Impact of Climate Changes and Air Quality on Asthma-Related Emergency Department Visits in a Tropical Country

Research Proposal

Executive Summary

This research proposal outlines a comprehensive approach to examine the relationship between climate variables, air quality parameters, and asthma-related emergency department visits in Singapore. Using data from 2016~2024, we will conduct time-series analyses to identify temporal patterns and correlations, with particular focus on monsoon seasonality, lag effects, and extreme weather conditions. The study will also evaluate the impact of the COVID-19 pandemic on these relationships, providing valuable insights for healthcare resource planning and public health interventions in tropical environments. Preliminary analysis indicates significant variation in asthma exacerbations tied to monsoon patterns and air quality fluctuations, underscoring the importance of this investigation despite Singapore's relatively stable climate.

Table of Contents

1. Background and Rationale

1.1 Asthma Burden in Singapore

Asthma represents a significant public health concern in Singapore, with 10.5% of the population diagnosed with lifetime asthma. Prevalence rates are particularly high among children (20%) compared to adults (5%) (Agency for Integrated Care, 2019), with one in three asthma patients reporting attacks in the past year. Singapore's asthma-related hospital admission rate is twice the OECD average, imposing a substantial economic burden of SGD 2.09 billion annually—approximately one-fifth of the national healthcare budget. Of this cost, approximately 80% stems from lost productivity rather than direct healthcare expenses (Lim et al., 2023).

Despite advances in asthma management, control remains suboptimal with only 38% of patients achieving well-controlled status, while 24% remain uncontrolled and 38% partially controlled. This poor disease control results in significant healthcare utilization, including emergency department visits and hospitalizations (Finkelstein et al., 2021).

1.2 Climate and Environmental Factors in Asthma Exacerbation

While Singapore experiences relatively consistent year-round temperatures compared to temperate regions, important climate variations do occur that appear to impact respiratory health:

- **Monsoon seasons** with distinct patterns:
  - Northeast Monsoon (November-March): More frequent and heavier rainfall, particularly in December and January.
  - Southwest Monsoon (May-September): Relatively drier conditions compared to the rest of the year.
  - Inter-monsoon periods: Characterized by different wind patterns and rainfall intensity.
- **Transboundary haze events** typically occurring between August-November due to regional forest fires, leading to dramatic spikes in air pollution measures (PSI and PM2.5).
- **Day-to-day variations** in temperature, humidity, and air quality that may impact respiratory health through both direct physiological effects and behavioural changes.

Preliminary analysis from the SingHealth dataset shows significant monthly variations in severe asthma exacerbations (2016~2019), with:

- Lower rates during July-September (Southwest Monsoon).
- Higher rates during November-December (early Northeast Monsoon).
- This pattern closely aligns with Singapore's monsoon seasons, with our preliminary statistical analysis showing significant differences between these periods ($p<0.05$).

These findings suggest that even in Singapore's relatively stable climate, subtle environmental variations may have important clinical implications for asthma patients.

1.3 Knowledge Gap and Research Opportunity

The existing literature on climate-asthma relationships focuses primarily on temperate regions with pronounced seasonal variations. Singapore's equatorial climate, characterized by high year-round temperatures, consistent humidity, and distinct air pollution patterns (including transboundary haze), offers an important yet understudied context for examining these associations.

The consistent year-round temperature in Singapore actually provides a scientific advantage: it allows us to isolate the effects of other environmental variables (air quality, rainfall patterns, humidity) without the confounding effect of substantial seasonal temperature swings seen in temperate regions.

Furthermore, the COVID-19 pandemic provided a natural experiment to observe changes in both environmental conditions and healthcare utilization patterns. This unique circumstance allows us to examine how pandemic-related measures affected both environmental exposures and asthma outcomes.

The preliminary findings of significant monsoon-related variations in asthma exacerbations, coupled with the documented respiratory effects of transboundary haze in the region, provide a compelling foundation for a comprehensive investigation of environmental drivers of acute asthma care in Singapore.

2. Research Questions and Objectives

2.1 Primary Research Question

To what extent and through what mechanisms do specific climate variables (temperature, rainfall, humidity) and air quality parameters (PSI, PM2.5) drive temporal variations in asthma-related emergency department visits in Singapore?

