



**GUIDANCE FOR MONITORING
THE EFFECT OF THERAPY WITH
MEDICINAL PRODUCTS IN REAL
MEDICAL PRACTICE,
PURSUANT TO THE LAW ON
MEDICINAL PRODUCTS IN
HUMAN MEDICINE**

GLOSSARY:

Direct comparison: comparison of treatments either by means of a single comparative study or a pairwise meta-analysis or other method for synthesis of comparative studies without indirect comparisons.

Effectiveness: describes how well a treatment works in patients; includes efficacy and safety.

Exchangeability: if patients from one treatment group were substituted into another, the same treatment effect is expected; contains the components similarity, homogeneity, and, in the case of indirect comparisons, consistency.

Indirect comparison: a broad term to refer to any evidence synthesis in which treatment groups from different studies are compared. This includes evidence synthesis in which inference about the relative effectiveness of two treatments is made without the use of trials comparing both treatments head-to-head; indirect comparisons are also made when more general methods of network meta-analysis are applied, even when head-to-head studies for the comparison of interest are available.

Meta-analysis: the synthesis of two or more comparative studies with a common intervention and comparator, to produce a pooled estimate of the relative treatment effect. Sometimes referred to as pairwise meta-analysis to distinguish from network meta-analysis.

Network meta-analysis (NMA): generalisation of meta-analysis to evidence networks consisting of more than two treatments, which can include both direct evidence and indirect evidence. NMA incorporates other terms used in the literature to describe the synthesis of both direct and indirect evidence, such as mixed treatment comparisons and indirect treatment comparisons.

Population-adjusted method for indirect comparisons: method for indirect comparisons in which a mix of individual patient data (IPD) from one or more trials, and aggregate data from other trials, are used to adjust for relevant population characteristics that differ between studies in order to estimate a treatment effect.

I. Introduction

Real-world evidence (RWE) is clinical evidence on a medical product's safety and efficacy that is generated using real-world data (RWD) resulting from routine healthcare delivery. Data quality is key to health technology assessment. Ensuring quality data requires dealing with challenges arising from the complexity and diversity of resources, study designs and analytical methods that are used. Therefore, in order to optimize the usefulness and transparency of RWE studies to HTA and regulatory authorities, it is important to develop a set of common principles and basic reporting standards:

- to ensure that HTA regulators and agencies have sufficient information to evaluate a study for appropriateness of use in decision-making;
- to provide basic reporting standards for RWE studies that are in line with global standards;
- to prioritize transparency in reporting while taking into account the practical challenges associated with RWD and RWE.

Use of Real-World Evidence across different regulators

National Institute for Health Care Excellence (NICE) in England, the Canadian Agency for Drug and Technology in Health (CADTH), Haute Autorité de santé (HAS) in France, and Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen/Der Gemeinsame Bundesausschuss (IQWiG/G-BA) in Germany, are moving toward optimal use of RWE in their HTA processes. Other HTA and reimbursement agencies in Asian countries including Japan, Taiwan, and South Korea rely on RWE to adjust prices and reassess funded technologies. Some other countries like India, China, and Singapore make reimbursement decisions based on RWE several years after market entry. Recently, some HTA bodies, such as HAS, NICE, CADTH and the Institut national d'excellence en santé et en services sociaux (INESSS) have initiated developing RWE frameworks to broaden the use of RWE in HTA. HAS developed guidelines on using RWE in HTA in 2021, while NICE released its RWE framework as a guide for good practice in 2022. CADTH has also developed a national framework to optimize the use of RWE collaboratively with INESSS and the Canadian regulatory body, Health Canada, in 2022. However, standard frameworks for the use of RWE from various emerging sources in HTA routine practice worldwide are lacking.

The European Medicines Agency (EMA) established a coordination centre Data Analysis and Real World Interrogation Network (DARWIN EU®)(EMA, 2021 r.). DARWIN EU aims to provide access to timely and reliable RWE on diseases, populations and effectiveness of medicines for human use from real world healthcare databases across the European Union (EU). The establishment of initiatives such as the DARWIN EU platform and the European Health Data Space-EHDS are evidence of the growing importance of using real-world data in healthcare regulation. The aim of these initiatives is to collect as much real-world evidence (RWE) as possible from EU Member States for secondary use and to generate evidence that can subsequently guide regulatory decision-making and support the evaluation of health technologies. In this respect, the European regulation aims at creating a framework with rules and standards to ensure the reliability, quality and security of the collected health

data used subsequently in decision-making - Towards European Health Data Space (TEHDAS), which is a project of the European Commission, and guidance on data quality of the European Medicines Agency - Heads of Medicines Agencies (EMA-HMA).

