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

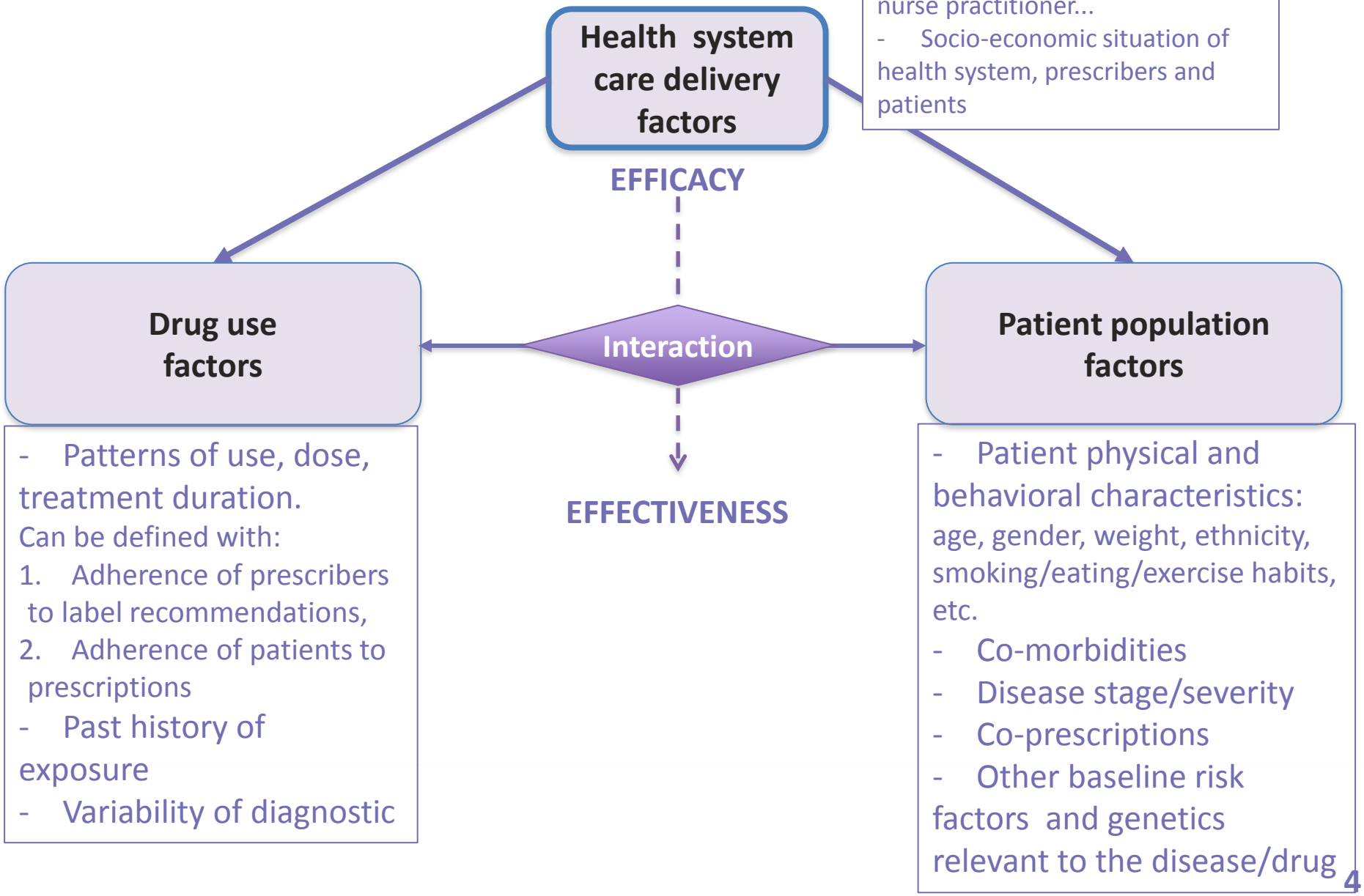
- Learning about a new drug's effectiveness is essential to appraising its benefit/risk in each real-life care setting/country
- Double challenge
  - Effectiveness is rarely measured pre-authorization
    - Clinical development focuses on uncovering efficacy
    - Pragmatic trials are usually not implemented 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.

Factors that drive 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

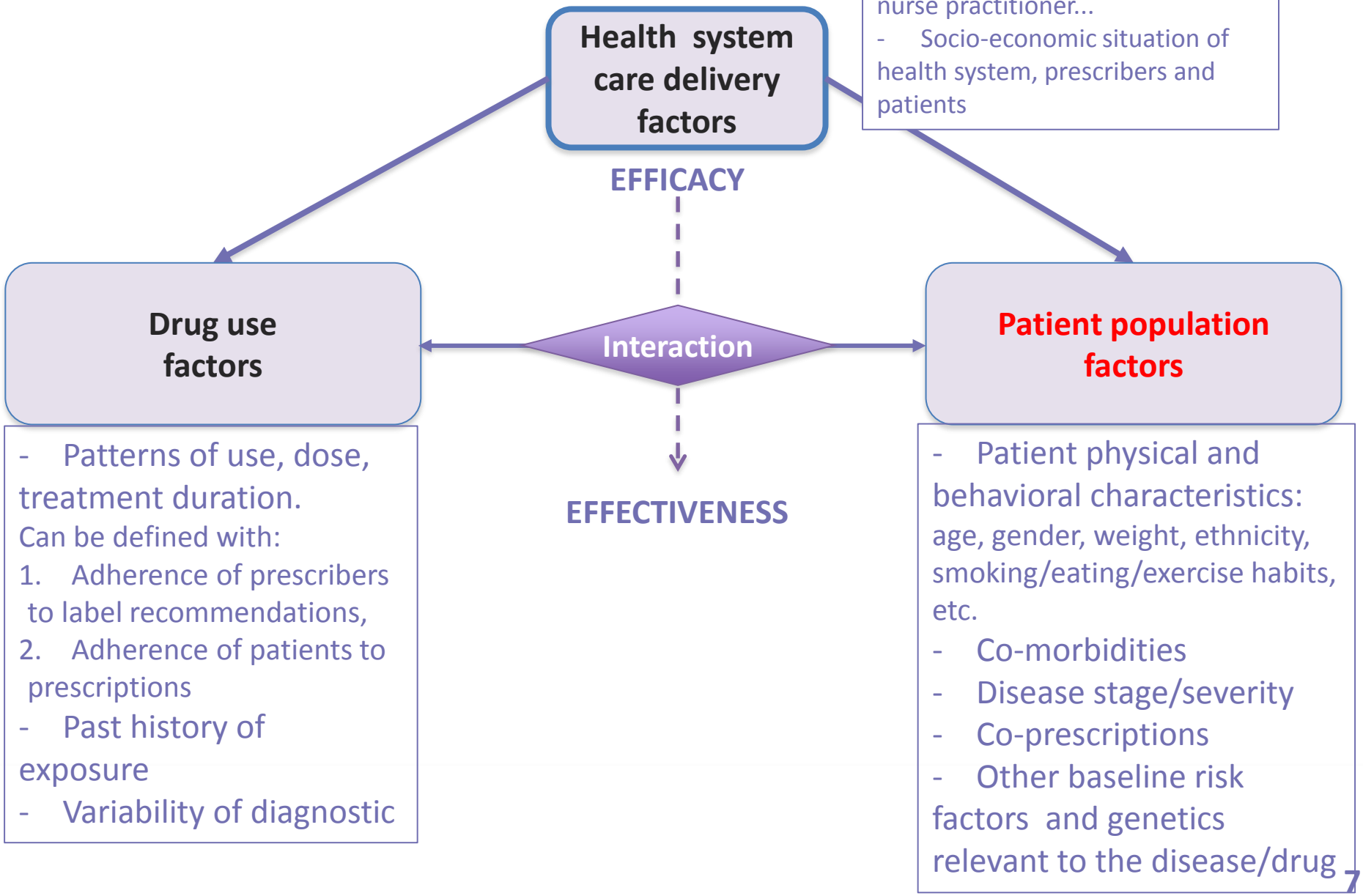
* Karcher, Nordon, Neumann, Nikodem, Zyla, Chevrou-Severac, Jimenez, Bala, Abenhaim. Methods to Evaluate Real-World Effectiveness in Pre-Authorization Trials SLR. HTAi 2015.

