation metrics:

RMSE: $\sqrt{(median(v) - truth)^2}$

Log Score: *logs_sample(y = truth, dat = v)*

CRPS: *crps_sample(y = truth, dat = v)*

Prediction Interval (PI) Coverage:

75% PI: check if truth is within the 12.5th and 87.5th percentiles of v

90% PI: check if truth is within the 5th and 95th percentiles

95% PI: check if truth is within the 2.5th and 97.5th percentiles

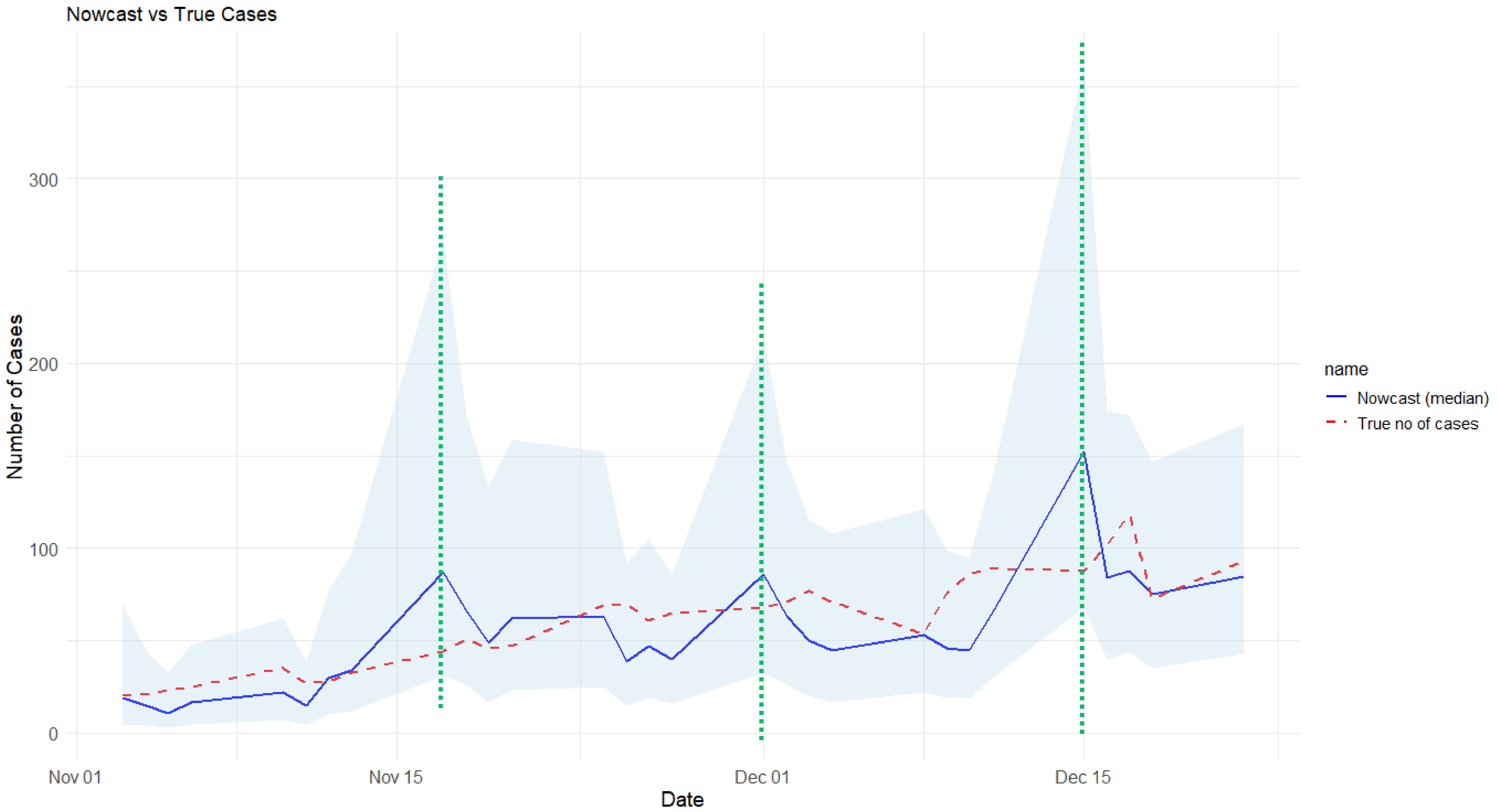
Return: list(res_rmse, res_logs, res_crps, pi_75, pi_90, pi_95)

