Methods to identify drivers of effectiveness (D2.2)

Context

Drivers of effectiveness (DoE) are characteristics related to the patients, the disease or the healthcare system which interact with the drug’s pharmacological effect. If drivers of effectiveness are not correctly taken into account during the drug development research (e.g., in randomized-controlled trials), the effectiveness of the drug measured in routine clinical practice may differ from what is expected based on studies done during the drug development (“efficacy-effectiveness gap” – EEG).

To counter this, drivers of effectiveness should be investigated early during the drug development process. Several methods may be used – separately or in combination – to identify important drivers of effectiveness.

How to use this document

The methods described were used in the realm of GetReal/WP2 to identify drivers of effectiveness in several case studies exploring different therapeutic areas (Hodgkin’s Lymphoma, schizophrenia and diabetes). The full-reports of the case studies can be found elsewhere. The strengths and limitations of these methods are explained, based on the experience of research teams while conducting these case studies.

Key methods

- Identification of DOE using information from the literature (systematic/focused)
- Identification of DOE using information from clinical experts
- Identification of DOE using information from actual data (patient-level/aggregate; observational/trial data)
1. Methods based on literature reviews

General principles
Structured literature reviews should be used systematically to identify drivers of effectiveness related to the patients/disease, the actual use of the drug or the healthcare system. Depending on the time available, a structured Focused Literature Review (FLR) or a Systematic Literature Review (SLR) may be chosen.

A SLR is a literature review based on a clearly formulated question and using systematic and explicit methods to identify, select, critically appraise relevant research in a comprehensive manner, and to qualitatively analyse and interpret the data from the studies included in the review. In a second step, meta-analysis techniques may be used to quantitatively analyse and summarise the results. A FLR (sometimes also referred to “Targeted Literature Review”) is a review based on a clearly formulated question using explicit methods to identify, select, critically appraise, qualitatively analyse and interpret key relevant research. The key differences between FLR and SLR are detailed in the “Focus Point” section.

PICOS framework
Both literature reviews approaches share the same PICOS framework to formulate the question and structure the search:

- **Population**: Disease/Indication of interest; if the disease is rare, it might be worth exploring diseases sharing similarities in potential drivers or effectiveness (e.g. all haematological malignancies instead of Hodgkin’s Lymphoma alone).
- **Intervention**: the drug(s) of interest, or class(es) of drug of interest with a same mode of action (e.g., antipsychotic drugs). In rare diseases or diseases with very few treatment options, the search may encompass all available treatments.
- **Comparator**: if an active comparator will be obviously used in RCTs, it is worth using it in the search (e.g., insulin). Sometimes, the search is meant to be broad and any comparator is worth to be included in the search.
  - One should be aware whether it is more relevant to explore relative effects (effects of the drug compared to another drug) or absolute effects (no comparison).
- **Outcome**: one or several outcomes of interest may be defined such as those used as a measure of effectiveness (e.g. overall survival in haematology or oncology) and/or those potentially impacted by DoEs (e.g. HbA1c in diabetes).
- **Study type**: Randomized Clinical Trials (RCTs) and/or observational studies depending on the objective.
  - Three approaches may be considered depending on the objective of the search and are detailed below.

A first approach is to focus on studies that specifically explored effect-modification and interaction, be it an RCT or an observational study. The objective is to understand if response to drugs may differ in certain sub-groups of patients in general and not necessarily in relation to the impact this might have in terms of EEG.

A second approach is to focus on the exploration of a gap in evidence between design characteristics of RCTs and of observational studies such as inclusion/exclusion criteria or decisions on treatments (e.g. treatment withdrawals) mandated by study protocols. The objective is to identify articles that
explored this gap (e.g. studies comparing data in RCTs and observational studies) and/or identified or suggested characteristics which are DoE or may account for an EEG (e.g. older patients excluded from RCTs when age is a potential or identified DoE).

A third approach is to review the RCTs and observational studies published and hereby investigate directly whether an EEG exists. This may be combined with a meta-analysis of the effect estimates from the identified RCTs and observational studies (see aggregate data analyses below). Efficacy results from RCTs are compared to effectiveness results of observational studies using the same exposure and outcome to the extent possible; a potential gap is thus identified and patient-related or disease-related characteristics potentially impacting the outcome are compared across study designs (RCTs and observational studies).

