UNIVERSITY
of York

RESEARCH

CHE
Centre For Health Economics

**A Comparison of NHS Digital
and CHE Algorithms for Spells
and CIPS Construction**

**Anastasia Arabadzhyan,
Adriana Castelli, James Gaughan,
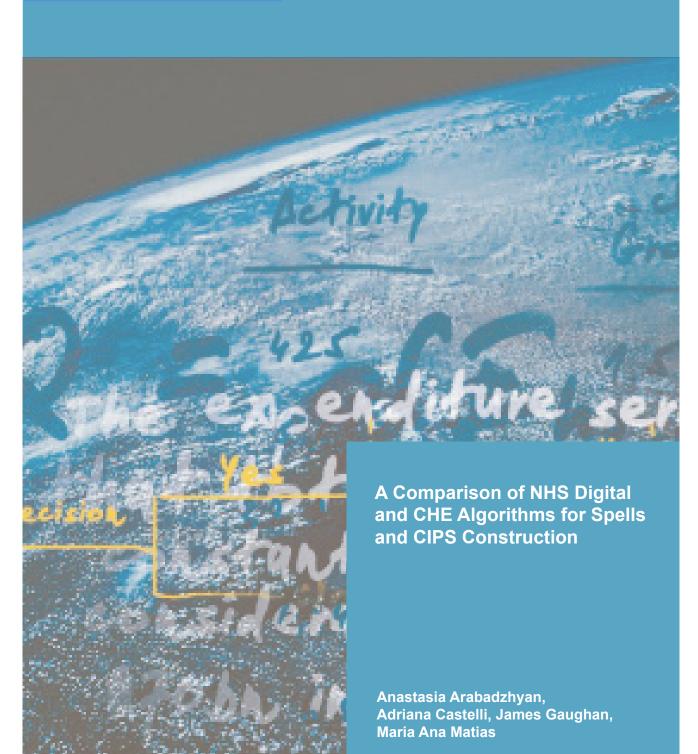Maria Ana Matias**

A comparison of NHS Digital and CHE algorithms for spells and CIPS construction

Supplementary material to CHE Research Paper 182
Productivity of the English National Health Service: 2018/19 update

Anastasia Arabadzhyan
Adriana Castelli
James Gaughan
Maria Ana Matias

Centre for Health Economics, University of York, UK

April 2021

Background to series

CHE Discussion Papers (DPs) began publication in 1983 as a means of making current research material more widely available to health economists and other potential users. So as to speed up the dissemination process, papers were originally published by CHE and distributed by post to a worldwide readership.

The CHE Research Paper series takes over that function and provides access to current research output via web-based publication, although hard copy will continue to be available (but subject to charge).

Further copies

Only the latest electronic copy of our reports should be cited. Copies of this paper are freely available to download from the CHE website www.york.ac.uk/che/publications/. Access to downloaded material is provided on the understanding that it is intended for personal use. Copies of downloaded papers may be distributed to third parties subject to the proviso that the CHE publication source is properly acknowledged and that such distribution is not subject to any payment.

Printed copies are available on request at a charge of £5.00 per copy. Please contact the CHE Publications Office, email che-pub@york.ac.uk, telephone 01904 321405 for further details.

Centre for Health Economics
Alcuin College
University of York
York,
YO10 5DD, UK
www.york.ac.uk/che

In this document we set out to compare the algorithms developed by NHS Digital and the Centre for Health Economics (CHE), University of York, to construct NHS hospital provider spells and Continuous Inpatient Spells (CIPS) of hospital care using the Hospital Episode Statistics Admitted Patient Care (HES APC) dataset.

HES APC contains data on all admissions to NHS hospitals in England, along with those to private providers where the care is funded by the NHS. Each record of the dataset corresponds to a single episode of care under one consultant in a given hospital. Each record contains the patient's socio-demographic characteristics, episode start and end dates, diagnoses and procedures, among other variables. In most applications, to avoid double counting, the pool of episodes is restricted to Finished Consultant Episodes (FCEs), i.e. episodes under the care of a consultant in a single hospital and which were concluded within a financial year (yearly data releases cover all inpatient care episodes from 1989/90 financial year onwards). Since 1997/98, it has also been possible to follow patients over time and across hospitals, using a pseudonymised patient identifier.