To evaluate the predictive performance of the nowcasting models, we developed the calculate_scores() function to quantify accuracy and uncertainty metrics based on reporting delays and specific dates. This function compares posterior predictive samples from the models to observed values, assessing model performance.

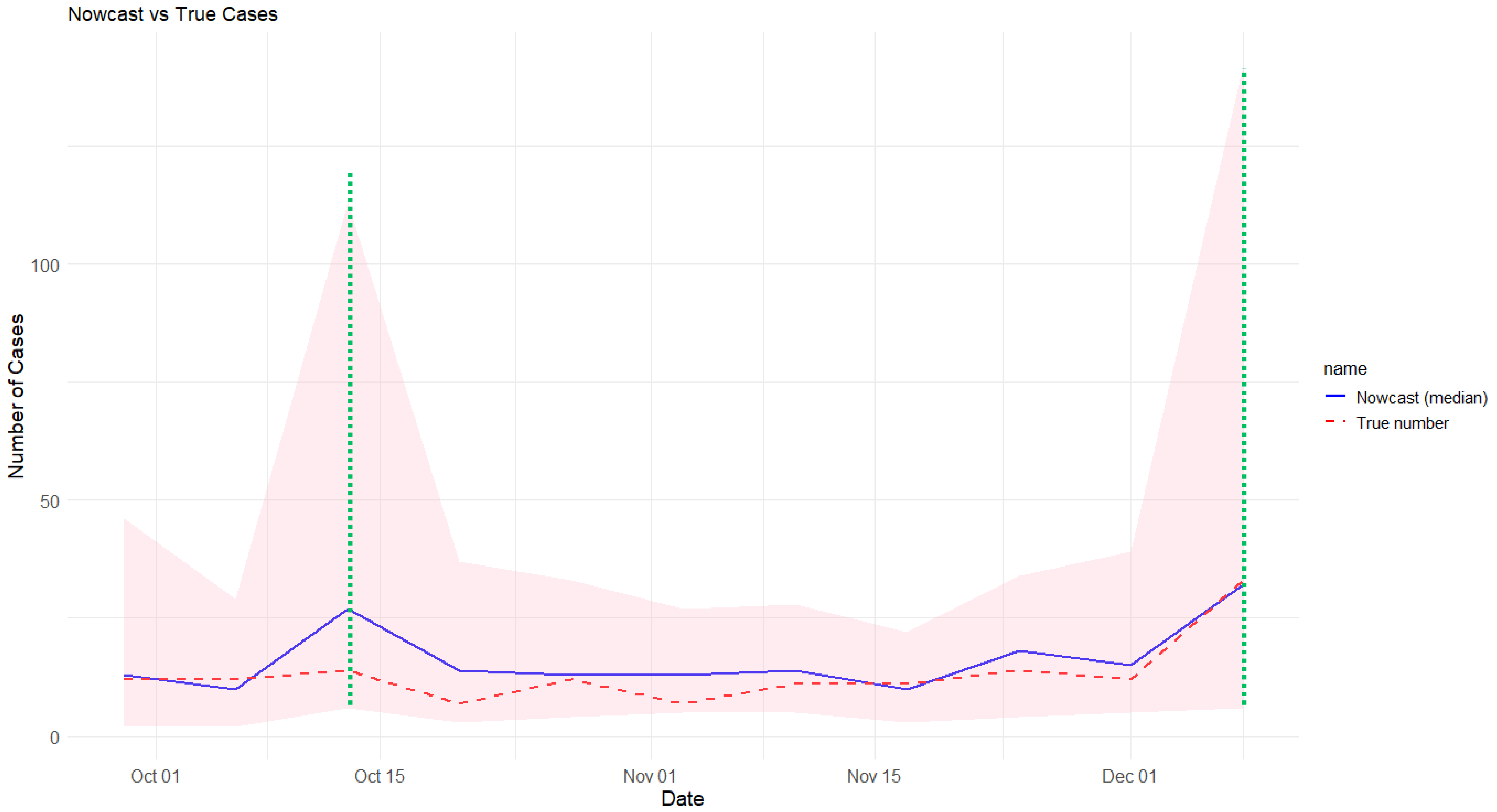
The function iterates through posterior samples, extracting predictions based on reporting delays to understand their impact on forecast accuracy. It computes key metrics such as root mean square error (RMSE), which measures the average distance between predictions and observed values, and the logarithmic score and continuous ranked probability score (CRPS), which evaluate the quality of probability forecasts.

