Comprehensive Analysis of the ongoing Epidemic at Plateau de Saclay: A Medical and Epidemiological Report

Date: 10/05/2030

Contributors:

Dr. Dominique Bennett, MD, PhD

Pauline Jensen, Biostatistician, Intern

Loïc Blackburn, Medical Engineer

Confidentiality Preamble

This document and the information contained herein are strictly confidential and intended solely for the use of authorized personnel involved in the clinical study. The contents of this report are protected by privacy laws and regulations, and any unauthorized disclosure, copying, distribution, or use of any part of this report is strictly prohibited and may result in legal action.

This report contains sensitive information regarding a clinical study conducted on the patients having been infected by the disease, as of April 10. The data, findings, and conclusions are to be used exclusively for the purposes of this study and related medical and scientific research.

Access to this report is restricted to individuals who have been granted explicit permission by the overseeing authorities. All readers are expected to uphold the highest standards of confidentiality and integrity in handling this sensitive information.

By proceeding to read this document, you acknowledge your understanding of these confidentiality requirements and your commitment to abide by them.

Disclaimer

It is important to note that the data and analysis are subject to change as new information becomes available. The findings herein are based on the dataset available as of April 10th and may not fully represent the ongoing dynamics of the epidemic.

Abstract

This report presents a comprehensive statistical analysis of an epidemic dataset, providing key insights into the spread and impact of the disease across various demographics and locations. Our analysis includes a detailed examination of infection rates, recovery times, and hospitalization probabilities, stratified by factors such as age, sex, and campus location. A notable aspect of this report is the investigation into the characteristics of 'Patient Zero', which offers crucial understanding of the initial outbreak and subsequent transmission dynamics. Furthermore, we delve into the prevalence and characteristics of asymptomatic cases, largely uncovered through the development and application of a novel antigenic test. This test played a pivotal role in identifying silent carriers, contributing significantly to our understanding of the disease's spread. The report combines traditional epidemiological approaches with innovative testing strategies, presenting a multifaceted view of the epidemic. Our findings underscore the complexity of disease transmission and highlight the importance of tailored public health strategies in controlling such outbreaks.

1) Patient zero

The first reported case in the Plateau de Saclay epidemic was a 22-year-old male student from HEC, named Zakariya Borras. His initial consultation with the school physician on April 28 was prompted by high fever, which soon progressed to a cluster of additional symptoms, including skin rash, fatigue, headaches, and swollen lymph glands. Despite these symptoms, the possibility of quarantine was not immediately considered due to the lack of clear evidence for human-to-human transmission at the time, and sexually transmitted diseases were ruled out. Remarkably, Borras was not placed on medical leave following his consultation, a decision that may have had significant implications in the context of epidemic spread. Subsequent mild delirium, possibly induced by persistent fever, was observed when Borras was found aimlessly wandering in the corridors of CentraleSupélec, inadvertently distributing pages from his personal journal. These pages later proved critical in tracing the origins of his infection, leading to the identification of a mosquito bite on April 23 as the likely initial transmission event.

Borras' condition remained stable for a few days but eventually necessitated hospitalization on May 2 due to the onset of neurological complications. As of this report, he remains hospitalized, and his case underscores the complexities of diagnosing emerging infectious diseases. The spread of the virus highlights the importance of rapid response in infectious disease management. This case has become a pivotal study in the ongoing efforts to understand and contain the epidemic at Plateau de Saclay, offering vital insights into the nature of the disease, its transmission vectors, and effective response strategies.

2) Viral test

In the midst of the ongoing epidemic, a unique turn of events has led to a significant breakthrough in antigenic testing. During this crisis, one of the students from HEC, who had to be hospitalized due to the infection, turned out to be the son of a high-ranking director of a renowned pharmaceutical group. While we maintain confidentiality regarding the identity of this individual and the pharmaceutical group, this connection has facilitated the rapid mobilization of substantial resources. Leveraging both financial backing and expert manpower, a reliable and quick antigenic test was developed in response to the epidemic. This fortuitous development allowed for the execution of an extensive testing campaign, comprising 1000 tests conducted in two phases of 500 each. This strategic initiative has been instrumental in our efforts to understand and combat the spread of the virus in this unique epidemic scenario.

Biological Protocol for Viral Testing

Objective: To detect the presence of a specific virus in individuals, potentially transmitted via mosquito bites.

Materials and Equipment:

1. Sample Collection Kits: Including sterile swabs for nasal or throat swabbing.

2. Personal Protective Equipment (PPE): Gloves, masks, protective eyewear, and gowns.

3. Viral Transport Media (VTM): For safe transport of the samples to the laboratory.

4. RNA Extraction Kits: To extract viral RNA from the samples.

5. Reverse Transcription Polymerase Chain Reaction (RT-PCR) Kits: For amplifying and detecting viral RNA.

6. Real-Time PCR Machine: For conducting RT-PCR.

7. Biosafety Cabinets: For sample handling to prevent contamination.

8. Incubators and Freezers: For storing samples and reagents.

9. Data Recording System: For tracking samples and recording results.

Procedure:

1. Sample Collection:

- Use PPE to ensure the safety of the healthcare provider.
- Collect nasal or throat swabs from the individuals.
- Place swabs in VTM immediately after collection and seal securely.

2. Transport and Storage:

- Transport the samples to the laboratory under recommended conditions.
- Store the samples at 2-8°C if testing within 48 hours, or freeze at -80°C for longer storage.

3. RNA Extraction:

- Extract viral RNA from the samples using the RNA extraction kit, following the manufacturer's instructions.
- Perform this step in a biosafety cabinet to prevent contamination.

4. RT-PCR Setup:

- Prepare the RT-PCR mix with specific primers and probes for the target virus.
- Add the extracted RNA to the RT-PCR mix.

5. Amplification and Detection:

- Place the RT-PCR setup in the real-time PCR machine.
- Run the PCR cycle according to the protocol specified for the virus.

- The machine will amplify the viral RNA and detect it through fluorescence signals.