II. Purpose

In Bulgaria, the institution responsible for collecting and analyzing real-world data related to the use of innovative medicines is the National Council on Prices and Reimbursement of Medicinal Products, and the process developed for this purpose is the monitoring of the effect of therapy. The process is fully organized and led by the government institution.

When a new medicinal product is launched on the market, a much wider group of patients have access to treatment compared to the clinical trials. These patients may have comorbidities and be taking other medicines that could influence the effect of the new drug therapy. The observation of these patients in real therapeutic practice makes it possible to report on the results and effect of the innovative therapy, and to ensure rational drug use.

The purpose of monitoring the effect of therapy is to compare the therapeutic efficacy and safety of the new drug in real-world settings with efficacy and safety data provided by the clinical trials.

The National Council on Prices and Reimbursement of Medicinal Products analyzes information collected from medical institutions. The Council provides this analyzed information to the National Health Insurance Fund and the Ministry of Health with a view to supporting reasoned decisions for the efficient and appropriate spending of public funds.

III. Scope

Randomized controlled trials (RCTs) are the gold standard for establishing the efficacy and safety of health technologies. However, trials often produce results for specific target populations and setting within controlled environments, thus limiting the generalizability of results to patients in real-world settings.

However, randomised controlled trials are not always available for several reasons, including randomization is considered unethical, for instance because of high unmet need; patients are unwilling to be allocated to one of the interventions in the trial; a small number of eligible patients; healthcare professionals are unwilling to randomize patients to an intervention which they consider less effective. For example, in some circumstances, the evaluation of drugs for rare diseases, is not always feasible through conducting randomized controlled trials which detect important clinical outcomes. Real-world evidence (RWE) can potentially provide more generalizable evidence that fills knowledge gaps on the effectiveness, safety, and cost effectiveness of medicinal products.

RWE are data relating to patient status and/or the delivery of health care collected from variety of sources, and can include electronic medical records, clinical and disease registries, and administrative databases.

RWD can also be drawn from other prospective sources, including pragmatic and hybrid trials. RWD can provide information about medical history, demographics, socioeconomics

factors, health behaviours, experiences, clinical and functional outcomes, resources use, and costs.

In Bulgaria, the scope of the process for monitoring of the effect of therapy includes:

- medicinal products that have undergone a health technology assessment but lack evidence of therapeutic effectiveness and/or have an ineffective cost-effectiveness ratio,
- extension of therapeutic indication of a medicinal product already included in the Positive Drug List but this extension have not been public funded to date, and lack evidence of therapeutic effectiveness and/or have an ineffective cost-effectiveness ratio.

The Council may also, ex officio, impose an obligation for the monitoring of the effect of therapy for medicinal products already included in the PHS when:

1. the medicinal product is a comparative alternative to a medicinal product designated for monitoring the effect of therapy;
2. a need has arisen for the respective paying institution to carry out an analysis of the efficient and appropriate use of public funds for a medicinal product.

IV. Participants in the process

- NCPR – The National Council on Prices and Reimbursement of medicinal Products
- Healthcare establishments



Figure 1.

V. Methodology

Monitoring of the effect of therapy is carried out by healthcare establishments that have structures aligned with the disease profile under certain conditions and criteria. These establishments ensure that their hospital information system is compatible with the Council's information system. The monitoring period lasts until the initial maintenance of the reimbursement status of the medicinal product or from 1 to 3 years, depending on the therapeutic regimen (duration of the therapeutic course of treatment) of the medicinal product or the follow-up period of the medicinal product with which it is compared.

The data are provided to the National Council for Prices and Reimbursement of Medicinal Products in the form of XML files, sent daily by hospitals.

Data processing is done using a technology solution for analysis of vast amounts of structured and unstructured data from real practice using different sources. This technology solution is a regularly changing data system as the information is processed in real time.