Additionally, it examines empirical coverage of prediction intervals at 75%, 90%, and 95% confidence levels, providing insight into the reliability of the forecasts.

DAILY DATA (2020) VS. WEEKLY DATA (2022)

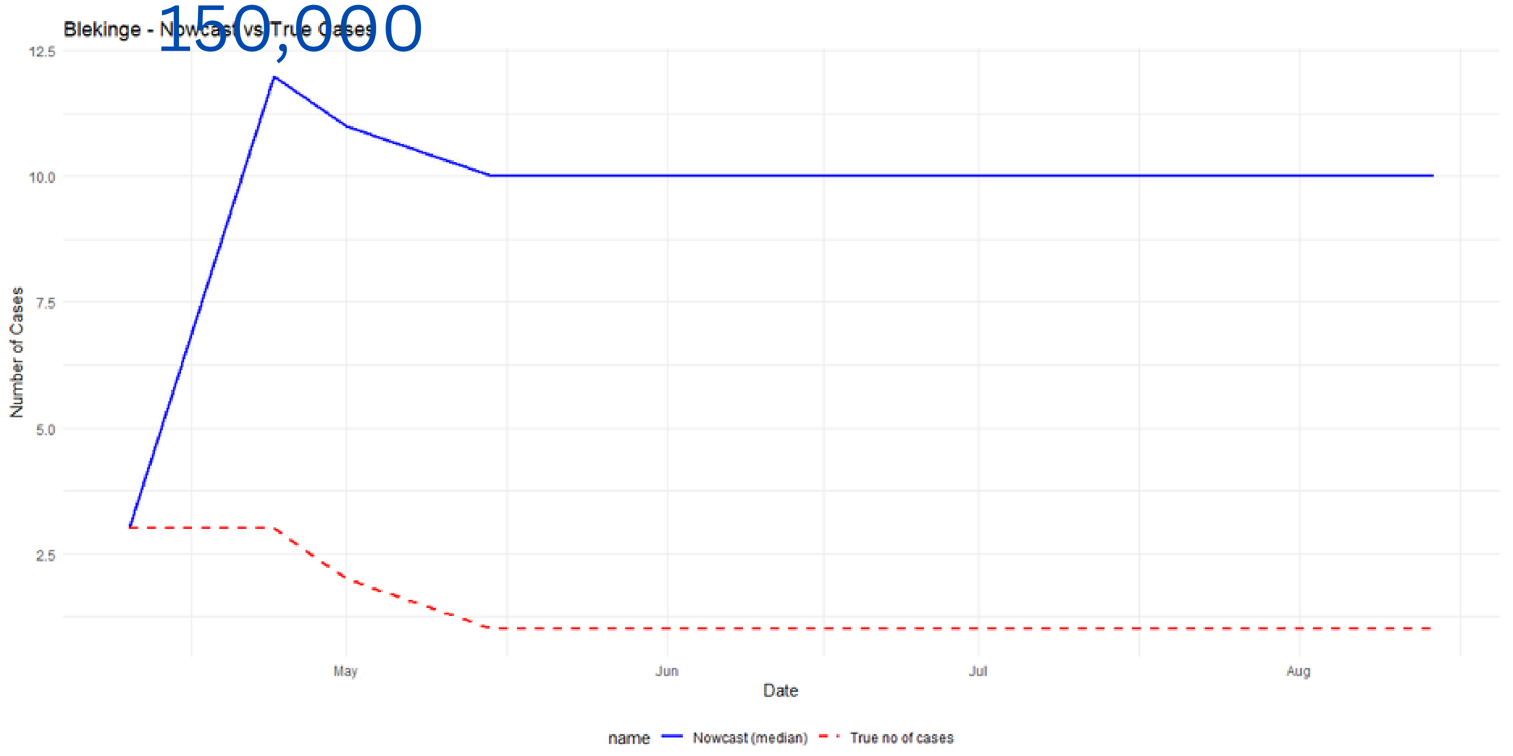


DAILY DATA (2020) VS. WEEKLY DATA (2022)

