

Multiple Decrement Modeling of BPJS Health Transitions Under Air Pollution Exposure in DKI Jakarta

Bernadeth Saskia Laudya Cintya¹, Leyton Raynaldo Ozzora^{1*}, Helena Margaretha¹, Ferry Vincenttius Ferdinand¹

¹ Mathematics Department, Faculty of Science and Technology, Universitas Pelita Harapan, Tangerang, Indonesia

*Corresponding email: leytonzzora25@gmail.com

Abstract

Air pollution represents a significant environmental health burden in Indonesia. Despite this, the quantitative relationship between ambient pollutant concentrations and population-level health state transitions within Indonesia's national health insurance system remains uncharacterized. This study estimates cause-specific forces of decrement for health state transitions within the BPJS Kesehatan administrative claims data from DKI Jakarta spanning 2022 to 2024, and examines the association between ambient air pollutant concentrations and estimated transition intensities through exploratory LASSO regression. The analytical dataset comprises 454,001 FKTP and 147,379 FKRTL episodes, stratified by five-year age bands and calendar quarter. Multiple decrements across both care settings are modeled, with the FKTP-to-FKRTL referral (*rujuk lanjut*) decrement serving as the primary focus. Findings provide preliminary evidence that ambient air pollution variables may contain useful information for exploratory stress-testing of BPJS health transition rates. NO₂ emerged as a candidate predictor in exploratory screening, appearing with a positive coefficient in 6 of 17 age-band-specific models, with cross-validation RMSE ranging from 5.28 to 38.17 per person-year and model-implied elasticities ranging from 0.063 to 0.544. Scenario analysis showed that a hypothetical 20% increase in ambient NO₂ corresponded to fitted referral intensity increases of between 1.3 and 10.9 percent across retained age bands.

Article History:

Submitted/Received: 15 May 2026

First Revised: 26 June 2026

Accepted: 20 July 2026

Published: 31 July 2026

Keywords:

Air Pollution, BPJS Kesehatan, Force of Decrement, FKTP, FKRTL

1. INTRODUCTION

Air pollution represents one of the most significant environmental health risks globally, responsible for an estimated 9 million premature deaths annually [1]. Long-term epidemiological evidence demonstrates that chronic exposure to particulate matter (PM_{2.5}) and traffic-related nitrogen dioxide (NO₂) consistently increases the risk of respiratory, cardiovascular, and all-cause mortality [2], with the adverse effects of NO₂ shown to be independent of concurrent particulate exposure [3]. In Indonesia, the burden is particularly acute in its capital: PM_{2.5} concentrations in Jakarta consistently exceed the World Health Organization annual guideline of 5 µg/m³ by a factor of six to eleven, with annual average levels across the greater metropolitan area ranging from 30 to 55 µg/m³ throughout the study period [4], while NO₂, which is the primary marker of traffic-related combustion in a city of over ten million motor vehicles, represents a further and distinct exposure pathway with documented associations with respiratory morbidity and healthcare utilization [3]. The annual health and economic cost of this exposure in Jakarta alone has been estimated at approximately USD 2.9 billion [5]. Despite this documented burden, the quantitative relationship between ambient pollutant concentrations and population-level health transitions within Indonesia's formal healthcare system remains poorly characterized.

Indonesia's national health insurance program, Jaminan Kesehatan Nasional (JKN), administered by BPJS (Badan Penyelenggara Jaminan Sosial) Kesehatan, provides a unique opportunity to examine these transitions at scale, covering approximately 95% of Indonesia's population as of 2023. Under BPJS, enrolled individuals access care through a tiered referral system: primary care facilities (Fasilitas Kesehatan Tingkat Pertama,

FKTP) serve as first points of contact, with referrals to advanced care facilities (Fasilitas Kesehatan Rujukan Tingkat Lanjutan, FKRTL) occurring when conditions exceed primary care capacity. Each referral represents not only a clinical escalation but also a substantial increase in claim expenditure for BPJS. In the DKI Jakarta dataset analyzed in this study, FKTP facilities recorded 454,001 episodes across 31,590 unique enrollees during 2022–2024, generating 34,626 referrals to FKRTL.

The actuarial literature has established multiple decrement models as a rigorous framework for quantifying competing causes of exit from a given state, with transition intensities estimated empirically from administrative claims data [6]. In the domain of health insurance, these models have been applied to disability income insurance, multi-state morbidity transitions, and social insurance systems, using individual-level covariates to construct decrement tables [7, 6]. Separately, the environmental epidemiology literature has produced substantial evidence linking ambient air pollutant concentrations to elevated all-cause and cause-specific mortality and morbidity, with consistent findings across cohort studies in Asia, North America, and Europe [8, 9, 10]. Within the Indonesian context, studies leveraging BPJS Kesehatan administrative data have characterized healthcare utilization patterns and referral dynamics primarily through sociodemographic and clinical lenses, documenting disparities in access, facility readiness, and chronic disease burden [11, 12, 13].

Despite substantial progress across these three strands of literature, no study has incorporated ambient environmental exposures as time-varying covariates within an actuarial multiple decrement framework applied to national health insurance claims data. Environmental epidemiology establishes the health burden of pollution without translating findings into actuarial transition quantities, BPJS utilization studies exploit rich administrative data while treating the exposure environment as fixed, and actuarial health state models rely exclusively on static demographic inputs, leaving environmental drivers of claim transitions unquantified. If pollution systematically elevates referral intensity, decrement tables constructed under this omission will underestimate transition volatility, with direct consequences for premium adequacy and reserve sufficiency [14, 15]. This study addresses these gaps using BPJS administrative claims data linked to city-wide ambient air quality measurements across six pollutants. To formalize the objective of quantifying the association between environmental exposures and referral intensities, this research seeks to answer the following questions:

1. What are the estimated forces of decrement across BPJS health transitions?
2. Which pollutants are most strongly associated with referral intensities?
3. Can environmental variables contribute to actuarial stress-testing frameworks?

2. METHODS

This study proceeds in two sequential stages. In the first stage, cause-specific forces of decrement are estimated from episode records for FKTP and FKRTL separately, stratified by five-year age bands and calendar quarter. In the second stage, the estimated transition intensities are regressed against quarterly ambient air pollution concentrations using LASSO, and scenario analyses are conducted to quantify the actuarial implications of hypothetical pollution shifts.

2.1. Data Sources and Collection

Episode-based data were obtained from BPJS Kesehatan website through formal data access process [16]. To ensure strict data governance, all administrative records were anonymized by the provider prior to access. Sensitive personal identifiers were removed, and patients were assigned dummy IDs. These dummy IDs enable navigation through different BPJS files and across available years while ensuring that no personally identifiable information was accessed [11]. Furthermore, data access is strictly governed; researchers are required to register, undergo verification, and sign an officially stamped Integrity Pact to legally bind them to data security protocols.

