

## A machine-learning approach to identify cerebral palsy cases using primary care database

Fan, H<sup>1</sup>, Li, L<sup>1</sup>, Wijlaars, L<sup>2</sup>, and Gilbert, R<sup>1</sup>

<sup>1</sup>Institute of Child Health, University College London

<sup>2</sup>University College London

### Introduction

Cerebral palsy (CP) is a complex condition that can manifest in different ways and is likely to be under-recorded in primary care databases. As identification based only on diagnostic codes will underestimate the incidence of cerebral palsy, an approach is needed to use all available information to identify more cases.

### Objectives and Approach

We used a machine-learning approach to identify likely CP cases in the Clinical Practice Research Datalink (CPRD), a UK primary-care database: i) we selected potential codes (symptoms, therapies and managements) to identify CP cases by comparing relative frequencies of codes associated with known cases versus non-cases; ii) codes with higher discriminative ability were further selected using a Random Forest method based on a resampled, balanced population; iii) we calculated the predictive score for each patient employing the discriminative ability of selected codes; and iv) manually reviewed the full medical records of likely cases (i.e. patients with higher predictive score).

### Results

Primary care records were available for 343,199 young people aged 2000 were identified as likely cases (n=365, 0.1% of previously non-cases). Their full medical records in CPRD were manually reviewed by experts and 85 children (23.3% of likely cases) were validated as CP cases additional to the 341 initially identified by diagnostic codes, resulting in a 20% increase of CP cases.

### Conclusion/Implications

Data-driven schemes, such as the machine learning methods applied here, have the potential of identifying the most informative predictors in a cost-effective and rapid way to identify likely CP cases or other complex medical conditions in primary care database.

