The big data and information technology revolution in injury prevention: are we lagging behind?

Dr. Birgitte M. Blatter, Head unit Monitoring and Research, Consumer Safety Institute, Netherlands
BIG DATA
Background

- ICT revolution in daily life, mobile technology changing rapidly
  - Sensors leading to smart housing, cars, clothing
  - Wearable devices, physical activity trackers
  - Implantables
  - Big data, machine learning
- What about the use of information technology in health research in general and injury research in particular?
Goal and content

Are we using and taking advantage of these developments in injury research already?

- Innovative data collection
- Innovative data processing
- Literature on health research
- Practical examples from Consumer Safety Institute
- Successes and difficulties

The big data and information technology revolution in injury prevention: are we lagging behind?
Data collection

The big data and information technology revolution in injury prevention: are we lagging behind?
Review of literature I

- Successful application in environmental studies (smartphone use: photo’s, GPS, video’s): standing water, fast food, trash
- Infectious diseases outbreak studies, crowdsourced data collection

The big data and information technology revolution in injury prevention: are we lagging behind?
Review of literature II

- Widely used in health studies, sports: physical activity but also routes taken
- Guidelines on how to use, analyze and interpret data is needed to prevent researchers to make same mistakes

The big data and information technology revolution in injury prevention: are we lagging behind?
Evaluation of multifactorial intervention among elderly

- TOM pilots Veen vd Eurosafe 2019: balance, nutrition, social support, mobility:
  - wearable device to measure physical activity (steps and sit-stand transitions): ultimate goal to predict falls
  - not enough data, not sufficiently used
    - wearing comfort too low
    - no immediate feedback
    - too strict requirements for ‘daily average’ or ‘weekly average’

The big data and information technology revolution in injury prevention: are we lagging behind?
I-pads and online questionnaires instead of paper at hospital / ED

- Who still uses paper questionnaires?
- In hospitals in NL no e-mail addresses of emergency patients available, letters are sent → low response on online questionnaire
- Flyers and I-pads available in waiting room but low response → better if research personnel is present?
- ED visit for injuries: no long-term relationship, but one-off visit, not always in patients’ ‘own’ hospital
Apps on smartphone

● Cohort studies: apps can be installed to measure behaviour:
  • (type of) physical activity
  • dangerous use of smartphone (music, whatsapp, sms)
  • sleeping patterns
  • car or cycling speed
  • weather conditions
  • dangerous locations

● Injury research:
  • patient / case control studies
  • extract data afterwards?
  • privacy ? consent?

The big data and information technology revolution in injury prevention: are we lagging behind?
Data processing

The big data and information technology revolution in injury prevention: are we lagging behind?
Many data collected in hospital information systems (open text fields)

However, difficult to extract data

Automation of processes is difficult: ICT changes, no priority
Analyze open text fields by automatic text recognition software in DISS

- Administrative burden of our IDB (DISS) too high for ED personnel to register by means of questionnaires (n=14)
- Demand from local authorities occupied with road safety (n=7), ambulance data
- Merely extraction of open text fields in medical file
- Validity of automatic coding?
Can automatic text recognition software for coding injuries replace manual coding?

*Nijman S, Blatter B, Safety 2018, Bangkok*

“Car driver, collision against tree, high speed accident”

- Desired output:
  - Injury mechanism: contact with object
  - Products involved: car, tree

“Patient found intoxicated, used alcohol and speed”

- Desired output:
  - Injury mechanism: chemical mechanism
  - Products involved: alcohol, speed
Validation study on automatic text recognition II

• We taught the system (IBM Modeler) from scratch how to code information on accidents and injuries by means of different nodes

• All possible words were classified into libraries and the system was taught how to interpret sentences

• Data from one year, 100% manually checked

• Comparison: IBM Modeler out – manual check out
Nodes with coding instructions

The big data and information technology revolution in injury prevention: are we lagging behind?
## Results of validation study

