

Predictive use by Cambs Institute of Criminology/GLA

"The data requested will be used to create a model to predict exclusion based solely on so-called routine administrative data for London schools and schoolchildren."

The University of Cambridge is working with the Greater London Authority and the Education Endowment Foundation (EEF) to set up a randomised control trial in London schools in the coming months. The project is funded by the Council of Europe and the EEF and aims to assess the effectiveness of an intervention for children at risk for fixed term exclusion from school and is set to begin in October 2013 with roughly 800 children in Years 9 and 10 in roughly 40 selected schools.

The project aims to assess whether the proposed intervention is associated with multiple outcomes, such as: improved classroom behaviour; improved performance on educational assessments; reduction in the likelihood of being NEET after completion of compulsory schooling; improved attendance; better than predicted grades at GCSE.

The data requested will be used to create a model to predict exclusion based solely on so-called routine administrative data for London schools and schoolchildren. This can be achieved with an anonymous dataset (i.e. no pupil/school ID numbers needed, but indicators of which pupils attend which school would be needed to account for clustering of data) - the sole purpose is to create the best predictive model for exclusion. As it's not clear what the best predictors would be, we (necessarily) require as much information as possible on the pupils, short of knowing their names and addresses, hence the breadth of our request. We would also require as much information as possible about the schools in terms of school level measures of ethnicity, gender, academic performance, type of school (e.g. secondary, academy, specialist disability school) and so on. Once we have this model, we will then repeat the exercise with data from the 40 schools in the study. This will be a separate NPD application and/or may be collected directly from the participating schools. We will also use the current NPD request to explore the possibility of creating a brief predictive tool which schools can use to identify pupils at risk of exclusion, as an aid to informal teacher assessments.

Our specific request relates to individual and school level data on all current Year 10 and 11 pupils in London. This allows us to examine the predictive validity of the model against later exclusion (i.e. we can use the Year 9 data for the current Year 10 pupils to predict exclusion in Year 10; and Year 9/10 data for current Year 11 pupils to do the same). We understand that there is some time-lag for the exclusion data, so we would welcome advice from the NPD team in the event that what we propose would be hampered by this (e.g. would it more straightforward to just use the current Year 11 data rather than both Yr10/11?). We are also particularly keen to obtain pupil level data on school absence as this is likely to be a key corollary of later exclusion. The only additional 'flag' we request is an indicator of whether the pupil is currently Yr10 or Yr11.

Tier 1: pupil level, identifying and highly sensitive for use between September 2013 and 2019

Their statement " This can be achieved with an anonymous database" should not be taken to mean anonymous data are used. The data they requested is identifying and appears to be everything about the child except name and home address.

"As it's not clear what the best predictors would be, we (necessarily) require as much information as possible on the pupils, short of knowing their names and addresses, hence the breadth of our request."

"We would also require as much information as possible about the schools in terms of school level measures of ethnicity, gender, academic performance, type of school (e.g. secondary, academy, specialist disability school) and so on. Once we have this model, we will then repeat the exercise with data from the 40 schools in the study. This will be a separate NPD application and/or may be collected directly from the participating schools."

"We will also use the current NPD request to explore the possibility of creating a brief predictive tool which schools can use to identify pupils at risk of exclusion, as an aid to informal teacher assessments."

Download original requests :

Data application - The Department for Education was unable to find and provide this document via FOI. The only one of 20 requested.

FOI DfE response: "g. 5.9.2013 University of Cambridge Institute of Criminology - we have been unable to retrieve this documentation dating back to 2013"

source: https://www.whatdotheyknow.com/request/pupil_data_licensing_agreements_2#incomin-g-1079063