To characterise a patient's stay in a given hospital from admission to discharge, FCEs are aggregated into spells, which can be defined as the uninterrupted period of care delivered by a single provider (NHS Digital, 2016). When a patient is seen by multiple consultants during the same hospital stay, the linked FCEs are said to belong to the same spell. However, if a transfer (to another hospital) occurs, a new spell begins.

In the case of a spell ending with a transfer, when the patient begins a new spell within 2 days, two (or more) consecutive spells can be linked into a Continuous Inpatient Spell (CIPS). A CIPS is an uninterrupted period of care within the NHS hospital sector. A CIPS starts when a patient is admitted to a hospital and ends when a patient is discharged (not transferred) or dies.

Below, we compare two approaches for the construction of spells and CIPS, developed by NHS Digital and the Centre for Health Economics respectively. We describe three consecutive steps: data preparation (step1), spell construction (step 2) and CIPS construction (step 3).

Step 1: Data preparation

The first step entails defining the pool of episodes and sorting them in a logical and chronological order. The NHS Digital algorithm includes FCEs (episode status (*epistat*[1]) equal to 3) with: a valid provider code (*procodet*); a valid admission date, i.e. between 01/01/1930 and the HES data period end date (31/03 of the year). The CHE method restricts the pool of episodes to those with: valid provider code (*procodet*); admission date (*admidate*)*;* episode identifier (*epikey*); episode order (*epiorder*); episode start date (*epistart*), episode end date (*epiend*) dates[2]; episode duration (*epidur*); and valid or missing discharge date (*disdate*)*.*

The CHE approach includes two further cleaning steps. First, a transit indicator is constructed to chronologically order FCEs occurring on the same day. It is also used to remove duplicates and to sort episodes before constructing the spells. Whilst the NHS Digital approach also constructs and uses a transit indicator, its use is limited to sorting episodes prior to CIPS construction. The CHE transit indicator (*transit_che*) takes the value 1 when an episode did not begin as a transfer but ended in a hospital transfer, takes the value 2 when an episode began and ended in a hospital transfer, and the value 3 when the episode began as a hospital transfer but did not end as a transfer (e.g. a patient was discharged, died or is still in hospital), and 0 otherwise. The NHS Digital transit indicator (*transit_nhsd*) is conceptually defined in the same way but differs slightly in that it considers different admission sources and discharge destinations. While NHS Digital defines hospital transfers as those occurring between any type of provider, the CHE algorithm considers transfers as only those between NHS providers other than medium secure units.

---

[1] HES APC variable names and indicators derived from HES APC variables are presented in Italic.
[2] Valid episode end date implies that the episode is finished.

Secondly, an additional procedure for the detection of duplicates and their exclusion is applied.[3] Records are ordered by patient id (*extract_hesid*), episode start date (*epistart*), episode order (*epiorder*), episode end date (*epiend*), and transit indicator (*transit_che*). A single episode for each combination of patient ID, episode start and end dates, episode order, and transit indicator value is kept and the others dropped as duplicates.

Table 1 provides a summary of the data preparation steps described above, comparing the NHS Digital and CHE algorithms.

Table 1: Data preparation steps

NHS Digital	CHE algorithm
Criteria for selecting episodes	
Finished Consultant Episodes (FCEs): *epistat* = 3 Episodes with a valid provider code (*procodet*) and *admidate* (between 01/01/1930 and HES data period end date, i.e. 31/03/XX)	Episodes with valid *procodet, admidate, epikey, epiorder, epistart, epiend, epidur,* and valid or missing *disdate.*
Transit indicator: used to sort when constructing CIPS	Transit indicator: used to sort when constructing spells and CIPS, and to detect duplicates
transit_nhsd = 1 if {*admisorc* is NOT in [49-53, 87[4]] OR *admimeth* is NOT in [81, 2B]} AND *disdest* is in [49-53, 84, 87] transit_nhsd = 2 if {*admisorc* is in [49-53, 87] OR *admimeth* is in [81, 2B]} AND *disdest* is in [49-53, 84, 87] transit_nhsd = 3 if {*admisorc* is in [49-53, 87] OR *admimeth* is in [81, 2B]} AND *disdest* is NOT in [49-53, 84, 87]	transit_che = 1 if *admisorc* is NOT in [51-53] AND *admimeth* is NOT in [81, 2B] AND *disdest* is in [51-53] transit_che = 2 if {*admisorc* is in [51-53] OR *admimeth* is in [81, 2B]} AND *disdest* is in [51-53] transit_che = 3 if {*admisorc* is in [51-53] OR *admimeth* is in [81, 2B]} AND *disdest* is NOT in [51-53]
	Duplicates removal
	1. Sort episodes by *extract_hesid, epistart, epiorder, epiend, transit_che.* 2. Drop duplicates defined as records with the same *extract_hesid, epistart, epiorder, epiend*, and *transit_che.*

