UNDERSTANDING FOLKHALSOMYNDIGHETEN'S COVID- 19 DATA UPDATES

OLAYEMI MORRISON

27 Jan, 2025

AGENDA OVERVIEW

01

INTRODUCTION

02

DATA DESCRIPTION

03

RESEARCH QUESTIONS

04

METHODOLOGY

05

RELATED PAPERS

INTRODUCTION

This thesis focuses on analysing and proposing solutions to improve the accuracy of real-time COVID-19 case reporting using nowcasting techniques. It addresses challenges from delays and inconsistencies in reporting by assessing the quality of historical data and identifying regions with unusual reporting patterns. The study proposes dynamic correction methods using nowcasting and hierarchical Bayesian models to tackle common issues of underreporting and lag. By comparing these methods to existing approaches like Tom Britton's nowcasting algorithm, it aims to enhance predictive accuracy and uncertainty quantification. The findings will help develop more reliable public health tools for managing pandemics.

DATA DESCRIPTION

The dataset consists of COVID-19-related statistics provided by Folkhälsomyndigheten (FHM), Sweden's Public Health Agency. The data includes weekly reported cases, hospitalizations, and deaths for various population groups and regions across Sweden. Each row corresponds to a specific week, and the dataset spans from early 2020 to late 2024.

år	veckonummer	Antal_fall_vecka	Antal_fall_100000inv_vecka	Antal_fall_	Kum_anta	Kum_fall_	Antal_nya	Kum_anta	Antal_avli	Antal_avlidna_milj_inv_vecka	Kum_anta	Kum_antal_avlidna_milj_inv
2020	6	1	0	0	1	0	0	0	0	0	0	0
2020	7	0	0	0	1	0	0	0	0	0	0	0
2020	8	0	0	0	1	0	0	0	0	0	0	0
2020	9	13	0	0	14	0	0	0	0	0	0	0
2020	10	211	2	2	225	2	3	3	0	0	0	0
2020	11	838	8	10	1063	10	16	19	5	1	5	1
2020	12	924	9	17	1987	19	88	107	44	4	49	5
2020	13	1957	19	28	3944	38	238	345	190	18	239	23
2020	14	3229	31	50	7173	69	277	622	451	44	690	67
2020	15	3740	36	67	10913	106	289	911	664	64	1354	131
2020	16	3752	36	72	14665	142	250	1161	657	64	2011	195
2020	17	4203	41	77	18868	183	238	1399	546	53	2557	248
2020	18	3832	37	78	22700	220	188	1587	543	53	3100	300

DATA DESCRIPTION

KEY VARIABLES INCLUDE:

- **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.