### Injury mechanism - after manual check autotext

<table>
<thead>
<tr>
<th>Injury mechanism autotext</th>
<th>Fall</th>
<th>Contact with object</th>
<th>Contact with person or animal</th>
<th>Foreign body</th>
<th>Threat to breathing</th>
<th>Chemical mechanism</th>
<th>Thermal mechanism</th>
<th>Electricity, radiation, explosion</th>
<th>Physical over-exertion</th>
<th>Unspecified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall</td>
<td>42.500</td>
<td>1.009</td>
<td>289</td>
<td>12</td>
<td>8</td>
<td>33</td>
<td>16</td>
<td>3</td>
<td>216</td>
<td>147</td>
</tr>
<tr>
<td>Contact with object</td>
<td>1.979</td>
<td>13.370</td>
<td>182</td>
<td>104</td>
<td>8</td>
<td>44</td>
<td>27</td>
<td>9</td>
<td>85</td>
<td>140</td>
</tr>
<tr>
<td>Contact with person or animal</td>
<td>779</td>
<td>366</td>
<td>3.219</td>
<td>18</td>
<td>0</td>
<td>15</td>
<td>1</td>
<td>1</td>
<td>66</td>
<td>78</td>
</tr>
<tr>
<td>Foreign body</td>
<td>41</td>
<td>88</td>
<td>32</td>
<td>790</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>18</td>
</tr>
<tr>
<td>Threat to breathing</td>
<td>10</td>
<td>8</td>
<td>6</td>
<td>13</td>
<td>68</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Chemical mechanism</td>
<td>156</td>
<td>68</td>
<td>8</td>
<td>10</td>
<td>0</td>
<td>1.469</td>
<td>3</td>
<td>1</td>
<td>14</td>
<td>38</td>
</tr>
<tr>
<td>Thermal mechanism</td>
<td>35</td>
<td>30</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>85</td>
<td>401</td>
<td>4</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>Electricity, radiation, explosion</td>
<td>5</td>
<td>10</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>69</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Physical over-exertion</td>
<td>182</td>
<td>64</td>
<td>45</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>868</td>
<td>67</td>
</tr>
<tr>
<td>Unspecified</td>
<td>2.544</td>
<td>2.711</td>
<td>864</td>
<td>167</td>
<td>19</td>
<td>1.221</td>
<td>89</td>
<td>14</td>
<td>726</td>
<td>3.720</td>
</tr>
<tr>
<td>Total</td>
<td>48.231</td>
<td>17.724</td>
<td>4.647</td>
<td>1.119</td>
<td>106</td>
<td>2.950</td>
<td>550</td>
<td>103</td>
<td>1.981</td>
<td>4.229</td>
</tr>
</tbody>
</table>

- **% true**: 88% 75% 69% 71% 64% 50% 73% 67% 44% 88%
- **% unknown**: 5% 15% 19% 15% 18% 41% 16% 14% 37%
- **% false**: 7% 9% 12% 14% 18% 9% 11% 19% 20% 12%

**Total true**: 81% of the cases was coded correctly for injury mechanism by autotext recognition software

### Table 1

<table>
<thead>
<tr>
<th>Injury mechanism autotext</th>
<th>Fall</th>
<th>Contact with object</th>
<th>Contact with person or animal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall</td>
<td>42.500</td>
<td>1.009</td>
<td>289</td>
</tr>
<tr>
<td>Contact with object</td>
<td>1.979</td>
<td>13.370</td>
<td>182</td>
</tr>
<tr>
<td>Contact with person or animal</td>
<td>779</td>
<td>366</td>
<td>3.219</td>
</tr>
</tbody>
</table>
Machine learning

- Already satisfactory but 100% manual check
- Switch to new version of IBM Modeler is much work
- Is machine learning an alternative for nodes with coding instructions? Many commercial providers

*scientific study of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions*, relying on patterns and inference instead

- Are we still in control?
- New / changed categories?

Machine learning not the same as automatic coding
Conclusions

● Endless possibilities and opportunities of apps, sensors
● Use in injury research still limited
● More cohort studies (and case control studies) in injury research
● Share (innovative) methods more across research fields, open source applications make it possible
● Privacy and user friendliness are impeding factors
● Make use of abundance of existing data: experiment!

If not.. YES we will lag behind