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[18,25]
2. Extensive cost of running such trials due to larger sample size required[14]
3. Operational difficulties in recruiting certain populations, and in minimising measurements/study visits[30,31]
4. Uncertainty in reactions from regulatory bodies[30,32]

* Karcher, Nordon, Neumann, Nikodem, Zyla, Chevrou-Séverac, Jimenez, Bala, Abenhaim. Methods to Evaluate Real-World Effectiveness in Pre-Authorization Trials SLR. HTAi 2015.

Factors that drive 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*):

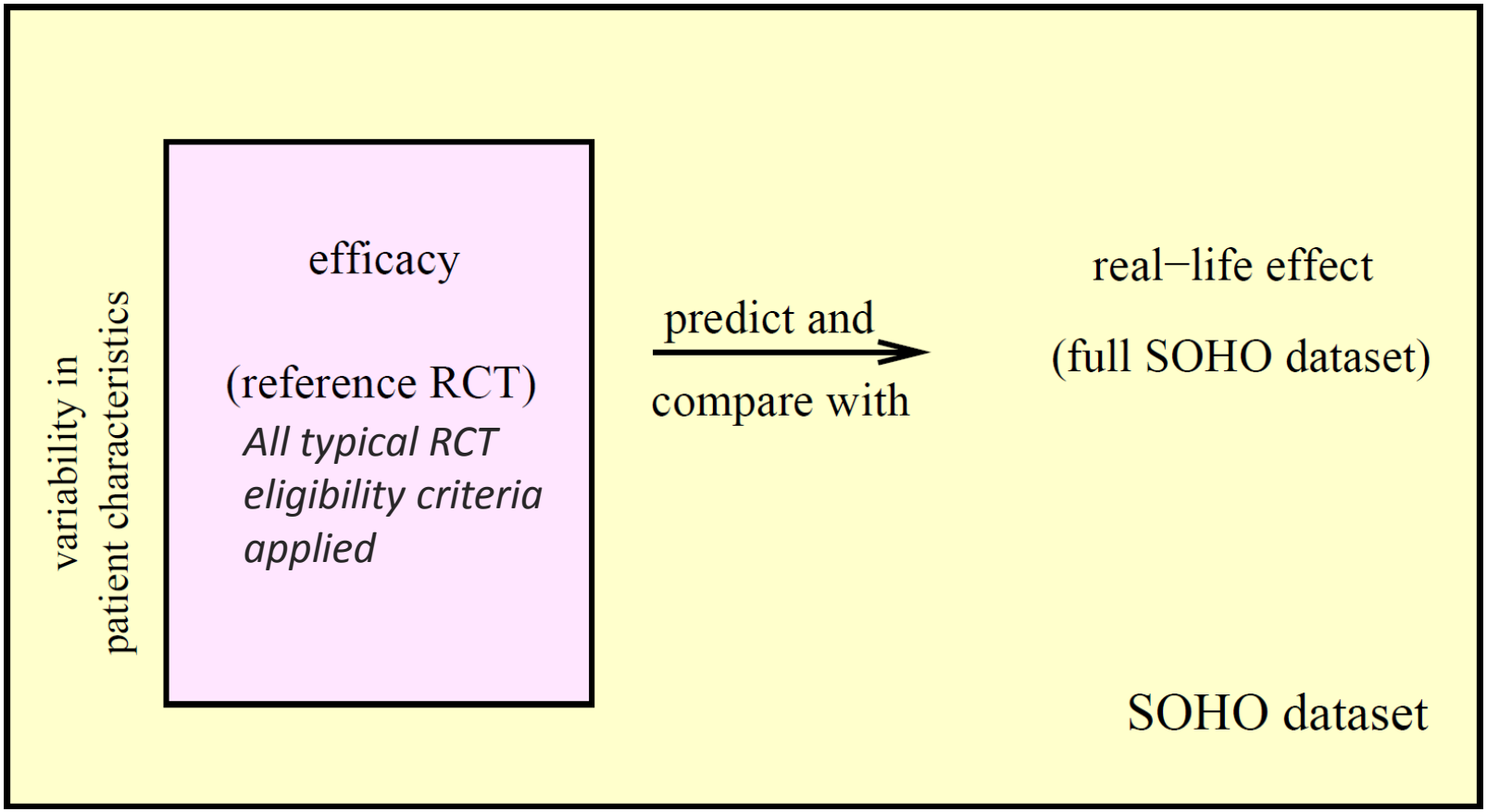
Eligibility criteria applied to create a reference RCT

