Using geospatial approaches and machine learning for asthma and COPD outcomes: a systematic review

**Enriching OMOP CDM**

**PRESENTER:** Daniel Jeannetot
d.jeannetot@erasmusmc.nl

**INTRO:**
Asthma & COPD are major contributors to morbidity and mortality worldwide. OMOP CDM databases provide a unique opportunity to enrich Electronic health records with geospatial data and machine learning approaches to improve patient-level predictions. This systematic review shows that this is still an untapped approach with large potential for exploration.

**METHODS**
1. Systematic review following PRISMA guideline
2. 4 databases queried
3. 3 reviewers involved in full text review
4. 12 specific characteristics for data extraction including type of models (ML/non-ML), spatial scale, spatial approach.

**RESULTS**
- 1805 papers screened.
- 669 Included (Title)
- 1,044 Embase
- 887 Medline
- 598 Web of Science
- 67 Google Scholar top 100
- 32 Cochrane Central Register
- 1,136 Excluded (Title)
- 2,628 Identification
- 1,805 1st Round screening
- 123 Full text review
- 546 Excluded (Abstract)
- 72 Excluded (Full review)
- FINAL SELECTION: 51

The type of scale used varied greatly, with most papers using a local administrative level (e.g. counties, neighbourhoods), thus local, but hard to compare or generalize. Only a few used grid-based spatial data, and even then the resolutions ranged from 5m grid to 1km grid and beyond, leading to widely disparate estimates and areas.

OHDSI provides a coherent and readily available infrastructure to help Asthma/COPD research leverage observational data, machine learning, and geospatial approaches for very large-scale analyses.

**Inclusion criteria**
- Has modelling/prediction methods
- Has geospatial/geostatistical approaches
- Explanatory variables include geographical/environmental (air pollution, green/blue space, etc.)
- Main outcome is COPD and/or Asthma related
- Population should be 18 years old or above.

**Search term categories**
1. Asthma and/or COPD AND
2. Prediction models (OR. Modelling, Machine Learning etc.) AND
3. Spatial (OR geostatistical; geo*, etc.) AND
4. ADULT (NOT children, etc.)

**Key points**
- Population varied greatly in age groups and sample size (min=105, max=+50000)
- Scale greatly varied but generally local
- <10 papers used Machine learning algorithms
- Most geospatial approaches are 2 steps
- <10 papers used specific geostatistical tools
- Inconsistent quality and application of geospatial tools

*Daniel Jeannetot, Johnmary Arinze, Victor Pera, Peter Rijnbeek, Katia Verhamme*

List of included studies available in annex