2.2 Secondary Research Questions

1. What is the quantifiable impact of monsoon seasonality on asthma exacerbation rates, controlling for other environmental and patient factors?
2. What are the precise temporal lag relationships (0~7 days) between environmental exposures and asthma exacerbations, and do these relationships differ by pollutant type?
3. Do extreme environmental conditions (defined by 5^{th} and 95^{th} percentiles) have disproportionate effects on asthma ED visits compared to moderate variations, and what is the magnitude of this differential impact?
4. How did the COVID-19 pandemic specifically alter relationships between environmental parameters and asthma exacerbations, and have these relationships returned to pre-pandemic patterns?
5. Are there identifiable patient subgroups (by age, GINA step, comorbidity profile) who demonstrate heightened sensitivity to specific environmental triggers?

2.3 Research Objectives

1. To characterize temporal trends in asthma-related ED visits from 2016~2024, identifying monsoon-related patterns and long-term changes.
2. To quantify associations between climate variables, air quality parameters, and asthma-related ED visits at daily, weekly, and monthly scales.
3. To determine optimal time lags between environmental exposures and resultant healthcare utilization.
4. To assess the impact of extreme climate conditions on asthma exacerbations.
5. To compare environmental factor associations with asthma-related ED visits before, during, and after the COVID-19 pandemic.
6. To identify vulnerable patient subgroups with heightened sensitivity to environmental triggers.
7. To develop a predictive model for asthma-related ED visits based on environmental variables.

2.4 Hypothesis

Based on preliminary data analysis and literature review, we can propose the following specific, testable hypotheses:

- **H1: Monsoon Seasonality Hypothesis**

Asthma-related ED visits in Singapore demonstrate a consistent seasonal pattern with 15~25% higher rates during the Northeast Monsoon (Nov-Mar) compared to the Southwest Monsoon (May-Sep), independent of patient demographics and comorbidities.

- **H2: Air Quality Dose-Response Hypothesis**

For every 10-unit increase in 24-hour PSI or 10 μg/m³ increase in PM2.5, asthma-related ED visits increase by 5~8%, with effects persisting for 3~5 days after exposure, following a non-linear dose-response relationship.

- **H3: Extreme Weather Threshold Hypothesis**

Environmental conditions above the 95^{th} percentile (for temperature, rainfall, pollution) or below the 5^{th} percentile (for temperature) are associated with 30~40% higher rates of asthma exacerbations compared to median conditions (25^{th}~75^{th} percentiles), with distinct thresholds identifiable for each parameter.

- **H4: Humidity-Temperature Interaction Hypothesis**

The combination of high humidity (>85%) and rapid temperature changes (>3°C within 24 hours) has a synergistic effect on asthma exacerbations, increasing ED visits by 20~25% compared to either factor alone.

- **H5: COVID-19 Impact Hypothesis**

The relationship between environmental factors and asthma ED visits was fundamentally altered during the COVID-19 pandemic, with a 40~50% reduction in environment-attributable visits during lockdown periods and a different pattern of seasonal variation that has not fully returned to pre-pandemic trends as of 2023.

- **H6: Vulnerability Gradient Hypothesis**

Patients with higher GINA treatment steps (4~5) and those with multiple respiratory comorbidities demonstrate 2~3 times greater sensitivity to adverse environmental conditions compared to those with well-controlled asthma (GINA steps 1~2).

- **H7: Transboundary Haze Hypothesis**

During documented transboundary haze episodes (PSI>100), asthma-related ED visits increase by 30~40% compared to equivalent periods without haze, with effects persisting for 7~10 days after air quality improves.

3. Methodology

3.1 Study Design

This study will employ a retrospective time-series analysis design, examining associations between environmental factors and asthma-related ED visits over an 8-year period (2016~2024). This longitudinal approach allows for identification of temporal patterns, seasonal variations, and long-term trends. We will first conduct exploratory data analysis to visualize patterns, followed by formal statistical testing of our specified hypotheses.

3.2 Study Setting and Population

The study will be conducted using data from the SingHealth Regional Health System, encompassing Singapore General Hospital (SGH) and SingHealth Polyclinics (SHP). The combined cohort includes 36407 patients with asthma, comprising 8816 patients from SGH and 29574 from SHP. SGH has a nationwide catchment area, while SHP serves primarily eastern Singapore.