- 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

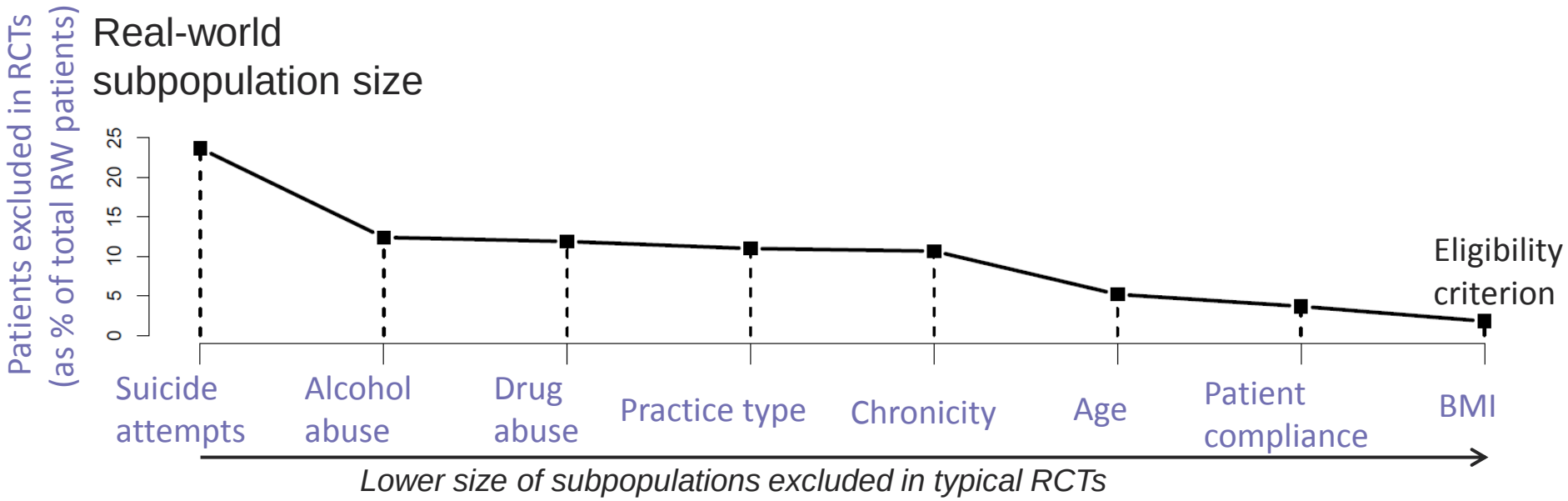
Then, within the synthetic reference RCT, restricted sets of patients who initiated a specific drug at baseline were obtained for our study.

* 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



Population differences in RCT and observational data



Different eligibility criteria are excluding different proportions of the RW population from RCTs.

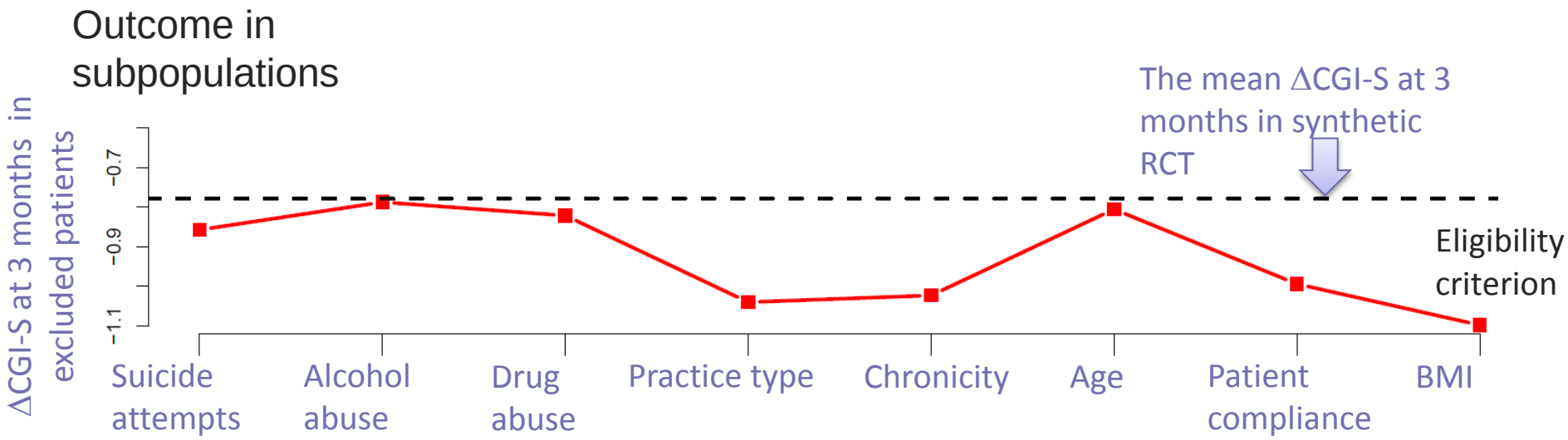
Drug effects in synthetic reference RCT vs SOHO

Patients taking drug:	Mean ΔCGI-S at 3 months (change from baseline)			
	Synthetic reference RCT		SOHO ("real-life")	
Drug R	n=1127	-0.78 (SD = 0.95)	n=5591	**-0.88** (SD = 0.82)
Drug AE	n=498	-0.66 (SD = 0.89)	n=2188	**-0.71** (SD = 0.94)
Drug AD	n=199	-0.54 (SD = 0.99)	n=847	**-0.60** (SD = 0.99)
Drug H	n=118	-0.77 (SD = 0.84)	n=456	**-0.87** (SD = 0.97)

- 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”)

