Predicting Asthma Attacks in Primary Care with Machine Learning

First published: 14/06/2022

Last updated: 03/01/2024

Study Planned

Administrative details

EU PAS number

EUPAS47718

Study ID

47719

DARWIN EU® study

No

Study countries

United Kingdom

Study description

Introduction: Most asthma attacks and subsequent deaths are potentially preventable. We aim to develop a prognostic tool for identifying patients at high risk of asthma attacks in primary care by leveraging advances in machine

learning. Methods and analysis: Current prognostic tools use logistic regression to develop a risk scoring model for asthma attacks. We propose to build on this by systematically applying various well-known machine learning techniques including deep learning techniques to a large longitudinal deidentified primary care database, the Optimum Patient Care Research Database, and comparatively evaluate their performance with the existing logistic regression model and against each other. Machine learning algorithms vary in their predictive abilities based on the dataset and the approach to analysis employed. We will undertake feature selection, classification (both one-class and two-class classifiers) and performance evaluation. Patients who have had actively treated clinician-diagnosed asthma, aged 8–80 years and with 3 years of continuous data, from 2016 to 2018, will be selected. Risk factors will be obtained from the first year, while the next 2 years will form the outcome period, in which the primary endpoint will be the occurrence of an asthma attack.

Study status

Planned

Research institutions and networks

Institutions

University of Edinburgh (UofE)

United Kingdom

First published: 23/11/2018

Last updated: 16/12/2024

Institution | Educational Institution | Hospital/Clinic/Other health care facility | ENCePP partner

Networks

Optimum Patient Care (OPC) Network

United Kingdom (Northern Ireland)

First published: 26/09/2015

Last updated: 16/06/2025

Network | ENCePP partner

Contact details

Study institution contact

Syed Ahmar Shah ahmar.shah@ed.ac.uk

Study contact

ahmar.shah@ed.ac.uk

Primary lead investigator

Syed Ahmar Shah

Primary lead investigator

Study timelines

Date when funding contract was signed
Planned: 02/11/2020

Study start date
Planned: 09/12/2021

Data analysis start date
Planned: 09/12/2021

Actual: 01/02/2022

Date of final study report
Planned: 02/12/2024

Sources of funding

- Non-for-profit organisation (e.g. charity)
- Other

More details on funding

Asthma UK, Chief Scientist Office Scotland

Regulatory

Was the study required by a regulatory body?

No

Is the study required by a Risk Management Plan (RMP)?

Not applicable

Methodological aspects

Study type

Study type:

Non-interventional study

Scope of the study:

Assessment of risk minimisation measure implementation or effectiveness

Disease epidemiology

Main study objective:

We aim to develop a prognostic tool for identifying patients at high risk of asthma attacks in primary care by leveraging advances in machine learning.

Study Design

Non-interventional study design

Cohort

Study drug and medical condition

Medical condition to be studied

Asthma

Population studied

Age groups

- Children (2 to < 12 years)
- Adolescents (12 to < 18 years)
- Adults (18 to < 46 years)

- Adults (46 to < 65 years)
- Adults (65 to < 75 years)
- Adults (75 to < 85 years)
- Adults (85 years and over)

Estimated number of subjects

100.000

Study design details

Outcomes

Asthma Attacks

Data analysis plan

We propose to apply various well-known machine learning techniques including deep learning to a large longitudinal deidentified primary care database, the Optimum Patient Care Research Database, and comparatively evaluate their performance with the existing logistic regression model and against each other. Machine learning algorithms vary in their predictive abilities based on the dataset and the approach to analysis employed. We will undertake feature selection, classification (both one-class and two-class classifiers) and performance evaluation. Patients who have had actively treated clinician diagnosed asthma, aged 8–80 years and with 3 years of continuous data, from 2016 to 2018, will be selected. Risk factors will be obtained from the first year, while the next 2 years will form the outcome period, in which the primary endpoint will be the occurrence of an asthma attack.

Data management

ENCePP Seal

The use of the ENCePP Seal has been discontinued since February 2025. The ENCePP Seal fields are retained in the display mode for transparency but are no longer maintained.

Data sources

Data source(s)

Optimum Patient Care Research Database

Data sources (types)

Electronic healthcare records (EHR)

Use of a Common Data Model (CDM)

CDM mapping

No

Data quality specifications

Check conformance

Unknown

Check completeness

Unknown

Check stability

Unknown

Check logical consistency

Unknown

Data characterisation

Data characterisation conducted

No