Extending the Critical Path
A presentation by the Critical Illness Definitions and Geographical Variations Working Party

17 February 2014

Agenda

• Background
• Hospital Episodes Statistics data set
• Methodology for aggregate population rates table CIBT08
• Example: Cancer
• CIBT08 composition by illness
• Comparisons with CIBT02 and ACL04
• Geodemographic variation investigations
• Discussion and questions

17 February 2014
Background

Evolution of a Working Party

2009
• Health & Care PEC launches member-led research initiative

2010
• CI Working Parties form late in 2010 and select two topics:
  • ABI+/non-ABI Definitions & Geographical Variations

2011
• HES data request submitted
• Desk-based research commences

2012
• HES data analysis begins
• Initial findings presented at Health & Care Conference

2013
• Working parties merge with revised aims, including CIBT08
• Final report and presentation December 2013
What is in the 2013 report?

- Data and methodology overview
- Illness-by-illness analysis
  - Cancer, Heart Attack, Stroke
  PLUS
  - 36 other CIs (in alphabetical order) of which:
    - 17 are ABI standard
    - 16 are non-ABI full payment
    - 3 are non-ABI partial benefits and not included in CIBT08
    - No TPD
- CIBT08 tables
  - Compared with CIBT02 and ACL04

17 February 2014
Report layout for each illness

• What is it?
• Symptoms & treatment
• Risk factors
• Insurance industry definitions
• Derived incidence rates
• Geodemographic analysis selected conditions only

Working Party members

• Peter Banthorpe
• Phil Cleverley
• Christine Fairall
• Adele Groyer
• Jennifer Loftus
• Ketiwe Nhende
• Christopher Reynolds
• Daniel Ryan
• Matthew Smith
• James Tait
• Neelish Tiwari

• Thanks also to
  – CACI & Experian
  – Health & Social Care Information Centre
  – Oracle for MySQL (free)
  – Institute & Faculty of Actuaries
  – Former Working Party Members Aaron Tindale & James Shattock
  – Neil Robjohns and authors of “Exploring the Critical Path”
HES data

Hospital Episodes Statistics data set

- Seriatim data of all finished consultant episodes in NHS hospitals
  - Inpatient and outpatient data
- Data years 1989/90 to 2009/10 received
  - 1997/98 to 2009/10 are coded with unique patient identifiers
- 18 million records for 2009/10 alone!
What the HES data looks like

<table>
<thead>
<tr>
<th>Patient Identifier</th>
<th>Unique identifier by patient – 47m of these</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic Patient Information</td>
<td>Age, gender</td>
</tr>
<tr>
<td>Basic Episode Information</td>
<td>Date started, date finished, admission method, current status etc</td>
</tr>
<tr>
<td>Diagnosis Information</td>
<td>Up to 20 different diagnoses</td>
</tr>
<tr>
<td>Procedure Information</td>
<td>Up to 20 different operations, with date of operation</td>
</tr>
<tr>
<td>Geographical Information</td>
<td>Postal district, Lower Super Output Area, IMD Rank, Mosaic Type, ACORN Type, Health ACORN type</td>
</tr>
</tbody>
</table>

Example data

- 20 Diagnosis codes
- Each record is an individual episode
- Each ICD code could appear multiple times as a primary diagnosis (DIAG_01) or in secondary diagnosis fields
Analysing the Data - a SQL algorithm

Single Life (HES_ID) → Derive complete episode history
20 diagnoses/operations for each episode → Define the "order"
Choose the first event of interest

<table>
<thead>
<tr>
<th></th>
<th>All Data</th>
<th>First Ever</th>
<th>First Ever CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Simply a sum of all episode counts that match the ICD/OPCS coding</td>
<td>A sum of all episode counts that match the ICD/OPCS coding where only 1 is permitted per life</td>
<td>A sum of all episode counts that match the ICD/OPCS coding assuming there is no relevant medical history that could qualify as a CI claim. Only 1 count is permitted per life.</td>
</tr>
</tbody>
</table>

Example to explain different HES counts

LEAD IN 1997 - 2006

INVESTIGATION 2007 - 2009

<table>
<thead>
<tr>
<th>Year</th>
<th>2008 Cancer</th>
<th>2009 Heart attack</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>First ever</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>First ever CI</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