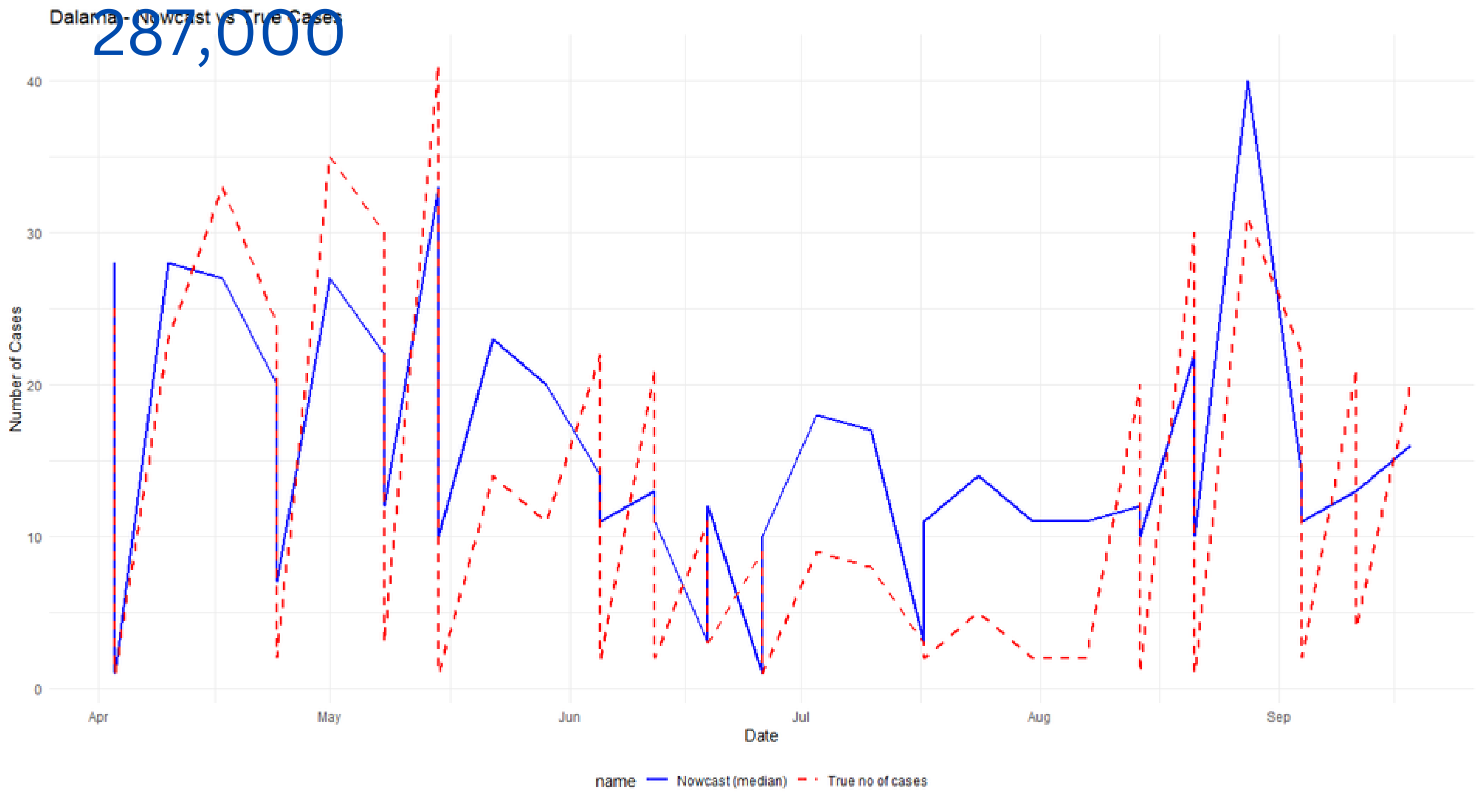


Regional (2023-10-01)