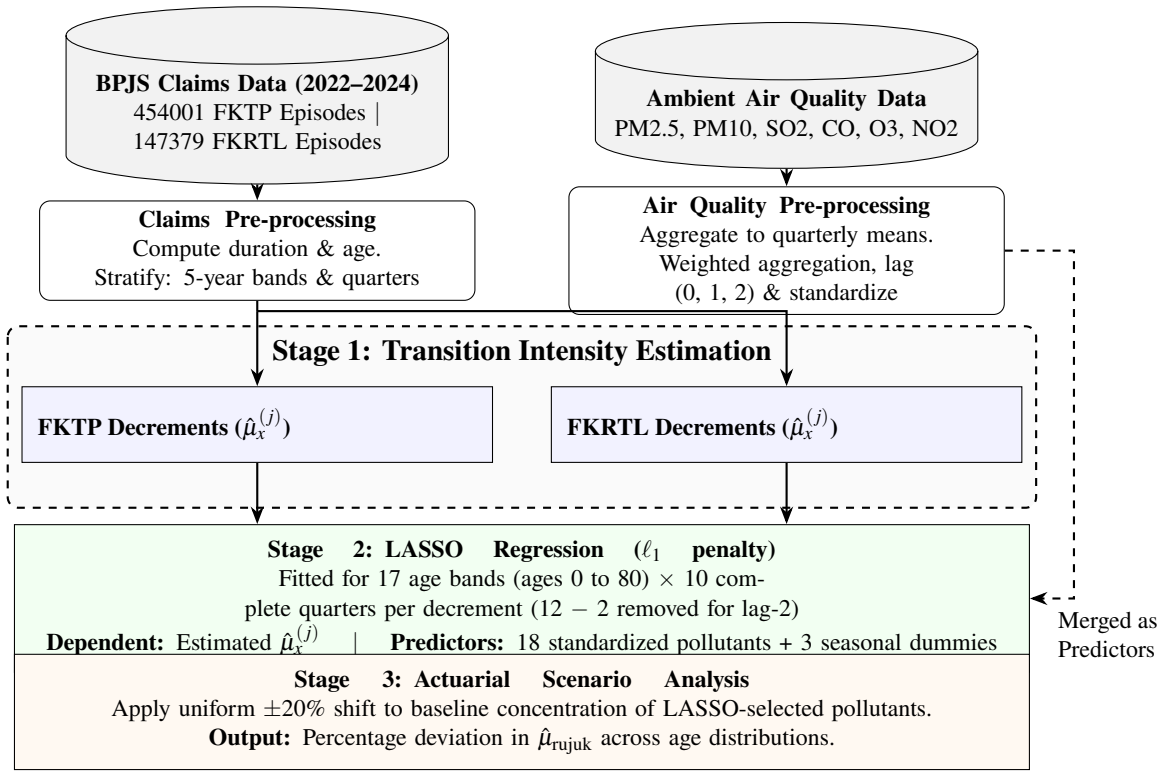


Figure 1. Sequential workflow mapping

By mid-2022, Indonesia’s weekly COVID-19 case and mortality counts had entered a sustained declining trend as the country transitioned toward endemicity [17]. Accordingly, this study focuses on 2022-2024 data in order to reflect post-pandemic conditions and to reduce the likelihood that pandemic era disruptions would systematically bias the estimated transition intensities. Each observation corresponds to a single episode of care. The final analytical dataset comprises 454,001 FKTP episodes and 147,379 FKRTL episodes from DKI Jakarta. Air quality data were obtained from the Satu Data Jakarta open data portal [18], covering six pollutants: PM2.5, PM10, SO2, CO, O3, and NO2.

2.2. Data Pre-processing

Prior to analysis, both claims datasets underwent a standardized preparation pipeline. For each care setting, episode records across 2022 through 2024 were merged into a single dataset, retaining IDs, entry and exit dates, exit status, primary diagnosis, and facility type. Episode duration was computed as the difference in days between the exit and entry dates and appended as an additional variable. Age at episode entry was obtained by linking each ID to the corresponding date of birth in the BPJS membership file, with age computed as the difference in years between date of birth and episode entry date. Episodes were then assigned to five-year age bands and calendar quarters to enable estimation of the forces of decrement. For the environmental covariates, city-wide averages were assigned to all individuals. Consequently, true individual exposure may differ substantially from these city-wide averages due to variations in residential location, commuting behavior, and localized occupational exposure.

For the air quality data, daily observations were first averaged across active stations to produce daily city-wide concentrations, then aggregated to monthly means. During the 36-month study period, 34 of 180 possible station-month observations (18.89 %) were missing due to partial station inactivity; these were accommodated by taking the simple average across the available active stations for that month. Where station readings were completely unavailable for a given month, which occurred in 2 of the 36 months (5.56 % of the data), values were linearly interpolated from adjacent periods to preserve continuity in the lagged predictor set. These interpolated

months were assigned a weight of zero in the quarterly aggregation step and therefore do not contribute to the quarterly means used as concurrent pollution predictors, ensuring they inform only the lag structure without entering the concurrent exposure estimates. Monthly values were subsequently aggregated to quarterly means using a weighted average, with weights proportional to the number of active stations in each month. Quarterly pollution concentrations were merged with the quarterly claims data, such that each episode-quarter observation is associated with the corresponding quarterly pollutant averages. Lagged pollution terms at one and two quarters prior were constructed and appended to form the predictor set used in the LASSO regression stage. All pollution variables were standardized to zero mean and unit variance prior to model fitting to ensure comparability of LASSO-penalized coefficients across pollutants with different measurement scales.

2.3. Force of Decrement Estimation

Let $\mu_x^{(j)}(t)$ denote the force of decrement due to cause j for an individual aged x at time t . Under the multiple decrement framework, individuals occupying a given care state are simultaneously subject to m competing causes of exit, each operating independently [6]. The probability that an individual exits due to cause j within a small interval dt is approximately $\mu_x^{(j)}(t) dt$, and the total force of exit is $\mu_x^{(\tau)}(t) = \sum_{j=1}^m \mu_x^{(j)}(t)$. This assumption of independence is a foundational requirement in standard actuarial multiple decrement modeling. By assuming independence, the model isolates the cause-specific transition intensities, effectively mirroring the construction of an associated single decrement table. This ensures that the empirical force of decrement, $\hat{\mu}_x^{(j)}$, captures the absolute operational intensity of a specific transition without being mathematically conflated by the probability distributions of competing exits like death or recovery. This formulation provides a practical actuarial approximation for estimating cause-specific transition intensities and for constructing modular stress-testing assumptions. However, it does not imply that the underlying health processes are biologically independent.

For most decrements, the force of decrement is estimated empirically using the episode-based exposure method. Let $d_x^{(j)}$ denote the observed number of exits due to cause j among individuals in age band $[x, x+5)$ during a given calendar quarter, and let $E_x^{(j)}$ denote the corresponding central exposed-to-risk, measured in person-years. The exposed-to-risk for each episode is computed as the portion of the episode duration falling within the age band. Same-day episodes are assigned a minimum duration of 0.5 days to prevent a zero denominator in the exposure calculation. Consistent with standard actuarial practice, this conservative floor acknowledges a partial contribution to at-risk person-time while avoiding the upward bias of assigning a full day. The empirical estimator is then:

$$\hat{\mu}_x^{(j)} = \frac{d_x^{(j)}}{E_x^{(j)}} \quad (1)$$

This estimator is the maximum likelihood estimator of the force of decrement under the assumption of a constant force within each age-quarter cell [6]. The resulting force of decrement estimates are expressed in units of person-years⁻¹ and should be interpreted accordingly. Because most FKTP and FKRTL episodes are short-duration, with the majority resolving within days rather than weeks, the central exposed-to-risk per episode is a small fraction of a person-year, and correspondingly the estimated forces of decrement for high-frequency exit types such as healthy visits and outpatient care take large values. A force of decrement of $\hat{\mu} = 400$ per person-year, for instance, is equivalent to an expected episode duration before that type of exit of approximately $400^{-1} \times 365 \approx 0.9$ days, which is consistent with the administrative episode lengths observed in the dataset. These values are not anomalous; they reflect the episode-based rather than calendar-year-based construction of the exposed-to-risk.

The treatment of each decrement varies by care setting and data availability, as summarized in **Table 1**. For FKTP, empirical forces of decrement are estimated for Kunjungan Sehat, Berobat Jalan, Rujuk Lanjut, and Sembuh. FKTP Meninggal and Lain-lain are assigned a force of decrement of zero due to insufficient observed

events across age-quarter cells. For FKRTL, empirical estimation is applied to Sehat and Rujuk, while Pulang Paksa is assigned zero. Tidak Tahu is treated as a residual cause and excluded from the regression stage.