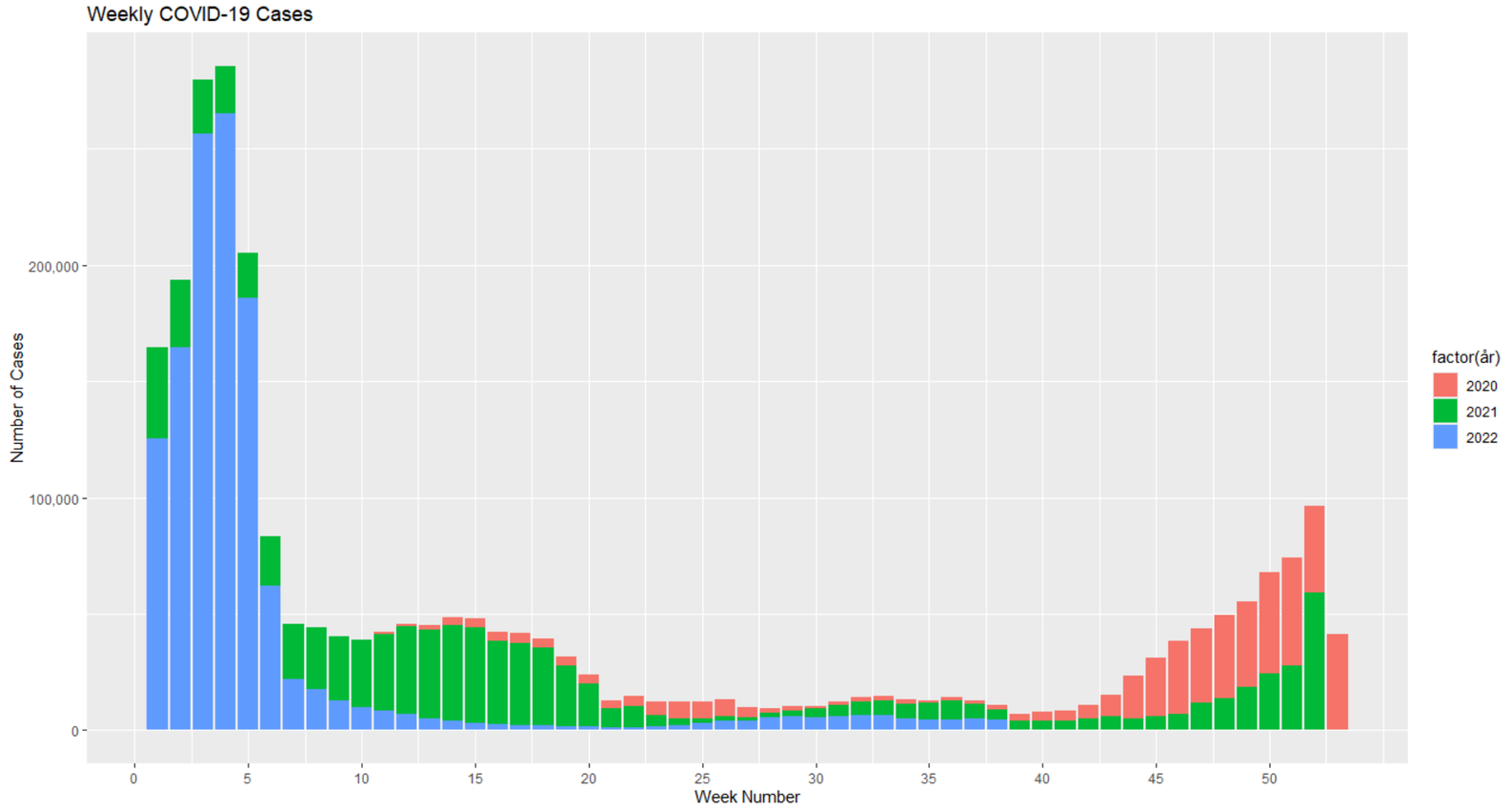
DATA DESCRIPTION

Data spans weekly reporting starting in 2020.

Early weeks (e.g., week 6-8) show minimal cases, likely pre-pandemic onset.

Cases and ICU admissions increase rapidly starting from week 9 onward.

Cumulative columns allow for tracking the total progression of cases, ICU admissions, and deaths over time.



RESEARCH QUESTIONS

The aim here is to evaluate the quality of the data over time and determine patterns in the reporting. To expand on this, consider these sub-questions and points to analyze:

01

TREND PATTERNS

How has the accuracy of reporting changed over the pandemic timeline? For instance, was early data (2020) less reliable than later data (2022)? Did the evolution of reporting policies (e.g., weekly instead of daily reports) impact quality?

02

REAL-TIME DATA SUITABILITY

For which periods was real-time data reliable enough for nowcasting? Are there specific phases when data quality worsened, making nowcasting unreliable? Can we identify patterns in underreporting during weekends, holidays, or specific timeframes.

03

RESEARCH QUESTION

Identify regions with consistently good or poor reporting. Evaluate data quality for different demographic groups(age, gender, etc.)

METHODOLOGY

1. FLEXIBLE BAYESIAN NOWCASTING MODEL:

- Builds on Bayesian inference to predict current cases based on historical delay distributions.
- Includes flexibility to account for time-varying changes in reporting delays.

Metrics to emphasize:

- RMSE: Evaluate predictive accuracy.
- Bias: Determine systematic over/underestimation.
- MAE: Check for consistent accuracy.

METHODOLOGY

2. HIERARCHICAL BAYESIAN MODEL

- Suitable for data with multiple levels (e.g., regions, populations).
- Able to incorporate data from different levels of aggregation (e.g., regional + national data) and borrow strength from similar groups.

Metrics to emphasize:

- Coverage of Credible Intervals: Ensure the model captures uncertainty.
- Posterior Predictive Checks: Verify model fit.
- Predictive Interval Width: Assess uncertainty in predictions.

RELATED PAPERS

**Flexible Bayesian Nowcasting – Application to COVID-19 fatalities in Sweden.
[Fanny Bergstrom, Felix Gunther, Michael Hohle, Tom Britton. January 20, 2022]

**Causal Inference for Time series Analysis: Problems, Methods and Evaluation
[Raha Moraffah, Paras Sheth, Mansooreh Karami, Anchit Bhattacharya, Qianru Wang, Anique Tahir, Adrienne Raglin, Huan Liu. February 2021].

**A hierarchical Bayesian model for estimating age-specific COVID-19 infection fatality rates in developing countries.
[Sierra Pugh, Andrew T. Levin. August 2023]

**Collaborative nowcasting of COVID-19 hospitalization incidences in Germany.
[Daniel Wolffram, Sam Abbott]

THANK YOU

27 Jan, 2025