Strengths and limitations
The strengths and limitations of the 3 approaches are described in the Table 1 below:

**Table 1. Strengths and limitations of the 3 possible approaches, for a Literature Review**

<table>
<thead>
<tr>
<th>Approach 1</th>
<th>Strengths</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>The search will target “interaction” regardless of the study type or the drug used; a lot of information can be retrieved if the disease has been explored for enough time; Uses scientifically recognized (SLR) or robust (FLR) methodology. A SLR can be published.</td>
<td>This approach is not possible in rare disease or diseases where “sub-groups” were not explored; Only characteristics which were explored can be identified</td>
<td>There is a potential for publication bias</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Approach 2</th>
<th>Strengths</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Directly retrieves the characteristics known or suspected to play a role in EEG and/or DoEs and allows to take those factors into account in subsequent data analysis Relatively easy and quick way (especially when a FLR is performed) to obtain information since it is already published. Uses scientifically recognized (SLR) or robust (FLR) methodology. A SLR can be published. It is generally a good starting point in order to evaluate existing knowledge in terms of EEG and DoEs, and the data gaps before embarking in further data generation.</td>
<td>Relies on available evidence and its quality which depends on the objectives and data presented by the authors. Interactions or correlations between factors cannot be explored if those explorations are not included in the publications. Explorations on unknown factors are usually limited.</td>
<td>There is a potential for publication bias</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Approach 3</th>
<th>Strengths</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uses scientifically recognized (SLR) or robust (FLR) methodology. A SLR can be published. The investigator have full control over what to investigate</td>
<td>Relies on available evidence and its quality which depends on the objectives and data presented in the identified studies. Only characteristics which were explored can be identified Explorations on unknown factors are usually limited.</td>
<td>There is a potential for publication bias</td>
</tr>
</tbody>
</table>
2. Methods based on experts interviews

General principles
Expert’s interviews can be used to generate hypothesis on potential drivers of effectiveness. Alternatively, they can be performed after literature review results to identify EEG or DoE not retrieved by the review and/or weigh the literature review results with clinical experience.

Clinicians with a strong experience in treating the disease should be considered and whenever possible clinicians who also have a “population-approach” (e.g., epidemiologists). This is important because their clinical experience should be provided with the population perspective in mind.

Preferably, several clinicians should be interviewed separately, using a structured interview guide, to follow the same flow of questions. A briefing package should be prepared and send to the experts in advance to summarize the background information and the objectives of the interviews. The method used to deal with inconsistent results should be planned in advance.

Strengths
Expert’s interviews should be considered

- to complement the literature review with clinical experience
- or replace the literature review when data analyses are planned afterwards, in order to generate hypothesis that can be validated during data analyses

Expert’s interviews are useful to explore the clinical relevance of the potential DOE identified, restrict the number of DOE to be tested in data analyses, or identify potential DOE that were missed during the LR.

Limitations
Expert’s interviews may be subject to opinion bias and subjectivity. Therefore, they should not be considered as the sole source of information and should always be combined with literature reviews and/or data analyses.
3. Methods based on data analyses

General principles
Data analysis is a very strong source of information to identify DOEs.

Which data should be used?
To identify drivers of effectiveness, the dataset used should include information on:

1. the “exposure” to drugs
2. the “outcome”, i.e., the drugs effect in patients
3. and key characteristics related to potential drivers of effectiveness (related to the actual use of drug, the patients/disease or the healthcare system)

Different sources of data can be used (electronic healthcare databases, observational studies, RCTs), but a disease registry for instance will not necessarily contain information on the exposure to drugs or on key outcomes. Note also that in RCTs, only a few covariates are usually collected/reported on patients (age, gender) or disease (duration of illness, baseline risk) and usually lack information on the healthcare system (healthcare provider). Also, the actual use of drugs is assumed to be standard (dose, duration, adherence of patients, etc.) and is thus not heterogeneous.

More heterogeneous data will offer more information on the variability of patients, disease characteristics, use of drugs in routine practice, etc. However, more heterogeneous data require more patients to allow enough statistical power for the analyses.

In terms of granularity and precision in the information, patient-level data will have the optimal level of granularity/precision while aggregate data are a summary of patient-level data from different sources and lack of precision.

Which analyses to run?
The first step is to set the scene: what are we talking about? What is the conceptual model?

Firstly, the association of interest between “exposure” to a certain drug and a certain “outcome” needs to be specified

- A drug can be considered alone (no comparison) and its absolute effect measured, or compared to another or several other drugs and its relative effect measured
  - this choice is important: identifying drivers of absolute effectiveness is different from identifying drivers of relative effectiveness, and depends on the question needing to be answered.
- Several outcomes may be considered: continuous (evolution of symptoms, biological parameters, etc.) or dichotomous (death, hospitalization, etc.)
  - this choice is important: identifying drivers of effectiveness using one outcome measure may not be replicated with another outcome measure
A driver of effectiveness interacts in the association between the drug use (exposure) and the drugs effect in patients (outcome), meaning that for different “levels/strata” of each driver of effectiveness, the association between exposure and outcome will be of different strength/direction.

