Research Article

BIOMEDICAL Journal of Scientific & Technical Research

ISSN: 2574 -1241 DOI: 10.26717/BJSTR.2026.66.010328

Supplementary Content

Section 1: Data Source

Overview: The Global Burden of Disease Study (GBD) 2021 is a systematic global initiative encompassing various diseases, injuries, and risk factors. Its primary objective is to assess the impact of health determinants on populations worldwide, offering a comprehensive analysis of health patterns and outcomes among diverse regions and demographics. In the GBD study, 204 countries or territories were categorized into 21 GBD regions based on geographical and socioeconomic criteria, and further classified into 5 quintiles according to their sociodemographic index (SDI) values. The SDI is a composite indicator that measures the social and economic determinants of health outcomes. It is calculated as the geometric mean of the total fertility rate for women under 25, the average educational level of individuals aged 15 and above, and the lagged distributive income per capita. The SDI serves as a robust measure of sociodemographic development and its influence on health outcomes, enhancing stability and interpretability. The index is classified into five groups according to its value: low (SDI: 0–0.466), low-middle (SDI: 0.466–0.619), middle (SDI: 0.619–0.712), high-middle (SDI: 0.712–0.810), and high (SDI: 0.810–1).

Case Definition: LRIs, diagnosed by clinicians as pneumonia or bronchiolitis, have corresponding ICD-10 codes A48.1, A70, B97.4-B97.6, J09-J15.8, J16-J16.9, J20-J21.9, J91.0, P23.0-P23.4, and U04-U04.9. ICD-9 codes for LRIs were 079.82, 466-469, 470.0, 480-481.9, 482.0-482.89, 483.0-483.9, 484.1-484.2, 484.6-484.7, 487-490.9, 510-511.9, and 513.0-513.9. Mycobacterium tuberculosis and SARS-CoV-2 were excluded from the LRI category and analyzed separately.

Fatal Burden of LRIs: The estimation of LRI relied on databases that included vital registration (VR), verbal autopsy (VA), sample vital registration (VR-S), surveillance, and minimally invasive tissue sample (MITS) diagnosis. In locations with a robust VR system, VR is the primary data source for mortality estimation, while in areas with incomplete VR coverage, other data types are used in conjunction. Sex and age categories (under 5 and 5 years and above) were separately modeled using the Cause of Death Ensemble Model (CODEm). Covariates were selected through four classes of models: linear mixed effects regression (LMER) models of the natural logarithm of the cause-specific death rate, LMER models of the logit of the cause fraction, spatiotemporal Gaussian process regression (ST-GPR) of the natural logarithm of the cause-specific death rate, and ST-GPR models of the logit of the cause fraction. Covariates were ranked based on the strength of evidence for a causal relationship with LRI mortality, ranging from 1 to 3. To mitigate multicollinearity, both statistical significance and plausibility of the covariates were assessed to prevent generating implausible results. Sampling variance, non-sampling variance, and garbage code redistribution variance were considered for data variance. Models underwent out-of-sample predictive validity tests for evaluation. Weighting of the component models was based on their predictive validity rank to establish their influence on the ensemble estimate. CoDCorrect was employed to adjust LRI mortality estimates by rescaling them in accordance with the uncertainty levels for internal consistency.

To calculate years of life lost (YLLs), the number of CoDCorrected deaths was multiplied by the predicted life expectancy, with a specific focus on premature mortality. The predicted life expectancy comprised two elements:

(1) A universal, gender-neutral, ideal life expectancy assuming equal life expectancy for males and females; and

(2) The mean age of death from life tables adjusted for shocks for each location, sex, age bracket, and year.

Non-Fatal Burden of LRIs: Incidence and prevalence data for LRIs were collected through systematic literature reviews, hospital inpatient and outpatient records, US claims data, and surveys. Non-representative surveys, focusing solely on refugees, prisoners, or individuals who inject drugs, were not included. Self-reported prevalence of LRI symptoms from population-representative surveys was also taken into account. The optimal case definition for LRIs involves the presence of cough, difficulty breathing, chest symptoms, and fever. To accommodate seasonal variations in LRI symptom prevalence, a generalized additive model was utilized. Prior to modeling, adjustments were made to hospital inpatient and US inpatient claims data to correct for readmissions and multiple diagnoses. A meta-regression approach was employed to forecast the adjustment factor for hospitalization data to ensure compatibility with the reference case definition. The average duration of LRIs, at 7.79 days, was used for the conversion of incidence data to prevalence. The distribution of moderate (85%) and severe (15%) LRIs is dictated by the health states of the acute infectious disease episode.

The Bayesian meta-regression modeling framework, DisMod-MR 2.1, was used to estimate the incidence and prevalence of LRI. Estimation in DisMod-MR 2.1 occurs at five levels: global, super-region, region, country, and subnational location. After the global fit, LRI was modeled by sex while retaining all years of data until the lowest level in the cascade. A log-Gaussian equation was used for the data likelihood. The overall estimates of LRI incidence and prevalence were adjusted for 2020 and 2021 to account for the reductions in influenza and RSV mortality related to COVID-19. Disability weights ranging from 0 (perfect health) to 1 (complete loss of health) were used to compute years of disability (YLDs) by multiplying the prevalence in each severity category. The sum of YLLs and YLDs determined the disability-adjusted life years (DALYs).

