Making Better Use of Registry Data in Designing Pragmatic Trials

Basel Biometric Society Novartis Campus, January 13, 2016

Heiner C. Bucher MD, MPH,
Basel Institute for Clinical Epidemiology & Biostatistics
University of Basel, University Hospital Basel Switzerland

heiner.bucher@usb.ch
www.ceb-institute.org
Overview

Innovative approaches for the use of routinely collected data:

• Using registry and cohort data for pragmatic trials

• System-wide resource use studies by privacy preserving anonymous probability-based data linkage of claim data with clinical cohorts
The future challenges of evidence-based medicine

• Evidence will be based on a more diverse family of data sources and methodologies than the conventional (RCT) study type

• Reshape towards comparative effectiveness analysis of head to head comparisons with real world data
  – Improved methods to (re-) analyse RCT and non-RCT studies (marginal structural models)
  – Data linkage of observational data, resource use data, genetic data, biobanks
  – Indirect comparison, MTA
Concept of registry-based randomized clinical trial (RRCT)

- High quality registries contain a large and comprehensive set of variables relevant for prognosis and patient outcome

- Comprehensive coverage of patients
  - SWEDEHEART registry of all hospitalized patients with heart problems (PCI, Valve replacement, ICD, etc.) in Sweden

- Patients admitted to hospital are asked to allow for randomisation rather than physician preference for treatment
Advantages of RRCTs (I)

- A large proportion of less selected patients are available
- Better identification of eligible patients by large scale screening of inclusion and exclusion criteria
- More rapid patient recruitment
- Less costly, relevant data is routinely & prospectively collected
Advantages of RRCTs (II)

- Higher external validity of RRCTs
- More balanced research questions (investigator & industry driven)
- More appropriate benefit / harm assessment due to larger number of and less selected patients
- Collection of better health resource use data for cost-effectiveness analysis using the parallel claim data registries
Disadvantages of RRCTs (I)

- High up front costs for data system development
- Registries may contain large amount of irrelevant data
- Time intensive search strategies for identification of patients and relevant patient parameters
- For drugs or medical devices that require comprehensive safety reporting and strictly defined endpoints the methodology is not different but data collection and monitoring requirements are very high in the context of routinely collected data
Disadvantages of RRCTs (II)

- There are still walls between research and the healthcare setting to enable life-span learning from real world data and shared clinical trial data

- Ethical considerations

- Unresolved issues of:
  - Patient consent
  - Ownership of data
  - Protection of personal data
  - Governance
Randomized trial of a computerized coronary heart disease risk assessment tool in HIV-infected patients receiving combination antiretroviral therapy

Bucher HC *Antivir Ther* 2010;15:3

- **Objective**
  - To investigate whether systematic provisions of CHD risk profiles and evidence-based guidelines to physicians improves in HIV-infected patients
    - Total cholesterol (primary endpoint)
    - Framingham risk scores, systolic & diastolic blood pressure (secondary endpoints)

- **Design**
  - Cluster RCT nested into the Swiss HIV Cohort Study
SHCS CHD risk profiles for charts

10 Years CHD Risk
- 25% >20% >20% >20% >20%

Glucose (mmol/L)
* = fasting state
(+10.0) 6.1 5.5 5.6 6.0 5.3 5.6 (+300)

Smoking (Cig/day)
9.0 1.9 0.4 0.4 0.5 0.7 1.1

Individual Targets
- LDL: < 2.6 mmol/L

CHD Family History: No

Events
PTCA: 08.2005
AMID: 08.2005

Blood pressure
Systolic/Diastolic
Age: 49
Results

165 physicians in SHCS data base taking care of HIV-patients

Randomised

- 80 physicians received updated CHD risk profiles & guidelines
  - 22 physicians with no patient assessments*
    - 1 physician not seeing eligible patients
- 85 physicians received guidelines
  - 23 physicians with no patient assessments*
    - 2 physicians not seeing eligible patients

- 57 physicians included into intervention group
- 60 physicians included into control group
Results