Our preliminary analysis from this dataset has identified approximately 2990 patients with 21654 severe exacerbation cases recorded between 2016 and 2020, with a median of 4 exacerbations per patient (range: 1-175). Severe exacerbation cases were defined as:

- Outpatient visit with rescue therapy (n = 10762)
- ED admission with asthma diagnosis (J459) (n = 3130)
- Inpatient visit with asthma diagnosis (J459) (n = 3490)
- Application of oral corticosteroids (n = 5438)

This rich dataset allows for comprehensive analysis of environmental effects on asthma exacerbations.

3.3 Data Sources and Variables

3.3.1 Clinical Data

- **SingHealth COPD and Asthma Data Mart (SCDM)**: This database provides comprehensive clinical information on asthma patients in the SingHealth system, as described in the SDG-CARE collaboration.
- **Emergency Department Records**: Asthma-related ED visits identified using ICD-10 code J459 from the "07 Asthma ED Admission.csv" dataset, containing information on patient demographics, visit dates, and diagnoses.
- **GSK Master Cohort Database**: Contains detailed patient information including demographics, clinical characteristics, medication usage, comorbidities, and healthcare utilization patterns.

3.3.2 Environmental Data

- **National Environment Agency (NEA) Data**: Daily PSI, PM2.5, and other pollutant measurements from the "NEA_PSI_20150211_20231211.csv" dataset. This includes hourly historical air quality data (PM2.5, PM10, O3, SO2, NO2, CO) from 5 regional stations (North, South, East, West, Central) and national averages.

- **Meteorological Data**: Daily temperature, rainfall, and wind speed measurements from 13 weather stations across Singapore, available in the "final_cleaned_weather_data.csv" and "nea_historical_13_stations_weather_2016_2021.csv" datasets.
- **Transboundary Haze Episodes**: Documented periods of transboundary haze (PSI>100) will be identified and coded as separate variables to test the Transboundary Haze Hypothesis.
- **Monsoon Season Data**: Each day will be categorized according to monsoon season (Northeast, Southwest, Inter-monsoon) based on meteorological definitions from the Singapore Meteorological Service.

3.3.3 Variables of Interest

Table 2: Key Variables and Definitions

Variable Category	Specific Variables	Definition	Data Source
Outcome Variables	Daily ED visits	Count of asthma-related emergency department visits per day.	Asthma ED Admission.csv
	Hospitalization rates	Proportion of ED visits resulting in admission.	GSK Master Cohort
	Repeat ED visits	Count of patients with multiple ED visits.	GSK Master Cohort
Climate Variables	Mean Temperature (°C)	Daily average temperature.	final_cleaned_weather_data.csv
	Daily Rainfall Total (mm)	Total precipitation in a 24-hour period.	final_cleaned_weather_data.csv
	Mean Wind Speed (km/h)	Daily average wind velocity.	final_cleaned_weather_data.csv
Air Quality Variables	PSI	Pollution Standards Index	NEA_PSI dataset
	PM2.5 (μg/m³)	Concentration of fine particulate matter.	NEA_PSI dataset
Temporal Variables	Day of week	Calendar day (Monday~Sunday)	Derived
	Month	Calendar month	Derived
	Pandemic period	Pre-COVID/Early-COVID/Post-COVID	Derived
Patient Variables	Age	Patient age at time of visit.	GSK Master Cohort
	Gender	Male/Female	GSK Master Cohort
	Asthma severity	Based on GINA step	GSK Master Cohort

3.4 Data Processing and Preparation
3.4.1 Clinical Data Processing
1. **Data Extraction**: Retrieve asthma-related ED visits using ICD-10 code J459.
2. **Data Cleaning**: Remove duplicate entries, reconcile inconsistencies, and handle missing values.
3. **Patient Identification**: Create unique patient identifiers while ensuring anonymization.
4. **Temporal Aggregation**: Calculate daily, weekly, and monthly counts of ED visits.
5. **Pandemic Period Classification**: Categorize data into pre-COVID (before February 2020), early-COVID (February-May 2020), and post-COVID (June 2020 onwards) periods.
6. **Severe Exacerbation Definition**: Standardize definition of severe exacerbations across datasets based on our preliminary analysis definition.