Table 1. Treatment of decrements by care setting

Setting	Decrement	Treatment
FKTP	Healthy Visit	Empirical $\hat{\mu}$
	Outpatient care	Empirical $\hat{\mu}$
	Referral to FKRTL	Empirical $\hat{\mu}$
	Recover	Empirical $\hat{\mu}$
	Others	$\hat{\mu} = 0$ (insufficient data)
	Death	$\hat{\mu} = 0$ (insufficient data)
FKRTL	Healthy Visit	Empirical $\hat{\mu}$
	Referral in FKRTL	Empirical $\hat{\mu}$
	Death	Gompertz MLE
	Self-discharge	$\hat{\mu} = 0$ (insufficient data)
	Unknown	Residual (excluded from regression)

For FKRTL Meninggal, the crude empirical estimator is unreliable due to sparse death counts: only 301 deaths were recorded across 17 age bands and 12 quarters, yielding a median of fewer than two events per age-quarter cell and precluding stable stratified estimation. Parametric graduation is therefore applied to smooth the mortality intensity across the age dimension. The Gompertz force of mortality is specified as:

$$\mu_x = Bc^x \quad (2)$$

where $B > 0$ is a scale parameter and $c > 1$ governs the rate of exponential increase with age [6]. The Gompertz law of mortality is selected on three grounds. First, it is the canonical parametric form for human mortality at adult ages and has a well-established theoretical derivation as the force of mortality implied by exponentially increasing biological deterioration with age [19, 20]. Second, its log-linear structure in age means that parameter estimation via maximum likelihood on individual episode records is computationally tractable even under sparse event counts, and a preliminary log-linear regression of age-band-aggregated crude rates on age band midpoints confirmed an approximately linear relationship between $\log \hat{\mu}_x$ and age, supporting the functional form empirically. Third, relative to the Makeham extension $\mu_x = A + Bc^x$, the additional parameter A representing age-independent background hazard could not be reliably identified with 301 total events; the simpler two-parameter Gompertz was therefore preferred on parsimony grounds [6].

Parameters are estimated by maximum likelihood on the full pooled individual-level episode record from FKRTL, without quarter stratification. The log-likelihood is:

$$\ell(B, c) = -\frac{B}{\log c} \sum_i c^{x_i} (c^{t_i} - 1) + \sum_i \delta_i (\log B + (x_i + t_i) \log c) \quad (3)$$

where x_i is the age at episode entry, t_i is the episode duration in years, and $\delta_i \in \{0, 1\}$ is the death indicator for episode i [6]. Optimization is performed via the Nelder-Mead algorithm with starting values obtained from a log-linear regression of age-band-aggregated crude rates on age band midpoints. The fitted intensity $\hat{\mu}_x = \hat{B}c^x$ is evaluated at age band midpoints and applied uniformly across quarters, assuming no seasonal variation. This temporal pooling is empirically necessary due to severe event sparsity: with only 301 total deaths, the median of fewer than two events per age-quarter cell cannot reliably support stratified estimation. While sacrificing temporal granularity, this approach ensures a stable age pattern.

2.4. Pollution Regression

To examine the association between ambient air pollution and estimated transition intensities, a separate LASSO regression is fitted for each combination of decrement and age band. Age bands [85, 90) and above are excluded from this analysis due to severe outcome sparsity, with more than 80% zero-cell counts in the observed decrement counts across quarters; the analytical range covers 17 age bands from [0, 5) to [80, 85). This yields 17 independent LASSO models per decrement, each fitted on 10 quarterly observations.

The dependent variable in each model is the empirical force of decrement $\hat{\mu}_x^{(j)}$ for the corresponding age band and decrement, as estimated in the preceding stage. The predictor set comprises 21 variables: 18 pollution terms constructed from six ambient pollutants (PM2.5, PM10, SO2, CO, O3, and NO2) at lag 0, lag 1, and lag 2 quarters, plus three seasonal dummy variables for Q1, Q2, and Q3, with Q4 as the baseline category. The LASSO estimator minimizes the penalized objective:

$$\hat{\beta} = \arg \min_{\beta} \left\{ \sum_{t=1}^T \left(\hat{\mu}_{x,t}^{(j)} - \beta_0 - \sum_{k=1}^K \beta_k z_{k,t} \right)^2 + \lambda \sum_{k=1}^K |\beta_k| \right\} \quad (4)$$

where $z_{k,t}$ denotes the k -th standardized predictor at quarter t , $T = 12$ is the number of quarterly observations, $K = 21$ is the number of predictors, and $\lambda \geq 0$ is the regularization parameter. The intercept β_0 is not penalized [21].

LASSO is selected over ordinary least squares on empirical grounds: variance inflation factors exceeding 60 for PM-related predictors confirm extreme multicollinearity among the pollution variables, rendering OLS estimates unstable [22]. LASSO addresses this by imposing an ℓ_1 penalty on the coefficient vector, shrinking redundant predictors to exactly zero and retaining only the variables with the strongest independent associations [23]. The regularization parameter λ is selected via 5-fold cross-validation using the `lambda.1se` criterion, which selects the most parsimonious model within one standard error of the minimum cross-validation error, further reducing the risk of overfitting given the limited sample size.

It must be acknowledged that the statistical setting for LASSO estimation in this study is severely constrained. After removing two quarters to construct lag-2 predictors, each age-band model is fitted on $n = 10$ quarterly observations with $K = 21$ predictors, yielding a predictor-to-observation ratio of approximately 2:1. Under ordinary least squares, this system would be underdetermined. While LASSO resolves this underdetermined system by shrinking most coefficients to zero, the resulting feature selection remains highly sensitive to small data perturbations. Because the short 2022–2024 observation window precludes robustness checks such as bootstrap stability analysis, the LASSO results must be interpreted solely as variable-screening outputs and not inferential evidence. A non-zero coefficient reflects retention under this specific specification, not necessarily a robust association that would persist across different lag structures or longer time series. This limitation is inherent to the study design and cannot be resolved without extending the observation period beyond the 2022–2024 window available from BPJS Kesehatan administrative records.

2.5. Scenario Analysis

To translate the regression results into actuarially interpretable quantities, scenario analyses are conducted for the pollutant most frequently retained by LASSO with nonzero coefficients across age bands. The analysis applies a uniform $\pm 20\%$ shift to the quarterly baseline concentration of the selected pollutant, computed as the mean observed concentration over 2022–2024, while applying the same proportional shift to lag-0, lag-1, and lag-2 values of the selected pollutant, and holding all remaining predictors at their observed values. The implied change in $\hat{\mu}_{\text{rujuk}}$ is computed for each age band and expressed as a percentage deviation from the baseline force of decrement, providing an elasticity-style summary of environmental sensitivity across the age distribution.

3. RESULT AND DISCUSSION

This section presents the empirical transition-intensity patterns and the exploratory relationship between ambient pollutants and the forces of decrement across modeled BPJS health state transitions. The analysis first summarizes all modeled decrements to provide a structural overview of the claims data. The main discussion then focuses on the FKTP-to-FKRTL referral decrement, which produced the clearest and most interpretable pollutant-related signal in the dataset. Results for the remaining decrements are reported as secondary findings, as their pollutant-related patterns were weaker, less stable, or insufficient for detailed interpretation.

3.1. Summary of Modeled Decrements

Table 2. presents a structural overview of all modeled decrements based on BPJS DKI Jakarta claims data from 2022 to 2024. The pollutant analysis was restricted to age bands [0,5) through [80,85); age bands [85,90) and above were excluded due to event sparsity, with zero quarterly counts observed in the majority of age-quarter cells for these bands, preventing reliable force of decrement estimation. All statistics in **Table 2.** reflect this restriction.

Table 2. Summary of modeled decrements based on BPJS claims data, DKI Jakarta 2022–2024. Analysis restricted to age bands [0,5) through [80,85)

Decrement	Level	Events	Mean $\hat{\mu}$	Highest band
Healthy Visit	FKTP	259,699	411.10	[10, 15)
Outpatient Care	FKTP	159,414	251.31	[0, 5)
Referral to FKRTL	FKTP	34,626	66.62	[75, 80)
Recovery	FKTP	254	0.34	[25, 30)
Healthy Visit	FKRTL	135,003	457.73	[70, 75)
Referral in FKRTL	FKRTL	1,991	6.82	[5, 10)
Death (Gompertz)	FKRTL	301	1.11	[80, 85)

The modeled decrements differ substantially in event frequency and average intensity. Healthy visits and outpatient care exhibited the highest estimated forces of decrement, with mean values of 411.10 and 251.31 per person-year respectively for FKTP, and 457.73 per person-year for healthy visits at FKRTL. These values are consistent with the episode-based exposure construction: because most episodes resolve within one to three days, the denominator of equation (1) is small in person-year terms, producing large intensity estimates that correctly reflect rapid episode turnover rather than constituting anomalous outliers. The high outpatient intensity in the [0,5) age band is consistent with documented patterns of elevated healthcare utilization among young children in Indonesian primary care settings [12]. In contrast, the referral, recovery, and mortality decrements exhibited substantially lower intensities, consistent with their relative rarity as episode outcomes.

