Methodology for Correction of Missing or Miscoded Episode Types in Primary Care Computerised Medical Records

N. Smith[1,*], V. Livina[1], R. Byford[2], A. Correa[2,3], F. Ferreira[2], I. Yonova[2,3], S. de Lusignan[2,3]

1. National Physical Laboratory, Hampton Road, Middlesex, TW11 0LW
2. Department of Clinical and Experimental Medicine, University of Surrey, Guildford, UK
3. Royal College of General Practitioners, London, UK

* Corresponding author: nadia.smith@npl.co.uk

1. Abstract

Identifying incident (first or new) episodes of illness is important for NHS England to inform about the seasonal onset of diseases and early warning of epidemics, as well as differentiating change health service utilisation from change in pattern of disease. The most reliable way of differentiating incident from prevalent cases is through the clinician assigning episode type to the patient's computerised medical record (CMR); however, this is not widely done in a consistent way. The objective of this collaborative study between Royal College of General Practitioners (**RCGP**) Research and Surveillance Centre (**RSC**), University of Surrey and the National Physical Laboratory (**NPL**) is to develop a methodology to correct missing or miscoded episode types. The data is gathered from the RCGP RSC network of over 200 practices, is analysed and poor episode typing corrected by disease type, using statistical methods.

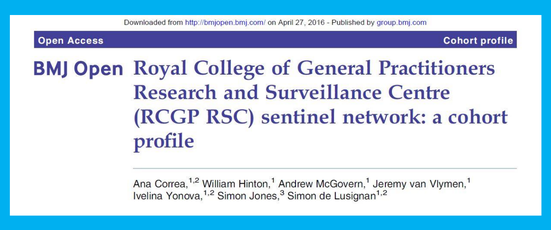
2. Data Science at NPL

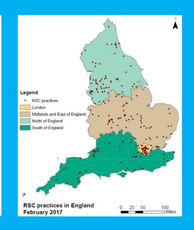
NPL is the UK's National Measurement Institute (NMI), is a world-leading centre of excellence in developing and applying the most accurate measurement standards. It does this to enable Governmental and Societal requirements to be met, which include Energy, Healthcare and Security as well as other sectors.

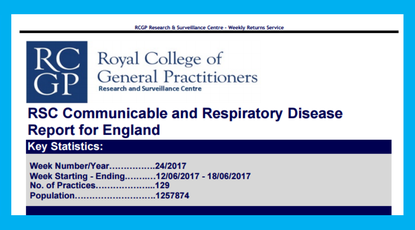
NPL is active in areas of mathematics and scientific computing which support measurement science and has years of experience in evaluating and understanding measurement uncertainty. NPL is currently active in applying this knowledge to the medical area, as well as building skills in the data provenance and data security areas.

3. RCGP RSC

The Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC) is an internationally renowned source of information, analysis and interpretation of primary care data. In 2015, it established a new data and analysis hub at the University of Surrey.







Established in 1957, the RSC is an active research and surveillance unit which collects and monitors data since 1967, in particular influenza and other respiratory diseases, from over 230 practices across England, and provides a Weekly Returns Service (WRS).

4. Episode types

The WRS monitors number of patients consulting with **incident (first or new) episodes** of illness classified by diagnosis in England, and are the key primary care element of the national disease monitoring systems run by Public Health England (PHE).

Identifying incident cases is important to predict **seasonal onset** of diseases and early warning of any **epidemic** at regional and national levels, and monitor vaccine uptake and effectiveness.

Episode type (first, new or ongoing) is assigned to the patient's Computerised Medical Record (CMR) by the clinician – it is the way to differentiate incident from prevalent cases.

5. Data quality problem

Poor quality data in primary care can have many **negative consequences** within the practice, the RCGP RSC's surveillance programme, and other epidemiological research.

The RCGP RSC provides **feedback** and **online training** on the importance of good quality entry of medical problems where episode typing is key.

The RCGP RSC runs data quality checks to ensure the data received every week is coherent. Around 15 to 20 practices are excluded due to **low numbers of incident events** (first or new episode types).

6. Methodology for correcting episode types

An algorithm in SQL has been developed by NPL that adds missing and updates incorrect episode types for the diseases monitored weekly by the RCGP RSC. The algorithm analyses data from a 6-week increment, searching for all events of the monitored diseases and calculates an episode using statistical methods and based on clinical knowledge.

	ILI and Acute Bronchitis		All monitored diseases	
Number of events	4030		9716	
Number of missing episode types	215	5.3 %	1362	14.02 %
Number of miscoded episode types	305	7.6 %	1166	12 %
Number of updated episode types	520	12.9 %	2528	26.02 %

Currently the methods are being tested in practices with good episode type data quality for validation. Once fully validated, the algorithm will be used to correct data weekly and historical data will be analysed retrospectively.

Department for Business, Energy & Industrial Strategy
FUNDED BY BEIS

Royal College of General Practitioners Research & Surveillance Centre

UNIVERSITY OF SURREY

NPL National Physical Laboratory

www.npl.co.uk