Cancer surveillance

TU/e: big data & health meetings
Contents

• IKNL
• Cancer surveillance program
• Tooling
• Demo
• Projects
Cancer in 2018: cause of death

Bron: www.cbs.nl
Cancer in 2018

Incidence: 116,500

Prevalence: 800,000

5-yr survival: 64%
Cancer 1990-2018

Bron: www.cijfersoverkanker.nl
Integraal Kankercentrum Nederland (IKNL)

- Independent knowledge institute
- Netherlands cancer registry (NCR)
- National coverage since 1989
- ~550 employees (500 FTE)
- Financing: 85% VWS, 15% subsidies
Netherlands comprehensive cancer organisation

**Mission**
- To reduce the impact of cancer

**Vision**
- Less cancer
- More cure or longer lives
- Better quality of live
- Improved societal participation
- Dying with dignity

**Values**
- Independent
- Data driven and innovative
- Resolute but flexible
- Ambitious and humble

*Letting Data Live*
Core activities

1. Registration & Abstraction
   - Population Based Registry
   - Patient data
   - Actionable Insight

2. Analysis & Knowledge generation
   - Knowledge

3. Reporting & Implementation
   - NKRP
Netherlands cancer registry

Pathology laboratories (PALGA)
Electronic Health Records
PROMS
GBA

IKNL data managers

0100 1011
0100 1011
0111101
1011110

R&D:
• Epidemiologists, clinical informats, data scientists…
• ±250 scientific publications/year
• ±350 data requests per year
• Trusted Advisors
• Multidisciplinary tumor teams
• NKR online
• Cijfersoverkanker.nl

2.5M patients
±200 items per patient

Patient journey

Disease – care – outcome

...
Netherlands comprehensive cancer organisation
Cancer surveillance

Cancer population monitor based on the NCR
Mesothelioma incidence in a Dutch shipyard

Stumphius J.

Asbest in een bedrijfsbevolking; een onderzoek naar het voorkomen van asbestlichaampjes en mesotheliomen op een scheepswerf en machinefabriek

Jan Stumphius

Proefschrift Universiteit van Amsterdam

Taal: Nederlands
Uitgever: Van Gorcum [en]: Assen
Voorzieningen: 1986
Kanszoveen: 260 pagina's, 23 cm, 11, 11
Aanklaming: Proefschrift Universiteit van Amsterdam

Stumphius saved lives!
Progress in cancer outcomes

Hunches → Data → Research → Intervention
But...

- What if no-one has an insight?
- We have the cancer registry (NCR)
- Can we let the NCR speak?
- Don’t have a moral obligation to do so?
Cancer monitor: a net of analyses

Statistics:
- Incidence
- Survival
- Care variation
- ...

Dimensions
- Regions
- Hospital
- Patient groups
- ...

incidence
standardized rate (ESR)
Postcode area
All cancers 2007-2016
Geographical analysis
2007-2016
Where are most new cancer cases?
All cancers - incidence

Statistic: incidence
Dimensions: province
Population - 2016
Where are relatively most new cancer cases?
All cancers – crude rate (cases/100,000 people)

Statistic: crude rate
Dimensions: province
Average age - 2016
And if we **standardize** for age?
All cancers – European standardized rate (RESR, 2013)

Statistic: standardized rate (RESR)
Dimensions: province
Finding surprising insights across tumors & regions
2007-2016
Finding surprising facts

Net:
- Alle tumor types (~300)
- 40 COROP region’s

Cause: Human herpesvirus (HHV8)
Risk factor: HIV

<table>
<thead>
<tr>
<th>Regio</th>
<th>diff_norm</th>
<th>Incidence</th>
<th>tumorsoort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overig Zeeland</td>
<td>2.253364</td>
<td>60.0</td>
<td>Mesotheliom van het peritoneum</td>
</tr>
<tr>
<td>Achterhoek</td>
<td>2.161902</td>
<td>196.0</td>
<td>19 Spijsverteringsorganen, niet nader omschreven</td>
</tr>
<tr>
<td>Arnhem/Nijmegen</td>
<td>2.133622</td>
<td>317.0</td>
<td>19 Spijsverteringsorganen, niet nader omschreven</td>
</tr>
</tbody>
</table>
Kaposi sarcoma (1989-2016)
ESR (1976) – NKR 1989-2018