Table 3. summarizes the LASSO screening results across all seven modeled decrements. The results show that decrements differ substantially not only in their empirical intensity levels but also in the degree to which ambient pollutant variables were retained as informative predictors.

Among the modeled decrements, the FKTP-to-FKRTL referral decrement had the largest number of age bands with non-zero LASSO coefficients (11 of 17) and the most interpretable pollutant-related signal, with CV-RMSE ranging from 5.28 to 38.17 per person-year. Outpatient care also retained non-zero coefficients in 11 of 17 age bands, but its pollutant pattern was dominated by SO₂ lag-2 with a negative sign that is difficult to attribute to a direct biological mechanism, as discussed further in Section 3.7. The recovery decrement could only be attempted in 9 of 17 age bands due to zero variance in estimated intensity for the remaining bands, and is treated as unmodelable. The death decrement at FKRTL, while technically fitted across 17 bands, produced CV-RMSE values up to 5.39 per person-year against a mean intensity of only 1.15, indicating that the

Table 3. Summary of LASSO screening results across all modeled decrements

Decrement	Level	Bands fitted	Non-zero bands	Mean $\hat{\mu}$	CV-RMSE range	Status
Referral to FKRTL	FKTP	17	11	66.62	5.28–38.17	Primary
Outpatient Care	FKTP	17	11	251.31	14.82–61.48	Secondary
Healthy Visit	FKTP	17	9	411.10	20.58–84.71	Secondary
Recovery	FKTP	9	1	0.34	0.15–0.70	Unmodelable
Healthy Visit	FKRTL	17	4	457.73	18.35–87.31	Secondary
Referral in FKRTL	FKRTL	17	4	6.82	0.98–3.61	Secondary
Death (Gompertz)	FKRTL	17	4	1.11	0.43–5.39	Inconclusive

crude intensity estimates were too noisy for reliable pollutant screening. These results collectively support the decision to focus the primary pollutant analysis on the FKTP-to-FKRTL referral decrement. Secondary results for the remaining decrements are reported in Section 3.7.

3.2. Empirical Referral Intensity by Age Band

The estimated FKTP-to-FKRTL referral intensity showed a general increase with age, though the pattern was not strictly monotonic across all age bands. The highest average referral intensities were observed in the [75,80) and [70,75) age bands, with mean values of approximately 141.1 and 122.9 per person-year, respectively. Younger age bands generally exhibited lower and more stable intensities. The [80,85) age band showed a relatively high standard deviation (36.0) relative to its mean (96.5), indicating that referral intensity was particularly volatile in this group over the study period.

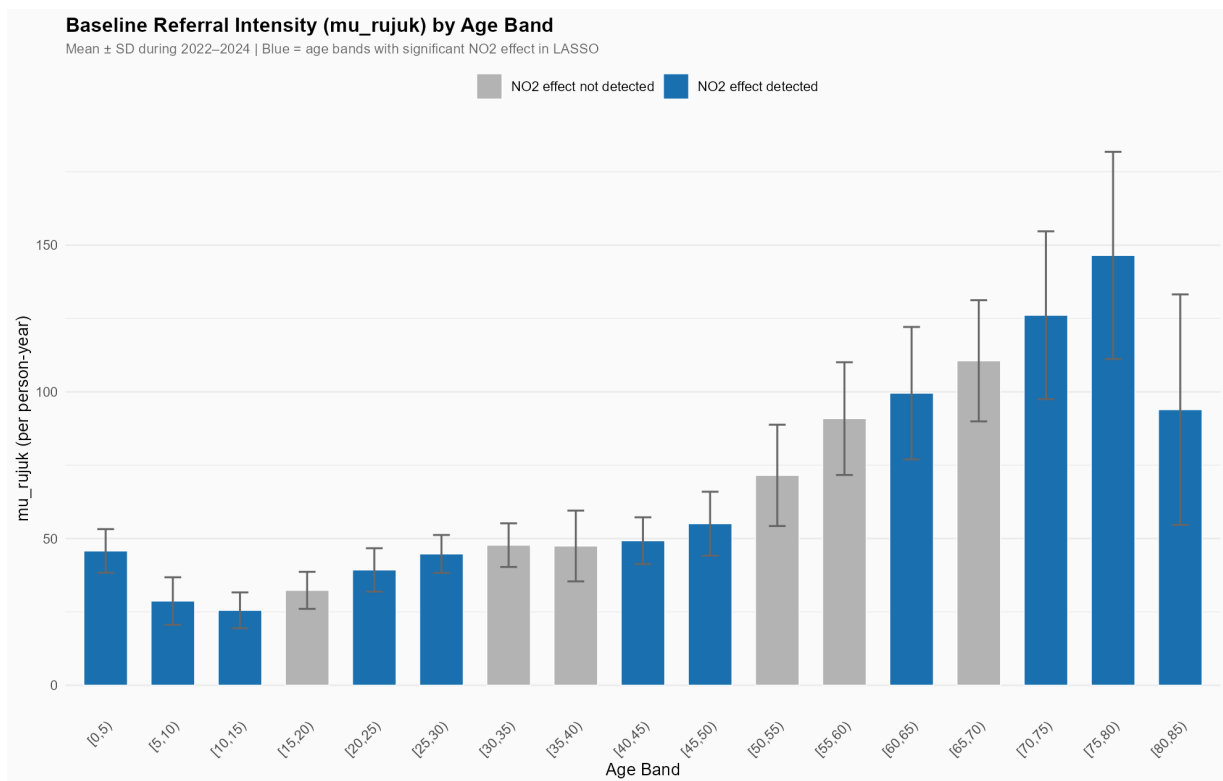


Figure 2. Mean referral intensity ($\hat{\mu}_{rujuk}$) from FKTP to FKRTL by age band, with standard deviation bars, 2022–2024. Blue bars indicate age bands where NO₂ was retained as a positive predictor by LASSO; grey bars indicate age bands where no pollutant was selected. Age bands [85,90) and above are excluded

Figure 2. illustrates the age-varying structure of the referral decrement. The clear heterogeneity in both

mean intensity and variability across age bands motivates the age-stratified approach used in this study: a single pooled model would obscure the different pollutant-intensity relationships that may operate within distinct age groups. Pollutant screening was therefore conducted separately for each age band using individual LASSO models.

3.3. Exploratory LASSO Screening of Pollutant Variables

A separate LASSO model was fitted for each of the 17 included age bands within the FKTP-to-FKRTL referral decrement, yielding 17 age-specific models. Each model treated the quarterly estimated force of decrement ($\hat{\mu}_{ru,juk}$) as the outcome and included 21 predictors: 18 pollutant variables (six pollutants at concurrent lag, one-quarter lag, and two-quarter lag) and three quarterly seasonal dummy variables (Q1, Q2, Q3, with Q4 as the reference category). Seasonal dummy variables were included to reduce the possibility that retained pollutant predictors merely reflect shared seasonality between referral utilization and pollutant concentrations, rather than a decrement-specific association.

Given severe multicollinearity among pollutant predictors (variance inflation factors exceeding 60 for PM-related predictors), ordinary least squares estimation was not appropriate. LASSO regression was selected because it simultaneously performs variable selection and coefficient shrinkage, retaining only predictors that contribute sufficiently to explaining outcome variation under penalization. The regularization parameter was selected via five-fold cross-validation using the one-standard-error rule (λ_{1se}). Although leave-one-out cross-validation (LOOCV) is often considered for small samples, five-fold cross-validation was adopted because LOOCV estimates can suffer from high variance, as the training sets across iterations are nearly identical and produce highly correlated error estimates [24, 25]. Five-fold cross-validation, which withholds 20% of observations per fold, provides a more stable basis for selecting λ in this setting, which is consistent with the conservative, sparser model selection adopted throughout this study given the limited number of observations per band ($n = 10$). It should be noted that this cross-validation procedure is used solely for regularization parameter selection, not forecasting validation; the folds are not temporally ordered, and the objective is penalty level selection rather than assessment of out-of-sample predictive performance.