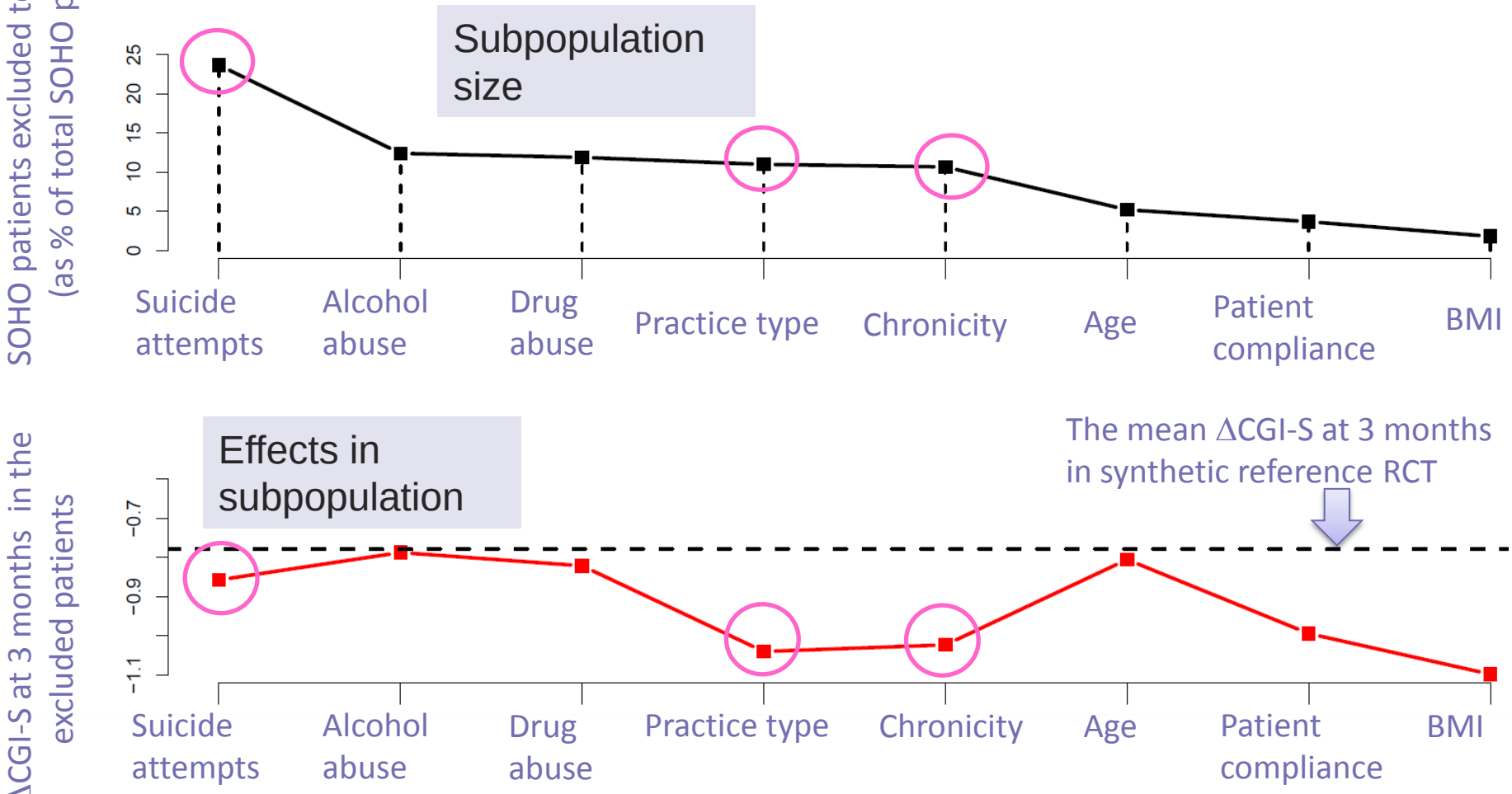


The more negative ΔCGI-S at 3 months, the better patient responding to drug, the more influential the exclusion criterion

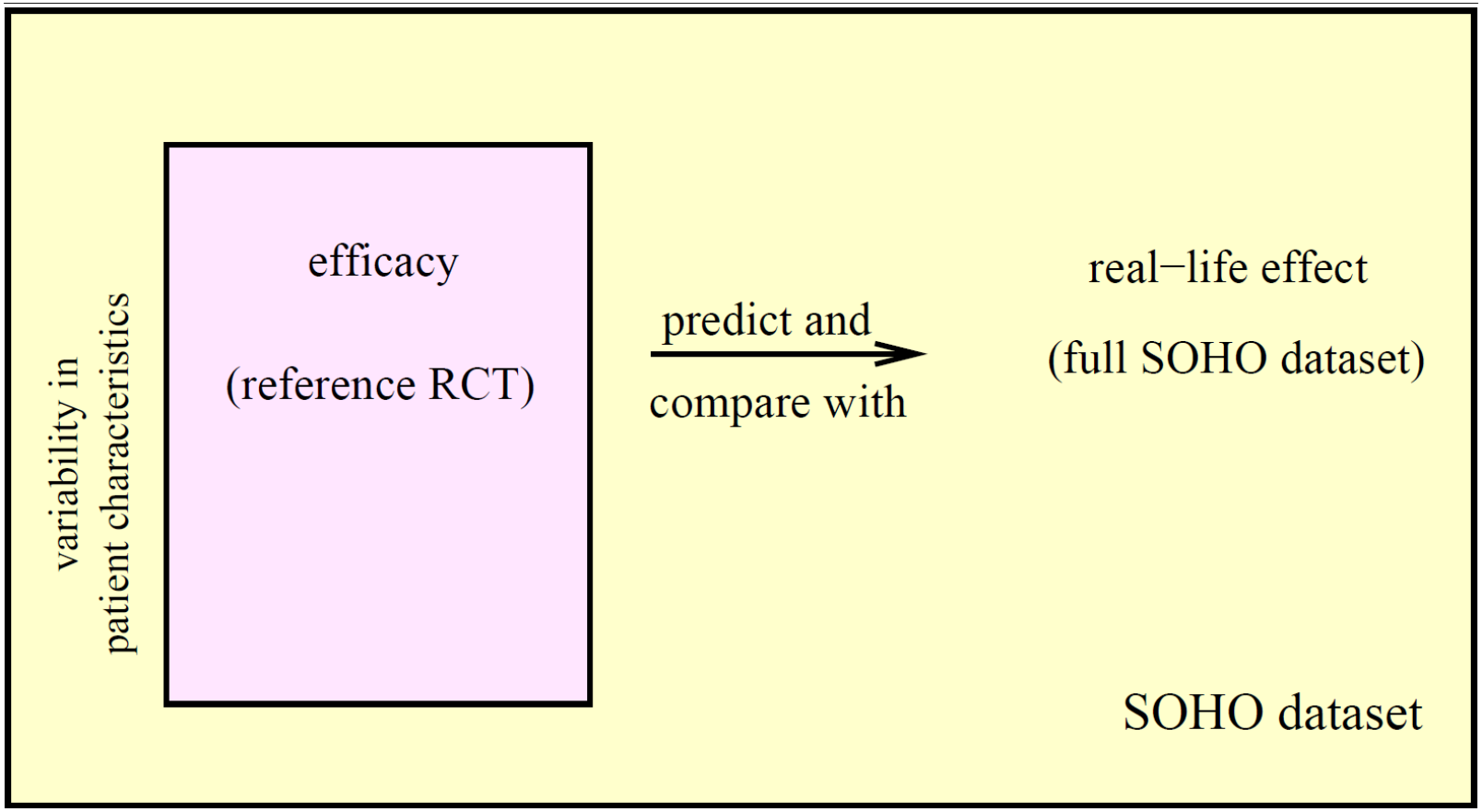
Different eligibility criteria are excluding RW subpopulations who have different outcomes than RCT populations.