2941 eligible patients
- 903 excluded
  - 16 pregnancies
  - 411 cART naive
  - 298 not on cART
  - 178 incomplete data
  - 16 other reason
- 2094 patients included
  - Patients with intermediate assessments
    - 0: 34
    - 1: 1188
    - ≥2: 288
  - 584 did not attend final assessment
    - 500 failed to attend
      - 31 withdrew from cohort
      - 24 lost to follow up
      - 29 died
  - 1510 patients with final assessments
    - 1468 patients providing primary outcomes
2841 eligible patients
- 921 excluded
  - 10 pregnancies
  - 382 cART naive
  - 330 not on cART
  - 199 incomplete data
- 1955 patients included
  - Patients with intermediate assessments
    - 0: 21
    - 1: 1108
    - ≥2: 332
  - 534 did not attend final assessment
    - 473 failed to attend
      - 17 withdrew from cohort
      - 23 lost to follow up
      - 21 died
  - 1461 patients with final assessments
    - 1413 patients providing primary outcomes
### Baseline characteristics of patients

<table>
<thead>
<tr>
<th></th>
<th>Intervention</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n=2094</td>
<td>n=1995</td>
</tr>
<tr>
<td>Female (%)</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>Current smokers (%)</td>
<td>45</td>
<td>46</td>
</tr>
<tr>
<td>Median systolic blood pressure [IQR]</td>
<td>125 [115-135]</td>
<td>121 [112-133]</td>
</tr>
<tr>
<td>On antihypertensive medication (%)</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>Median total cholesterol [IQR]</td>
<td>4.9 [4.2-5.7]</td>
<td>5.0 [4.3-5.7]</td>
</tr>
<tr>
<td>Diagnosed as diabetic (%)</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Family history of CVD (%)</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Framingham risk ≥ 10% (%)</td>
<td>26</td>
<td>25</td>
</tr>
</tbody>
</table>
Effects of the intervention on primary & secondary endpoints

<table>
<thead>
<tr>
<th></th>
<th>Unweighted analysis</th>
<th></th>
<th>Weighted analysis</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n (%)</td>
<td>Estimate</td>
<td>95% CI</td>
<td>Estimate</td>
</tr>
<tr>
<td><strong>Primary outcome</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total cholesterol mmol/l</td>
<td>2881 (70)</td>
<td>-0.02</td>
<td>-0.10 to 0.06</td>
<td>-0.03</td>
</tr>
<tr>
<td><strong>Secondary outcomes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Systolic blood pressure mm Hg</td>
<td>2935 (72)</td>
<td>-0.5</td>
<td>-1.8 to 0.8</td>
<td>-0.4</td>
</tr>
<tr>
<td>Diastolic blood pressure mm Hg</td>
<td>2935 (72)</td>
<td>-0.7</td>
<td>-1.8 to 0.5</td>
<td>-0.4</td>
</tr>
<tr>
<td>Framingham risk score %</td>
<td>2829 (69)</td>
<td>-0.1</td>
<td>-0.3 to 0.1</td>
<td>-0.1</td>
</tr>
</tbody>
</table>
Change in drug management and CV events in patients with VL < 50 copies/ml at baseline

<table>
<thead>
<tr>
<th></th>
<th>Framingham risk &lt; 10% at baseline</th>
<th>Framingham risk &gt; 10% at baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intervention n = 772</td>
<td>Control n = 675</td>
</tr>
<tr>
<td>Started new cART component (%)</td>
<td>33</td>
<td>39</td>
</tr>
<tr>
<td>Started abacavir (%)</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Started atazanavir (%)</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Started any drug that reduces CV risk (%)</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>Stopped any PI (%)</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>Experienced CV event (%)</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Benchmark trial to lower antibiotic prescription nationwide in primary care in Switzerland

• Population
  – 2400 Board certified primary care physicians in Switzerland with high antibiotic prescription rates (above 75th percentile of antibiotic prescriptions)

• Intervention
  – Quarterly feedback on prescription rates (web-based, email reminder and letters) for 24 months
  – Evidence-based guidelines on the use of antibiotics in primary care

• Control
  – No intervention (not informed about trial)
Outline Benchmark Trial

• Outcome
  – Primary: antibiotic prescription rate

• Data
  – Nationwide reimbursement data of health insurers (Tarifpool, SASIS Santésuisse)
Selection of physicians and intervention
Benchmark Trial

Primary care physicians with ZSR number in Santéuisse TarifPool

Physicians with AB prescription rate in upper 75th percentile

Random sample of physicians

Intervention (50%)
Control group (50%)

Physicians with at least one access to the web-interface during the first 12 months

Selected sample of interested physicians
AB prescriptions per condition

Secondary outcome: Validation of EACS criteria after 24 months

Primary outcome: AB prescriptions after 12 months

Primary outcome: AB prescriptions after 24 months
Prototype Web-Application Benchmark Trial