The potential drivers of effectiveness to be identified through data analyses should be hypothesised prior running the analyses to avoid:

- random findings, or data-driven results
- multiple testing

A graph is useful to construct and visualize the conceptual model (see Figure 1). The graph aims at eliciting what exposure is considered, which outcome is considered and the hypothesized drivers of effectiveness.

**Figure 1.** Example of a conceptual model

In the example above, the exposure is Drug A (vs. drug B – relative effectiveness), the outcome is the evolution of schizophrenia symptoms (time frame to be defined) and 2 potential drivers of effectiveness are explicated: (1) adherence to medication and (2) cannabis use/abuse. The figure also shows the fact that adherence and cannabis use/abuse may be correlated.
Patient-level data

Outline

Data are derived from RCTs or preferably from observational cohort studies primarily aiming at assessing the efficacy or effectiveness of drugs. Observational cohorts contain more information on healthcare settings and patients’ characteristics.

The available dataset will be used to:
- replicate the estimation of the drug’s effect (“core association model”)
- explore effect-modification and interaction on this core association

The objectives may be:
1) to test any effect-modification of potential DoE, as identified through Literature Review and/or Expert’s interviews (as detailed above)
2) to simulate the impact on drug’s effect estimate, of excluding or not patients with specific characteristics

Plan of analyses

Prior to the analyses, a statistical plan should be outlined, using the conceptual model as described above. Following the example above, the core association model is the association between exposure to drug A (as opposed to drug B – we are in a case of comparative effectiveness!) and the outcome (evolution of schizophrenia symptoms between baseline and 3 months):

- $Y$ represents the outcome
  - this is a continuous outcome; a linear regression model is thus used here
- $X_1$ represents the exposure to the drug: when drug A is used, $X_1=0$ and when drug B is used, $X_1=1$
- $\beta_1$ is the regression parameter associated to $X_1$; it represents the strength of association between $Y$ and $X_1$
- $\varepsilon$ represents the residual variance, not explained by the parameters (“noise”)

To ensure internal validity (i.e., the outcome is due to the exposure – there is a causality) and control for confounding bias (i.e., the association is not due to imbalance in baseline risk), the core model may be adjusted on confounders (several methods exist, e.g. propensity score adjustment).

- this will be noted $\Sigma \beta_i X_i$

Core association model: $Y = \alpha + \beta_1 X_1 + \Sigma \beta_i X_i + \varepsilon$

Drivers of effectiveness at a patient-level

Measuring effect modification of patient-level characteristics (e.g., age, disease severity score) may be done in many ways, including (but not restricted to):

- testing interaction of a potential DoE (denoted $Z$), on the core association model
  - $Y = \alpha + \beta_1 X_1 + \beta_2 X_1 \times Z + \Sigma \beta_i X_i + \varepsilon$
  - The test is: is different from 0?
- fitting the core model in the complete population and then in a restricted population (representing an RCT) and comparing the values of $\beta_1$
Drivers of effectiveness at a healthcare system-level

Measuring effect modification of healthcare system related characteristics (e.g., coverage system, number of specialists per 100,000 people in the country) may be done in many ways, including (but not restricted to):

- testing interaction of a potential DoE (denoted Z), on the core association model
  \[
  Y = \alpha + \beta_1 X_1 + \beta_2 X_1\cdot Z + \sum \beta_i X_i + \varepsilon
  \]
  - The test is: is different from 0?
- fitting a hierarchical model (mixed linear regression model) with a random intercept at a healthcare system level.

Aggregate data

Outline

Data are derived from RCTs and observational studies exploring the comparative/relative efficacy/effectiveness of one particular drug. As in meta-analyses, articles will be identified through a Systematic Literature Review (described above as the third approach under Literature Review), relevant data will be extracted from the studies and then analysed.

The objectives are:

3) to identify a potential EEG, defined as a difference in terms of drug’s effect estimates between RCTs and observational studies (drug’s effect estimates derived from RCTs being referred to as “efficacy” and drug’s effect estimates derived from observational studies being referred to as “effectiveness”)

4) to identify DoE, defined as study population characteristics that could explain this EEG, if any

Extraction of data

Firstly, as in meta-analyses it is recommended to assess the quality of each study identified through the SLR, and the validity of the drug’s effect estimate. More information can be found elsewhere (e.g. Higgins JPT, Green S, Cochrane Collaboration. Cochrane handbook for systematic reviews of interventions. Chichester, England ; Hoboken, NJ: Wiley-Blackwell; 2008).