3.4.2 Environmental Data Processing
1. **Data Integration**: Merge data from multiple weather stations using geospatial weighting based on population density.
2. **Missing Data Imputation**: Use appropriate methods (e.g., multiple imputation, nearest neighbor interpolation) to handle missing environmental data.
3. **Data Validation**: Cross-check with publicly available reports for consistency and validate against historical records.
4. **Extreme Value Identification**: Define 5th and 95th percentiles for each environmental variable to identify extreme conditions.

5. **Temporal Alignment**: Ensure alignment of environmental data with clinical outcomes for lag analysis.
6. **Monsoon Classification**: Classify each day according to monsoon season based on meteorological definitions.

3.4.3 Geospatial Data Processing

1. **Geocoding**: Assign geographical coordinates to patient residences based on postal codes.
2. **Environmental Exposure Assignment**: Link patients to nearest weather and air quality monitoring stations.
3. **Spatial Clustering**: Identify high-density areas of asthma exacerbations.
4. **Area-Level Variables**: Incorporate area-level environmental exposure data.

Table 3: Data Processing Steps and Methods

Processing Stage	Steps	Methods/Tools
Data Cleaning	Remove duplicates.	pandas.drop_duplicates()
	Handle missing values.	Multiple imputation, KNN imputation.
	Validate data consistency.	Cross-reference with published reports.
Temporal Processing	Aggregate by time periods.	Daily, weekly, monthly summaries.
	Create seasonal indicators.	Monsoon period flags.
	Create pandemic period indicators.	Pre/During/Post COVID categorization.
Spatial Processing	Assign spatial coordinates.	Geocoding with OneMap API.
	Calculate distances to pollution sources.	GIS spatial analysis.
	Define catchment areas.	Voronoi polygons around weather stations.
Environmental Processing	Calculate extreme thresholds.	5th and 95th percentiles.
	Create heatwave indicators.	≥3 consecutive days above 95th percentile.
	Create haze indicators.	PSI > 100 for ≥24 hours.

Table 4: Air Quality Classification Based on PSI Values

PSI Value	Air Quality Descriptor
0 ~ 50	Good
51 ~ 100	Moderate
101 ~ 200	Unhealthy
201 ~ 300	Very unhealthy
Above 300	Hazardous

(National Environment Agency, 2014)

3.5 Analytical Methods

3.5.1 Descriptive Analysis

1. **Temporal Trends**: Visualize and characterize patterns in ED visits, environmental variables, and their relationships over time, with specific focus on monsoon-related patterns.
2. **Summary Statistics**: Calculate means, medians, standard deviations, and ranges for all variables of interest, stratified by monsoon season and pandemic period.
3. **Extreme Event Characterization**: Identify and describe periods of extreme weather or air quality conditions and quantify associated changes in asthma ED visits.
4. **Exploratory Visualizations**: Create time series plots, heat maps, and other visualizations to identify patterns that inform hypothesis testing.

3.5.2 Time Series Analysis

1. **Time Series Decomposition**: Separate trends, seasonality, and irregular components in both ED visits and environmental variables using statistical decomposition methods.
2. **Correlation Analysis**: Perform Spearman's correlation analysis to identify associations between environmental factors and ED visits at various time scales (daily, weekly, monthly).
3. **Lag Analysis**: Conduct cross-correlation analyses with lags of 0-7 days to identify delayed effects of environmental exposures on ED visits, with specific analysis for each pollutant type.

4. **Monsoon Season Analysis**: Compare asthma exacerbation rates between monsoon seasons using appropriate statistical tests, controlling for other environmental factors.

3.5.3 Regression Modeling

1. **Poisson Regression**: Model daily counts of ED visits as a function of environmental variables while adjusting for confounders: ED Visits ~ $\beta_0 + \beta_1(\text{Temp}) + \beta_2(\text{Rain}) + \beta_3(\text{PM2.5}) + \beta_4(\text{Season})$ + covariates.
2. **Negative Binomial Models**: Use if overdispersion is detected in the count data.
3. **Distributed Lag Non-linear Models (DLNM)**: Apply to capture complex non-linear and delayed relationships between environmental exposures and health outcomes: ED Visits ~ β_0 + s(Temp, lag) + s(PM2.5, lag) + s(Rain, lag) + covariates Where s() represents smooth functions of the variables and their lags.
4. **Interaction Models**: Test for synergistic effects between environmental variables: ED Visits ~ $\beta_0 + \beta_1(\text{Temp}) + \beta_2(\text{Humidity}) + \beta_3(\text{Temp}\times\text{Humidity})$ + covariates
5. **Stratified Analysis**: Conduct separate analyses for different patient subgroups (age, gender, GINA step, comorbidities) and time periods (pre-, during, and post-COVID).