A threshold of a minimum of 30 patients treated with the respective drug was set for reporting the effect of drug therapy and comparisons with alternatives. With a smaller number of patients, the sample population is considered too small to draw statistically significant conclusions. A minimum of 30 events (progression or death) in the specific population in real-world practice are needed for comparisons between the results from real-world practice and the results from clinical trials.

The interpretation of the results is summarised in an annual report.

V.1.Types of data:

- Epidemiological data

An epidemiological study measures a parameter of occurrence (generally incidence, prevalence or risk or rate ratio) of a health phenomenon (e.g., a disease) in a specified population and with a specified time reference (time point or time period). Epidemiological studies may be descriptive or analytic. Descriptive studies do not aim to evaluate a causal relationship between a population characteristic and the occurrence parameter and generally do not include formal comparisons between population groups. Analytic studies (also called causal inference studies), in contrast, use study populations assembled by the investigators to assess relationships that may be interpreted in causal terms. In pharmacoepidemiology, analytic studies generally aim to quantify the association between exposure to a medicine and a health phenomenon, and test the hypothesis of a causal relationship.

Studies can be interventional or non-interventional (observational). In interventional studies, the subjects are assigned by the investigator to be either exposed or unexposed. Most often, in these studies, exposure is assigned randomly and are known as randomised clinical trials (RCTs), and are typically conducted to test the efficacy of treatments such as new medications. In RCTs, randomisation is used with the intention that the only difference between the exposed and unexposed groups will be the treatment itself. Thus, any differences in the outcome can be attributed to the effect of such treatment. In contrast to experimental studies where exposure is assigned by the investigator, in observational studies the investigator plays no role with regards to

which subjects are exposed and which are unexposed. The exposures are either chosen by, or are characteristics of, the subjects themselves

- Data from RCTs

Randomised controlled trials are the preferred study design for estimating comparative effects. Non-randomised evidence may add value if randomised controlled trials are absent, not directly relevant to the research question or of poor quality. They can also complement trial evidence to answer a broader range of questions. The preferred study design will be context dependent. It may depend on whether variation in the exposure is within individuals over time, between individuals, or between other groups such as healthcare providers. In general, confidence in non-randomised study results is strengthened if results are replicated using different study designs or analytical methods, known as triangulation. In interventional studies, individuals (or groups of individuals) are allocated to 1 or more interventions according to a protocol. Allocation to interventions can be random, quasi-random or non-random. In observational studies, interventions are not determined by a protocol but instead according to the preferences of health and social care professionals and patients. Hybrid studies may make use of both interventional and observational data. Both interventional and observational studies can be uncontrolled. Uncontrolled studies are appropriate only in rare cases, in which the natural course of the disease is well understood and highly predictable and the treatment effect is very large. A comparison group is needed to generate reliable and informative estimates of treatment effects. Controlled studies can make use of variation in exposures and outcomes across individuals (or groups), within individuals (or groups) over time, or both.

- Cohort studies

In cohort studies, individuals are identified based on their exposures and outcomes compared during follow up. Cohort studies follow a group of individuals over time to investigate the relationship between an exposure and a disease outcome. Usually, cohort studies will compare individuals subject to different exposures from the same data source. However, they can also combine data from different sources including from interventional and observational data sources. In this case, the observational data is used to form an external control to the intervention used in the trial. The trial will often be an uncontrolled single-arm trial but could also be an arm from a controlled trial. External data can also be used to augment concurrent controls within a randomised controlled trial. External controls can also be formed from data from previous clinical trials.

- Self-controlled studies

Self-controlled, or 'within-subject', designs make use of variation in exposure status within individuals over time. Self-control case series studies compare the occurrence of an event in during a period when an individual is exposed to a specific risk factor to the occurrence of the same event during periods when the individual is not exposed. These include case-crossover, self-controlled case series, and variants of these designs. They are

most appropriate for transient exposures with acute-onset events. While primarily used in studies of adverse effects of medicines (including vaccines), they have been used to assess the effects of oncology medicines using the experiences of individuals on prior lines of therapy. This is most relevant if appropriate standard-of-care comparators are not available.

- Cross-sectional studies

Cross-sectional studies measure the prevalence of a disease and its associated risk factors at a particular point in time – information on current exposures and outcomes is collected at a single time point. These studies can provide information on the burden of disease in a population and help to identify risk factors for the disease. While they can be used to estimate intervention effects, they are less reliable than longitudinal studies (such as cohort studies) if there is need for a clear temporal separation of exposures and outcomes.