17 February 2014
HES data adjustment and limitations

- Remove day cases and unfinished episodes
- Limitations
  - England only
  - No private hospitalisations / non-hospitalised treatments
  - No sudden deaths *we used other data sources to adjust for this*
  - Coding changes over time
  - Incompleteness of older data, especially multiple diagnoses
  - Inconsistencies between individual coders
  - Multiple HES IDs for one patient

Methodology for CIBT08 rates
Summary of data inputs: CY 2007 - 2009

- HES first ever & first ever CI counts
- Sudden deaths (various research sources)
- Population count (ONS data, adjusted for 2011 census)
- Population prevalence of all* CIs (various research sources)
- 28-day mortality rates (various research sources)
  - for standalone rates only
- Proportion of deaths per illness (ONS Mortality in the 21st Century)
- Population death rate (ONS England Interim Life tables 2007-9)
- Graduated each of these using penalised B-splines

* Cancer, Heart Attack & Stroke only

17 February 2014

Rate calculation steps

A: Crude first ever incidence

\[
\text{Crude first ever incidence} = \text{First ever HES count PLUS Sudden deaths} \div \text{Gross population (ONS 2007 – 2009)}
\]

B: Crude first ever CI incidence = A x (1 - Overlap)

\[
\text{Overlap} = \frac{\text{First ever CI HES Count}}{\text{First Ever HES Count}}
\]

E: Derived incidence rate (Ix)

\[
\text{Derived incidence rate (Ix)} = \text{First ever CI HES count PLUS Sudden deaths} \div (\text{Gross population LESS “All” CI Prevalence})
\]

17 February 2014
Standalone CI rates

E: Derived incidence rate ($l_x$)

First ever CI HES count PLUS Sudden deaths Gross population LESS "All" CI Prevalence

G: Standalone CI incidence rate ($l'_x$)

$= l_x * (1 – 28\text{-day mortality rate})$

Convert from an "$m_x$-type" figure to $q_x$

Accelerated CI rates

J: Addition for Accelerated CI ($l_x - k_x.m_x$)

First ever CI HES count PLUS Sudden deaths Gross population LESS "All" CI Prevalence $k_x.m_x = \text{Proportion of population deaths in respect of that illness}$

Total accelerated rate all CIs

$= \Sigma_{\text{all CI}} l_x + (1 - \Sigma_{\text{all CI}} k_x.m_x) . m_x$

Convert from an "$m_x$-type" figure to $q_x$
Main differences vs CIBT02: Data

• HES only
• Different ICD-10 codes used in places
  – e.g. Benign Brain Tumour rates increased significantly because of this
• No adjustment for insurance definition severity unless HES data allows this

Main differences vs CIBT02: Method

• First ever calculated from data
  – not wider research
• Only adjustment to crude count is sudden death
• Remove all prior CI’s from hospitalisations (overlap)
• Population denominator is reduced by prevalence of all* CIs
  – Largely offsets effect of the change to all CI overlap

* Cancer, Heart attack & Stroke only
Example: Cancer

Insurance definition

Cancer – excluding less advanced cases (2011)

Any malignant tumour positively diagnosed with histological confirmation and characterised by the uncontrolled growth of malignant cells and invasion of tissue.

The term malignant tumour includes leukaemia, sarcoma and lymphoma except cutaneous lymphoma (lymphoma confined to the skin).

For the above definition, the following are not covered:

- All cancers which are histologically classified as any of the following:
  - pre-malignant;
  - non-invasive;
  - cancer in situ;
  - having borderline malignancy; or
  - having low malignant potential;

- All tumours of the prostate unless histologically classified as having a Gleason score greater than 6 or having progressed to at least clinical TNM classification T2N0M0.

- Chronic lymphocytic leukaemia unless histologically classified as having progressed to at least Binet Stage A.

- Any skin cancer (including cutaneous lymphoma) other than malignant melanoma that has been histologically classified as having caused invasion beyond the epidermis (outer layer of skin).

- All versions of ABI Statement of Best Practice definitions are reproduced.