3.5.4 Extreme Climate Analysis

1. **Percentile-Based Comparison**: Compare ED visits during extreme conditions (5^{th} and 95^{th} percentiles) with those during normal conditions (25^{th}~75^{th} percentiles).
2. **Threshold Identification**: Use segmented regression or change-point analysis to identify specific thresholds beyond which environmental factors have disproportionate effects.
3. **Mann-Whitney U Tests**: Evaluate statistical significance of differences in ED visit rates during extreme versus normal conditions.
4. **Risk Ratio Calculation**: Calculate and compare risk ratios for asthma exacerbations during extreme versus normal environmental conditions.

3.5.5 COVID-19 Impact Analysis

1. **Interrupted Time Series Analysis**: Assess changes in trends before, during, and after COVID-19 restrictions using segmented regression models.
2. **Comparative Modeling**: Develop separate models for pre-, during, and post-COVID periods to identify changes in environmental factor coefficients.
3. **Chi-Square Tests**: Compare patient profiles across pandemic periods.
4. **Environmental Change Analysis**: Quantify changes in air quality and other environmental factors during lockdown periods and relate to changes in asthma exacerbations.

Table 5: Analytical Methods and Their Applications

Analysis Type	Specific Methods	Primary Purpose	Output Metrics
Descriptive Analysis	summary statistics	characterize distributions	means, medians, IQR, ranges
	time plots	visualize temporal patterns	trend graphs
	heat maps	visualize spatial patterns	spatial distribution maps
Correlation Analysis	Pearson/Spearman correlation	assess linear associations	correlation coefficients, p-values
	cross-correlation	identify lag relationships	lag correlation coefficients
	partial correlation	control for confounders	adjusted correlation coefficients
Regression Modeling	Poisson regression	model count data	rate ratios, 95% CI
	negative binomial	handle overdispersion	rate ratios, 95% CI
	DLNM	model complex lag structures	cumulative exposure effects
Comparative Analysis	Mann-Whitney U	compare non-normal distributions	p-values, effect sizes
	chi-square	compare categorical variables	p-values, proportions
	interrupted time series	assess intervention effects	level and slope changes

3.6 Software and Technical Implementation

The analysis will be conducted using Python with the following key packages:

- **pandas** and **numpy**: Data manipulation and preprocessing.
- **matplotlib** and **seaborn**: Data visualization.
- **scipy.stats**: Statistical testing.
- **statsmodels**: Time series analysis and regression modeling.
- **scikit-learn**: Machine learning algorithms for predictive modeling.
- **dlnm** (Python implementation): Distributed lag non-linear models.
- **geopandas** and **folium**: Geospatial analysis and visualization.

Table 6: Software Packages and Their Functions

Software Package	Version	Primary Functions
pandas	1.5.0+	data manipulation, cleaning, aggregation
numpy	1.22.0+	numerical operations, array handling
matplotlib/seaborn	3.5.0+/0.11.0+	data visualization, plotting
scipy.stats	1.8.0+	statistical testing, p-value calculation
statsmodels	0.13.0+	time series analysis, regression modeling
scikit-learn	1.0.0+	machine learning algorithms, cross-validation
geopandas	0.10.0+	geospatial data handling and analysis
folium	0.12.0+	interactive map creation

4. Expected Outcomes and Significance

4.1 Anticipated Results

Monsoon Seasonality Effect: A 15~25% increase in asthma-related ED visits during the Northeast Monsoon (Nov-Mar) compared to the Southwest Monsoon (May-Sep), with statistical significance after controlling for other environmental and patient factors.

Air Quality Impact: A dose-response relationship between air pollutants and asthma exacerbations, with a 5~8% increase in ED visits for each 10-unit increase in PSI or 10 μg/m³ increase in PM2.5.