- Case-control studies

In case-control studies individuals are selected based on outcomes, and odds of exposures are compared. Case-control studies conducted within existing database studies are generally not recommended because they use less information than cohort studies. Case-control studies are most useful for rare outcomes or if there is a need to collect further information on exposures, for example, from manual medical record review or primary data collection.

- Quasi-experimental studies

Quasi-experimental studies and natural experiments exploit external variation in exposure across people or over time (an 'instrument') that is otherwise unrelated to the outcome to estimate causal effects. Common quasi-experimental methods include instrumental variable analysis, regression discontinuity, interrupted time series and difference-in-difference estimation. They are frequently used in public health settings when randomisation is not always feasible but have also been used in medical technologies evaluations. Instrument-based approaches may be useful if confounding because of unknown or poorly measured confounders is expected or an appropriate instrument is available that is associated with the exposure of interest and does not affect the outcome except through the exposure. Examples of instruments that have been used in healthcare applications include variation in physician treatment preferences or hospital formularies, genes, distance to healthcare providers or geographic treatment rates, arbitrary thresholds for treatment access, or time (for example, time of change to clinical guidelines that have immediate and substantial impacts on care patterns).

- Real-World Data

Real-world evidence contrasts strongly with evidence generated from RCTs. In fact, RWE provides answers to many of the well-known disadvantages of RCTs. Generation of

RWE from RWD can solve many of the above problems. Below are some advantages of RWE:

- No strict eligibility criteria, and thus fewer chances of no exclusions based on concomitant medications and comorbidities.
- Quicker, cost-effective: less time required for patient recruitment/enrolment and completing the research.
- Possibility of undertaking research that cannot be done with RCT, such as that on high-risk groups like pregnant women and children.
- Ability to track real-world patient behaviour.
- Rapid and more straightforward retrieval of and access to data.
- Large sample size facilitates sub-population analyses and less common effects.
- Large sample size facilitates better generalisability and modelling.

Thus, while RCTs still remain the gold standard for assessing safety and efficacy of drugs and medical products and the evidence from RCT represents the outcome of a ‘standardised’ intervention used in an ‘idealised’ setting, RWE represents the outcome of ‘variable’ treatment patterns in the ‘real world’. Therefore, RWE complements the RCT findings and can contribute to enhanced evidence generation.

Health data from real practice can help achieve more effective, higher quality, safer and more personalized health care. Health data and data science could dramatically transform public health and improve healthcare systems. Health data can also play a critical role in accelerating the development of new drug treatments for patients who need them most.

V.2. Data collection:

- Criteria by which data are collected

Generating RWE is performed through data collection in real settings from different sources such as health care databases, patient registries, databases connected with mobile devices, social media and platforms for patients.

Source data that meets basic analysis requirements should cover at least the following criteria:

- The database is active and data is available

The database shall be continuously active during the study period and the recorded data shall be accessible, i.e. have access to the data and may be evaluated by third parties, in particular regulatory authorities.

- Data usage meets ethical and security requirements

The use of source data should comply with ethical review regulations and with relevant data security and privacy requirements.

- The coverage of key variables

Source data is usually incomplete, but should have some coverage, including at least outcome variables, exposure/intervention variables, demographic variables, and important covariates related to the purpose of the study.

- The sample size is sufficient

Data governance should be fully considered and prejudged in order to ensure the sample size required for statistical analysis.

- Data collection methods

The data that is available for evidence generation goes through a process that is specific to the type of data, and to the processes and organizations that produce it. The stages of the common lifecycle include: defining the data requirements, collecting or generating the data, managing and processing the data, publishing the data, procuring and aggregating the data, testing and accepting it, delivering for consumption. It is not necessary that all phases are present in each process for processing and using data.

The proposal for the Regulation of the European Parliament and of the Council on the European Health Data Area and the TEHDAS and EMA-HMA guidance on data quality address primary and secondary data use.

The primary use of electronic health data is the processing of personal electronic health data to provide health services for the assessment, maintenance, and restoration of the state of health of a patient to whom they relate, including the prescription and dispensing of drugs and medical devices, as well as social insurance, administrative, and reimbursement services.