6. Data Analysis and Interpretation:

- Analyze the PCR data to determine the presence of the virus.

- A positive result indicates the presence of viral RNA.

- A negative result indicates the absence of detectable viral RNA.

- Record and report the results appropriately.

7. Quality Control:

- Include positive and negative controls in each batch of tests to ensure the reliability of the results.

8. Safety and Waste Disposal:

- Follow biosafety guidelines for handling and disposing of biological samples and contaminated materials.

Quality control

The specific antrigenic test has very good performance, with no risk of having false positive and only a very low probability of false negative (0.02).

Protocol for Inclusion of Individuals in Viral Testing Study

Objective:

To conduct a viral testing study on a group of individuals, to monitor and understand the spread and characteristics of a virus, potentially transmitted via mosquito bites.

1. Participant Selection and Enrollment:

- **Criteria**: Ensure a diverse group in terms of age, sex, and other demographics. Only individuals from the HEC campus with no disease symptoms on May 6th.
- **Size**: 500 individuals are selected.
- **Informed Consent:** Obtain informed consent from all participants. Explain the purpose, procedure, potential risks, and benefits of the study.

2. Pre-Test Counseling and Data Collection:

- **Counseling**: Provide information about the testing process and what the results may imply.
- **Baseline Data Collection:** Collect baseline data including age, sex, medical history, and potential exposure to mosquitoes. Each individual will be tested twice, first on May 6th, then on May 10th.

3. Sample Collection:

- Follow the sample collection procedure outlined in the Biological Protocol for Viral Testing.

4. Testing and Monitoring:

- Perform the viral testing as per the biological protocol.
- Schedule follow-up tests as required (e.g., after 4 days).

5. Data Recording and Management:

- Recording Results: Record the test results alongside the participant's demographic data.
- Database Management: Store the data in a secure database, maintaining confidentiality and integrity.

6. Post-Test Counseling:

- Provide results to the participants along with appropriate counseling.
- For those who test positive, offer guidance on medical care.
- For those who test negative, provide information on preventive measures.

7. Follow-Up:

- Conduct follow-up tests as required (e.g., as indicated, tests were conducted on May 6 and May 10).
- Monitor the development of symptoms and additional health parameters.

8. Data Analysis:

- Analyze the data for patterns in virus spread, incubation period, symptom development, etc.
- Maintain the anonymity of participants in any external reports or publications resulting from the study.

9. Ethical Considerations:

- Ensure the study adheres to ethical guidelines for research involving human subjects.
- Review and approval by an Institutional Review Board (IRB) or equivalent.

10. Safety and Compliance:

- Follow all local health and safety guidelines.
- Ensure proper training for all staff involved in the study.

Data Analysis

The results of the viral test can be found in the file "test_HEC_6_10_may.csv". The dataset contains information about individuals, including their last name, first name, sex, age, campus affiliation, and several columns related to virus detection and exposure to a mosquito bite. The relevant columns for the analysis we present in the following are:

- `detection_6may`: Indicates whether individuals were detected with the virus (1) or not (0) on May 6.

- `time_since_exposition_6may`: Shows the time elapsed since the probable mosquito bite that caused the infection, as of May 6.

- `detection_10may`: Indicates the virus detection status on May 10, similar to `detection_6may`.

- `time_since_exposition_10may`: The time since exposure, as of May 10.

- `symptoms_10_may`: Indicates whether individuals have symptoms on May 10.

To analyze the data, we look into the following:

1. The proportion of individuals who were virus-free on May 6 but tested positive on May 10.

2. The number of individuals who remained virus-free between May 6 and May 10.

3. The average time since exposure for those who tested positive on May 10 but were negative on May 6.

4. Analysis of symptom development among those who were initially virus-free but tested positive later.

We estimate that:

- **Prevalence on May 6:** The prevalence of the virus among the individuals on May 6 is 2.4%. This is calculated as the proportion of individuals who tested positive for the virus on May 6 relative to the total number of individuals in the dataset. Note that the individuals tested were necessarily asymptomatic (it was a selection criterion), such that this result is biased.
- **Prevalence on May 10**: The prevalence on May 10 is 5.4%. This indicates an increase in the number of positive cases within the 4-day period between the two testing dates.
- **Proportion of Asymptomatic Individuals on May 10**: Among those who tested positive on May 10, 77.8% were asymptomatic (see Fig. 1). This high proportion of asymptomatic cases is particularly significant given that all individuals included in the May 6 testing were asymptomatic.

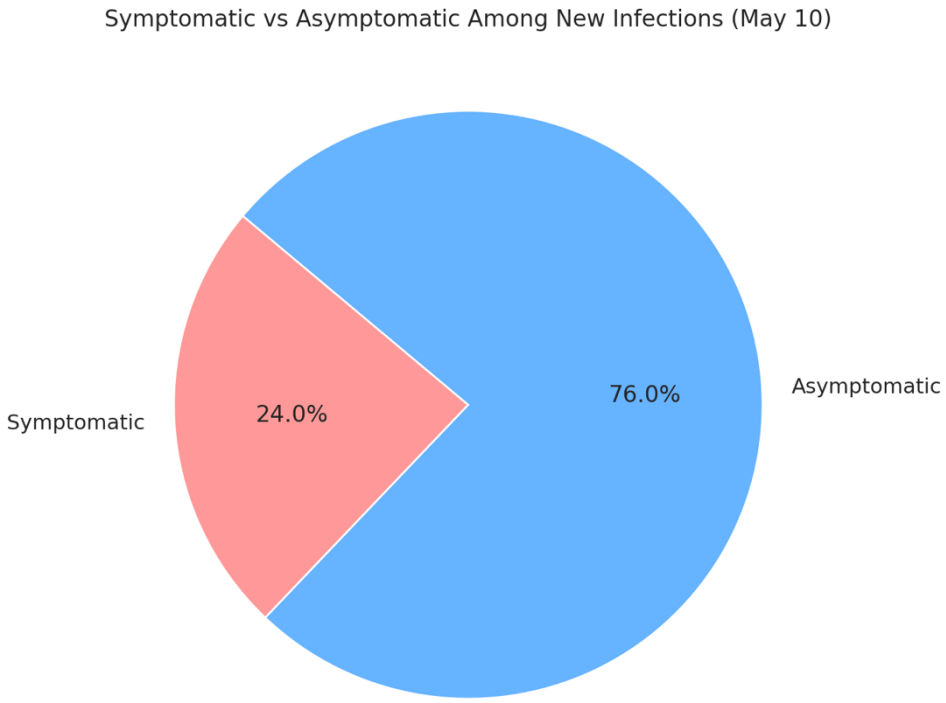

Figure 1. Pie Chart of Symptomatic vs Asymptomatic Among New Infections: This chart illustrates the proportion of symptomatic versus asymptomatic individuals among those who were newly infected by May 10. It offers a clear visual representation of the symptom development in this group.