Lag Effects: Peak effects of environmental exposures on asthma exacerbations occurring 3~5 days after exposure, with different lag patterns for different pollutant types.

Threshold Effects: Disproportionately higher impacts of extreme environmental conditions (above 95th or below 5th percentiles) compared to moderate variations, with a 30~40% increase in ED visits during extreme conditions.

Pandemic Impact: A marked reduction (40~50%) in environment-attributable asthma ED visits during the COVID-19 lockdown period, with altered patterns of seasonal variation that have not fully returned to pre-pandemic trends.

Vulnerability Gradient: Identification of vulnerable subgroups (GINA steps 4~5, multiple comorbidities) with 2~3 times greater sensitivity to adverse environmental conditions.

Transboundary Haze Effect: A 30~40% increase in asthma-related ED visits during and after documented transboundary haze episodes, with effects persisting for 7~10 days.

Table 7: Expected Findings Based on Preliminary Analysis

Environmental Factor	Expected Direction of Association	Expected Lag Period	Hypothesized Mechanism
PSI (overall)	positive correlation	3~5 days	airway inflammation from pollutant exposure
PM2.5	positive correlation	2~4 days	fine particle penetration into lower airways
Temperature (extreme high)	positive correlation	0~2 days	direct thermal stress on respiratory system
Temperature (extreme low)	positive correlation	1~3 days	cold-induced bronchospasm
Rainfall (heavy)	positive correlation	1~2 days	increased humidity, mold/fungal spore release
Wind speed (high)	variable	0~1 days	pollutant dispersion or concentration depending on conditions

4.2 Public Health and Clinical Implications

The findings from this study will have several important implications:

Enhanced Prediction Systems: Development of early warning systems for high-risk periods based on environmental forecasts, particularly during monsoon transitions and haze events.

Resource Allocation: Improved planning for healthcare resource distribution during predicted high-demand periods, especially during the Northeast Monsoon.

Targeted Interventions: Identification of high-risk locations and vulnerable patient populations for focused preventive measures.

Patient Education: Development of targeted guidance for asthma patients on managing environmental triggers specific to Singapore's climate patterns.

Medication Management: Evidence-based recommendations for prophylactic medication adjustment during high-risk environmental periods.

This research follows a cyclical "Identify-Intervene-Review" approach to population health management, with potential interventions including:

- Asthma action plan modifications for monsoon season transitions
- Air quality alert systems with specific thresholds for patient notification
- Preventive medication escalation protocols during high-risk periods

4.3 Contribution to Knowledge

This study will make several novel contributions to the scientific literature:

1. Provide the first comprehensive analysis of climate-asthma relationships in a tropical urban setting over an extended timeframe, with specific focus on monsoon patterns.
2. Establish precise dose-response relationships between environmental factors and asthma exacerbations in a tropical climate.
3. Identify specific thresholds at which environmental factors have disproportionate effects on asthma outcomes.
4. Document the impact of pandemic-related disruptions on both environmental exposures and healthcare utilization patterns.
5. Demonstrate the utility of integrated healthcare and environmental datasets for population health management in tropical settings.

4.4 Technological Integration and Innovation

Findings from this research may inform:

1. **Right-siting and Resource Allocation**: Optimal placement of healthcare resources based on geospatial analysis of asthma burden and environmental risk factors.
2. **Singapore Geospatial Master Plan 2024-2033**: Contribution to healthcare planning through geospatial insights on environment-health relationships.
3. **Preventive Healthcare Initiatives**: Support for preventive approaches through better understanding of environmental triggers and development of prediction systems.
4. **Digital Health Applications**: Development of patient-facing mobile applications that integrate real-time environmental data with personalized risk assessments.

5. Feasibility Assessment

5.1 Data Availability and Access

All required datasets have been secured through the SDG-CARE collaboration:

- SingHealth clinical data access is approved (Ref No. 2017/2950).
- NEA and Meteorological Service data are available from public databases:
  Historical weather: https://www.weather.gov.sg/climate-historical-daily/
  Legacy Pollutant Standards Index:

https://data.gov.sg/datasets/d_8a7850dc3993dc45f1620b9972c58d4d/view#tag/default/GET/environment/psi

- Data linkage protocols have been established and validated.