No all exlusion criteria impact effectiveness with the same magnitude (patients under drug “R”)

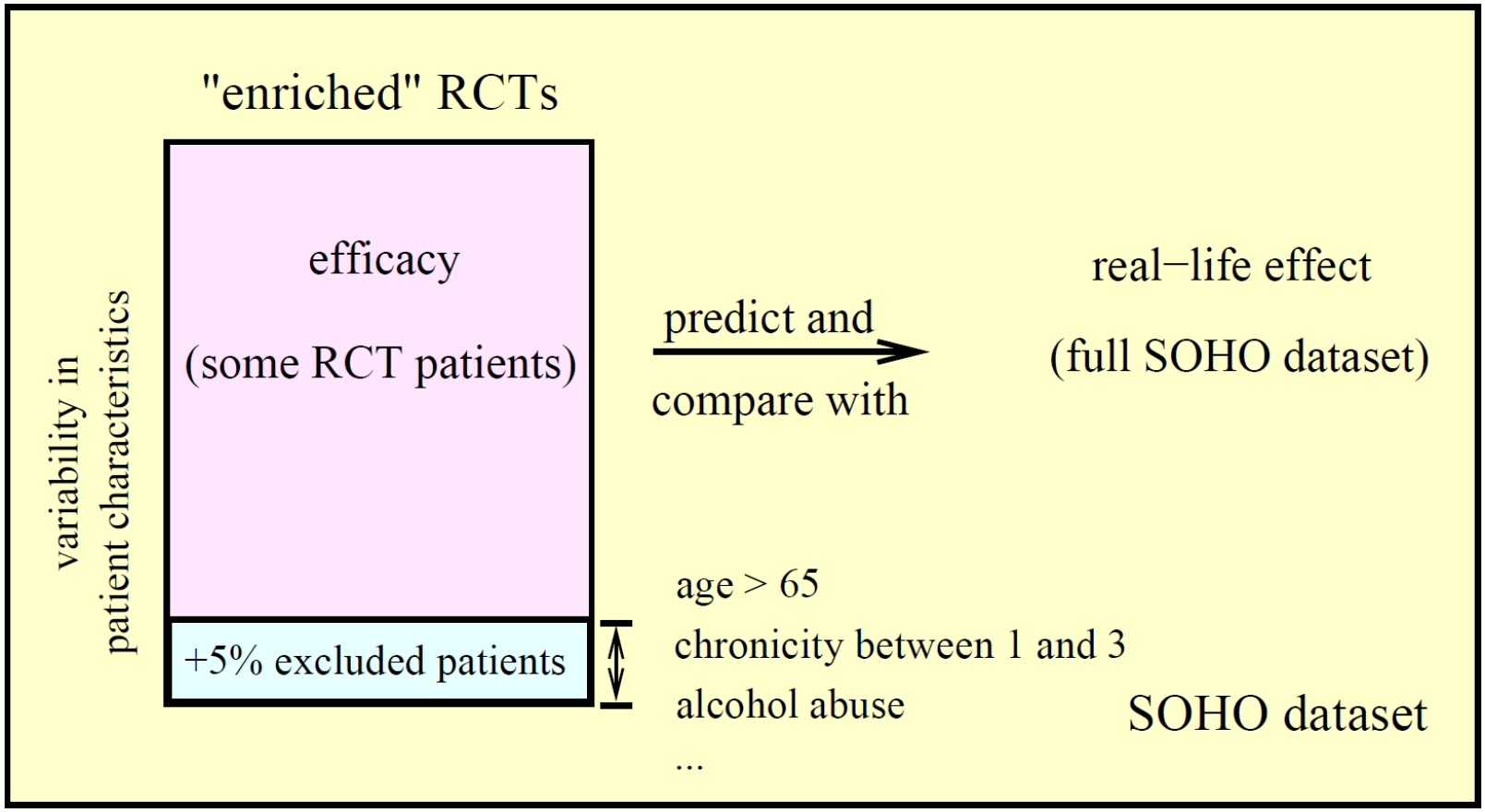
comparison of synthetic RCTs & SOHO observational data (10,000 patients)



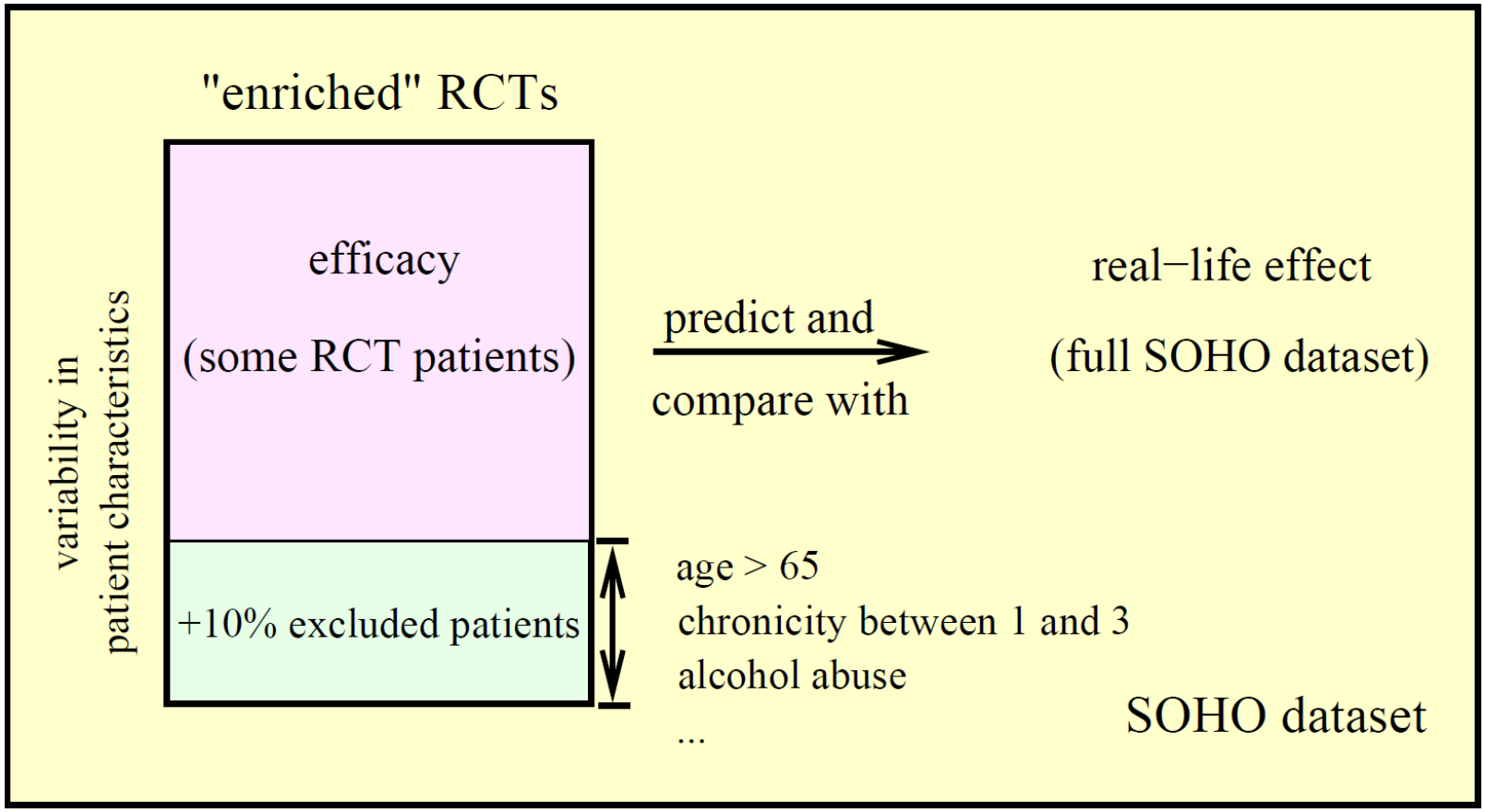
Predicting drug effects using a single drug group



Enriching RCTs to improve predictions



Enriching RCTs to improve predictions



Prediction of real-life effect in SOHO using only data from synthetic reference RCT or enriched RCTs

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.