Greater-Amsterdam
<table>
<thead>
<tr>
<th>omschrijving</th>
<th>Regio</th>
<th>Incidence</th>
<th>diff_standardized</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prostaatcarcinoom</td>
<td>Flevoland</td>
<td>2330.0</td>
<td>3.010453</td>
</tr>
<tr>
<td>Kaposi sarcoma</td>
<td>Noord-Holland</td>
<td>168.0</td>
<td>2.995154</td>
</tr>
</tbody>
</table>
Prostate carcinoma in Flevoland

Statistic: standardized rate

Dimension: province

*: p<=0.05
**: p<=0.01
Prostate carcinoma
Cluster analysis
Cancer cluster analysis (guideline 2012)

Receives signals from worried citizens and doctors (± 40 per year) → communication

Provides data from the Netherlands Cancer Registry

Can investigate further, e.g. environmental links

Reactive
## Info berekening

<table>
<thead>
<tr>
<th>Locatie:</th>
<th>Bereken verwachte incidentie</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ref. nummer:</td>
<td></td>
</tr>
<tr>
<td>Berekenende periode: 10 jaar</td>
<td></td>
</tr>
</tbody>
</table>

### Periode

<table>
<thead>
<tr>
<th>Vanaf</th>
<th>T/m</th>
<th>Kankersoort</th>
<th>Men</th>
<th>Vrouwen</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>2010</td>
<td>Luchtpijp &amp; Long</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Verwachte aantal op basis van landelijke incidentie in 10 jaar tijd

<table>
<thead>
<tr>
<th></th>
<th>Ondergrens</th>
<th>Boven grens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mannen</td>
<td>7</td>
<td>21</td>
</tr>
<tr>
<td>Vrouwen</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>Mannen-vrouwen</td>
<td>11</td>
<td>26</td>
</tr>
</tbody>
</table>

### Aantal inwoners in gebied

<table>
<thead>
<tr>
<th>Leeftijd</th>
<th>Mannen</th>
<th>Vrouwen</th>
<th>Totaal</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-4</td>
<td>60,95238</td>
<td>59,28571</td>
<td>120,238095</td>
</tr>
<tr>
<td>5-9</td>
<td>64,04752</td>
<td>65,47519</td>
<td>129,522731</td>
</tr>
<tr>
<td>10-14</td>
<td>66,42857</td>
<td>62,14286</td>
<td>128,571429</td>
</tr>
<tr>
<td>15-19</td>
<td>63,57143</td>
<td>56,47519</td>
<td>120,04663</td>
</tr>
<tr>
<td>20-24</td>
<td>48,80952</td>
<td>43,57143</td>
<td>92,3809526</td>
</tr>
<tr>
<td>25-29</td>
<td>50,71429</td>
<td>48,09624</td>
<td>98,8105238</td>
</tr>
<tr>
<td>30-34</td>
<td>62,14286</td>
<td>63,33333</td>
<td>125,47619</td>
</tr>
<tr>
<td>35-39</td>
<td>76,19048</td>
<td>76,42857</td>
<td>152,619048</td>
</tr>
<tr>
<td>40-44</td>
<td>85,47619</td>
<td>83,80952</td>
<td>169,285714</td>
</tr>
</tbody>
</table>

### Betrouwbaarheidsinterval

<table>
<thead>
<tr>
<th>Betrouwbaarheidsinterval</th>
</tr>
</thead>
<tbody>
<tr>
<td>95</td>
</tr>
</tbody>
</table>

