IE3100R Systems Design Project AY2024/25

DEVELOPING ENVIRONMENTAL INDICATORS FOR ASTHMA EXACERBATION MANAGEMENT IN SINGAPORE

Group 13 Team Members: Cheyenne Chia, Lee Yng-Yng, Ryan Quek Wei Heng, Tang Yee Teng
NUS Advisor: A/P Zhang Junyu | **Industry Supervisors:** Sean Lam; Adam Quek; See Wei Qiang; Wu Juntian

Introduction

Problem Description

1 in 10 Singaporeans suffer from **Lifetime Asthma**, resulting in an estimated **annual economic burden** of **$2.09 billion**

Asthma outcomes are multifaceted, a complex interplay of medical and **non-medical factors** — collectively known as **Social Determinants of Health (SDOHs)**, revealed to **account for up to 80% of health outcomes.**


Income


Education


Employment


Healthcare

and many more ...

SDOHs: "the conditions in which people are born, grow, work, live, and age, and the wider set of forces and systems shaping the conditions of daily life" (World Health Organization, 2013)

Project Aims & Objectives

To (i) **synthesise** a **representation** of environmental SDOHs impacting asthma patients in Singapore, and (ii) generate **evidence-based insights** to inform **more precise** and **targeted asthma management interventions**, by:

(i) Utilising **open-source** data to **develop** and **validate methodological approaches** for **quantifying environmental SDOH factors** that influence asthma exacerbation risk; and

(ii) Conduct **statistical analysis** to **identify key environmental risks**, thereby establishing the foundation for targeted interventions.

Data Sources & Characteristics

HOUSING & DEVELOPMENT BOARD

Public housing postal codes were extracted exclusively from the public housing dataset to form a **master open-source database.**

National Environment Agency
Safeguard · Nurture · Cherish

Daily PSI and pollutant readings of each region (North, South, East, West, Central) were **aggregated from 2017 to 2018.**

URBAN REDEVELOPMENT AUTHORITY

2019 LandUse MasterPlan used to calculate the **nearest Euclidean distance** between **each postal code** and **land use areas** (Roads, Green spaces, Business 1, Business 2, Religion).

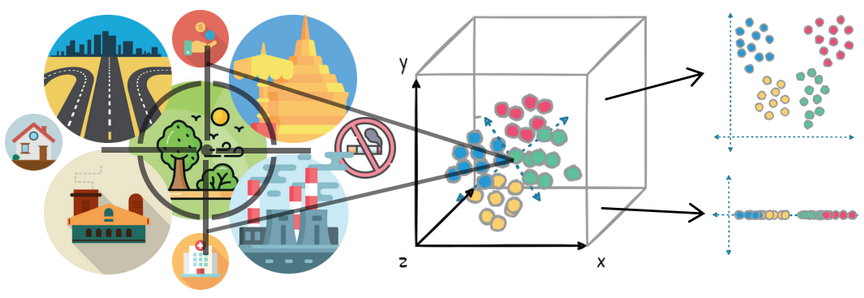
SingHealth
Defining Tomorrow's Medicine

Data of **21,215** patients and their 2019 exacerbation counts, filtered to **15,214** patients residing in public housing.

Methodology I: Indicator Creation

Principal Component Analysis (PCA)

Dimensionality Reduction (DR) technique that transforms a set of correlated features into a new set of uncorrelated variables, known as **Principal Components (PCs)**, while preserving maximal variance.



E.g. **100 or more** SDOH factors, of unknown individual importance, **all compressed** into a mere few **Principal Components (PCs)**

Methodology II: Indicator Validation

RIGOUR . RELEVANCE. UTILITY

(a) Construct Validation

Validate if the PCs obtained are able to **represent** the **selected environmental SDOH variables**.

01 Select Reference Standard
Construct concordance with the Pollutants Standard Index (PSI)

02 Geospatial Reconcilation
Mapping of postal codes to regions (East, West, North, South, Central)

03 Correlation Analysis
Use Pearson Correlation to examine exposure-response relationships

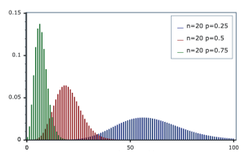
04 Methodological Adaptation
There are only **five** PSI (NEA) regions, hence, **high data interdependency**
- Deprioritised use of p-value
- Prioritised use of r: ($|r| \geq 0.75$) ≈ strong

Extracted Example of Mapped Patient Dataset

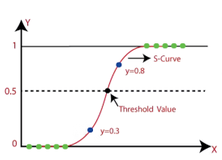
Postal	District	...	distance_temple	...	PSI_Region	PC1_Score_Rev	PC2_Score	PC3_Score
520351	East		0.95441299		79	1.715537	2.317094	3.752001
750493	North		1.80400358		94	-4.048429	-0.252622	-1.133357
...	...	...	...	...	...	...	...	...