Then, from each study, the following information is extracted:

- the drug’s effect estimate (e.g. difference between intervention arms, of change in HbA1c from baseline to follow-up point)
- the uncertainty related to this estimation (e.g. 95% confidence interval)
- characteristics of study populations and of study designs

It may very well be that not all relevant information is provided; in such case, the researcher then has to do some calculation. More information on this can be found elsewhere (Ref: tools from Cochrane: observational studies: www.riskofbias.info, RCTs: Higgins JPT, Altman DG. Assessing risk of bias in included studies. In: Higgins JPT, Green S, eds. Cochrane handbook for systematic reviews of interventions. Wiley, 2008:187-241).
In the comparison across study designs, the reviewer should also consider difference in length of study duration/follow-up and whether a meta-analysis within each study design is feasible.

**Analyses of data**

Differences in the characteristics study populations and/or characteristics of studies may lead to potential drives of effectiveness. If for example the RCTs include only young individuals, while the observational studies include both young and older individuals and if age is a drug’s effect modifier, then an identified efficacy-effectiveness gap may be due to different effect related to age. A difference in characteristics of the studies may also lead to potential drivers of effectiveness, due to the fact that they do not measure the same thing. For example this could be related to the length of study duration/follow-up. If for example the RCTs have a shorter duration, while the observational studies have a longer duration, then an identified efficacy-effectiveness gap may be due to different effect depending of exposure time.

In theory there is no limit to which characteristics of the study population or characteristics of study to be extracted and compared across study designs. However, in practice the comparison is limited to information available in both the identified RCTs and observational studies. Characteristics of the study population includes, but is not limited to, socio-demographic (age, gender, BMI, social class etc.) and disease severity and burden. Characteristics of the study includes, but is not limited to, study duration, level of exposure (e.g. dose, regime, etc.), washout period, analysis (intention-to-treat, per-protocol, as-treated, etc.).
Annex

Focus point on Literature Reviews

The overall process of a structured literature review (SLR or FLR) is summarized in the following figure:

![Literature Review Process Diagram](image)

Table 2. Key differences between Systematic and Focused Literature Reviews

<table>
<thead>
<tr>
<th></th>
<th>Systematic Literature Review</th>
<th>Focused Literature Review</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Search</strong></td>
<td>Exhaustive search of primary studies</td>
<td>Search of key relevant studies</td>
</tr>
<tr>
<td></td>
<td>It is mandatory to use at least 2 different search databases (usually at least Medline through Pubmed, and Embase)</td>
<td>Evaluate the need to use different search databases. The use of a single database (Pubmed or Embase) is possible</td>
</tr>
<tr>
<td></td>
<td>It is usually strongly recommended that search is also performed in grey literature (e.g. congress abstracts)</td>
<td>Evaluate the need to search in grey literature. If data retrieved in published articles is deemed sufficient and recent enough, the search in grey literature can be omitted</td>
</tr>
<tr>
<td></td>
<td>SLRs or meta-analysis already published on the topic are used for the screening of their references only. They are not part of the review itself</td>
<td>A stepwise approach starting with published SLRs or meta-analysis on the topic is a good approach to quickly identify and summarize the results. Those SLRs are included in the review itself. Their references can also be useful for study screening purposes</td>
</tr>
<tr>
<td><strong>Screening of Titles and Abstracts</strong></td>
<td>Should be performed by two independent reviewers (i.e. double independent review)</td>
<td>A single reviewer is acceptable</td>
</tr>
<tr>
<td><strong>Full text review</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Critical Appraisal</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Data extraction</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Approximate time of the review</strong></td>
<td>6 to 9 months (depends on the disease, and the available resources to perform the review)</td>
<td>2 to 3 months (depends on the disease, and the available resources to perform the review)</td>
</tr>
<tr>
<td><strong>Publication</strong></td>
<td>SLRs should follow state of the art international guidelines for literature reviews (e.g. Cochrane guidelines) and can be published in high impact journals. Their publication is strongly recommended as they can be useful for future FLRs.</td>
<td>FLRs are deviating from international guidelines for literature reviews and therefore publication in high impact journals may be challenging. FLRs are more fit for internal use of the information or internal decisions</td>
</tr>
</tbody>
</table>