Step 2: Spells construction

In the NHS Digital approach, episodes are sorted by patient ID (*HESID_mapped*), admission date (*admidate*), provider code (*procodet_mapped*) and spell ID (*provspnops*). If two (or more) consecutive episodes have the same *HESID_mapped*, *admidate*, *procodet_mapped* and *provspnops*, they are considered to be within the same spell. Next, to identify the correct order of episodes within the newly created spells, episodes within the spells are sorted by episode start date (*epistart*), episode order (*epiorder*), episode end date (*epiend*), and episode number (*epikey*). Spell level admission and discharge variables (needed for the CIPS construction) are defined as the values of the corresponding variables of the first and the last episode in the spell respectively.

Unfortunately, not all the variables used by NHS Digital for the construction of the provider spell are available to CHE. The variable *provspnops* is considered as a sensitive data field by NHS Digital and it is not available to all HES APC data users. The closest analogue to *provspnops* which can be used by CHE to

---

[3] Prior to receipt, HES data undergoes a cleaning procedure that removes some duplicate records, but some records remain, hence the need for an additional cleaning step.
[4] The encodings of the variables are listed in the Annex.

reconstruct spells using the NHS Digital approach is the anonymised *susspellid* (which is a SUS[5]-generated spell identifier). In addition, *HESID_mapped* is also not identical to the patient ID variable available to CHE (*extract_hesid*). In addition, the provider identifier (*procodet*) undergoes a mapping procedure[6], which generates the *procodet_mapped* variable. Therefore, CHE's algorithm cannot perfectly replicate the structure of spells and CIPS as constructed with the NHS Digital algorithm. However, the discrepancies are very minor. The CHE's algorithm follows a similar logic to NHS Digital's, but uses a slightly different set of variables when sorting episodes and assigning them to provider spells.

The main difference between the two approaches is that the CHE algorithm does not use the spell identifiers available in the HES APC dataset, for two reasons: (1) following a change in access rights, from 2010 the *provspno* spell identifier available to CHE was replaced by *susspellid*, which is not identical to other spell ID variables and (2) ready-to-use spell ID variables may have missing or invalid values. Therefore, the CHE method for the construction of spells does not rely on any external spell identifier(s) but uses both logical and chronological ordering to define provider spells. This is done as follows. First, episodes are ordered by: person identifier (*extract_hesid);* admission date *(admidate);* episode start date *(epistart);* episode order *(epiorder);* episode end date *(epiend);* transit indicator (*transit_che*); discharge date *(disdate);* provider code *(procodet);* and episode number *(epikey).* Consecutive episodes with the same *extract_hesid*, *procode3* (3-digit provider code), admission date (*admidate)* and discharge date *(disdate)* are considered to form a spell. If the discharge date in the first of the two consecutive episodes is missing and the second of the two consecutive episodes has a valid discharge date (indicating that the patient is still in hospital after the first episode), they are also considered to be in the same spell[7].

Table 2 summarises the main steps of the NHS Digital and CHE algorithms for the construction of provider spells.

Table 2: NHS Digital and CHE 'spell construction' algorithms

NHS Digital	CHE algorithm
Two (or more) consecutive episodes which have the same *HESID_mapped*, *admidate*, *procodet_mapped* and *provspnops* are considered to be in the same provider spell[8]. The order of episodes within the spell is determined by sorting on *epistart*, *epiorder*, *epiend*, *epikey*. Spell admission and discharge variables are defined as the values of corresponding variables of the first and the last episode in the spell respectively.	Episodes are ordered by *extract_hesid, admidate, epistart, epiorder, epiend, transit, disdate, procodet, epikey*. Two (or more) consecutive episodes are in the same spell if: They have the same *extract_hesid, procode3, admidate*, and *disdate* OR *disdate* of the previous episode is missing (i.e. the patient was not discharged).

Step 3: CIPS construction

Moving to the CIPS construction, both NHS Digital's and CHE's algorithms follow the same logic and assign spells to a single CIPS if a hospital transfer took place. NHS Digital's method starts with sorting episodes by: patient ID (*HESID_mapped*); spell admission date (*spell_admidate*); spell discharge date

---

[5] Secondary Uses Service.

