下是对你要分析的空气污染物（PM2.5、NO₂、O₃）与年度哮喘（asthma）发病数据之间匹配的建议：

**✅ 结论（简明版）：**

| **污染物** | **建议使用的年度指标** | **原因说明** |
| --- | --- | --- |
| **PM2.5** | 年度平均值（annual mean） | 长期暴露关联明确，平均值代表慢性影响 |
| **NO₂** | 年度平均值（annual mean） | 代表交通相关长期暴露，慢性呼吸系统效应 |
| **O₃** | 年度高百分位或**年中最大8小时平均** | 对呼吸系统急性影响显著，通常用“夏季最大值”或高峰暴露衡量 |

**🧠 背景解析（详细说明）**

**1. 📌 时间匹配原则**

* **Asthma incidence** 是按 **年度** 汇总的慢性健康数据。
* 所以污染物也应处理为 **年度尺度的代表性指标**，以避免时间尺度不匹配带来的偏误。

**2. ☁️ PM2.5**

* 是一个长期暴露性污染物，广泛用于慢性病（如哮喘、心血管病）研究。
* **建议使用**：
  + **年度均值（μg/m³）**
* 也可以分季节看变化，但主分析建议用年度平均。

**3. 🚗 NO₂**

* 来源以交通尾气为主，对儿童哮喘有广泛研究支持。
* **建议使用**：
  + **年度平均值**
* 可选补充指标（用于敏感性分析）：冬季平均值（供暖/交通活跃）

**4. 🌞 O₃**

* 是一种次生污染物，浓度在夏季高，易造成呼吸道急性刺激。
* **不建议仅用年度平均**，因为它会稀释高暴露时段的影响。
* **建议使用**：
  + **最大8小时平均浓度的年最高值**（annual max 8h avg）
  + 或 **夏季（5–9月）平均最大8小时值**

**🧪 实践参考（研究文献中的常见做法）：**

* WHO、EPA、Health Canada 建议在流行病学研究中：
  + PM2.5、NO₂：用**年平均浓度**
  + O₃：用**年最大8小时移动平均的第4高值**（或第95百分位）

**🧰 如果你做的是空间分析（如LISA, GWR, 空间回归）：**

* **污染物值应该以年为单位聚合**，并匹配健康数据的空间单位（如PHU、Census Division）
* 如果做的是时空面板模型（如 spatio-temporal GAM），则也可以按季节或月度建模，但需有健康数据时间序列支持。

**✅ Summary Recommendation (Concise Version):**

| **Pollutant** | **Recommended Annual Metric** | **Rationale** |
| --- | --- | --- |
| **PM2.5** | Annual mean concentration (μg/m³) | Strongly associated with long-term exposure; mean represents chronic effects |
| **NO₂** | Annual mean concentration | Reflects traffic-related long-term exposure; relevant for respiratory outcomes |
| **O₃** | High-percentile values or annual max 8-hour average | Strong acute respiratory effects; typically assessed via summer peaks or maximum exposures |

**🧠 Detailed Background Explanation**

**1. 📌 Temporal Matching Principle**

Since asthma incidence is aggregated on an **annual** basis, pollutant exposure metrics should also be summarized on an **annual scale**.  
This alignment avoids temporal mismatch and ensures appropriate inference in exposure–response modeling.

**2. ☁️ PM2.5**

* A key long-term exposure pollutant commonly studied in relation to chronic diseases such as asthma and cardiovascular conditions.
* **Recommendation**: Use **annual mean concentration (μg/m³)**.
* Seasonal means can be explored, but annual mean is preferred for primary analysis.

**3. 🚗 NO₂**

* Mainly traffic-related; extensively linked to childhood asthma in the literature.
* **Recommendation**: Use **annual mean concentration**.
* **Optional**: For sensitivity analysis, consider winter averages (reflecting heating and traffic activity).

**4. 🌞 O₃ (Ozone)**

* A secondary pollutant with higher concentrations in summer; causes acute respiratory irritation.
* **Caution**: Do **not** rely solely on annual averages, as they dilute high-exposure periods.
* **Recommendation**:
  + Use **annual maximum 8-hour average concentration**, or
  + **Summer-season (May–September)** 8-hour daily max average

**🧪 Common Practices (From Scientific Literature):**

* Organizations like **WHO**, **EPA**, and **Health Canada** recommend:
  + For **PM2.5** and **NO₂**: use **annual mean concentrations**
  + For **O₃**: use **the 4th highest 8-hour daily maximum** or **95th percentile** as a metric for acute risk exposure

**🧰 If You're Conducting Spatial Analysis (e.g., LISA, GWR, Spatial Regression):**

* Pollutant values should be aggregated **annually** and aligned with the spatial unit of health data (e.g., PHU, Census Division).
* For spatio-temporal models (e.g., **spatio-temporal GAM**), seasonal or monthly pollutant data can be modeled **if** asthma data are available over time at matching resolution.

要计算 **臭氧（O₃）最大8小时平均浓度的年最高值（Annual Max 8-hour Average Ozone Concentration）**，你需要从**小时级别的 O₃ 数据**中，按如下步骤进行：

**✅ 计算步骤（Annual Max 8h Avg O₃）：**

**1. 准备原始数据**

* 数据单位：**每小时的 O₃ 浓度（通常以 ppb 或 µg/m³）**
* 时间范围：**某一整年**（例如 2023-01-01 到 2023-12-31）
* 一列应为时间戳（例如 "2023-06-15 14:00"），一列为 O₃ 浓度

**2. 滑动计算 8 小时平均浓度**

对每一天，从凌晨 0 点开始，计算连续 8 小时的平均值，得到所谓的 **"8-hour rolling average"**。

例如，2023-06-15：

* 第1个窗口：00:00–07:00 的平均浓度
* 第2个窗口：01:00–08:00 的平均浓度  
  ...
* 第17个窗口：16:00–23:00 的平均浓度  
  （一天可得 17 个值）

**3. 对每一天，提取当天的最大 8 小时平均值**

* 即对每天的 17 个滑动窗口值，保留 **最大的一个**。

这样你会得到 **365 个 "daily max 8h avg" 值**（闰年为 366）。

**4. 从全年中选出最大值**

* 从上述 365 个每日最大值中，选出全年 **最高的一个值**：

python

CopyEdit

annual\_max\_8h\_avg = max(daily\_max\_8h\_avgs)

这就是所谓的：

**Annual Max 8-Hour Average Ozone Concentration**

**📌 可选替代指标（用于健康研究）：**

* **第4高值**（EPA 标准推荐，用于稳定异常极值）
* **95th percentile**（Health Canada 也接受）
* **夏季（5–9月）平均最大8小时值**

**Annual Mean of Daily Max 3-Hour Average O₃**

| **用途** | **指标** |
| --- | --- |
| 主要暴露指标 | Annual mean PM2.5 / NO₂ |
| 补充敏感性指标 | Annual mean of daily max 3h avg |
| O₃ 替代暴露指标 | Summer mean daily max 8h avg |
|  |  |

asthma\_rate ~ annual\_mean\_PM2.5 + annual\_mean\_daily\_max3h\_PM2.5

asthma\_rate ~ annual\_mean\_NO2 + annual\_mean\_daily\_max3h\_NO2