ANTIBIOTIKA VERSCHREIBUNG IN DER GRUNDVERSORGUNG

MONATLICHE VERSCHREIBUNGSRATE PRO 100 KONSULTATIONEN

JUNI 2011 (TOTAL 31.89)

MAKROLIDE 16.32
ANDERE BETALAKTAME 23.91
TRIMETHOPRIM & SULFOMETHASOLE 23.13
AMOXICILLINE 24.53
DOXYCYCLINE 23.12
FLUOROQUINOLONE 21.25

VERTEILUNG DER MONATS-RATEN IN DER VERGLEICHSSREGION

Klicken Sie auf einen Monat oder benutzen Sie die Pflegetafeln zur Anzeige von Details in den unterstellenden Feldern.
Wählen Sie eine Antibioticagruppe um Details zu dieser Gruppe zu sehen.
Opportunities
Benchmark Trial

• Generalisability
  – High external validity through inclusion of primary care physicians

• Potential impact
  – Directed at high prescribers nationwide

• Novelty
  – Little evidence whether benchmarking, monitoring and guideline provision reduces physicians’ prescription behaviour
  – Use of routinely collected health care data for an intervention trial at population scale in Switzerland
Privacy preserving probability-based data linkage of claim data with prospectively collected cohort data

- An example from
- The Swiss HIV Cohort Study
- Helsana, largest health insurer in Switzerland covering 20% of the Swiss population
Data structure of the Swiss HIV Cohort Study

Claim data resource use costs

National cohort death registries

Cancer registry

Clinical variables
Demographic data
Provider data

Laboratory CD4 cells, HIV VL
Chemistry, Hematology

Repository

Drug resistance database

Genetics HGS

Comorbidity studies
CCT, Dexa, Neurocognitive assessment

Behavior variables
Psychosocial variables

Privacy preserving anonymous linkage
Costs of HIV and non-HIV related comorbidity

• Late presentation (CD4 cells < 350 cells/µL) is the most important reason for HIV related morbidity

• HIV infected individuals are at higher risk of CVD, end stage renal disease, liver related comorbidity (HCV, HBV) and cancer than non-HIV infected individuals

• Resource use and costs of late presentation and HIV and non-HIV related comorbidity are not well known

• Information on non ART drug use in the SHCS is limited
Goals of the pilot study

• Evaluate the feasibility and the validity of privacy preserving anonymous matching in the SHCS for claim data
• Collection of resource use data for future cost-effectiveness analyses
• Evaluate possibility for pharmacoepidemiological studies
• Evaluate whether pilot can be extended to
  – To include more health insurers
  – Other cohorts (Swiss Transplant Cohort Study)

• Evaluate whether routine annually mergers can be established
Set-up of a pilot study for a merger of prospective clinical and cost claim data: Swiss HIV Cohort Study

- Individuals in SHCS with follow-up by December 31st 2013 (n=8888)
  Data set with no date of birth
- Individuals in SHCS on ART in 2012-2013 (n=8607) plus added encrypted date of birth
  Anonymised SHCS data subset with new patient ID 'A'
- Insured population of Helsana (n=1.5 Mio. 15% of Swiss population)
- Individuals with any ART in 2012-2013 with encrypted date of birth (n=2660)
  Anonymised data subset with new ID 'B'

SHCS Data centre (ISP/IM Lausanne independent data manager)
Probability matching based on encrypted date of birth, gender, ART drug
Generation of a new anonymised data set: Replace encrypted date of birth with year of birth. Generate new patient ID 'C' by random generator

Study data set with new unique patient ID 'C' given to SHCS investigators for present study

ART: Antiretroviral therapy
SHCS: Swiss HIV Cohort Study

Destroy data set after job by SHCS data centre
Privacy preserving probabilistic record linkage

Data masking, encrypting and probabilistic linkage

Conclusions:

• RRCTs are an interesting option for head to head comparisons in settings with registry data of high quality exist
  – May reduce cost
  – Facilitate rapid recruitment

• Privacy preserving anonymised matching may allow to considerable enrich observational data research
  – Matching is resource intense
  – Possibility of routine linkage
  – Interesting possibilities for monitoring, health economic studies
Thanks to:

**Collaborators:**
Jim Young, Juliane Schäfer, Ramon Saccilotto, Lars Hemkens, Selene Leon Reyes, Dominik Glinz

**Organisations:**
Swiss HIV Cohort Study
Helsana

**Funders:**
Stiftung Institut für klinische Epidemiologie
Swiss National Science Foundation, Gottfried und Julia Bangerter-Rhyner Stiftung