A practical guide to adding patient heterogeneity into Phase III trials

Case study in schizophrenia

Hélène Karcher, PhD
LASER Analytica – WP2 GetReal consortium

BBS Seminar: Disease and Product Registries
Jan 13th, 2016 - Basel
Context

> Learning about a new drug’s effectiveness is essential to appraising its benefit/risk in each real-life care setting/country

> Double challenge

  • Effectiveness is rarely measured pre-authorization
    > Clinical development focuses on uncovering efficacy
    > Pragmatic trials are usually not implemented early

  • Effectiveness is not universal
    > Effectiveness is specific to a care setting / country!
Two ways to improve learning about effectiveness early in clinical development

1. More pragmatic design, i.e., any aspect of study design: population, type of randomization, blinding, monitoring, etc.

2. Better “analyses tools”, i.e., any aspect of data analyses: statistical or model-based analyses, predictive models, etc.

Clinical development trial
Factors that drive effectiveness

**Drug use factors**
- Patterns of use, dose, treatment duration.
- Can be defined with:
  1. Adherence of prescribers to label recommendations,
  2. Adherence of patients to prescriptions
  - Past history of exposure
  - Variability of diagnostic

**Health system care delivery factors**
- Type of setting for care delivery (e.g., hospital, home)
- Type of prescriber: general physician, specialist, oncologist, nurse practitioner...
- Socio-economic situation of health system, prescribers and patients

**Patient population factors**
- Patient physical and behavioral characteristics: age, gender, weight, ethnicity, smoking/eating/exercise habits, etc.
- Co-morbidities
- Disease stage/severity
- Co-prescriptions
- Other baseline risk factors and genetics relevant to the disease/drug

**Interaction**

**Efficacy**

**Effectiveness**
Systematic Review of methods to incorporate pragmatism pre-authorization: results*

1. Many (39) methodological papers were identified that recommend how to relax trial features to make them more pragmatic, and to adapt analyses

2. However, this does not translate into many actual Phase 2-3 trials with pragmatic elements – due to scientific and operational hurdles
   • Systematic review only identified 18 pre-authorization trials with pragmatic elements
   • Typically only 1-2 selected features are pragmatic
     > Features required to conduct the trial for authorization
     > Features that could demonstrate a benefit not present in an RCT setting

Hurdles to incorporating effectiveness before authorization* (review of 39 articles)

1. Known and unknown confounders in real-world trials may lead to inconclusive effect sizes\textsuperscript{18,25}
2. Extensive cost of running such trials due to larger sample size required\textsuperscript{14}
3. Operational difficulties in recruiting certain populations, and in minimising measurements/study visits\textsuperscript{30,31}
4. Uncertainty in reactions from regulatory bodies\textsuperscript{30,32}

Factors that drive effectiveness

**Health system care delivery factors**

- Type of setting for care delivery (e.g., hospital, home)
- Type of prescriber: general physician, specialist, oncologist, nurse practitioner...
- Socio-economic situation of health system, prescribers and patients

**Drug use factors**

- Patterns of use, dose, treatment duration.
Can be defined with:
1. Adherence of prescribers to label recommendations,
2. Adherence of patients to prescriptions
- Past history of exposure
- Variability of diagnostic

**Patient population factors**

- Patient physical and behavioral characteristics: age, gender, weight, ethnicity, smoking/eating/exercise habits, etc.
- Co-morbidities
- Disease stage/severity
- Co-prescriptions
- Other baseline risk factors and genetics relevant to the disease/drug

**Interaction**

**EFFICACY**

**EFFECTIVENESS**
Case study: Test broadening of eligibility criteria in schizophrenia pre-authorization RCTs
Simulation study to test eligibility criteria in schizophrenia

> Objective

• Explore how to mitigate strict eligibility criteria in Phase 3 with real-life population heterogeneity

