Examining missing data and chance variation in public reporting of cancer stage at diagnosis

Supporting cancer data indicator development and use

Matthew Barclay, Georgios Lyratzopoulos, David Greenberg, Gary Abel

Acknowledgements
Early stage indicators

- Concerns about potentially unwarranted variation have led to the creation of performance indicators
- **Stage**: Only possible to consider as performance indicator in recent years, because of major and ongoing improvements in quality and completeness of data
- Statistical properties of indicators not always formally assessed
- Innovation comes with challenges: No (inter)national standard or precedent to guide us on early stage indicators – we need to find out ourselves, now

Aims

1. Investigate the bias in different approaches to handling missing stage data in measuring and comparing organisational performance
2. Investigate the impact of chance variation on the indicator
   - Investigate the Spearman-Brown reliability of the early-stage indicator
   - Quantify likely misclassification rates for CCG pay-for-performance targets due to chance variation
   - Estimate sample sizes allowing for reliable reporting
Dataset

Ten sites, chosen to match current stage indicators
- breast
- lung
- prostate
- colorectal
- melanoma
- NHL
- bladder
- kidney
- endometrial
- ovarian

English, national cancer data for study year 2013 (208,112 diagnoses)
18% missing stage information

Potential bias due to missing stage information
The ‘Gold Standard’

- Multiple imputation used to create data with our best estimate of the complete stage information
- Imputation models (stratified by cancer site) included extensive auxiliary information
  - CCG
  - Region
  - Sex
  - Age
  - Deprivation group
  - Ethnicity of patient
  - Screening detection status
  - Tumour grade
  - Nelson-Aalen estimate of cumulative hazard
  - Death within 30 days
  - Death within 365 days
  - Basis of diagnosis

Rubin D Wiley 1987
van Buuren S et al Stat Med 1999

Two ‘non-Gold Standard’ approaches to consider

- Missing-is-late? (current approach)
- Complete-case?
Gold standard early stage against % of tumours with missing stage, CCGs 2013

Missing-is-late early stage against % of tumours with missing stage, CCGs 2013
Complete-case early stage against % of tumours with missing stage, CCGs 2013

Bias in *missing-is-late / complete-case* approach against % of tumours with missing stage, CCGs 2013

- Bias not strongly associated with missing data
- Missing data leads to high bias
Bias in *missing-is-late* / *complete-case* approach against % of tumours with missing stage, CCGs 2013

(Current approach)

**Missing-is-late approach**

**Complete-case approach**

Bias compared with Gold Standard, % points

% of all tumours which have missing stage

Bias compared with Gold Standard, % points

% of all tumours which have missing stage
Currently...

- The complete-case approach should be adopted for calculating early-stage indicators
- “All missing are late stage” is not currently accurate
- The impact of missing data on the accuracy of the indicator should be monitored

Aims

1. Investigate the bias in different approaches to handling missing stage data in measuring and comparing organisational performance
2. Investigate the impact of chance variation on (putative complete-case) indicator
   - Investigate the Spearman-Brown reliability of the early-stage indicator
   - Quantify likely misclassification rates for CCG pay-for-performance targets due to chance variation
   - Estimate sample sizes allowing for reliable reporting
Reliability

All observed variation due to chance

Roland M et al. BMJ 2009
Lyratzopoulos G et al. Med Care 2011

Reliability

All observed variation due to real differences between organisations

Roland M et al. BMJ 2009
Lyratzopoulos G et al. Med Care 2011
Reliability

Minimum reliability considered useful
(want this for the PHOF indicator)

Roland M et al BMJ 2009
Lyratzopoulos G et al Med Care 2011

Reliability

Required level for high stakes applications
(want this for the CCG Quality Premium indicator)

Roland M et al BMJ 2009
Lyratzopoulos G et al Med Care 2011
Median reliability of the putative complete-case early-stage indicator (1-year of data)

Required for pay-for-performance: 0.9
Required for public reporting: 0.7

0.66 – CCGs
0.60 – Local authorities

We estimated notable underlying variation in proportion early-stage between CCGs (despite low reliability)

Modelled distribution of CCG performance

51.8
95% of CCGs
62.3

Performance on complete-case early stage indicator (%)
Misclassification of CCGs in respect of our putative complete-case indicator (1-year of data)

CCGs will receive a financial incentive for performance on the early stage indicator by:

- 60% of tumours diagnosed at an early stage
- Improving by 4 or more percentage points from the previous year

Incentive is £1 per head of population ~ £250k for an average CCG

Ran simulations based on diagnosis counts and a model of performance in 2013 to assess misclassification rates on both targets

<table>
<thead>
<tr>
<th>Misclassification of CCGs – 60% target (1-year of data; putative complete case indicator)</th>
</tr>
</thead>
</table>
| **True early stage ≥ 60%** | **True early stage < 60%** | Pos.
| Positive predictive value | 53% (37% to 68%) |
| Sensitivity | 73% (56% to 89%) |
| Specificity | 89% (85% to 94%) |
| Negative predictive value | 95% (92% to 98%) |
| Observed early stage ≥ 60% | 21 true positives | 19 false positives |
| Observed early stage < 60% | 8 false negatives | 161 true negatives |
### Misclassification of CCGs – 60% target
(1-year of data; putative complete case indicator)

<table>
<thead>
<tr>
<th>Observed early stage ≥ 60%</th>
<th>True early stage ≥ 60%</th>
<th>True early stage &lt; 60%</th>
</tr>
</thead>
<tbody>
<tr>
<td>21 true positives</td>
<td>19 false positives</td>
<td></td>
</tr>
<tr>
<td>8 false negatives</td>
<td>161 true negatives</td>
<td></td>
</tr>
</tbody>
</table>

**Sensitivity**
- 73% (56% to 89%)

**Specificity**
- 89% (85% to 94%)

**Positive predictive value**
- 53% (37% to 68%)

**Negative predictive value**
- 95% (92% to 98%)

---

### How many years of data are needed for better reliability?

<table>
<thead>
<tr>
<th>Years of data required for reliable indicators for 90% of organisations for both public reporting and pay-for-performance standards</th>
</tr>
</thead>
<tbody>
<tr>
<td>Years</td>
</tr>
<tr>
<td>CCGs</td>
</tr>
<tr>
<td>Public reporting (Reliability ≥ 0.7)</td>
</tr>
<tr>
<td>2.3</td>
</tr>
<tr>
<td>Pay-for-performance (Reliability ≥ 0.9)</td>
</tr>
<tr>
<td>8.7</td>
</tr>
<tr>
<td>LAs</td>
</tr>
<tr>
<td>Public reporting (Reliability ≥ 0.7)</td>
</tr>
<tr>
<td>2.7</td>
</tr>
<tr>
<td>Pay-for-performance (Reliability ≥ 0.9)</td>
</tr>
<tr>
<td>10.5</td>
</tr>
</tbody>
</table>
Implications

• Evidence suggests the **missing-is-late specification** of the early-stage indicator leads to biased comparisons of CCG performance

• The **complete-case specification** of the early-stage indicator is much less biased for 2013, and likely will continue to be less biased in the most recent data

• A putative complete-case indicator
  – will have low reliability when based on one year of data
  – will have adequate reliability for use in public reporting schemes when based on three years of data
  – will not be suitable for pay-for-performance use

Implications

• Careful examination of the properties of indicators can support evidence-based policy in this area

• Appropriate statistical techniques can give insight into the validity and reliability of indicators – and how they can be improved….
  – ….so that we can extract maximum value from high quality cancer staging data; making the excellent data ‘work harder’

• We look forward to further work with policy-makers and other colleagues to help optimise indicator specification