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 contributor 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 prediction**. This systematic review shows that this is still an untapped approach which large potential for exploration

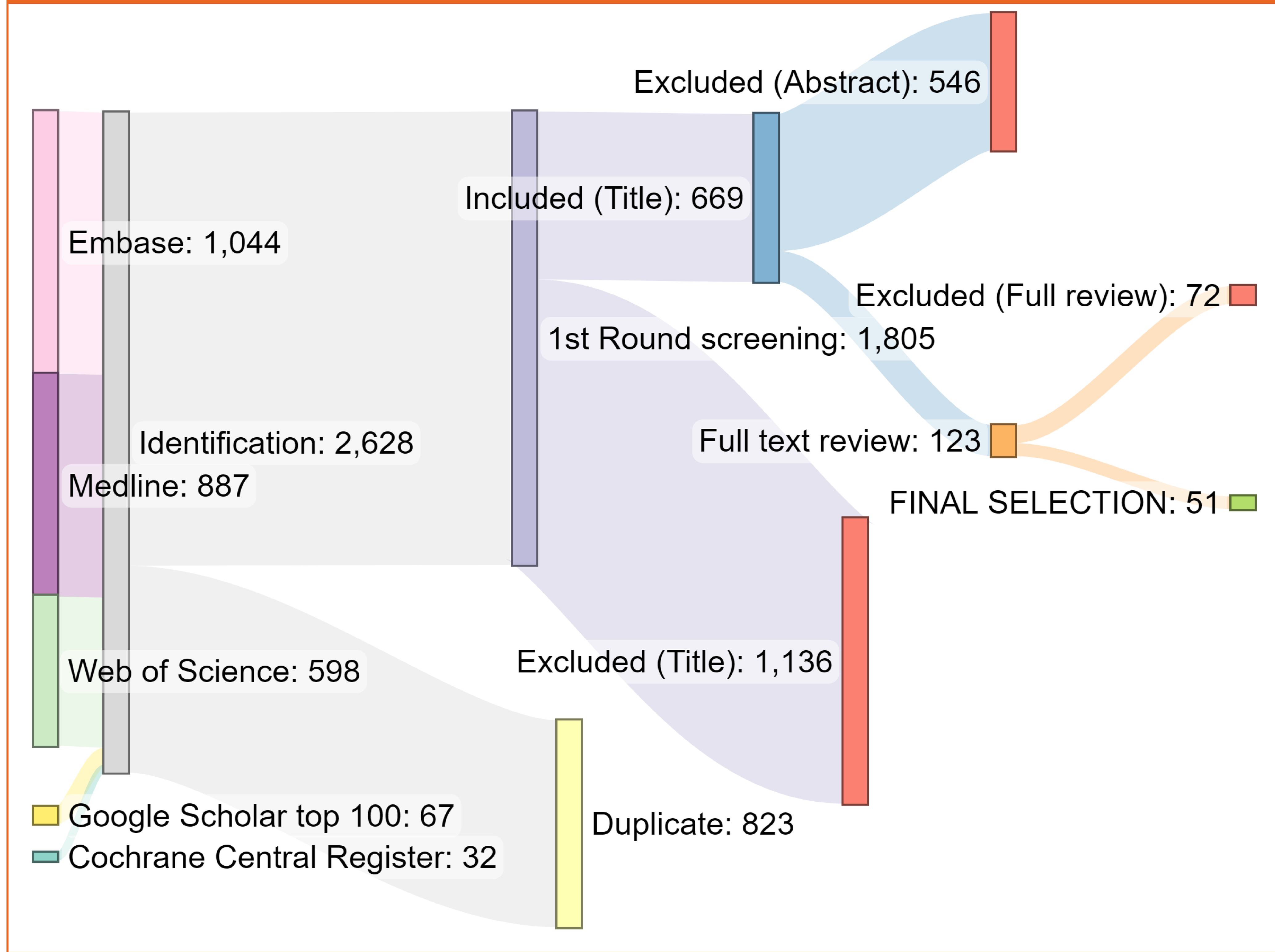
METHODS

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.
- 123 Papers fully reviewed
- 51 Articles making use of geospatial approach and modelling to measure and predict asthma/related outcomes.

Asthma/COPD research has a lot of potential to benefit from machine learning algorithms and geospatial approaches, especially if combined with observational data.



Most papers approached spatial information in two steps: use of geospatial models to estimates an environmental factor exposure (e.g. specific air pollutants values at specific location) to then integrate values in non-spatial regressions models (.e.g linear, multivariate, etc.), removing specific geospatial and geographical processes information.

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

List of included studies available in annex

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

Erasmus MC Universitair Medisch Centrum Rotterdam

OHDSI