It Takes One Line (of SAS) to Analyze Your Data, but it Takes Many Lifelines to See it (Through)

Shawn Murphy, Victor Castro, Wannapa Mahamaneerat, and Taowei David Wang
tdwang@partners.org

Partners HealthCare
October 30th, 2012
Does Glargine Increase Cancer Risk?

• A large German study on diabetes database suggests so
  – Controversial, but very large sample
  – Other Countries did the same study
    • Sweden, UK, Scotland
  – Conflicting results

• What about our own data?
  – Focus on glargine and pancreatic cancer
  – Performed 2 studies
Cohort Study from EHR

70,772 Patients
1980-2010
Cohort Study from EHR

70,772 Patients

2135 Glargine only
51177 Other insulin only
17460 Both
All Patients with some form of Insulin Therapy

- Glargine
- Other Insulin
- Both

Cohort Study from EHR

Statistical Significance
Cohort Study from EHR

But in Lifelines2, we see that Order of Cancer and Glargine is not enforced.
All Patients with some form of Insulin Therapy

Glargine

Other Insulin

Both

Statistical Significance

But in Lifelines2, we see that

Order of Cancer and Glargine is not enforced.

Nor is the time taken from Glargine to Cancer taken into account.
Cohort Study from EHR

All Patients with some form of Insulin Therapy

All Patients with Pancreatic Cancer after Exposure to Insulin Therapy

Glargine

Other Insulin

Both
All Patients with some form of Insulin Therapy

All Patients with Pancreatic Cancer after Exposure to Insulin Therapy

Glargine

Other Insulin

Both

Much smaller Statistical Significance!
Case-Control Study in EHR

8000+ pancreatic cancer patients (cases)
Case-Control Study in EHR

- All Patients with Pancreatic Cancer
  - 8000+ pancreatic cancer patients (cases)

- Matched Controls: Patients without Pancreatic Cancer
  - 25000+ pancreatic cancer patients (controls)
    - Matched by Age, Gender, Fact Counts, Start and End Dates
Case-Control Study in EHR

8000+ pancreatic cancer patients (cases)

Fact Counts
But… Fact distribution can be very different
Case-Control Study in EHR

8000+ pancreatic cancer patients (cases)

Fact Counts
But… Fact distribution can be very different

Case

Control
Case-Control Study in EHR

8000+ pancreatic cancer patients (cases)

Fact Counts
But… Fact distribution can be very different
We used Fact Count/Year.

No Statistical Significance
Summary

• Running statistical analysis is easy
• Ensuring our data is
  – Clean, representative, and suitable for the type of analysis is not
• Lifelines2 is helpful
  – Check if our data is clean
  – See if matching criteria are appropriate
  – Present patient data to colleagues
References and Links

- Lifelines2 in i2b2 (TimeAlign) https://community.i2b2.org/wiki/display/timealign/TimeAlign
Backup Slides
Cohort

• Start with people with Glargine and people with other insulin and follow them longitudinally to see who develops cancer.
• Relative risk.
Case-Control

- Focus only on patients who have pancreatic cancer (case) and select similar patients to compare with (control)
- Odds-ratio
- Better-suited because pancreatic cancer is a relatively rare condition. This way we narrow the scope
Relative Risk

\[
\text{Relative Risk} = \frac{a/(a+b)}{c/(c+d)}
\]

<table>
<thead>
<tr>
<th></th>
<th>Pancreatic Cancer</th>
<th>No Pancreatic Cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glargine</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>No Glargine</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>