Based on the analysis of the dataset, we also find:

1. **Proportion of New Infections**: 5% of the individuals were virus-free on May 6 but tested positive on May 10.

2. **Individuals Remaining Virus-Free**: 463 individuals remained virus-free between May 6 and May 10.

3. **Average Time Since Exposure for New Infections**: Among those who tested positive on May 10 but were negative on May 6, the average time since exposure was approximately 6.12 days.

4. **Symptom Development in New Infections**: 24% of the individuals who were initially virus-free but tested positive later developed symptoms by May 10.

This data provides insights into the infection spread and symptom development over the 4-day period between the two tests.

We can approximate the incubation time for those who were not detected with the virus on May 6 but tested positive and showed symptoms on May 10. This incubation period would be the time since their probable exposure (as of May 10) minus the four days between the two tests (see Fig. 2).

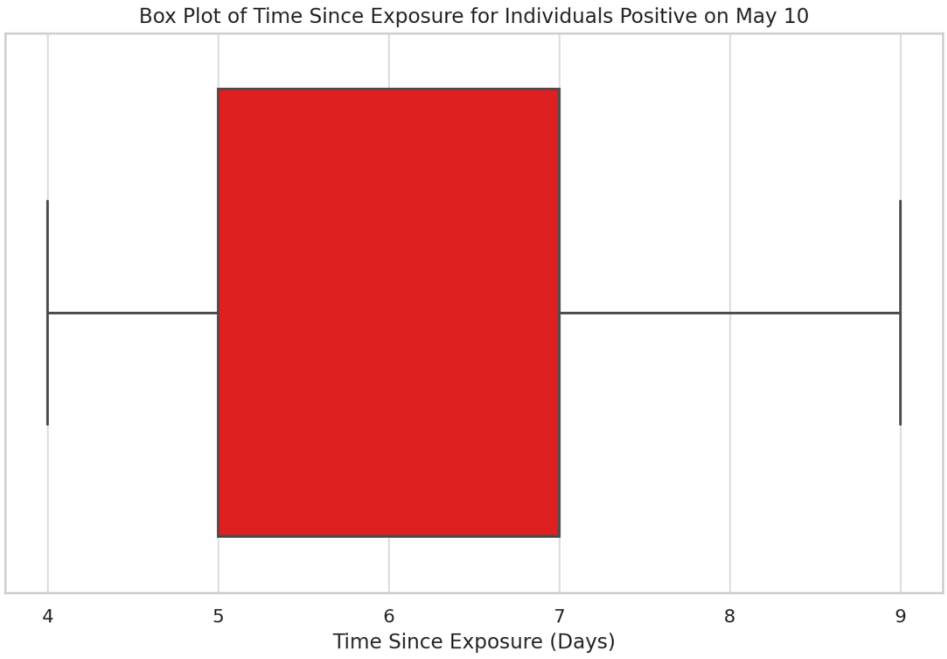

Figure 2: box plot showing the distribution of the time since exposure for individuals who tested positive on May 10. This plot focuses on the subset of individuals who had a positive test result on May 10, providing insights into the exposure time for this specific group. Similar to the previous box plot, this visualization includes the interquartile range, median, and potential outliers, specifically for those who were virus-positive as of May 10.

3)Statistical Analysis

Introduction

Note: The statistical analysis presented here has been conducted by the biostatistician Ms. Jensen, currently doing his L3 internship in our team.

The following paragraphs present an analysis of data from the Plateau de Saclay epidemic, focusing on the distribution of infected patients, the prevalence of symptoms, and the timelines of infection and recovery. The dataset as of April 10th includes information on patient demographics, reasons for consultation, and dates of infection, recovery, and hospitalization.

This comprehensive analysis provides valuable insights into the epidemic's impact, spread, and dynamics. Understanding these aspects is crucial for effective epidemic management and response planning.

Demographic Distribution

The first aspect of the analysis examined the age and gender distribution of the infected patients. The age distribution (Figure 3) shows a wide range of affected individuals, with a notable concentration in certain age groups, but probably reflecting the population distribution on the Plateau de Saclay, with a high proportion of young adults living there. The gender distribution (Figure 4) provides insights into the epidemic's impact on different genders. There is a well balance of infected between male and female.