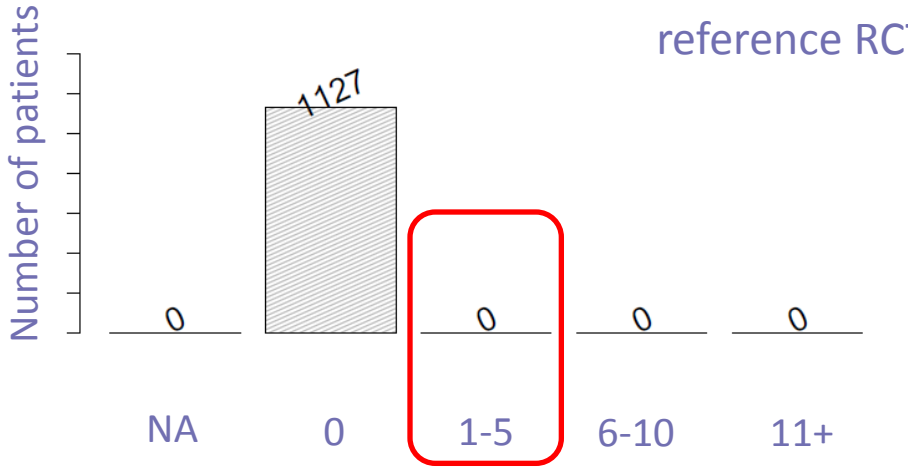
Evaluation of prediction accuracy

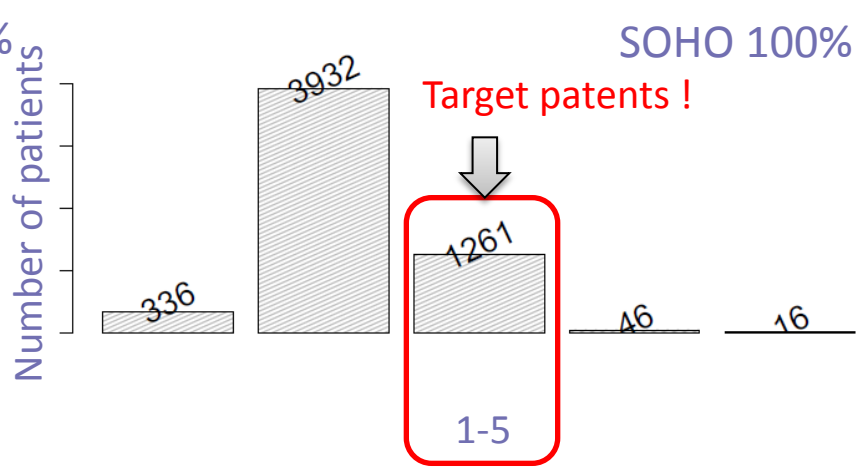
- The accuracy of prediction has been measured by Mean squared error (MSE)

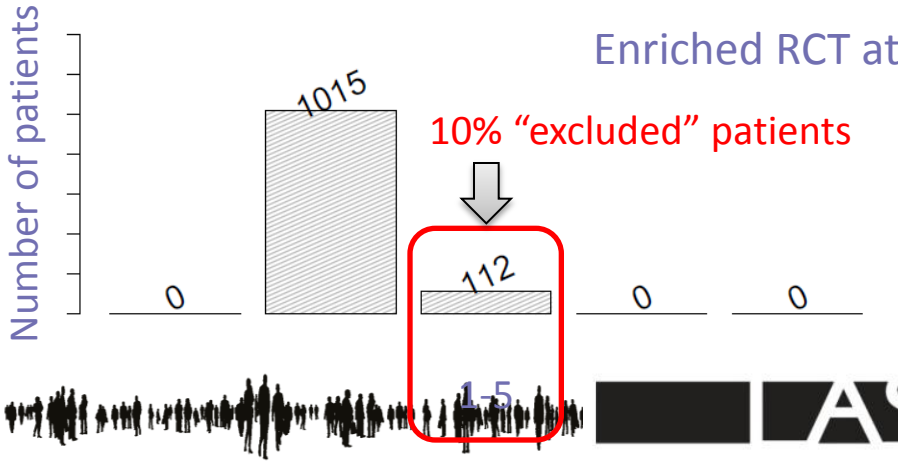
MSE = mean((predicted ΔCGI-S at 3 months – real-life observed ΔCGI-S at 3 months)^2)

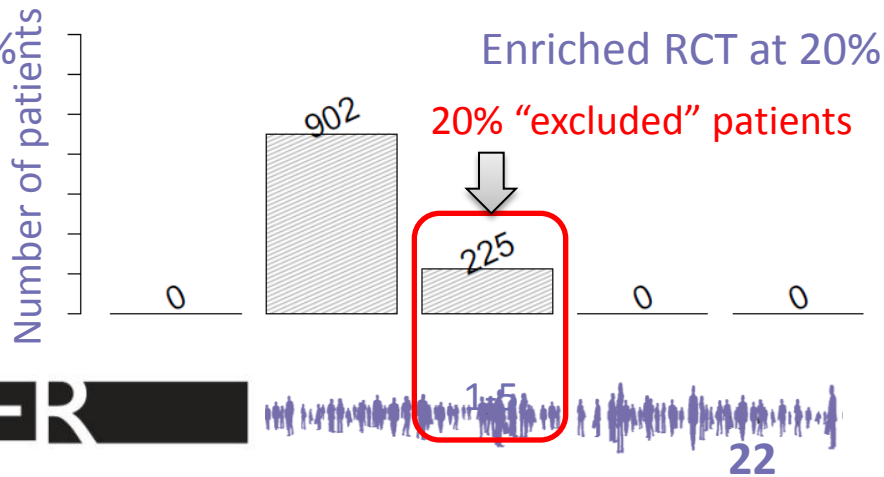
- As each enrichment requires *random* replacement of patients, 100 independent repetitions were performed. They provided the mean squared distance between the prediction and real-life observation (MSE), the standard deviation and the derived confidence interval (CI).
- Several enrichment factors which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