(b) End-Use Validation

Apply output-validated indicator to **SingHealth patient data**; check for **over-dispersion** and **excess-zero counts**; fit to **regression models** to **validate** relationship between **indicator scores & exacerbation counts.**



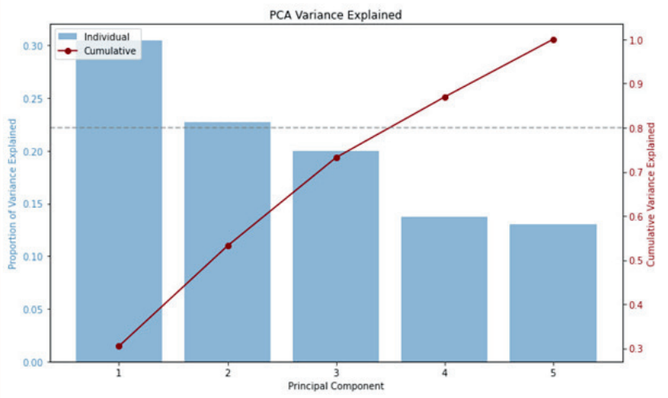
Negative Binomial
- Uses **top 5%-winsorised** exacerbation counts (2019) of **values [0, 7]**
- Accounts for **over-dispersion**
- Yields **Incident Rate Ratios (IRRs)**



Logistic
- Uses **binary** counts of **values 0 or 1**
- Accounts for **excess-zero counts**
- Yields **Odds Ratios (ORs)**

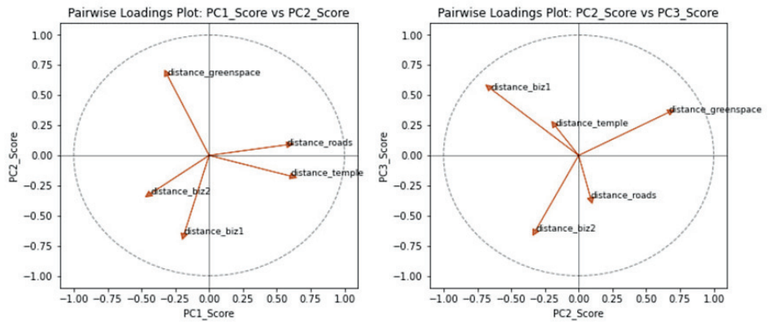
Results & Discussion

PC Selection & Indicators Creation



Three PCs retained using Kaiser criterion (eigenvalue > 1):
- To ensure consistent interpretability of PC scores, the coefficients of PC1 loadings were **reversed**

Feat. / PC	PC1	PC2	PC3
Distance to nearest temple	**0.5987**	-0.1711	0.2419
Distance to nearest CAT1 road	**0.5722**	0.0880	-0.3511
Distance to nearest "Business_1" labelled zone	-0.1841	**-0.6503**	0.5536
Distance to nearest "Business_2" labelled zone	-0.4310	-0.3162	**-0.6205**
Distance to nearest greenspace	-0.3074	**0.6634**	0.3559
Individual PC Eigenvalue	1.5250	1.1356	1.0004



PCs validate the hypothesised **inverse relationship** between **green spaces and pollution sources.**

Validation Against PSI & Its Constituents

Reversed PC1 (Urban_Pollution)
- Strongly correlated with SO_2 (r=0.86), attributed to vehicular emission and religious burning

PC2 (Greenspace_Mitigation)
- Strongly correlated with CO (r=0.76) & SO_2 (r=-0.79)
- Partially aligned expectations, as high CO levels could be attributed to combustion patterns from light industrial PM, but contradict expectations where SO_2 levels are low despite increasing distance from green spaces

PC3 (Industrial_Pollution)
- Strongly correlated with PM10 (r=-0.98) & PM2.5 (r=-0.77)
- Activities within the "Business_2" zones (labelled by URA), does not effectively capture correlations against environmental pollutants such as PM10 & PM2.5

Validation Against Patients' Exacerbation Counts

Negative Binomial Regression Results

Parameter	Coeff	IRR	90% CI	p-value
Urban_Pollution	0.0366	1.0373	1.0211 — 1.0538	**1.28E-04**
Greenspace_Mitigation	0.0033	1.0033	0.9857 — 1.0214	0.7615
Industrial_Pollution	-0.0397	0.9610	0.9458 — 0.9766	**4.84E-05**

Logistic Regression Results

Parameter	Coeff	OR	90% CI	p-value
Urban_Pollution	0.0682	1.0707	1.0424 — 1.7862	**2.83E-05**
Greenspace_Mitigation	-0.0162	0.9839	0.9544 — 1.0147	0.3846
Industrial_Pollution	-0.0048	0.9952	0.9682 — 1.0234	0.7774

Significant p-values (<0.05) demonstrate **indicators' success** in quantifying the **relationship against patients' asthma exacerbation** counts.

Conclusion

Key Findings

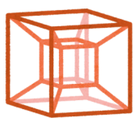
- **Urban_Pollution: IRR=1.037, OR=1.07 —** Each unit increase **raises** exacerbation **counts** by **3.7%** and **odds** of experiencing exacerbation (counts ≥ 1) **by 7%**.
- **Greenspace_Mitigation:** No significant relationship with asthma exacerbation.
- **Industrial_Pollution: IRR=0.9610 —** Each unit increase **reduces** exacerbation **counts** by **3.9%**

Significance of Findings

- Strategic urban environmental interventions should be implemented
- Measures, beyond green spaces, should be explored to find a better representation of the mitigating factors of pollutants
- Asthma exacerbation has a more complex and nuanced relationship with "Business_2" labelled zones, necessitating further studies

Limitations & Future Work

Regionally limited patient data: SingHealth facilities concentrated in the **East** and **North-East** regions
- Collaborate with additional healthcare providers to construct a more balanced distribution of patients across different regions



Limited scope of SDOH: only environmental factors explored despite multiple facets/dimensions of SDOHs of Asthma
- Incorporate other dimensions (i.e. socioeconomic, behavioural etc.)

Temporal limitations: only single-year, single-valued patient exacerbation data are available
- Longitudinal cohort studies of patients spanning multiple years, even decades, provides more robust temporal insights

KEY SKILLS & TAKEAWAYS:


Python | Data Wrangling | Web-Scraping | Machine Learning


Data Analytics | Geospatial Data Utilization


Project Management | Systems Thinking


Group Collaboration | Communication