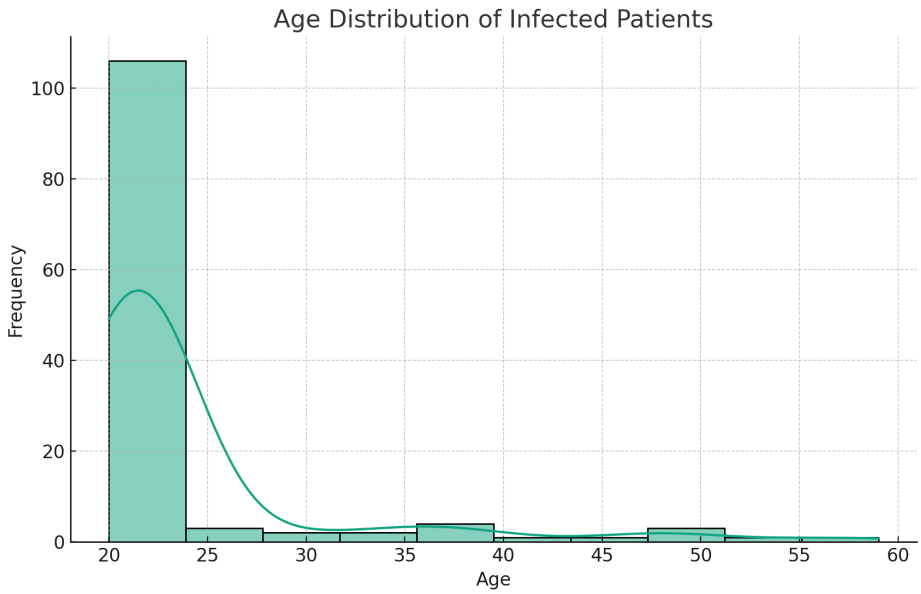

Figure 3: Age distribution of Infected patients

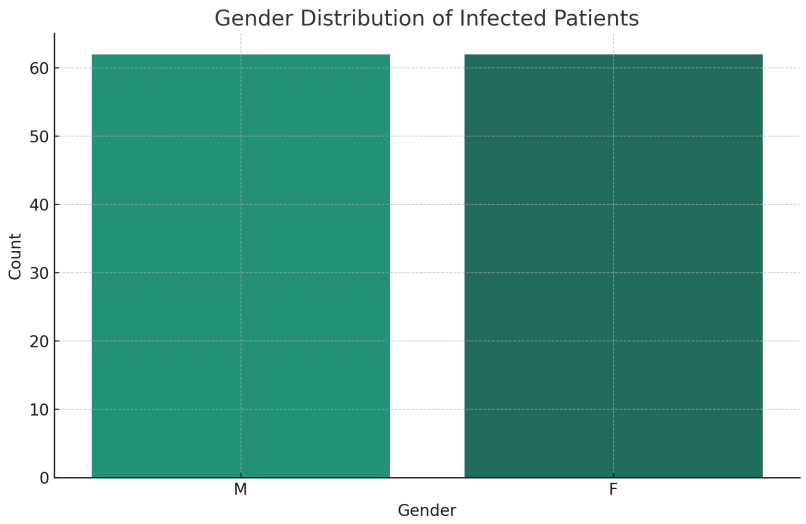

Figure 4: Gender distribution

Symptom Prevalence

An analysis of symptoms leading to consultation revealed a varied set of reasons, with some symptoms being more prevalent than others (Figure 5). The symptoms presented here correspond to the first reason for consultation. Symptomatic individuals rapidly exhibit more than one symptom. This information is crucial in understanding the clinical presentation of the epidemic. Any individual presenting one of these symptoms might be susceptible to being infected by the disease and should consult a physician immediately.

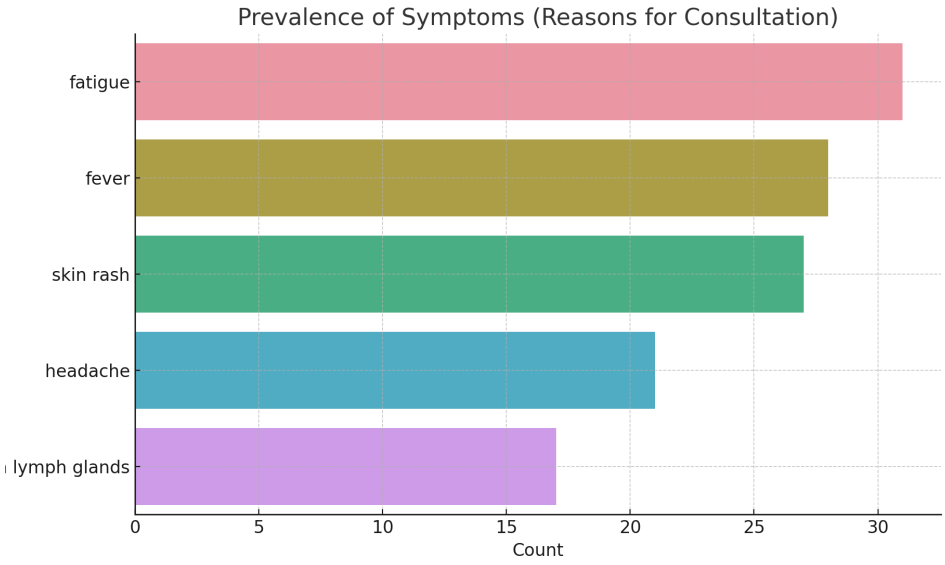

Figure 5: Prevalence of Symptoms

Infection Timeline

The infection timeline (Figure 6) illustrates the spread of the infection over time, highlighting days with peak new reported infections (asymptomatic individuals, therefore, are not seen here). This timeline is essential for understanding the dynamics of the epidemic's spread.