> Method*: use real-world data to optimize clinical trials

1. Study patient characteristics and interplay between factors and outcome in a real-life schizophrenia population (SOHO)
2. Define the subpopulation eligible for a typical pre-authorization trial “reference RCT”
3. Re-include in this “reference RCT” a minimal subset of patients who would usually be excluded (=broaden the eligibility criteria)
   > Method of quotas (stratification) for patient inclusion in trials
   > Combined with predictive modeling of the outcome in the RW population
4. Evaluate how “efficient” each re-inclusion is

* “Reverse” of the method used in Schneeweiss et al. Increasing Levels of Restriction in Pharmacoepidemiologic Database Studies of Elderly and Comparison With Randomized Trial Results. Med Care. 2007
Data source : SOHO study

> A prospective, observational study of 10,218 schizophrenia patients
  • from 10 European countries
  • followed over 3 years
  • who received antipsychotic treatment

Outcome: CGI-S score
  • Clinical Global Impression-Severity
  • Assesses severity of the patient’s mental illness at time of rating with one question
  • 7-point scale: from 1 (not at all ill) to 7 (extremely ill)
  • We used mean ΔCGI-S at 3 months (change from baseline) as outcome.
Create a synthetic reference RCT within SOHO

Out of 10,218 SOHO patients, 2,132 patients were selected to define a “synthetic reference RCT” with the following eligibility criteria (taken from a meta-analysis of 212 trials*):

Eligibility criteria applied to create a reference RCT
- Age between 18 and 65 years old
- Duration of illness superior to 3 years
- BMI between 17 and 40
- No history of alcohol or drug abuse
- Patient with compliance to prescribed antipsychotic therapy
- Patient without suicide attempt in the past 6 months
- Patient included in public or combined practices

Then, within the synthetic reference RCT, restricted sets of patients who initiated a specific drug at baseline were obtained for our study.

* Leucht et al. Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis. Lancet. 2013
Create a synthetic reference RCT within SOHO

- Variability in patient characteristics
  - Efficacy (reference RCT)
    - All typical RCT eligibility criteria applied
  - Predict and compare with
  - Real-life effect (full SOHO dataset)

SOHO dataset
Population differences in RCT and observational data

Different eligibility criteria are excluding different proportions of the RW population from RCTs.
Drug effects in synthetic reference RCT vs SOHO

<table>
<thead>
<tr>
<th>Patients taking drug:</th>
<th>Mean $\Delta$CGI-S at 3 months (change from baseline)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Synthetic reference RCT</td>
</tr>
<tr>
<td>Drug R</td>
<td>n=1127 -0.78 (SD = 0.95)</td>
</tr>
<tr>
<td>Drug AE</td>
<td>n=498 -0.66 (SD = 0.89)</td>
</tr>
<tr>
<td>Drug AD</td>
<td>n=199 -0.54 (SD = 0.99)</td>
</tr>
<tr>
<td>Drug H</td>
<td>n=118 -0.77 (SD = 0.84)</td>
</tr>
</tbody>
</table>

> Real-life effect is slightly better than effect in RCT under all 4 drugs
  - Excluded patients respond better to drug (trend)
> Cannot (yet) compare between drugs since patients under different drugs may intersect and may have uncomparable characteristics

Note: “R”, “AE”, “AD”, “H” are the most popular drugs in the SOHO study.
Outcomes comparison in RCT and observational data (patients under drug “R”)

The more negative ΔCGI-S at 3 months, the better patient responding to drug, the more influential the exclusion criterion

Different eligibility criteria are excluding RW subpopulations who have different outcomes than RCT populations.
No all exclusion criteria impact effectiveness with the same magnitude (patients under drug “R”).

Comparison of synthetic RCTs & SOHO observational data (10,000 patients).

Subpopulation size:

SOHO patients excluded to define RCT (as % of total SOHO patients):

- Suicide attempts
- Alcohol abuse
- Drug abuse
- Practice type
- Chronicity
- Age
- Patient compliance
- BMI

Effects in subpopulation:

The mean ΔCGI-S at 3 months in synthetic reference RCT:

ΔCGI-S at 3 months in the excluded patients:

- Suicide attempts
- Alcohol abuse
- Drug abuse
- Practice type
- Chronicity
- Age
- Patient compliance
- BMI
Predicting drug effects using a single drug group

- Variability in patient characteristics
- Efficacy (reference RCT)
- Predict and compare with
- Real-life effect (full SOHO dataset)
- SOHO dataset
Enriching RCTs to improve predictions

"enriched" RCTs

- efficacy
- (some RCT patients)

+5% excluded patients

Variability in patient characteristics

predict and compare with

real-life effect

(full SOHO dataset)

age > 65

chronicity between 1 and 3

alcohol abuse...