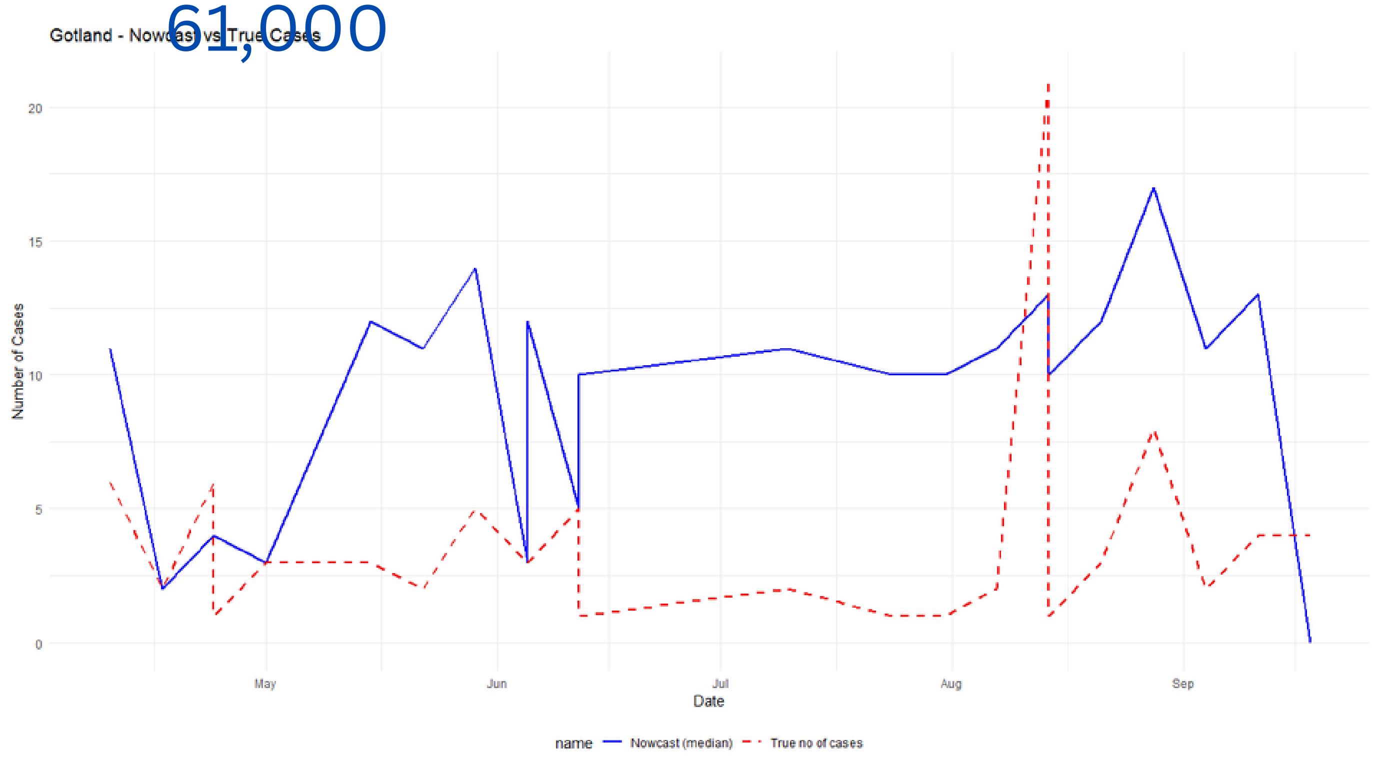
Blekinge ~ 150,000



Dalarna ~ 287,000



Gotland ~ 61,000



Uppsala ~ 430,000

Conclusion

This study aimed to evaluate the quality of COVID-19 reporting across Sweden and identify patterns in the data that could inform future nowcasting models. By examining the accuracy of reporting over time and across regions, several key insights emerged.

Conclusion

This study aimed to evaluate the quality of COVID-19 reporting across Sweden and identify patterns in the data that could inform future nowcasting models. By examining the accuracy of reporting over time and across regions, several key insights emerged.

Changes in Reporting Accuracy Over Time: Our analysis showed that reporting accuracy improved as the pandemic progressed and the reporting infrastructure evolved. Early data from 2020 was unreliable due to gaps and inconsistencies. From 2022 onwards, data became more consistent, reflecting advancements in collection processes. The change from daily to weekly reports impacted accuracy, as weekly data was often more refined but could delay real-time case trend capture.

Conclusion

This study aimed to evaluate the quality of COVID-19 reporting across Sweden and identify patterns in the data that could inform future nowcasting models. By examining the accuracy of reporting over time and across regions, several key insights emerged.

Changes in Reporting Accuracy Over Time: Our analysis showed that reporting accuracy improved as the pandemic progressed and the reporting infrastructure evolved. Early data from 2020 was unreliable due to gaps and inconsistencies. From 2022 onwards, data became more consistent, reflecting advancements in collection processes. The change from daily to weekly reports impacted accuracy, as weekly data was often more refined but could delay real-time case trend capture.