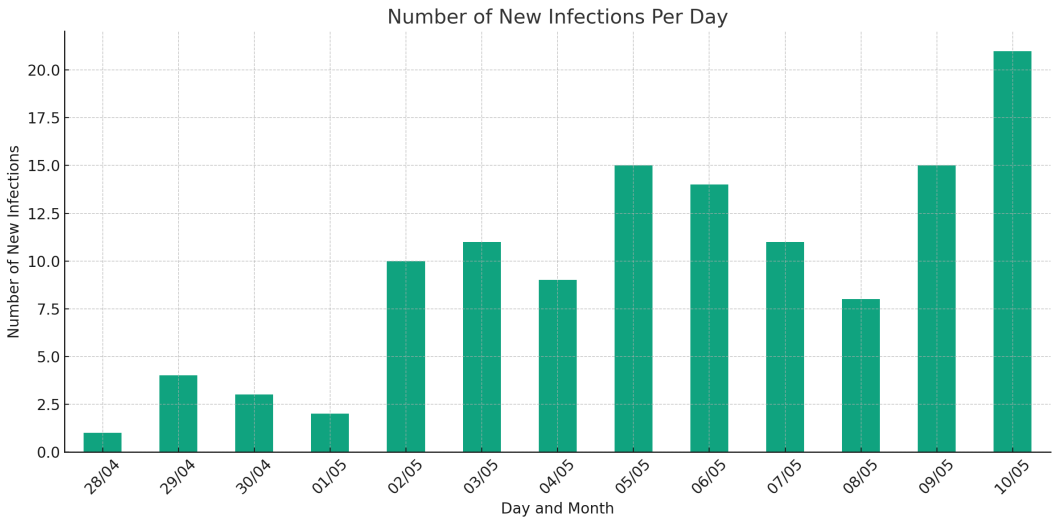

Figure 6: Number of new infections per day

Stratification of the new infections

A key element of our analysis involved stratifying the new infections (at least, those symptomatic) by various demographic and geographical factors, such as age, sex, and campus location. This stratification provided deeper insights into the spread of the epidemic among different segments of the population. Notably, the infection rates varied significantly between campuses, indicating potential differences in exposure or susceptibility. These stratified views are instrumental in understanding the nuanced dynamics of the epidemic and in forming targeted intervention strategies.

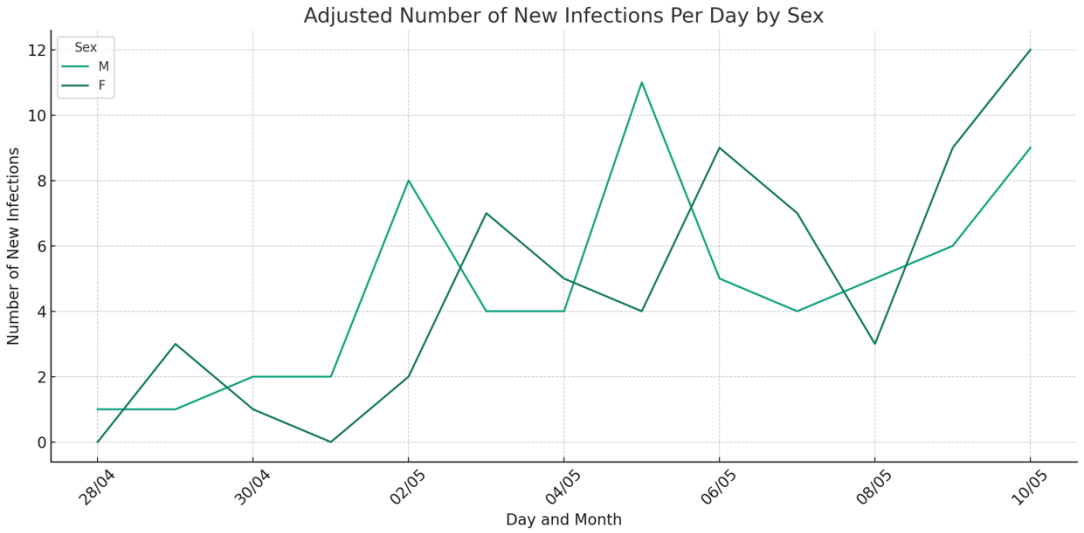

Figure 7: Stratifying the new reported infections by sex: Male (M) or Female (F)

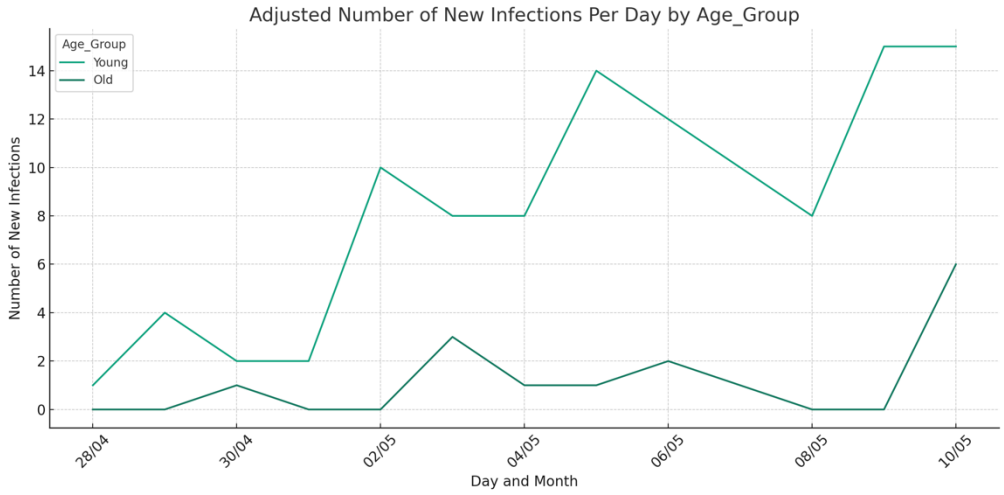

Figure 8: Stratifying the new daily infections according to whether the patients are older or younger than 30

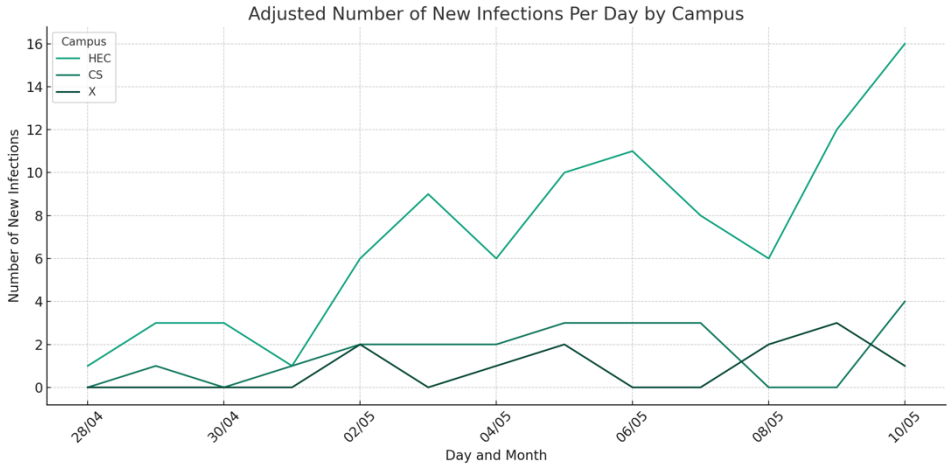

Figure 9: Comparing the new infections according to the residence site (campus)

Given the total population sizes for each campus (5,400 in HEC, 4,320 in X, and 5,700 in CS), we can conduct additional analyses to provide more context to the epidemic data.