The LASSO results should be interpreted as exploratory screening outcomes. A non-zero coefficient indicates that a predictor was retained under the penalized specification; it does not constitute independent causal identification. Because pollutant variables and their lagged values are correlated, LASSO may retain one variable as a representative predictor from a correlated group while shrinking others to zero. This behavior is a known property of LASSO under multicollinearity and does not imply that excluded variables have no relationship with the outcome.

Figure 3 summarizes the direction of retained LASSO coefficients across all pollutant variables and age bands. Among all predictors, NO₂ was the most consistently retained with a positive coefficient, appearing in 6 of 17 age bands: [25,30), [40,45), [45,50), [60,65), [70,75), and [80,85). This pattern is consistent with NO₂ being the primary marker of traffic-related combustion emissions in Jakarta, which may carry information relevant to health deterioration requiring referral [3]. However, the signal was not uniform across the full age profile: 11 of 17 age bands did not retain NO₂, and no pollutant was retained in 5 of 17 age bands. The absence of a retained signal in these bands reflects the limited statistical power available with $n = 10$ observations and should not be interpreted as evidence that NO₂ has no association with referral intensity in those age groups.

SO₂ at a two-quarter lag was frequently retained with a negative coefficient across multiple age bands. This pattern is difficult to attribute to a direct biological mechanism. A more plausible explanation is that SO₂ lag-2 captures residual seasonal structure not fully absorbed by the quarterly dummy variables, or that its negative sign reflects correlation with other retained predictors [23]. PM_{2.5} and PM₁₀ were not retained consistently, which is consistent with their high correlation with other pollutants in the Jakarta dataset.

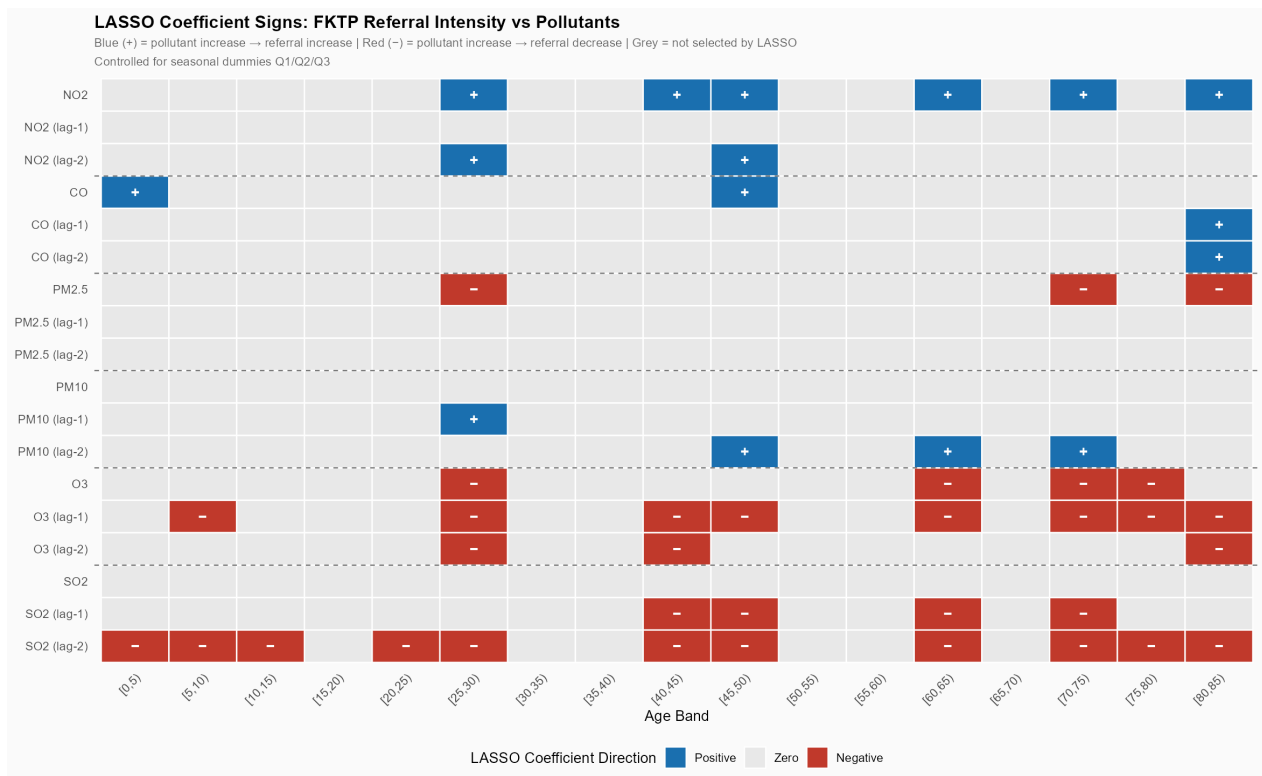


Figure 3. Direction of retained LASSO coefficients for the FKTP-to-FKRTL referral decrement, by pollutant variable and age band. Blue cells indicate positive retained coefficients; red cells indicate negative retained coefficients; grey cells indicate variables shrunk to zero by LASSO. Seasonal dummy variables are included in the model as controls but are not displayed here

3.4. Model-Implied Magnitude of NO2-Related Variation

For age bands in which NO2 was retained by LASSO, model-implied elasticities were computed to describe the magnitude of the fitted association on a scale that is comparable across age bands with different baseline intensities. Each elasticity value represents the percentage change in fitted referral intensity associated with a one percent change in NO2, evaluated at the observed mean values for each age band. These values are conditional on the fitted LASSO model and should not be interpreted as population-wide biological effect sizes.

Figure 4. shows that the largest model-implied elasticity was observed in the [25,30) age band (0.544), followed by [80,85) (0.452) and [45,50) (0.423). The [60,65) and [70,75) bands showed moderate elasticities of 0.351 and 0.335, respectively. The [40,45) band had the lowest elasticity among retained bands (0.063), indicating that variable retention by LASSO and practical magnitude of the fitted association are not equivalent. The elasticity pattern does not follow a strictly monotonic age gradient, which may reflect genuine heterogeneity in the relationship between NO2 exposure and referral behavior across age groups, or may partly reflect instability in the fitted coefficients arising from the small number of observations per age band.

3.5. Model-Based Scenario Analysis

A scenario analysis was conducted to express the model-implied sensitivity of referral intensity to hypothetical NO2 shifts in units that are interpretable for actuarial stress-testing purposes. The magnitude of the hypothetical shift was set at $\pm 20\%$, calibrated to the observed historical variability of NO2 in the dataset: the quarterly weighted mean NO2 concentration during 2022–2024 had a standard deviation of $5.15 \mu\text{g}/\text{m}^3$ around a mean of $23.98 \mu\text{g}/\text{m}^3$, corresponding to a coefficient of variation of approximately 21.5%. The $\pm 20\%$ scenario magnitude therefore approximates one historical standard deviation of quarterly NO2 fluctuation observed in Jakarta over the study period, rather than representing an arbitrary or regulatory threshold. Two hypothetical