- Examples of ABI+ versions are provided.
  - e.g. all skin cancers that have metastasised

- For other CIs examples of non-standard ABI definitions are provided.
Cancer population CI incidence (males)

<table>
<thead>
<tr>
<th>Age Band</th>
<th>20-39</th>
<th>40-59</th>
<th>60-79</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoothed, Interpolated Crude Rate</td>
<td>5.92</td>
<td>31.25</td>
<td>166.96</td>
</tr>
<tr>
<td>Adjustment for Overlap</td>
<td>-6.9%</td>
<td>-9.8%</td>
<td>-21.5%</td>
</tr>
<tr>
<td>Prevalence Rate</td>
<td>0.7%</td>
<td>5.5%</td>
<td>25.3%</td>
</tr>
<tr>
<td>Derived Incidence Rate $I_x$</td>
<td>5.55</td>
<td>30.11</td>
<td>180.57</td>
</tr>
<tr>
<td>28 Day Mortality Rates</td>
<td>-0.6%</td>
<td>-0.9%</td>
<td>-1.1%</td>
</tr>
<tr>
<td>Stand Alone Rates $I'_x$</td>
<td>5.51</td>
<td>29.83</td>
<td>178.42</td>
</tr>
<tr>
<td>Mortality Rates</td>
<td>9.15</td>
<td>36.28</td>
<td>219.94</td>
</tr>
<tr>
<td>Proportions of Deaths $k_x$</td>
<td>10.1%</td>
<td>26.0%</td>
<td>37.2%</td>
</tr>
<tr>
<td>Addition for Accelerated Rates $I_x - k_xq_x$</td>
<td>4.57</td>
<td>19.30</td>
<td>103.74</td>
</tr>
</tbody>
</table>

Rates are per 10,000 population

Cancer changes vs CIBT02

- Now includes overlap with all other prior CIs
- Additional ICD-10 codes D45, D46 and D47 (3%)
- Updated prevalence statistics (to reduce population count denominator)
- Updated survival rates (used in 28-day mortality adjustment)
Cancer CIBT08 vs CIBT02 standalone rates

Cancer sources of change vs CIBT02
CIBT08 composition by illness

CIBT08 Composition: Males

17 February 2014
CIBT08 Composition: Females

Combined CIBT08 comparison with other tables
Comparison with CIBT02

Observations vs ACL04

Residual line very much lower suggesting we are materially over-estimating some CIs which are also not well exposed in historic CMI data

Smoker rates are higher and >100% for heart attack
Other observations vs ACL04

ACL04 (by Cause) vs CIBT08 (Female Non-Smokers)

MS appears high, especially at younger ages
HES data not ideal because of outpatient treatment

17 February 2014
Geodemographic Segmentation

Geodemographic segmentation systems

Mosaic UK
acorn

Mosaic © Experian 2013. All rights reserved.
Acorn © CACI 1979-2013. All rights reserved.
These are some of the most affluent people in the UK. They live in wealthy high-status suburban, rural and semi-rural areas of the country…

These people tend to be older empty nesters and retired couples. Many live in rural towns and villages, often in areas where tourism is important…

These are the most prosperous people living in our main cities. They are very well educated and tend to be employed in senior managerial and professional…
Index of Multiple Deprivation

- IMD provides a relative measure of deprivation at small area level across England (LSOA)
- Deprivation domains used include:
  - Income;
  - Employment;
  - Health and disability
  - Education;
  - Crime;
  - Barriers to housing and services
  - Living Environment.
- We consider quintiles I1 – I5 where I1 is the most deprived.
- Our HES dataset includes the 2004 version of IMD.

17 February 2014
Cancer Sites – Bottom-Up

Cancer Sites – Top-Down
What conditions have we considered?

- See the full paper for geodemographic analysis for:
  - Aorta Graft Surgery;
  - Benign Brain Tumour;
  - **Cancer** (All, Breast, Lung, Melanoma)
  - Heart Valve Replacement and Repair;
  - Kidney Failure;
  - Major Organ Transplant;
  - **Heart Attack**;
  - Multiple Sclerosis;
  - **Stroke**.

- Summary results follow:
Expressions of individual views by members of the Institute and Faculty of Actuaries and its staff are encouraged.

The views expressed in this presentation are those of the presenter.