Reliability of Real-Time Data for Nowcasting: The reliability of real-time data for nowcasting varied significantly. During stable reporting periods, the data was accurate; however, during surges or public holidays, reporting delays and inaccuracies made it unreliable for predictions. These challenges emphasized the need for timely and consistent reporting for effective nowcasting.

Conclusion

This study aimed to evaluate the quality of COVID-19 reporting across Sweden and identify patterns in the data that could inform future nowcasting models. By examining the accuracy of reporting over time and across regions, several key insights emerged.

Changes in Reporting Accuracy Over Time: Our analysis showed that reporting accuracy improved as the pandemic progressed and the reporting infrastructure evolved. Early data from 2020 was unreliable due to gaps and inconsistencies. From 2022 onwards, data became more consistent, reflecting advancements in collection processes. The change from daily to weekly reports impacted accuracy, as weekly data was often more refined but could delay real-time case trend capture.

Reliability of Real-Time Data for Nowcasting: The reliability of real-time data for nowcasting varied significantly. During stable reporting periods, the data was accurate; however, during surges or public holidays, reporting delays and inaccuracies made it unreliable for predictions. These challenges emphasized the need for timely and consistent reporting for effective nowcasting.

Identifying Regions with Accurate and Inaccurate Reporting: The analysis of regional data showed significant differences in reporting quality across Sweden. Some regions maintained accurate reporting with high data integrity during the pandemic, while others experienced unreliable reporting with fluctuations in case counts and delays. Contributing factors included regional healthcare capacity, local infrastructure, and the focus of authorities on accurate data collection. Regions with fewer resources tended to report inconsistently, while those with stronger systems had higher accuracy.

Olayemi Morrison

Thank You

5th June, 2025