### Aantekeningen
Calculation

- **Population** by age group from CBS
- **Crude rate** by age group from cijfersoverkanker.nl
- **Incidence** from IKNL (data request)

Expected incidence

Poisson(incidence, expected) → P-value
Which tumor – PC4 combinations are potential cluster candidates in the period 2008-2017?
Medullary thyroid cancer

- Hereditary condition Multiple endocrine neoplasia type 2 (MEN2)

- “If a child inherits the mutated RET gene from an affected parent, there is almost a 100% chance of developing medullary thyroid cancer”

- Patient’s turned out to be family members
Cancer surveillance program (2019)
Summary cancer surveillance program

Data → Cancer monitor → Research → Intervention

Contact:
Harm Buisman
h.buisman@iknl.nl
Cancer surveillance

Tooling
An interactive explorer for the Netherlands Cancer Registry

Prostate - SIR - 2006-2017

Demo!
The tools

- **NCR** in datawarehouse

- **Back-end:**
  - **Python**: pandas, geopandas, scipy, …
  - Object-oriented setup (statistics, regions, tumor classifications)

- **Front-end:**
  - **Bokeh**
  - JavaScript/jquery/CSS/HTML
Welcome to Bokeh

Bokeh is an interactive visualization library that targets modern web browsers for presentation. Its goal is to provide elegant, concise construction of versatile graphics, and to extend this capability with high-performance interactivity over very large or streaming datasets. Bokeh can help anyone who would like to quickly and easily create interactive plots, dashboards, and data applications.

To get started using Bokeh to make your visualizations, start with the User Guide.

For examples of how you might use Bokeh with your own data, check out the Gallery.

For detailed information about specific Bokeh components, consult the Reference Guide.

If you are interested in contributing to Bokeh, or extending the library, see the Developer Guide.
Bokeh

- Package for Python
- Visualization library focused on interactivity
- Embedding using jinja2 package
- Fast prototyping using Bokeh server

- But…
  - Small userbase, but active developers
  - Reference material is minimal
  - Steep learning curve
The circle of life

- Calculate new statistics in python
- User interacts with visualizations and changes parameters
- Prepare figure and content using Bokeh
- Template is styled and external scripts are loaded
- Inject elements into jinja2 template

g.out@iknl.nl
Cancer surveillance

Current activities
Circular scan statistics

For each circle.

- Obtain actual and expected number of cases inside and outside the circle and calculate the Likelihood.

Compare Circles:

- Pick circle with highest likelihood function as Most Likely Cluster.
• Cluster detection irrespectively of administrative boundaries, and without assumptions about cluster size or location

• Adjusts for multiple testing
Local pattern detection from cancer registry data: a descriptive approach

**Model = target concept = incidence distribution**

### Overall Population*

- Prostate: 20.1%
- Breast: 26.4%
- Colon: 12.6%
- Lung: 11.9%
- Lymphatic and leukemia: 9.1%
- Bladder and urothelial cancer: 3.6%
- Skin cancer: 3.6%
- Rectum: 3.6%
- Nasal cavity and sinuses: 2.6%
- Mouth and pharynx: 2.6%
- Brain and other NOS: 1.8%

### Lateralization left, < 25yo, Noord-Brabant*

- Prostate: 20.1%
- Breast: 26.4%
- Colon: 12.6%
- Lung: 11.9%
- Lymphatic and leukemia: 9.1%
- Bladder and urothelial cancer: 3.6%
- Skin cancer: 3.6%
- Rectum: 3.6%
- Nasal cavity and sinuses: 2.6%
- Mouth and pharynx: 2.6%
- Brain and other NOS: 1.8%

---

*Interestingness (mock data)*

c.attanasio@iknl.nl
Other activities

- Time patterns
- Prognosis
- Data portal
- Interactive dashboard
- Care
- Knowledge system
Help is needed!

We welcome collaborations and interns!
Summary cancer surveillance program

Data → Cancer monitor

Thank you for your attention!

Contact:
Harm Buisman
h.buisman@iknl.nl