Quality of the national upper gastrointestinal cancer dataset

NCIN Data Briefing

Background

A report on the data quality of the national upper gastrointestinal (UGI) cancer dataset was produced for the UGI cancer Site Specific Clinical Reference Group.

It is important to analyse the quality of these data as large proportions of missing or poor quality information might lead to inaccurate conclusions being drawn. Analysis on more detailed subgroups may also be difficult. Various indicators of data quality are routinely used in cancer registration. These include Death Certificate Only (DCO) registrations, and the proportion with a definite anatomical subsite.

DCO registrations: Some cancer registrations are initiated by the patient’s death certificate. Registrations from a death certificate are followed up in hospital systems or using the Hospital Episode Statistics dataset. Many cases are found and their details are updated to form a complete registration. Those with no further information remain as DCO registrations. These registrations will have limited information and therefore need to be excluded from some analyses.

Anatomical site: Some cancers can be divided into anatomical subsites which include cancers in different parts of the organ for example, the upper, middle or lower third of the oesophagus. Other cancers, such as liver cancer, can only be divided into morphological subsites which include cancers in different types of cells within that organ. Of the upper GI tumours, cancers of the oesophagus, stomach, biliary tract and pancreas can be divided into anatomical subsites.

Data were extracted from the National Cancer Data Repository on 234,668 patients diagnosed with upper gastrointestinal cancers between 1999 and 2008.

Table 1: Number of registrations by upper gastrointestinal cancer type, England, 1999-2008

<table>
<thead>
<tr>
<th>Cancer type (ICD10 code)</th>
<th>Number of registrations</th>
<th>Death certificate only registrations</th>
<th>Number of registrations (excluding DCOs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oesophageal cancer (C15)</td>
<td>62,752</td>
<td>1,629</td>
<td>61,123</td>
</tr>
<tr>
<td>Stomach cancer (C16)</td>
<td>69,801</td>
<td>2,533</td>
<td>67,268</td>
</tr>
<tr>
<td>Duodenal cancer (C17.0)</td>
<td>2,851</td>
<td>83</td>
<td>2,768</td>
</tr>
<tr>
<td>Liver cancer (C22)</td>
<td>24,423</td>
<td>2,241</td>
<td>22,182</td>
</tr>
<tr>
<td>Gallbladder cancer (C23)</td>
<td>4,680</td>
<td>336</td>
<td>4,344</td>
</tr>
<tr>
<td>Biliary tract cancer (C24)</td>
<td>6,642</td>
<td>215</td>
<td>6,427</td>
</tr>
<tr>
<td>Pancreatic cancer (C25)</td>
<td>63,519</td>
<td>5,352</td>
<td>58,167</td>
</tr>
<tr>
<td>All cancer types</td>
<td>234,668</td>
<td>12,389</td>
<td>222,279</td>
</tr>
</tbody>
</table>

**KEY MESSAGE:**
Encouragingly, the proportion of death certificate only registrations decreased between 1999 and 2008.

More precise anatomical classification for oesophageal, stomach and pancreatic cancers is needed to allow specific subgroups to be defined and analysed.
Results

The proportion of DCO registrations is low and has decreased between 1999 and 2008 for all UGI cancer types (Figure 1). Figure 2 shows that only around 50% of oesophageal, stomach and pancreatic cancer registrations had a known anatomical subsite. The recording of this information was better for biliary tract cancers with around 90% of registrations having a known subsite.

Recommenations: Encouragingly, the proportion of death certificate only registrations was generally low and was decreasing between 1999 and 2008. These registrations would have to be excluded from survival analysis and may indicate incomplete case ascertainment; both factors which could potentially bias survival estimates. Therefore, work should continue to reduce the proportion of DCO registrations further. It is important that where applicable the precise anatomical subsite is recorded in medical records and that it is passed to the cancer registries. This will allow more specific subgroups to be defined and analysed.

Final note

The full report also investigates several other data quality measures including the proportion of registrations that were microscopically verified, had a linked Hospital Episode Statistics (HES) record, and had a known ethnicity. It also attempts to ascertain the completeness of the UGI dataset. For further information, this report is available on the NCIN website at http://www.ncin.org.uk/cancer_type_and_topic_specific_work/cancer_type_specific_work/upper_gi.aspx

FIND OUT MORE:

Thames Cancer Registry
Thames Cancer Registry is the lead cancer registry for upper gastrointestinal cancers
http://www.tcr.org.uk

Other useful resources within the NCIN partnership:

Cancer Research UK CancerStats – Key facts and detailed statistics for health professionals
http://info.cancerresearchuk.org/cancerstats/

The National Cancer Intelligence Network is a UK-wide initiative, working to drive improvements in standards of cancer care and clinical outcomes by improving and using the information collected about cancer patients for analysis, publication and research. Sitting within the National Cancer Research Institute (NCRI), the NCIN works closely with cancer services in England, Scotland, Wales and Northern Ireland. In England, the NCIN is part of the National Cancer Programme.