SOHO dataset
Enriching RCTs to improve predictions

- "enriched" RCTs
  - efficacy
    - (some RCT patients)
  - variability in patient characteristics
    - +10% excluded patients

- real-life effect
  - predict and compare with
  - (full SOHO dataset)
    - age > 65
    - chronicity between 1 and 3
    - alcohol abuse
    - ...

- SOHO dataset
The following linear regression model was used:

ΔCGI-S at 3 months ~
(age + chronicity + gender + BMI + hospitalization +
number of admissions in hospital + depression score + QOL score + patient compliance + country + work status +
housing condition + social activities + relationship +
negative symptom at baseline + positive symptom at baseline + cognitive symptom at baseline + dosage DDDeq)

I{if initiated the drug at baseline}

> The above covariates have been chosen through a Chi-square test for independence.
> The model was fitted in synthetic reference RCT and enriched RCTs, then used to predict the real-life drug effect in SOHO.
Evaluation of prediction accuracy

- The accuracy of prediction has been measured by Mean squared error (MSE)

\[
\text{MSE} = \text{mean}(\text{predicted } \Delta \text{CGI-S at 3 months} - \text{real-life observed } \Delta \text{CGI-S at 3 months })^2
\]

- As each enrichment requires \textit{random} replacement of patients, 100 independent repetitions were performed. They provided the mean squared distance between the prediction and real-life observation (MSE), the standard deviation and the derived confidence interval (CI).

- Several enrichment factors which were also exclusion criteria have been studied: suicide attempts, duration of illness (chronicity), practice type, alcohol abuse, drug abuse and age.
Distribution of number of suicide attempts in reference RCT, SOHO and 2 enriched RCTs under drug “R”

**Reference RCT 0%**
- Number of patients: 1127
- 0: NA
- 1-5: 0
- 6-10: 0
- 11+: 0

**SOHO 100%**
- Number of patients: 3362
- 1-5: 16
- 6-10: 16
- 11+: 0

**Enriched RCT at 10%**
- Number of patients: 1015
- 10% “excluded” patients: 112
- 1-5: 15
- 6-10: 0
- 11+: 0

**Enriched RCT at 20%**
- Number of patients: 902
- 20% “excluded” patients: 225
- 1-5: 15
- 6-10: 0
- 11+: 0
Predicted error using different RCTs enriched with few “suicide attempts”

Enriched RCTs -> better prediction

Reference RCT -> less robust prediction

Real-life % of between 1 to 5 suicide attempts

Mean squared error (predicted Vs. observed in SOHO)
Distribution of duration of illness in reference RCT, SOHO and 2 enriched RCTs under drug “R”

**Reference RCT 0%**

- Number of patients: 0 (NA), 0 (0-1), 0 (1-3), 646 (3-15), 295 (15-25), 147 (35-45), 36 (55-65), 3 (75-85), 0 (90+)

**SOHO 100%**

- Number of patients: 428 (0-1), 1136 (1-3), 1876 (1-3), 931 (1-3), 431 (1-3), 155 (1-3), 32 (1-3), 6 (1-3)

**Enriched RCT 10%**

- Number of patients: 112 (0-1), 266 (1-3), 132 (1-3), 32 (1-3), 3 (1-3), 0 (1-3)

**Enriched RCT 15%**

- Number of patients: 169 (0-1), 251 (1-3), 125 (1-3), 31 (1-3), 3 (1-3), 0 (1-3)

**Target patients**

- Reference RCT: 0%
- SOHO: 100%
- Enriched RCT 10%: 15%
- Enriched RCT 15%: 24%
Predicted accuracy using different RCTs enriched with shorter “duration of illness”