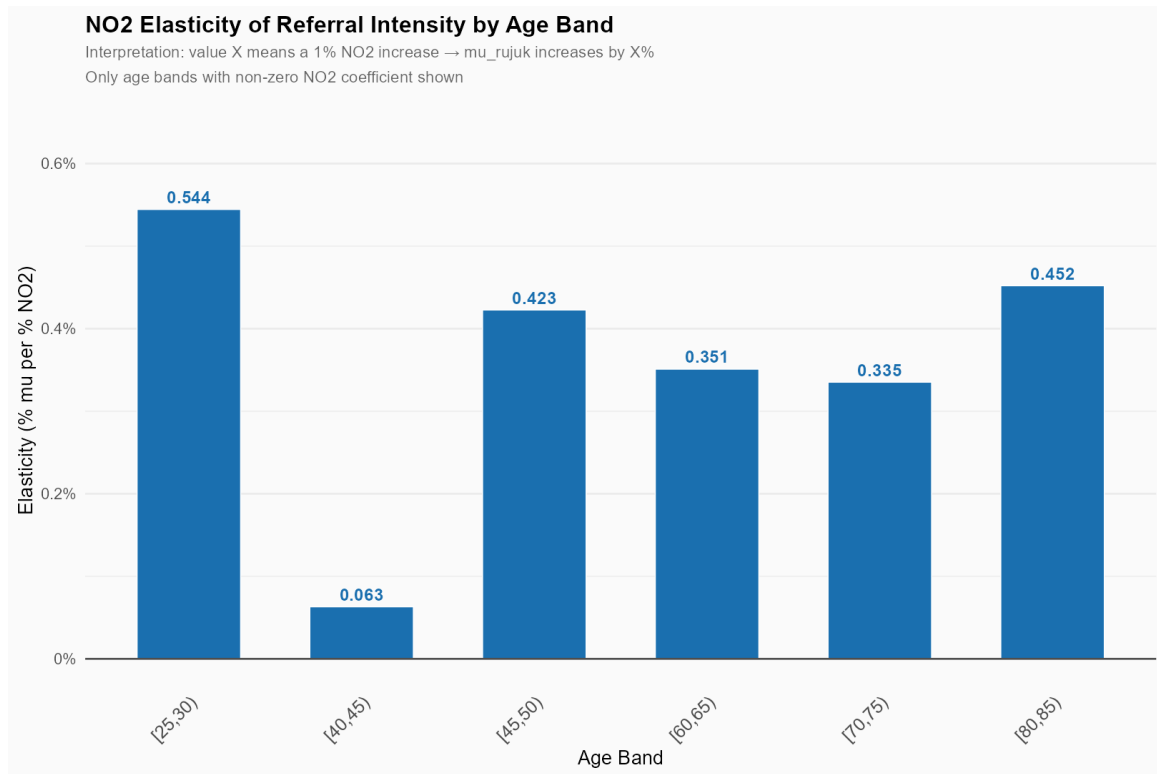


Figure 4. Model-implied NO2 elasticity of referral intensity for the six age bands in which NO2 was retained by LASSO. Each value represents the percentage change in fitted $\hat{\mu}_{rujuk}$ per one percent change in NO2 concentration, computed at observed mean values. Values are point estimates; no uncertainty intervals are reported as LASSO does not produce standard errors.

scenarios were defined relative to the 2022–2024 quarterly mean NO2 baseline: a 20 % increase in NO2, representing a deterioration in ambient air quality, and a 20 % decrease in NO2, representing an improvement scenario.

Figure 5. shows that under a 20 % increase in NO2, the fitted referral intensity increased by approximately 1.3 % in the [40,45) age band and up to 10.9 % in the [25,30) age band. The [45,50) and [80,85) age bands showed model-implied changes of 8.5 % and 9.0 %, respectively, while [60,65) and [70,75) showed changes of 7.0 % and 6.7 %. The 20 % decrease scenario produced changes of equal magnitude in the opposite direction, which follows directly from the linear model specification and does not imply that real-world referral responses would necessarily be symmetric. For actuarial purposes, the range spanned by the two scenarios may serve as an illustrative stress interval for sensitivity testing of claim transition assumptions, subject to the limitations of the underlying exploratory screening.

3.6. Visual Consistency Check Using Quarterly Time Series

Figure 6. presents normalized quarterly time series of NO2 and referral intensity for the six age bands in which NO2 was retained. Both variables were normalized to a [0, 1] scale within each age band to allow visual comparison of directional movements across time. This visualization serves only as a consistency check for the LASSO results and does not establish temporal causality or control for other time-varying factors.

In several age bands, particularly [45,50), [70,75), and [80,85), NO2 and referral intensity moved in broadly similar directions during parts of the observation window, which is visually consistent with the positive LASSO coefficients retained for those bands. The co-movement was not exact across all quarters, and some bands showed periods of divergence. This is expected, as referral intensity is also influenced by disease seasonality, healthcare utilization behavior, and operational referral practices not captured by the model. The time series

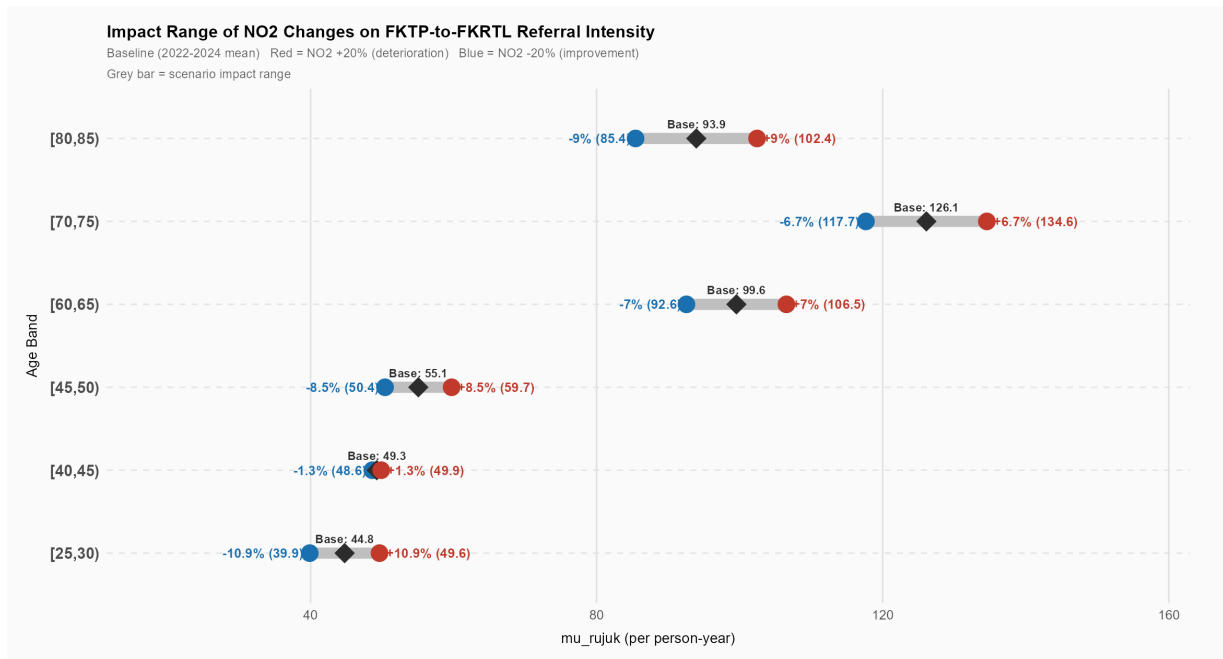


Figure 5. Model-based scenario analysis of hypothetical NO2 changes on fitted FKTP-to-FKRTL referral intensity for age bands in which NO2 was retained by LASSO. The diamond marker indicates baseline mean referral intensity; red circles show fitted intensity under a 20 % NO2 increase; blue circles show fitted intensity under a 20 % NO2 decrease. Labels show percentage change and absolute fitted value. Scenario bounds are point predictions only; no uncertainty intervals are shown as LASSO does not produce standard errors.