To begin, let's calculate the infection rate for each campus using the total population provided above. We'll also compute the standardized infection density per 1,000 people for each campus.

Based on the total population sizes for each campus and the number of infected patients, here are the infection rates and standardized infection densities:

Infection Rate per Campus:

- HEC: Approximately 1.70%
- CS: Approximately 0.37%
- X: Approximately 0.25%

Standardized Infection Density per 1,000 People:

- HEC: Approximately 17.04 infections per 1,000 people
- CS: Approximately 3.68 infections per 1,000 people
- X: Approximately 2.55 infections per 1,000 people

These metrics provide a clearer understanding of how the epidemic affected each campus relative to its population size. The HEC campus has a notably higher infection rate and density compared to the other two campuses, indicating a more significant impact of the epidemic there.

This kind of analysis is crucial for assessing the relative severity of the epidemic across different population centers and can guide targeted response strategies.

Recovery Timeline

The recovery timeline (Figure 11) provides insights into the duration of the infection and recovery rates, essential for understanding the burden on healthcare systems.

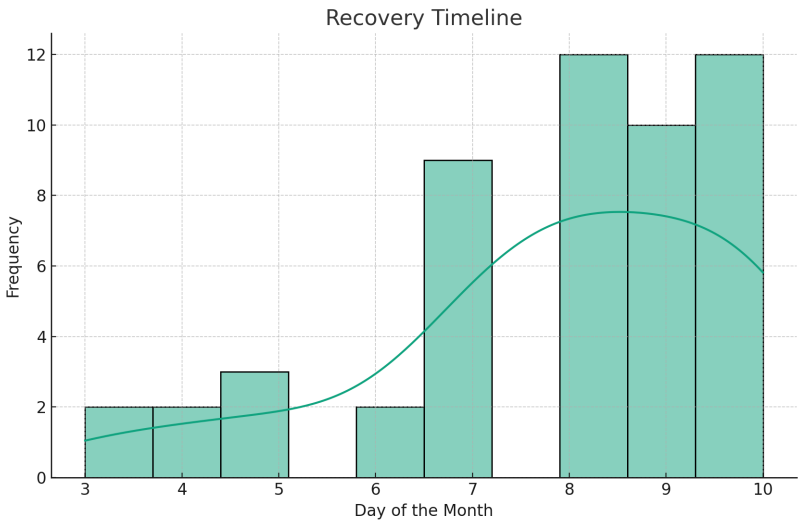

Figure 11: Recovery timeline

Prevalence of Active Infections

A crucial aspect of understanding the epidemic's impact is knowing the prevalence of active infections over time. The plot of daily prevalence (Figure 12) shows the number of people infected (only those who are symptomatic) and not yet recovered or hospitalized on each day, highlighting the active burden of the epidemic.

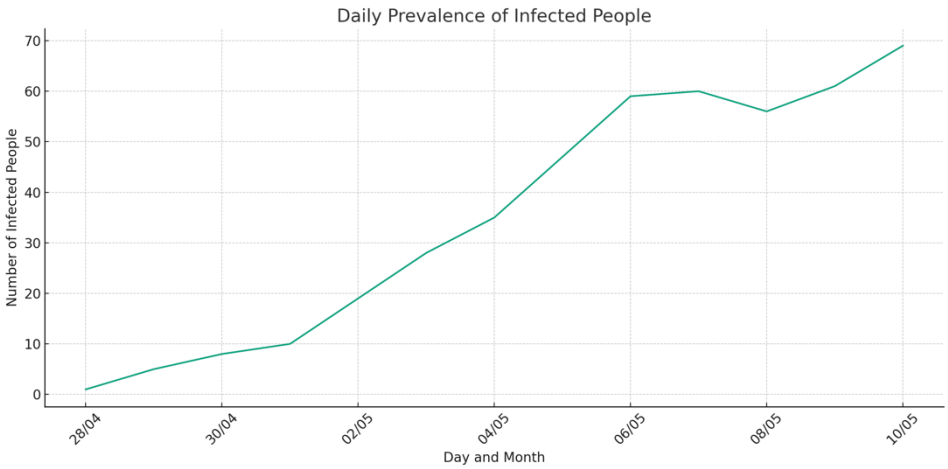

Figure 12: Prevalence of the disease over time

Stratification of the actual number of infected

The stratification of the epidemic data revealed significant insights into the prevalence of the infection across different demographics and locations. By examining the data across various strata such as age groups, sex, and campuses, we were able to uncover specific patterns of prevalence. For instance, certain campuses exhibited higher rates of infection, potentially indicating varying levels of exposure or effectiveness of containment measures. Understanding these patterns of prevalence is vital for developing targeted public health strategies and for efficient allocation of medical resources. It also aids in identifying high-risk groups and areas that require more focused attention and intervention.