- Reference RCT -> less robust prediction
- Enriched RCTs -> better prediction

Real-life % of chronicity between 1 and 3

enrichment of reference RCT in patients with chronicity between 1 and 3 (%)
Comparison of enrichment factors (patients under drug “R”)

<table>
<thead>
<tr>
<th>Enrichment factors</th>
<th>Excluded patient size</th>
<th>Mean ΔCGI-S at 3 months</th>
<th>Optimal enrichment percentage (real-life %)</th>
<th>Mean squared error of prediction</th>
<th>Actual coverage (expected coverage 0.95)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suicide attempts between 1 and 5</td>
<td>1323</td>
<td>-0.857</td>
<td>25%</td>
<td>0.807</td>
<td>0.877</td>
</tr>
<tr>
<td>Chronicity between 1 and 3</td>
<td>596</td>
<td>-1.023</td>
<td>15%</td>
<td>0.814</td>
<td>0.875</td>
</tr>
<tr>
<td>Private practice</td>
<td>614</td>
<td><strong>-1.040</strong></td>
<td>15%</td>
<td>0.815</td>
<td>0.873</td>
</tr>
<tr>
<td>Alcohol abuse</td>
<td>693</td>
<td>-0.787</td>
<td>15%</td>
<td>0.817</td>
<td>0.874</td>
</tr>
<tr>
<td>Age &gt; 65</td>
<td>290</td>
<td>-0.805</td>
<td>5%</td>
<td>0.828</td>
<td>0.872</td>
</tr>
<tr>
<td>Synthetic reference RCT</td>
<td>1127</td>
<td>-0.778</td>
<td>/</td>
<td>0.851</td>
<td>0.868</td>
</tr>
</tbody>
</table>
Validation of reference RCT from literature

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CGI-S at baseline</td>
<td>4.39±1.02</td>
<td>4.0±1.0</td>
<td>5.10±1.01</td>
<td>4.8±0.8</td>
<td>4.4±0.7</td>
<td>4.6±0.9</td>
<td>4.6±0.8</td>
</tr>
<tr>
<td>ΔCGI-S at 3 months</td>
<td>-0.78±0.95</td>
<td>-0.4</td>
<td>-1.52</td>
<td>-2.0</td>
<td>36.6% patients improved</td>
<td>-0.7±1.2</td>
<td>-1.6±0.9</td>
</tr>
</tbody>
</table>
Conclusion – methods

> We used a disease registry to guide addition of patient heterogeneity to standard Phase 3 trials in schizophrenia.

> The impact of the following trial design changes was assessed:
  > • Relax a few, selected exclusion criteria in a controlled way
  > • Quantify the gain in effectiveness prediction
  > • Keep sample size and measure improvement in outcome
Conclusion – results

> The best choice of enrichment factor to predict real-life effects was found to be driven by:

  • **Size of the excluded real-life population.** Excluding “number of past suicide attempts > 1” left out the greatest schizophrenia population from Phase 3 trials.

  • **Change in outcome in patients with this factor.** Patients with a practice type “private” and disease chronicity between 1 and 3 years had the most different outcome from typical Phase 3 patients.

> **Enriching typical Phase 3 with selected factors** improved the representability of real-life and as a result, it improved predictions of the real-life effects of the investigated drug.
Next steps

• Test how the variability of the effect size is modified through enrichment
  – Deduce the % successful Phase III after enrichment
• Build different types of prediction models from reference RCT to SOHO for both horizontal and longitudinal predictions.
• Combine different enrichment factors to generalize the analysis and accelerate collection of patients of interest
Acknowledgments

> LASER
  • Shuai Fu, Clementine Nordon, Lucien Abenhaim

> University of Ioannina
  • Orestis Efthimiou

> Harvard Medical School
  • Sebastian Schneeweiss

> Eli Lilly
  • Mark Belger, Iris Goetz

> Members of the GetReal WP2
Questions?

Hélène Karcher, PhD
LASER Analytica

Helene.Karcher@la-ser.com