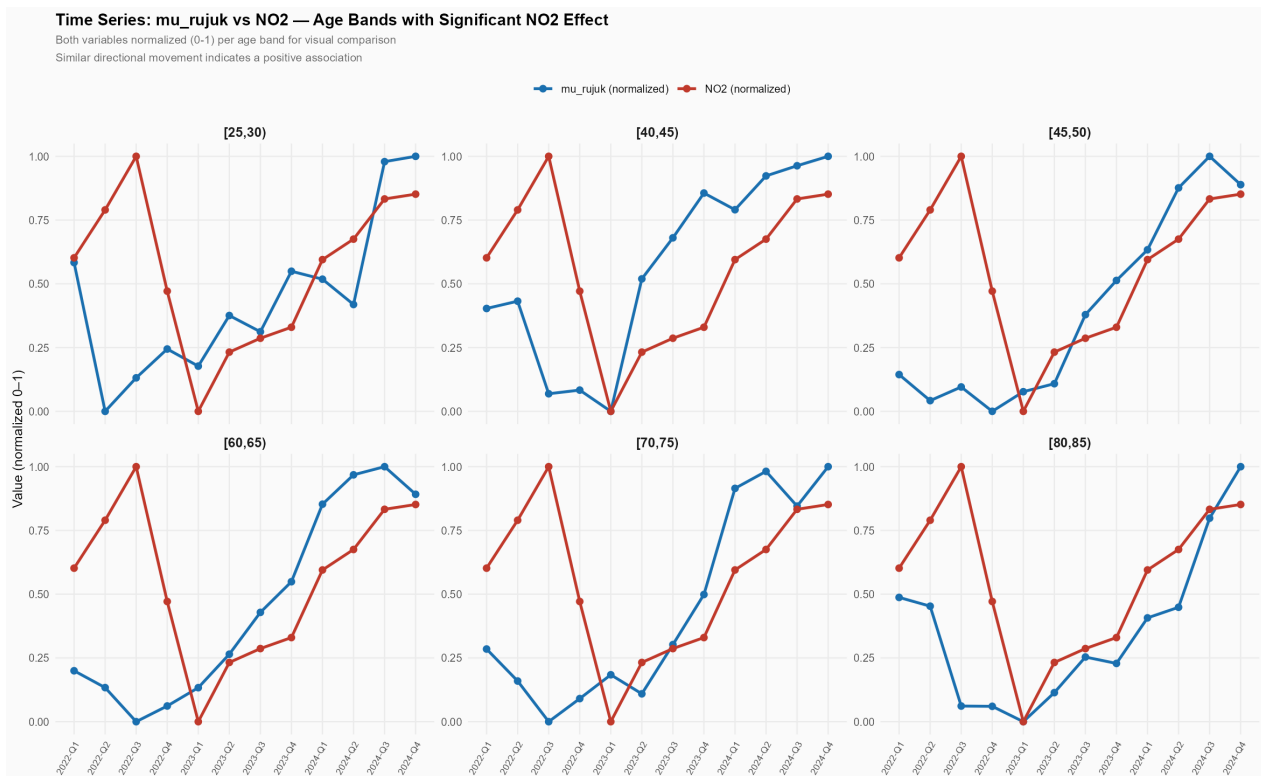


Figure 6. Normalized quarterly time series of NO2 and referral intensity ($\hat{\mu}_{rujuk}$) for the six age bands in which NO2 was retained by LASSO. Both series are normalized to [0, 1] within each age band. Co-movement should be interpreted as a visual consistency check only and does not establish causality

plot should therefore be read as limited corroborating evidence only.

3.7. Secondary Pollutant Results for Other Decrements

The same LASSO specification was applied to the remaining five modeled decrements to assess whether consistent pollutant-related signals appeared across the broader transition system. These results are treated as secondary findings because the retained pollutant patterns were weaker, less stable, or more difficult to interpret than those observed for the FKTP-to-FKRTL referral decrement.

For outpatient care at FKTP, SO₂ at a two-quarter lag was frequently retained with a negative coefficient across multiple age bands. As noted for the referral decrement, this pattern is difficult to attribute to a specific biological mechanism and may reflect residual seasonal structure. For healthy visits at FKTP, SO₂ lag-2 appeared with a positive coefficient in several age bands, indicating that the direction of the SO₂-related signal varies across different decrement definitions. For healthy visits at FKRTL and referral within FKRTL, the retained pollutant patterns were sparse and did not form a coherent interpretable structure. For the death decrement at FKRTL, the crude intensity estimates were too noisy for reliable pollutant screening given only 301 observed events distributed across 17 age bands and 12 quarters. The recovery decrement at FKTP could not be modeled due to insufficient temporal variation in estimated intensity: with 254 total events, the majority of age bands produced constant or near-zero quarterly values, preventing LASSO estimation.

These secondary results collectively reinforce the decision to center the main pollutant analysis on the FKTP-to-FKRTL referral decrement, which provided the most stable and interpretable pollutant-related signal in this dataset.

3.8. Interpretation and Limitations

The results indicate that the FKTP-to-FKRTL referral decrement provides the clearest pollutant-related signal among the modeled transitions. NO₂ was the most consistently retained pollutant in the exploratory LASSO screening, appearing as a positive predictor in 6 of 17 age bands. The model-implied elasticities were largest in the [25,30) and [80,85) age bands, though the evidence was concentrated in selected age bands and should not be generalized across the full age profile or to other decrement types.

These findings are broadly consistent with international evidence linking NO₂ exposure to increased use of primary-level health services. In a multi-municipality case-crossover study in southern Sweden, Taj et al. [26] reported a 1.85 % increase in primary health care visits per 10 $\mu\text{g}/\text{m}^3$ increase in NO₂ in one municipality, a magnitude broadly comparable to the elasticities estimated here, where a 1 % increase in NO₂ was associated with referral intensity increases ranging from 0.06 % to 0.54 % across the retained age bands. The age-heterogeneous pattern observed in this study, with stronger associations concentrated in older age bands, is also consistent with evidence from a national study in China, where Jin and Li [27] found that PM_{2.5} significantly increased hospitalization probability among adults aged 60 and above but found no significant effect for outpatient care use or for younger and middle-aged adults. Although that study examined PM_{2.5} rather than NO₂ and used a different healthcare utilization measure, the shared finding that older populations exhibit greater sensitivity to ambient air pollution lends external support to the age gradient identified in the present analysis.

The primary contribution of the pollutant component is not to establish a causal relationship between ambient NO₂ and referral rates. Rather, it demonstrates that environmental covariates may contain useful information for explaining variation in decrement intensity at the quarterly level. In an actuarial context, this type of analysis supports sensitivity analysis and stress testing of claim transition assumptions under environmental uncertainty, which is directly relevant for pricing and reserving in health insurance programs such as BPJS Kesehatan [6].

This carries several concrete points of entry into actuarial practice. In pricing, where a benefit is triggered by FKTP-to-FKRTL referral, such as a critical illness rider or a referral-linked indemnity within a BPJS-supplementary product, the baseline force of decrement $\hat{\mu}_{rujuk}$ feeds directly into the expected claim frequency

for that benefit; the model-implied elasticity range (0.063 to 0.544) and the $\pm 20\%$ scenario bounds (1.3 % to 10.9 % by age band) offer one way to construct an explicit pollution loading on top of the historical decrement table, particularly for cohorts in age bands where the NO₂ association was retained. In reserving, the same scenario bounds can be incorporated as an additional stress alongside conventional demographic and lapse shocks, applying the age-band-specific percentage shift to the referral decrement assumption before propagating it through the standard multiple decrement reserving recursion to assess the resulting movement in reserve adequacy. For BPJS Kesehatan's own program-level financial sustainability monitoring, the age-stratified scenario output could serve as a supplementary sensitivity disclosure analogous to standard mortality- or lapse-assumption sensitivity tables, flagging which age cohorts contribute most to referral volume volatility under deteriorating air quality. These applications are illustrative rather than validated, given the exploratory nature of the underlying LASSO screening where any operational use would require the robustness checks discussed below before being incorporated into a live pricing or reserving basis.

Several limitations are important. First, the analysis uses aggregated quarterly data with ten observations per age band after lag removal, which substantially limits statistical power and prevents individual-level covariate adjustment. Second, pollutant variables are highly correlated in the Jakarta dataset, so LASSO selection may be sensitive to alternative model specifications or different regularization choices. Third, the quarterly seasonal dummy variables absorb broad recurring patterns but cannot fully account for disease seasonality, operational referral practices, or behavioral factors that may confound the pollutant-intensity relationship. Fourth, the assignment of aggregated city-wide pollutant averages introduces ecological exposure error. Because individuals move throughout the city, their actual exposure may differ substantially from the modeled averages because of specific residential locations, commuting behaviors, and varying occupational exposures. Fifth, the model relies on the standard actuarial assumption of independent competing risks. While an individual's underlying frailty may simultaneously influence multiple health transitions, modeling dependent competing risks via continuous-time joint frailty models was precluded by severe event sparsity. Minority decrements, such as recovery and mortality, contained insufficient event counts to support stable individual-level estimation. Consequently, the data were systematically aggregated into discrete age-quarter cells to yield stable empirical forces of decrement suitable for LASSO regularization, representing a necessary methodological trade-off. Sixth, the scenario analysis is based on linear model predictions and does not quantify uncertainty around the implied intensity changes. These findings should therefore be treated as hypothesis-generating and validated using longer time series, finer spatial and temporal resolution, and richer individual-level data.

