Objectives Non-attendance of scheduled hospital appointments represents a major issue affecting service effectiveness, efficiency and quality of care costing the NHS over £1 billion annually. This impact is even more detrimental at a time where the NHS is experiencing record high waiting times in the peri-COVID-19 pandemic era.

Rather than a reactive model of discharging patients for nonattending their appointments, we propose a proactive model identifying patients at risk of not showing up and provide them with right support at the right time. This approach is especially important for vulnerable population including young people (YP) due to the complex interplay between developmental, socio-economic factors can impact significantly on their medical care.

The increasing use of electronic health record systems (EHRs) and data availability creates opportunities to develop risk scores for specific patient populations.

In this study, we aim to develop a machine learning approach to develop a complex, multi-dimensional predictive model to identify YP at risk of clinic nonattendance.

Methods University College London Hospital (UCLH) switched to a new EHRs in April 2019. We extracted data on outpatient Adolescent and Young Adult Rheumatology (AYAR) between 2019-2022.

Our primary outcome was nonattendance of a scheduled appointment.

Our Predictor variables were defined after literature review, consultation with clinical and operational teams. We extracted data on 67 predictors of nonattendance. These variables are broadly divided into demographics (e.g., Age, Sex, ethnicity) and index of multiple deprivation (IMD) extracted from office of national statistics (ONS) database. We also included service utilisation history (e.g., previous history of clinic non-attendance), appointment information (month, day, time, clinic codes), and patient engagement (e.g., active in MyChart [online patient portal]).

Using data from 11602 outpatient appointments in (AYAR) clinics at UCLH, we built and assessed the performance of a predictive model as to whether a YP would not attend a scheduled outpatient appointment. We used logistic regression analysis to fit and assess the Model built. We evaluated its fit based on discrimination and calibration.

Results We identified a total of 1517 clinic non-attendance out of total of 11602 (13.1%) appointment.

Female/male ratio was 2.03 in non attendance group as compared to 2.33 in total clinic population.

In terms of age group, 10% (606/5547) of clinics booked for YP aged 14–18 were not attended as compared to 15% (651/4282) in those aged [19–24].

Feature engineering analysis revealed that the most significant factors were IMD followed by distance, previous history of Non-attendance, age group and appointment hour.

Conclusions Aiming to identify YP at risk of Non-attendance, we used a step-by-step approach to build a model that can be applied using EHR and IMD data at the point of care. High proportion of YP nonattending their appointments were from deprived areas.

Accurate stratification of non-attendance risk can provide us with unique opportunity for preventative interventions, supporting to most vulnerable YP and improve the use of resources within the wider system.