Our preliminary analysis has already demonstrated the feasibility of integrating these datasets and identifying significant patterns in asthma exacerbations related to environmental factors.

5.2 Technical Expertise

The research project members brings complementary expertise in:

- Respiratory medicine and asthma management.
- Environmental health and climate science.
- Data science and machine learning.
- Geospatial analysis and visualization.
- Biostatistics and epidemiology.

5.3 Timeline

Table 8: Project Timeline

Phase	Activities	Timeline	Deliverables
Phase 1: Data Preparation	data cleaning and preprocessing descriptive statistics and visualization initial correlation analyses	weeks 1~2	clean datasets data quality report descriptive statistics summary
Phase 2: Time Series Analysis	time series decomposition lag analysis implementation Poisson regression modeling extreme climate analysis	weeks 3~5	time series models lag correlation results initial regression models
Phase 3: COVID-19 Impact Analysis	stratified analyses by pandemic period subgroup analyses by patient characteristics comparative modeling	weeks 6~7	COVID impact report stratified analysis results
Phase 4: Geospatial Analysis	mapping of asthma exacerbation hotspots integration with environmental exposure data	weeks 8~9	spatial distribution maps environmental exposure overlays
Phase 5: Synthesis and Reporting	model refinement and validation manuscript preparation stakeholder presentations	weeks 10~12	final models draft manuscript presentation slides

6. Ethical Considerations

Ethics board approval has been obtained as part of the SDG-CARE collaboration through SingHealth Centralized Institutional Review Board (Ref No. 2017/2950). Informed consent has been waived due to the use of anonymized data and minimal risk to patients.

Data privacy and security will be maintained through:

- De-identification of all patient data.
- Secure data storage with restricted access.
- Analysis of aggregate data rather than individual cases where possible.
- Compliance with relevant data protection regulations.

Regular ethics review will be conducted throughout the project to ensure continued compliance with ethical standards.

7. References

Agency for Integrated Care (2019) *Singapore National Asthma Programme (SNAP)*. Available at: https://www.primarycarepages.sg/schemes-and-programmes/singapore-national-asthma-programme (Accessed: 7 March 2025).

Finkelstein, E.A., Lau, E., Doble, B., Ong, B. and Koh, M.S. (2021) 'Economic burden of asthma in Singapore', *BMJ Open Respiratory Research*, 8(1), e000654. Available at: https://doi.org/10.1136/bmjresp-2020-000654 (Accessed: 7 March 2025).

Lim, L.H.M., Chen, W., Amegadzie, J.E. and Lim, H.F. (2023) 'The increasing burden of asthma acute care in Singapore: an update on 15-year population-level evidence', *BMC Pulmonary Medicine*, 23(1), p. 502. Available at: https://doi.org/10.1186/s12890-023-02797-7 (Accessed: 7 March 2025).

National Environment Agency (2014) *Computation of the Pollutant Standards Index (PSI)*. Available at: https://www.haze.gov.sg/docs/default-source/faq/computation-of-the-pollutant-standards-index-%28psi%29.pdf (Accessed: 7 March 2025).

Nanyang Technological University (n.d.) *Writing a research proposal*. Available at: https://www.ntu.edu.sg/docs/librariesprovider120/graduate-by-research/how-to-write-a-research-proposaldf0d7d5d-7ab0-444a-b00c-24793d7b9aa8.pdf (Accessed: 11 March 2025).

Singapore Land Authority and Maritime and Port Authority of Singapore (2024) *Singapore Geospatial Master Plan 2024–2033*. Available at: https://file.go.gov.sg/singapore-geospatial-master-plan-2024.pdf (Accessed: 11 March 2025).