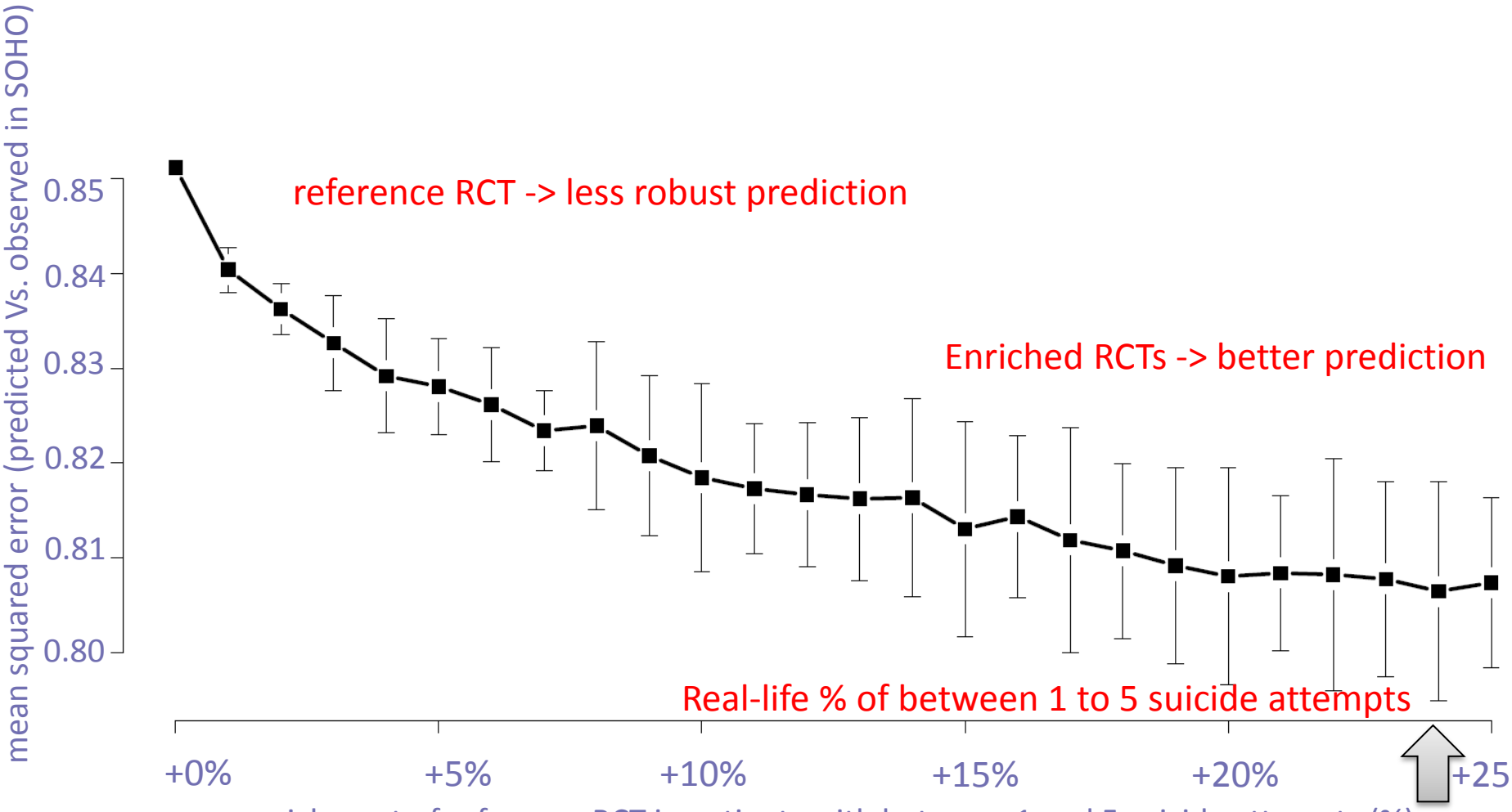




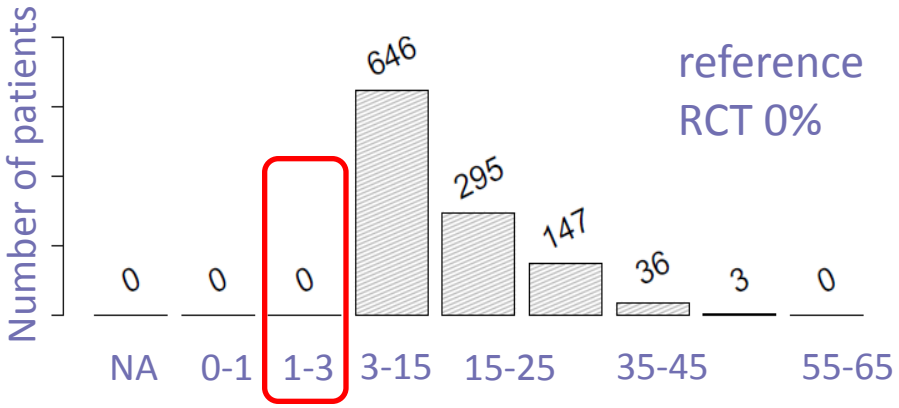


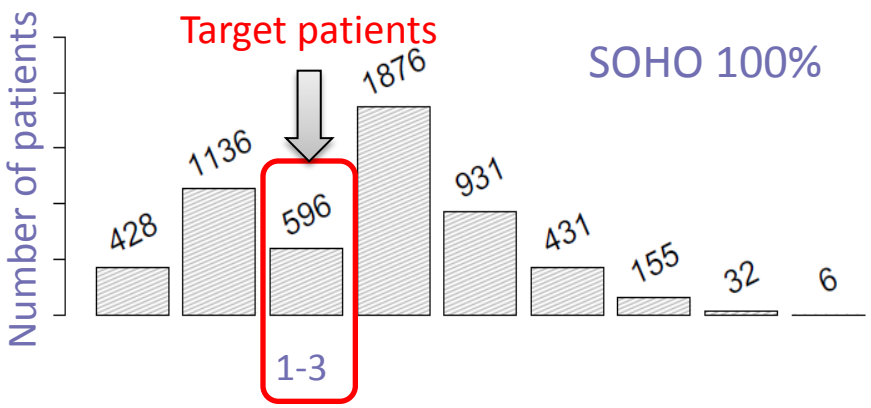


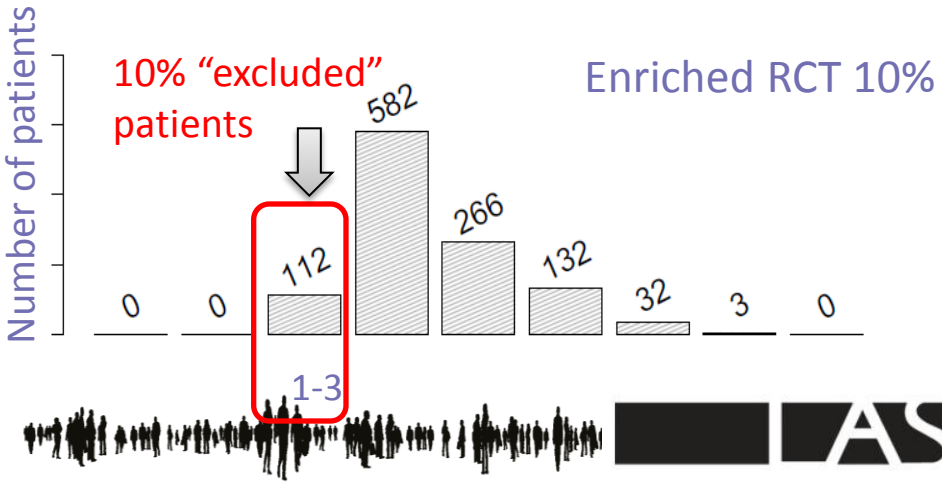
Predicted error using different RCTs enriched with few "suicide attempts"