[6] This implies that all activity is reported for the most recent provider codes (which matters if there have been any hospital mergers or splits).

[7] It is necessary to add this possibility because recording practices may differ across providers. E.g. a spell consisting of two episodes may have a missing discharge date recorded for the first episode in the spell, and a valid discharge date for the second episode. Alternatively, the first episode may also have a valid discharge date, which is identical to the one of the second episode.

[8] Regular attender episodes are considered as single episode spells and are not linked to other episodes.

(*spell_disdate*); transit indicator (*transit_nhsd*); and spell ID (*provspnops*). Then two consecutive episodes are assigned to the same CIPS if they belong to the same spell, or in the case of a transfer to another provider, when no more than two days have elapsed between discharge from the previous spell (provider 1) and admission to the next spell (provider 2). Cases where a patient is discharged to their usual place of residence and readmitted as an emergency to a ward for general patients/younger physically disabled patients or where a patient is taken to an A&E department and then admitted are not considered as being part of a single CIPS.

CHE's method is very similar but uses episodes' start and end dates (*epistart, epiend*) instead of admission and discharge dates, and a narrower range of discharge destinations and admission sources when defining a transfer. Table 3Table 3 summarises the algorithms and lays out the differences in more detail.

Table 3: NHS Digital and CHE 'CIPs construction' algorithms

NHS Digital	CHE algorithm
Episodes are ordered by *HESID_mapped, spell_admidate, spell_disdate, transit_nhsd, provspnops.*	Episodes are ordered by *extract_hesid, epistart, epiorder, epiend, transit_che, epikey.*
Two (or more) consecutive spells belong to the same CIP if they:	Two (or more) consecutive spells belong to the same CIP if they:
Belong to the same spell OR	Belong to the same spell OR
Have the same *HESID_mapped* AND not more than 2 days elapsed between discharge from the previous spell and admission to the next spell AND one of the following criteria is met, indicating that a transfer has taken place: {*disdest* of the previous provider spell is in [49-53, 84, 87] OR *admisorc* of next provider spell is in [49-53, 87] OR *admimeth* of next provider spell is in [81, 2B]} AND EXCLUDE the combination where *disdest* of the first spell is 19 AND *admisorc* of the next spell is 51 AND *admimeth* of the next spell is 21	Have the same *extract_hesid* AND the next FCE is the first in the spell (*epiorder* = 1) AND not more than 2 days elapsed between the end of the previous episode and the start of the next episode AND one of the following criteria is met, indicating that a transfer took place: {*disdest* of the previous episode is in [51-53] OR *admisorc* of the next episode is in [51-53] OR *admimeth* is in [81, 2B]}

Annex: episode status, admission sources and methods and discharge destination values used in the algorithms

Epistat (Episode status)
1 = Unfinished
3 = Finished
9 = Derived unfinished (not present on processed data)

Admisorc (Admission source)
49 = NHS other hospital provider: high security psychiatric accommodation in an NHS hospital provider (NHS trust)
50 = NHS other hospital provider: medium secure unit (1999-00 to 2006-07)
51 = NHS other hospital provider: ward for general patients or the younger physically disabled or A&E department
52 = NHS other hospital provider: ward for maternity patients or neonates
53 = NHS other hospital provider: ward for patients who are mentally ill or have learning disabilities
87 = Non-NHS run hospital

Admimeth (Admission method)
21 = Emergency: via Accident and Emergency (A&E) services, including the casualty department of the provider
2B = Transfer of an admitted patient from another hospital provider in an emergency
81 = Transfer of any admitted patient from another hospital provider other than in an emergency; this does not include admissions to high security psychiatric hospitals (HSPH)

Disdest (Discharge destination)
48 = High security psychiatric hospital, Scotland (from 1999-2000)
49 = NHS other hospital provider - high security psychiatric accommodation
50 = NHS other hospital provider - medium secure unit (from 1999-2000)
51 = NHS other hospital provider - ward for general patients or the younger physically disabled
52 = NHS other hospital provider - ward for maternity patients or neonates
53 = NHS other hospital provider - ward for patients who are mentally ill or have learning disabilities
84 = Non-NHS run hospital - medium secure unit (from 2003-04)
87 = Non-NHS run hospital

References:

NHS DIGITAL 2016. Methodology to create provider and CIP spells from HES APC data.
NHS DIGITAL 2017. HES Data Dictionary: Admitted Patient Care.