8. Appendices

Appendix A: Data Dictionary

Table A1: Clinical Dataset Variables

Dataset	**Variable**	**Description**	**Data Type**
07 Asthma ED Admission.csv	Patient.ID	Unique identifier for each patient	String
	Diagnosis.Institution.Code	Hospital code (SGH)	String
	Admit.Visit.Date	Date of ED visit	Date
	Diagnosis.Code	ICD-10 diagnosis code (J459)	String
	Diagnosis.Desc	"Asthma, unspecified"	String
GSK Master Cohort w Medications.csv	Patient ID	Unique identifier for each patient	String
	Age	Patient's age in years	Numeric
	Gender	Male/Female	Categorical
	Race	Chinese/Malay/Indian/Others	Categorical
	Final GINA STEP	Asthma treatment step (1-5)	Ordinal
	Exacerbation Counts	Annual exacerbation counts 2015-2024	Numeric
	No of ED Visits	Total emergency department visits	Numeric

Table A2: Environmental Dataset Variables

Dataset	**Variable**	**Description**	**Data Type**
NEA_PSI dataset	timestamp	Date and time of measurement	DateTime
	readings.psi_twenty_four_hourly.national	24-hour PSI	Numeric
	readings.pm25_twenty_four_hourly.national	24-hour PM2.5	Numeric
	readings.o3_eight_hour_max.national	8-hour maximum ozone	Numeric

final_cleaned_weather_data.csv	Station	Weather station location	String
	Date	Measurement date	Date
	Daily Rainfall Total (mm)	Daily precipitation	Numeric
	Mean Temperature (°C)	Average daily temperature	Numeric
	Mean Wind Speed (km/h)	Average wind velocity	Numeric

Appendix B: Analysis Code Templates

Time Series Decomposition:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from statsmodels.tsa.seasonal import seasonal_decompose

# load data
df = pd.read_csv('asthma_ed_visits.csv')
df['date'] = pd.to_datetime(df['date'])
df.set_index('date', inplace=True)

# resample to daily counts if needed
daily_visits = df.resample('D').count()

# apply seasonal decomposition
result = seasonal_decompose(daily_visits, model='additive', period=365)

# plot decomposition
fig, (ax1, ax2, ax3, ax4) = plt.subplots(4, 1, figsize=(12, 10))
result.observed.plot(ax=ax1)
ax1.set_title('Observed')
result.trend.plot(ax=ax2)
ax2.set_title('Trend')
result.seasonal.plot(ax=ax3)
ax3.set_title('Seasonality')
result.resid.plot(ax=ax4)
ax4.set_title('Residuals')
plt.tight_layout()
```

Lagged Correlation Analysis:

```
def lag_correlation(df, var1, var2, max_lag=7):
    """calculate correlation between var1 and var2 with lags 0 to max_lag"""
    correlations = []
    for lag in range(max_lag+1):
        correlation = df[var1].corr(df[var2].shift(lag))
        correlations.append((lag, correlation))
    return correlations

# calculate correlation between PSI and ED visits with lags 0~7
psi_correlations = lag_correlation(df, 'PSI', 'ed_visits', 7)

# plot results
lags, corrs = zip(*psi_correlations)
plt.figure(figsize=(10, 6))
plt.bar(lags, corrs)
plt.xlabel('Lag (days)')
plt.ylabel('Correlation coefficient')
plt.title('Correlation between PSI and ED visits with different lags')
```

Poisson Regression Model:

```python
import statsmodels.api as sm
import statsmodels.formula.api as smf

# prepare model data
model_data = df.copy()
model_data['dow'] = model_data.index.dayofweek  # day of week
model_data['month'] = model_data.index.month    # month

# fit Poisson regression model
model = smf.glm(
    formula='ed_visits ~ PSI + temperature + rainfall + windspeed + C(dow) + C(month)',
    data=model_data,
    family=sm.families.Poisson()
).fit()

# print summary
print(model.summary())
```

Extreme Climate Analysis:

```python
# define extreme temperature thresholds (5th and 95th percentiles)
temp_low = df['temperature'].quantile(0.05)
temp_high = df['temperature'].quantile(0.95)

# group data by extreme temperature days
extreme_low = df[df['temperature'] <= temp_low]
extreme_high = df[df['temperature'] >= temp_high]
normal = df[(df['temperature'] > temp_low) & (df['temperature'] < temp_high)]

# compare ED visits
from scipy.stats import mannwhitneyu

# test extreme low vs normal
low_test = mannwhitneyu(extreme_low['ed_visits'], normal['ed_visits'])
print(f"Extreme low temperature vs normal: p-value = {low_test.pvalue:.4f}")

# test extreme high vs normal
high_test = mannwhitneyu(extreme_high['ed_visits'], normal['ed_visits'])
print(f"Extreme high temperature vs normal: p-value = {high_test.pvalue:.4f}")
```