|  |  |
| --- | --- |
| **Project Title** | AAMOS-00 Study: Predicting Asthma Attacks Using Connected Mobile Devices and Machine Learning |
| **Creators** | Kevin C.H. Tsang, Hilary Pinnock, Andrew M. Wilson, Dario Salvi, Syed Ahmar Shah |
| **Identifier** | IRAS ID: 285505, REC reference: 20/EE/0286, Sponsor reference: AC20145 |
| **Description** | Monitoring asthma condition is essential to asthma self-management. However, traditional methods of monitoring require high levels of active engagement and patients may regard this level of monitoring as tedious. Passive monitoring with mobile health devices, especially when combined with machine learning, provides an avenue to dramatically reduce management burden. However, data for developing machine learning algorithms are scarce, and gathering new data is expensive. A few asthma mHealth datasets are publicly available, but lack objective and passively collected data which may enhance asthma attack prediction systems.  To fill this gap, we carried out the 2-phase, 7-month AAMOS-00 observational study to collect data about asthma status using three smart monitoring devices (smart peak flow meter, smart inhaler, smartwatch), and daily symptom questionnaires. Combined with localised weather, pollen, and air quality reports, we have collected a rich longitudinal dataset to explore the feasibility of passive monitoring and asthma attack prediction. Conducting phase 2 of device monitoring over 12 months, from June 2021 to June 2022 and during the COVID-19 pandemic, 22 participants across the UK provided 2,054 unique patient-days of data. This valuable anonymised dataset has been made publicly available with the consent of participants. |
| **Funders** | This work was supported by Asthma + Lung UK as part of the Asthma UK Centre for Applied Research grant number AUK-AC-2018-01. Support from Malmö University was co-funded by the Knowledge Foundation KK-stiftelsen. |
| **Rights** | Creative Commons Attribution 4.0 International Public License. Recommended text for attribution: authors who publish using this dataset are asked to acknowledge the research participants as data contributors and the study investigators as follows: “*These data were contributed by the participants of the AAMOS-00 study as part of Kevin Tsang’s PhD developed at the Usher Institute at the University of Edinburgh under the supervision of Dr Syed Ahmar Shah, Prof. Hilary Pinnock, and Prof. Andrew Wilson; and in collaboration with Dr Dario Salvi and the Mobistudy team based at Malmö University.*” |
| **Language** | English |
| **Dates** | Study ethical approval: 17th December 2020  Data start (phase 2): 24th June 2021  Data end (phase 2): 2nd June 2022 |
| **Location** | United Kingdom, Europe |
| **Methodology** | Protocol: "Predicting asthma attacks using connected mobile devices and machine learning; the AAMOS-00 observational study protocol" - BMJ Open, DOI: 10.1136/bmjopen-2022-064166 |
| **Study data pack** | aamos00\_anonym\_data\_documentation.docx  aamos00\_data\_dictionary.xlsx  anonym\_aamos00\_dailyquestionnaire.csv  anonym\_aamos00\_endquestionnaire.csv  anonym\_aamos00\_environment.csv  anonym\_aamos00\_patient\_info.csv  anonym\_aamos00\_peakflow.csv  anonym\_aamos00\_smartinhaler.csv  anonym\_aamos00\_smartwatch1.csv  anonym\_aamos00\_smartwatch2.csv  anonym\_aamos00\_smartwatch3.csv  anonym\_aamos00\_weeklyquestionnaire.csv |
| **File Formats** | CSV, XLSX |
| **Anonymising the study dataset** | The directly identifiable information fields were removed or replaced, these included name, dates (including date of birth and date of data entry), location, height, weight, medication used, and user key.   * The names of participants were removed. * The age in years was calculated at the end date of the study, which replaced the participant’s date of birth information. Furthermore, the age was reduced in granularity via the use of age ranges. * Likewise, only the age range of the age of asthma diagnosis was made available. The ranges were early childhood (0 to 6 years old), late childhood (7 to 11 years old), adolescence (12 to 18 years old), and late onset (19+ years old). * The daily locations of participants were not made available, but the local environmental data (weather, air quality, and pollen count) collected during the study were made available. A single location of participants at the UK region level was not identifiable and was made available. The information would be sufficient to link localised historic weather data. * BMI and theoretical maximum PEF were important measures which were calculated from height and weight. The BMI range and theoretical maximum PEF rounded to the nearest 5 were made available. * The list of medication used by patients were removed. * The participant user keys were replaced with a new random number between 100 and 999. * All dates of data entry in the dataset were removed. The dates of data entry were transformed to the number of days after each participant started phase 2.   However, patient sex and race were made available. They were indirectly identifiable information and considered to have a low risk of deanonymisation. |
| **Assessing the possibility of identification** | We considered what means could be used to identify individuals using this data, including the time, cost, technology, and ability to link datasets, we believe the risk of identification is minimal. Furthermore, the dataset has been reduced to contain the minimal number of variables to conduct analysis. |
| **Releasing the data** | Anonymised data is available via DataStore under open access (CC BY 4.0) |