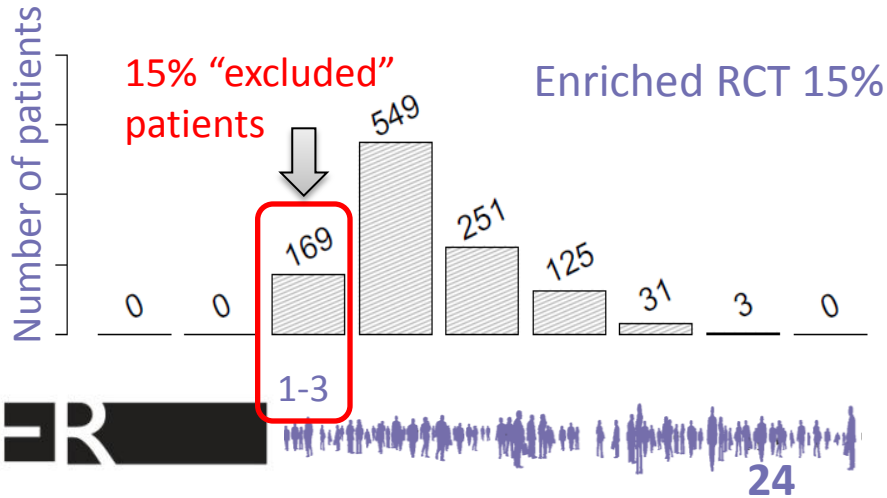


Distribution of duration of illness in reference RCT, SOHO and 2 enriched RCTs under drug “R”

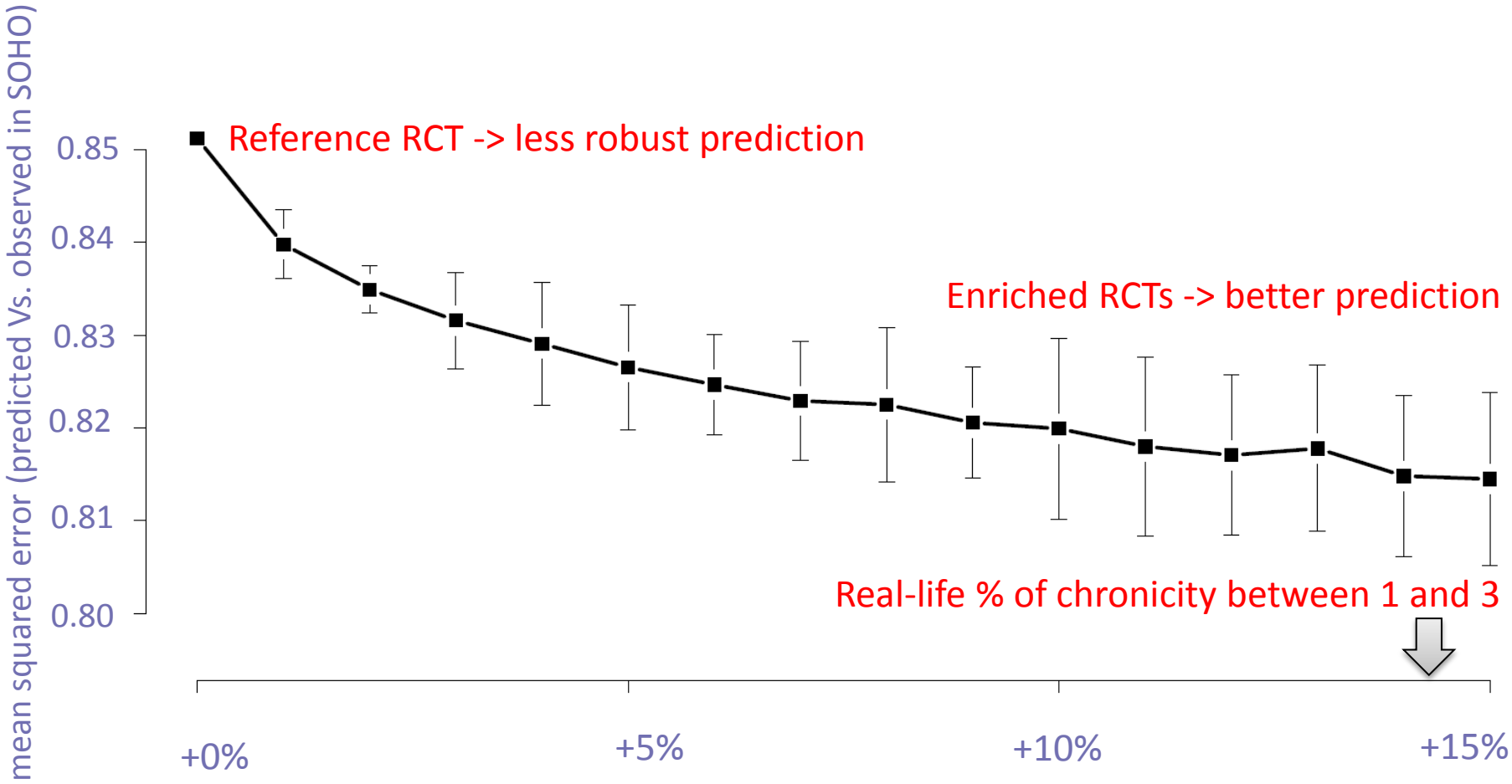








Predicted accuracy using different RCTs enriched with shorter "duration of illness"



Comparison of enrichment factors (patients under drug “R”)

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

Validation of reference RCT from literature

Study design	Synthetic reference RCT	CATIE study *(Lieberman et al, 2005)*	26-week prospective study in Korea (*Kwon et al, 2009*)	EUFEST study *(Kahn et al, 2008)*	Clinical PoC study *(Umbricht et al, 2014)*	COMET A study *(Cortesi et al, 2013)*	Two studies of OLZ treatment *(Lipkovich et al, 2007)*
CGI-S at baseline	4.39±1.02	4.0±1.0	5.10±1.01	4.8±0.8	4.4±0.7	4.6±0.9	4.6±0.8
ΔCGI-S at 3 months	-0.78±0.95	-0.4	-1.52	-2.0	36.6% patients improved	-0.7±1.2	-1.6±0.9

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