4. CONCLUSIONS

The FKTP-to-FKRTL referral decrement produced the strongest and most interpretable pollutant-related signal among all modeled transitions. Referral intensity exhibited a clear age gradient, rising from below 45 per person-year in younger bands to above 160 per person-year in the [75,80) band, with substantial quarterly volatility in older groups. NO₂ emerged as the most frequently retained predictor under the exploratory modeling framework, appearing with a positive coefficient in 6 of 17 age bands after seasonal adjustment. Model-implied elasticities ranged from 0.063 to 0.544 across retained bands, and scenario analysis indicated that a hypothetical 20 percent increase in ambient NO₂ corresponded to fitted referral intensity increases of between 1.3 and 10.9 percent. These findings suggest that ambient NO₂ concentration may carry useful information for explaining quarterly variation in referral rates, and that environmental covariates merit consideration as stress-testing inputs in actuarial models of claim transition rates under BPJS Kesehatan.

These results are hypothesis-generating. Each LASSO model was fitted on ten quarterly observations against 21 predictors, retained coefficients are sensitive to regularization choices, and pollutant variables are highly intercorrelated in the Jakarta monitoring network. Future research should prioritize extending the observation

window to improve statistical power, incorporating individual-level covariate adjustment through episode-level Cox models with clustered standard errors, linking enrollee records to facility-level characteristics and provider referral patterns, spatially disaggregating pollution exposure using station-level or interpolated surfaces, and propagating the fitted NO₂ sensitivity through a full multiple decrement table to produce actuarially actionable estimates of claim count and benefit present value changes under hypothetical pollution scenarios.

5. REFERENCES

- [1] R. Fuller, P. J. Landrigan, K. Balakrishnan, G. Bathan, S. Bose-O'Reilly, M. Brauer, N. Bruce, J. Caravanos, T. Chiles, A. Cohen, *et al.*, "Pollution and health: a progress update," *The Lancet Planetary Health*, vol. 6, no. 6, pp. e535–e547, 2022.
- [2] J. Chen and G. Hoek, "Long-term exposure to PM and all-cause and cause-specific mortality: A systematic review and meta-analysis," *Environment International*, vol. 143, p. 105974, 2020.
- [3] P. Huangfu and R. Atkinson, "Long-term exposure to NO₂ and O₃ and all-cause and respiratory mortality: A systematic review and meta-analysis," *Environment International*, vol. 144, p. 105998, 2020.
- [4] Centre for Research on Energy and Clean Air, "Indonesia air quality 2024: Jakarta's metro areas break WHO limit up to tenfold," tech. rep., CREA, 2025.
- [5] G. Syuhada *et al.*, "Impacts of air pollution on health and cost of illness in jakarta, indonesia," *International Journal of Environmental Research and Public Health*, vol. 20, no. 4, p. 2916, 2023.
- [6] D. C. M. Dickson, M. R. Hardy, and H. R. Waters, *Actuarial Mathematics for Life Contingent Risks*. Cambridge: Cambridge University Press, 2nd ed., 2013.
- [7] S. Haberman and E. Pitacco, *Actuarial Models for Disability Insurance*. Chapman & Hall/CRC, 1999.
- [8] P. Orellano, J. Reynoso, N. Quaranta, A. Bardach, and A. Ciapponi, "Short-term exposure to particulate matter (pm₁₀ and pm_{2.5}), nitrogen dioxide (no₂), and ozone (o₃) and all-cause and cause-specific mortality: Systematic review and meta-analysis," *Environment International*, vol. 142, p. 105876, 2020.
- [9] R. Chen, H. Kan, B. Chen, W. Huang, Z. Bai, G. Song, and G. Pan, "Association of particulate air pollution with daily mortality: the china air pollution and health effects study," *American Journal of Epidemiology*, vol. 175, no. 11, pp. 1173–1181, 2012.
- [10] D. L. Crouse *et al.*, "Ambient pm_{2.5}, o₃, and no₂ exposures and mortality," *Environmental Health Perspectives*, vol. 123, no. 11, 2015.
- [11] A. Fatimah, L. Anselmi, J. Gibson, A. Maharani, I. Irmansyah, S. Idaiani, D. Susilo, J. Wilkinson, M. Sutton, H. Susanti, H. Brooks, P. Bee, and H. Thabrany, "Indonesian national health insurance scheme longitudinal sample data 2015–2020: Overview and potential uses for health policy analysis," *BMC Health Services Research*, vol. 25, no. 1, p. 1593, 2025.
- [12] Q. Cheng, R. A. Fattah, D. Susilo, A. Satrya, M. Haemmerli, S. Kosen, D. Novitasari, G. C. Puteri, E. Adawiyah, A. Hayen, A. Mills, V. Tangcharoensathien, S. Jan, H. Thabrany, A. Asante, and V. Wiseman, "Determinants of healthcare utilization under the indonesian national health insurance system – a cross-sectional study," *BMC Health Services Research*, vol. 25, 2025.
- [13] J. Ardhan, R. A. Syakurah, H. Idris, and Y. Windusari, "Implementation of the bpjs health referral program in indonesia: Literature review," *Indonesian Journal of Global Health Research*, vol. 7, pp. 811–820, Oct. 2025.

- [14] R. Yi, M. Woldeyes, and G. Bradford, “The impact of wildfire-related and environmental air pollution on morbidity,” tech. rep., Society of Actuaries Research Institute, 2025.
- [15] Q. Zhao and Y. Wang, “Multilevel medical insurance mitigate health cost inequality due to air pollution: Evidence from China,” *International Journal for Equity in Health*, vol. 23, no. 1, p. 156, 2024.
- [16] BPJS Kesehatan, “Data sampel BPJS kesehatan.” <https://data.bpjs-kesehatan.go.id/bpjs-portal/action/blog.cbi?filter=data+sampel>, 2026. Accessed: 2026-03-25.
- [17] R. Sinto, “Covid-19 pandemic-to-endemic transition in indonesia: What does the future hold?,” *Acta Medica Indonesiana*, vol. 54, p. 159, Jul. 2022.
- [18] Pemerintah Provinsi DKI Jakarta, “Satu data jakarta.” <https://satudata.jakarta.go.id/>, 2026. Accessed: 2026-04-08.
- [19] W. M. Makeham, “On the law of mortality and the construction of annuity tables,” *Journal of the Institute of Actuaries*, vol. 8, pp. 301–310, 1860.
- [20] L. A. Gavrilov and N. S. Gavrilova, *The Biology of Life Span: A Quantitative Approach*. New York: Harwood Academic, 1991.
- [21] J. Friedman, T. Hastie, and R. Tibshirani, “Regularization paths for generalized linear models via coordinate descent,” *Journal of Statistical Software*, vol. 33, no. 1, pp. 1–22, 2010.
- [22] S. Altelbany, “Evaluation of ridge, elastic net and lasso regression methods in precedence of multi-collinearity problem: a simulation study,” *Journal of Applied Economics and Business Studies*, vol. 5, no. 1, pp. 131–142, 2021.
- [23] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.
- [24] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2nd ed., 2009.
- [25] L. Breiman and P. Spector, “Submodel selection and evaluation in regression: The x-random case,” *International Statistical Review*, vol. 60, no. 3, pp. 291–319, 1992.
- [26] T. Taj, E. Malmqvist, E. Stroh, D. Oudin Åström, K. Jakobsson, and A. Oudin, “Short-term associations between air pollution concentrations and respiratory health—comparing primary health care visits, hospital admissions, and emergency department visits in a multi-municipality study,” *International Journal of Environmental Research and Public Health*, vol. 14, no. 6, p. 587, 2017.
- [27] B. Jin and W. Li, “Air pollution, healthcare use, and inequality: Evidence from China,” *Economic Modelling*, vol. 141, p. 106913, 2024.