Risk Factor

Overview: The Comparative Risk Factor Assessment (CRA) framework developed by Murray and Lopez establishes a causal network hierarchy of risk factors influencing health outcomes, facilitating quantitative evaluation across various levels. Risks are categorized into four hierarchical levels based on evidence criteria. The CRA process involves six key steps:

1) Identifying and defining risk-outcome pairs;

2) Estimating relative risks (RR);

3) Modeling and quantifying exposure distribution;

4) Determining the Theoretical Minimum Risk Exposure Level (TMREL);

5) Calculating the Population Attributable Fraction (PAF); and

6) Accounting for potential mediation effects among different risk factors, if necessary.

Risk-Outcome Pairs: Risk-outcome pairs were included based on meeting convincing or probable evidence criteria. Convincing evidence are supported by multiple prospective observational studies and high-quality randomized controlled trials with consistent findings across populations. Probable evidence follows similar principles but is based on a smaller number of studies. Smoking is known to disrupt structural, functional, and immunologic defense mechanisms, making it widely recognized as a risk factor for LRIs. In the GBD study's risk factor classification, smoking is categorized at Level 3, falling under tobacco (Level 2) and behavioral risks (Level 1).

RR: The RR for mortality or morbidity based on exposure level or cause can be derived from both published and unpublished primary studies as well as secondary studies. Estimating the RR of smoking involves six key steps:

1) Inclusion of prospective cohort and case-control studies from January 1, 1970, to May 31, 2022, reporting effect sizes and uncertainties regarding the association between smoking intensity and outcomes.

2) Utilization of meta-regression—Bayesian, regularized, trimmed (MR-BRT) to determine the shape of the risk-outcome relationship and create nonparametric risk curves for smoking exposure.

3) Assessment and adjustment of systematic biases based on study characteristics.

4) Quantification of between-study heterogeneity using a study-specific random slope relative to the "signal" model.

5) Estimation of between-study heterogeneity, particularly for small studies, using the Fisher information matrix (FIM) to mitigate potential underestimation.

6) Generation of a burden of proof risk function (BPRF) using the estimated between-study heterogeneity and 95th quantiles. The BPRF represents the smallest harmful or protective effect at each exposure level consistent with the available evidence, while the risk-outcome score (ROS) is defined as the signed average log BPRF value. Egger's regression was used to detect publication bias.

Exposure Estimation: The prevalence of current and former smoking was estimated using data from cross-sectional nationally representative household surveys. Current smokers were defined as individuals who currently use any smoked tobacco product daily or occasionally, while former smokers were defined as individuals who have abstained from all smoked tobacco products for at least six months. Primary data were obtained from individual-level microdata and survey report tabulations. A total of 3603 sources on exposure to smoked tobacco from 204 countries or territories were included, covering 36 potential combinations of current, former, and/or ever smoked tobacco use, encompassing various frequencies of use and types of smoked tobacco products. Data from sources containing multiple case definitions were utilized to derive the adjustment coefficient, represented by the beta value from a linear model with one predictor and no intercept. This coefficient was

used to standardize alternative case definitions to align with the GBD case definition. The dataset was divided into a training subset, consisting of data already categorized into GBD sex-specific 5-year age groups, and a split subset, comprising data reported in aggregated age or sex. A Spatiotemporal Gaussian process regression (ST-GPR) approach was applied to estimate sex-geography-time-specific age patterns based on the training dataset.

The ST-GPR model incorporated a minimal age weight parameter to ensure that age patterns were data-driven without excessive smoothing, thereby preventing redundant enforcement of age-sex split patterns in the final exposure model. The prevalence of current and former smoking was assessed through ST-GPR. The mean function input to GPR consisted of a comprehensive time series of estimates derived from a mixed effects hierarchical linear model, along with weighted residuals that were smoothed over time, region, and age. Cigarettes per smoker per day and pack-years were assessed for current smokers, with one pack-year defined as smoking one pack of cigarettes daily for one year. Individual smoking histories were determined based on the age of initiation and quantity smoked, with distributions tailored to age, sex, and region. Years since cessation were considered for former smokers. The mean age of cessation was modeled using cross-sectional survey data.

TMREL: TMREL is the level of risk exposure that minimizes risk at the population level. For smoking, the TMREL is defined as 0 meaning no exposure.

PAF: PAF represents the percentage of cases or fatalities attributable to a particular disease that could be averted by transitioning the current exposure pattern to the TMREL. In the context of smoking, PAF is quantified as:

$$PAF = \frac{p(n) + p(f)\int \exp(x) * rr(x) + p(c)\int \exp(y) * rr(y) - 1}{p(n) + p(f)\int \exp(x) * rr(x) + p(c)\int \exp(y) * rr(y)}$$

The prevalence of never, former, and current smokers is denoted as $p(n), p(f)$, and $p(c)$, respectively. The distribution of years since quitting among former smokers is represented by $\exp(x)$, with $rr(x)$ indicating the relative risk for years since quitting. Additionally, $\exp(y)$ and $rr(y)$ signify the distribution and relative risk for cigarettes per smoker per day or pack-years.

ISSN: 2574-1241

DOI: 10.26717/BJSTR.2026.66.010328

Depeng Jiang . Biomed J Sci & Tech Res

Submission Link: https://biomedres.us/submit-manuscript.php