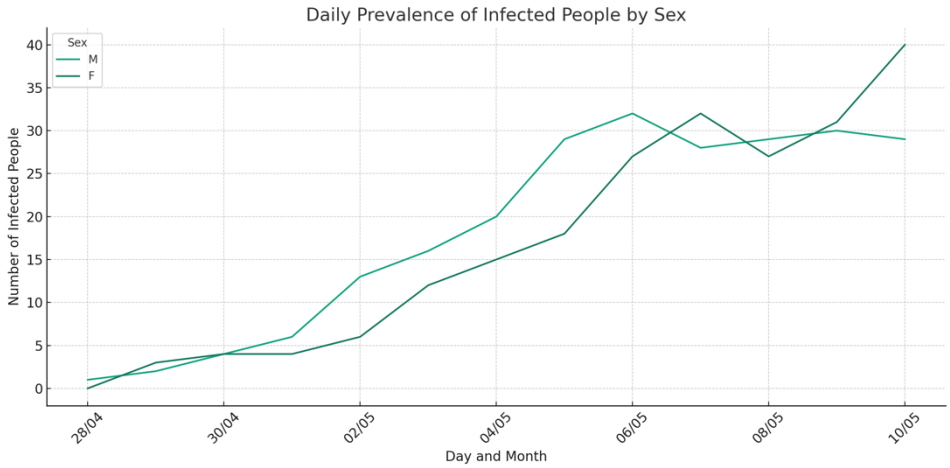

Figure 13: Prevalence by sex

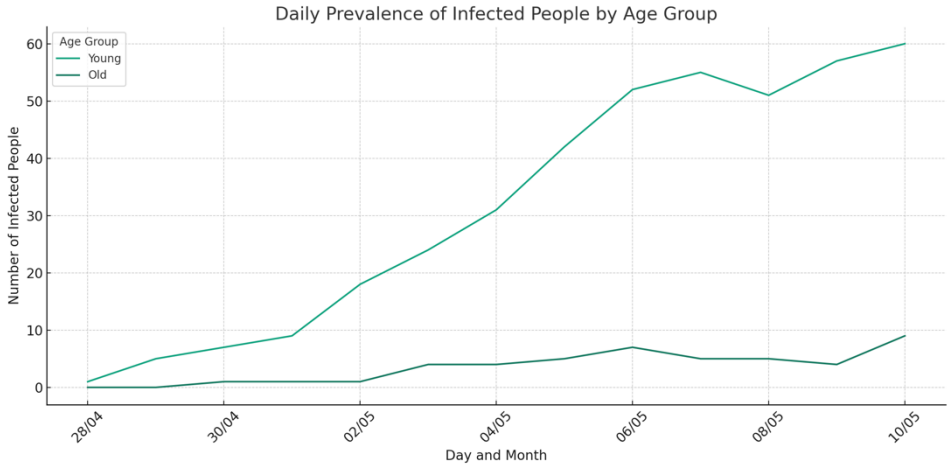

Figure 14: Prevalence by age

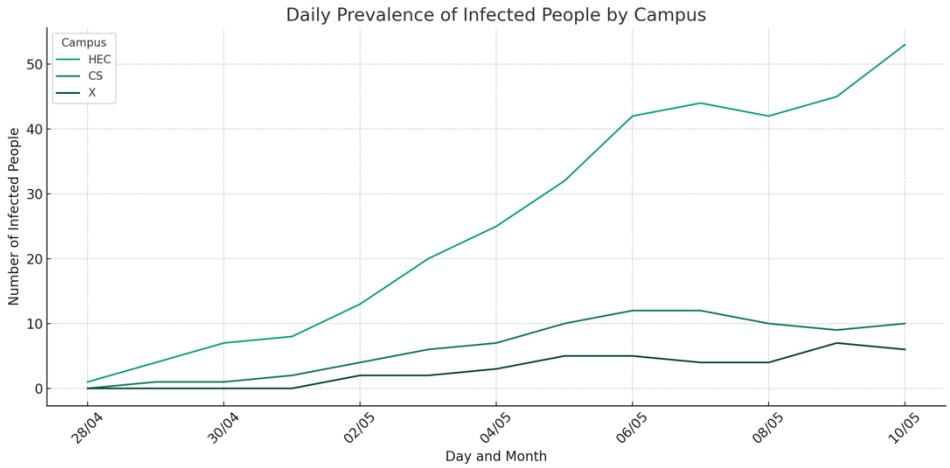

Figure 15: Prevalence by campus

Comparison of New Infections and Recoveries

A comparison of the daily new reported infections and new recoveries (Figure 16) was conducted to understand the balance between new cases and recovery rates.

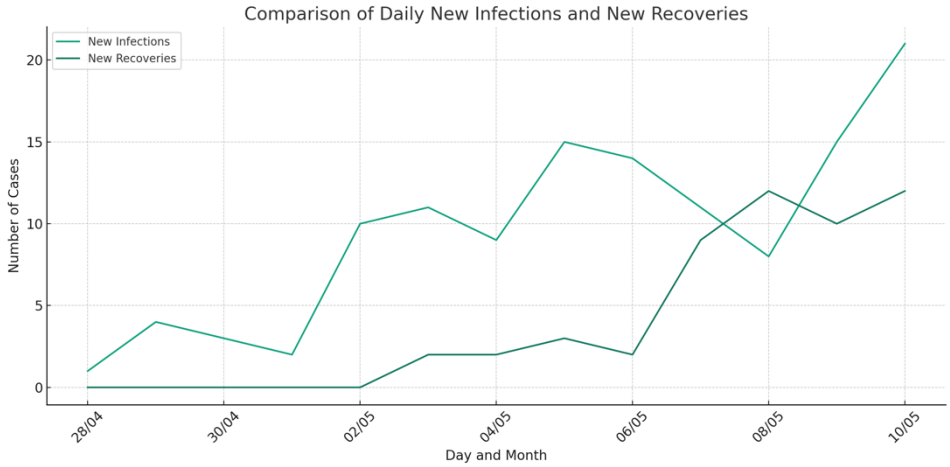

Figure 16: New infections vs. new recoveries

Length of Infection

The histogram of the length of infection (Figure 17) provides insights into the duration patients remained infected before recovery or hospitalization, a critical metric for healthcare planning.

