# Similarity-based approaches to identifying risk of future asthma attack using UK primary care data

**First published:** 22/04/2021

**Last updated:** 22/04/2021

Study Planned

## Administrative details

**EU PAS number**

EUPAS36059

**Study ID**

36060

**DARWIN EU® study**

No

**Study countries**

☐ United Kingdom

**Study description**

Our aim is to investigate similarity-based approaches to identify the risk of a patient having an asthma attack in the 1 year following an asthma review appointment in primary care. We will use primary care electronic health records to derive this model. The two similarity-based approaches that we will investigate are a neighbourhood approach (i.e. identifying the patients who are most similar to the patient of interest and training the model only on these patients) and a cluster approach (i.e. identifying groups of similar patients via cluster analysis and training a separate model for each group). The variables that we will use to quantify similarity and derive the risk score will be chosen based on similar risk-prediction studies in asthma.

---

**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**

### Asthma UK Centre for Applied Research

## Contact details


**Study institution contact**

Elspeth Horne elsie.horne@ed.ac.uk

**Study contact**

elsie.horne@ed.ac.uk

**Primary lead investigator**

Aziz Sheikh

**Primary lead investigator**


## Study timelines

**Date when funding contract was signed**
Planned: 18/08/2020

**Study start date**
Planned: 18/08/2020

**Data analysis start date**
Planned: 18/08/2020

**Date of interim report, if expected**
Planned: 30/06/2021

**Date of final study report**
Planned: 31/12/2021

## Sources of funding

- Other

## More details on funding

MRC (via PhD studentship)

## 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:**

Disease epidemiology

**Main study objective:**

To investigate similarity-based approaches to identifying the risk of asthma attack in the year following an asthma review appointment in primary care.

## Study Design

**Non-interventional study design**

Cohort

## Study drug and medical condition

**Medical condition to be studied**

Asthma

## Population studied

**Age groups**

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**

500000

## Study design details

**Outcomes**

risk of asthma attack in following 1 year

---

**Data analysis plan**

The primary analysis will be carried out using data from OPCRD. The approach used in OPCRD will be externally validated using data from the SAIL databank. Methods of analysis are summarise below: Traditional approach: We will use methods such as logistic regression to derive a score corresponding to the risk of asthma attack in the following year. Similarity-based approaches: We will use metrics such as the Euclidean distance to quantify the similarity between patients. We will then use this metric to identify a subsample using both a neighbourhood approach (i.e. all patients within a specified similarity threshold) and a cluster approach (i.e. identify subsamples using cluster analysis). We will use methods such as logistic regression to derive a risk prediction model from the identified subsample.

# 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.

**Conflicts of interest of investigators**

ENCePP_COI_EUPAS36059.pdf(64.99 KB)

**Composition of steering group and observers**

EUPAS36059-40675.pdf(26.03 KB)

# Data sources

**Data source(s)**

Optimum Patient Care Research Database

SAIL Databank

**Data source(s), other**

Optimum Patient Care Research Database (OPCRD), SAIL databank

**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