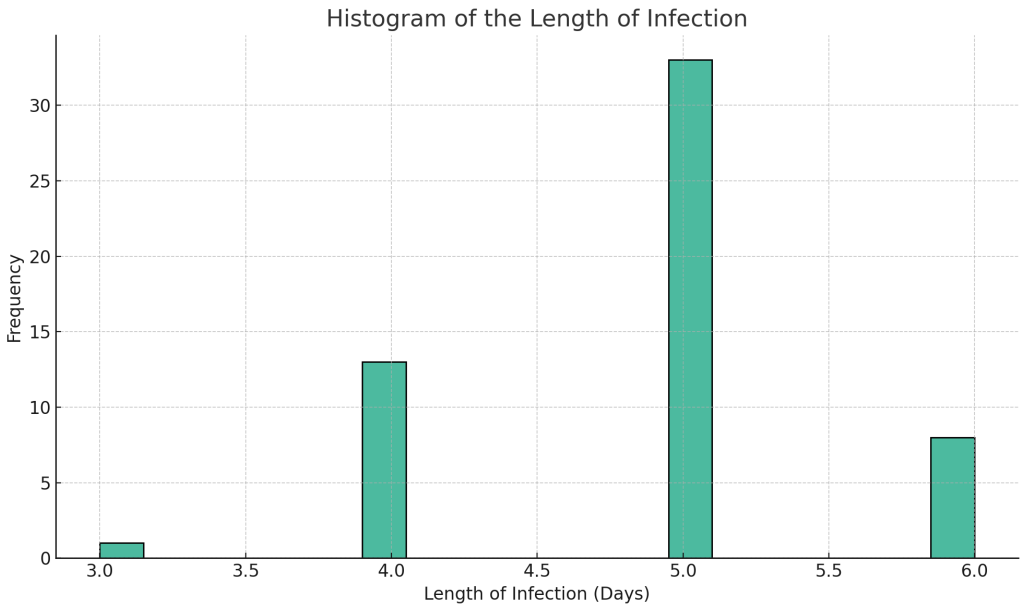

Figure 17: Time spent with disease symptoms (before being recovered or hospitalised)

Probability of Hospitalization

The analysis also calculated the probability of hospitalization at approximately 2.42%, with a 95% confidence interval ranging from 0.00% to 5.12%. This metric helps understand the severity of the epidemic.

Campus Comparison

A comparison between campuses in terms of the number of infected symptomatic patients (Figure 18) highlights the differential impact of the epidemic across various locations.

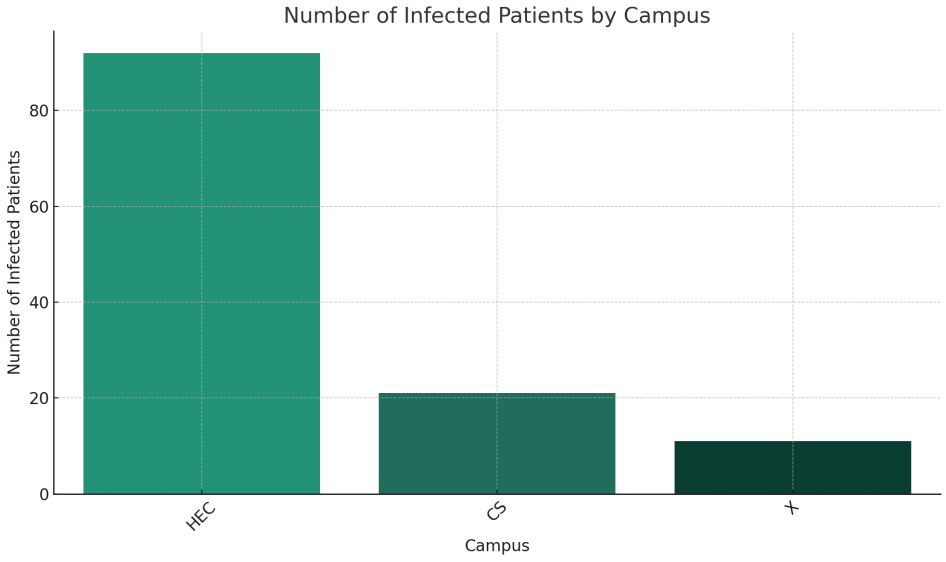

Figure 18: comparing the number of infected patients according to their living site

Correlation

The correlation analysis of the dataset reveals the following relationships between age, length of infection, and the binary variables for hospitalization and recovery:

1. Age:

- There is a weak negative correlation between age and length of infection (-0.21), suggesting younger patients may have slightly longer infection durations.

- Age shows a very weak positive correlation with hospitalization (0.02) and an almost negligible correlation with recovery (0.003).

2. Length of Infection:

- There is a moderate negative correlation between the length of infection and being hospitalized (-0.32), implying that longer infections might be less likely to result in hospitalization.

- Conversely, there is a moderate positive correlation between the length of infection and recovery (0.32), indicating longer infection durations might be associated with eventual recovery.

These correlations provide insights into the dynamics of the epidemic, particularly the relationships between patient age, infection duration, and outcomes such as hospitalization and recovery. However, it's important to note that correlation does not imply causation and these relationships might be influenced by other factors not captured in the dataset.

Relative Risk

The relative risk analysis comparing the risk of infection on each campus relative to the others yields the following insights:

1. Relative Risk for HEC Campus:

- Compared to CS: The risk of infection on the HEC campus is approximately 4.62 times higher than on the CS campus.

- Compared to X: The risk of infection on the HEC campus is approximately 6.69 times higher than on the X campus.

2. Relative Risk for CS Campus:

- Compared to HEC: The risk of infection on the CS campus is approximately 0.22 times (or 22%) that of the HEC campus.

- Compared to X: The risk of infection on the CS campus is approximately 1.45 times higher than on the X campus.

3. Relative Risk for X Campus:

- Compared to HEC: The risk of infection on the X campus is approximately 0.15 times (or 15%) that of the HEC campus.

- Compared to CS: The risk of infection on the X campus is approximately 0.69 times (or 69%) that of the CS campus.

These results indicate that the HEC campus had a significantly higher relative risk of infection compared to the other campuses. Conversely, the campus of Polytechnique (X) had the lowest relative risk in comparison to the others. This analysis is crucial for understanding the differential risk profiles across the campuses and could inform targeted health interventions.

Annex

In this document, we analyzed several dataset that can be found as supplementary material:

- “test_HEC_6_10_may.csv” related to the antiviral test
- “extract_infected_may10.csv” related to the infected individuals until May 10th