The secondary use of electronic health data aims at the provision of health or care or for public health, research, innovation, policy-making, official statistics, patient safety or regulatory purposes collected by entities and bodies in the health or care sectors, including public and private sector health or care providers, entities or bodies carrying out research in relation to those sectors, and the institutions, bodies, offices and agencies of the Union.

- Data quality

In the guidelines of the project of the European Commission "Towards the European Health Data Space" (TEHDAS) and of the European Medicines Agency - Heads of Medicines Agencies (EMA-HMA) on data quality, quality is defined as "fitness for purpose and users' needs in relation to the secondary use of data for health research, policy making and regulation, and that the data reflect the reality they are intended to represent."

The definition suggests an approach to data quality that includes elements of technical quality and utility. These are relevance, accuracy and reliability, coherence, coverage, completeness and timeliness. The approach is applied in two aspects: the first focuses on data quality and utility at the level of the dataset; and the second focuses on data quality management procedures at the level of the data owner.

Data quality management should be applied throughout the data lifecycle, focusing on (a) data collection, curation, storage, and staging, (b) data integration with relevant sources and systems, (c) data description and meta-data management (i.e., use of meta-data standards), (e) data quality assessment, profiling, and remediation, (d) data modeling, transformation, operationalization, and servicing.

To ensure the quality of the data collected, they undergo an assessment of the metrics of relevance - which of the data collected are relevant to monitoring the effect of therapy with a given medicine, accuracy and reliability - are the data correctly completed, coherence, coverage - do they cover all the necessary information, completeness and timeliness - are all the necessary data completed and in what time period. Data management processes

monitoring, incident detection and resolution and data enrichment procedures are also assessed.

The health data in monitoring the effect of medicines in Bulgaria corresponds to the steps that the life cycle of data, used to generate evidence, goes through:

Data lifecycle	Practice in Bulgaria
defining the necessary data	development of criteria for monitoring the effect of therapy
data collection/generation	from the hospital by transferring XML files from the hospital systems
data management and processing	NCPR/Danny Platform (Sqilline) - data extraction, anonymization and structuring
data procurement and aggregation	NCPR analyses the data
testing and acceptance	data validation, qualitative analysis
delivery for consumption	annual reports, publications

The National Council processes the data collected from the healthcare establishments designated to monitor the effect of the therapy of medicinal products. As a result of the processed information, the NCPR analyses the reported data (collected through the hospital information systems), which are essential for the correct collection and interpretation of the data defined for the monitoring of the effect of the therapy of medicinal products. Findings on the quality of the reported indicators and data provided by each of the healthcare establishments are summarized, including the specific facts. Where discrepancies are found, the relevant establishments shall be informed by letters of intent.

The most frequent gaps in the reported data concern: incorrectly reflected data on therapeutic lines or combined administration of medicinal products; missing or incorrectly filled information on genetic mutation or biomarker; determination of the type of therapy, etc. The Council's proactive approach and collaboration with medical establishments has led to the optimisation of the database of medicinal products for which the effect of therapy is monitored.

V.3. Data processing

- Statistical analysis of real-world data

Descriptive statistics and unadjusted analysis

Descriptive statistics are used to summarize and describe the basic features of the population, and can be used to assess imbalances between the study groups. These include measures of range, dispersion, and central tendency for continuous variables, number and percent for categorical variables, and plots for evaluating data distributions. The standardized

mean difference is often used to characterize the magnitude of differences in covariates between the exposure groups. The important first step in unadjusted analysis is to define a proper time scale and time origin for the data. A misspecification of the time origin can lead to biased estimates of all the outcome probabilities of interest. The denominator of this estimated probability must include participants who are at risk and not participants without potential for experiencing the event at the time.

Univariate or unadjusted analysis can be used to provide a preliminary assessment of which covariates are associated with exposure and/or study outcomes. Causal diagrams are also an important tool for identifying the role that covariates play given our understanding of the temporal and causal relationships among these measures, the exposure, and outcomes of interest.

- Estimation of absolute vs relative measures of effects

The reporting of relative effect estimates (e.g. hazard ratios, relative risks, and odds ratios) is routine and allows comparisons across settings with apparent ease. Relative measures can obscure potentially important differences when the background risk of the outcome varies between groups or settings. For example, when comparing a younger population with a low mortality rate (1/1000 person-years) to an older population with a higher mortality rate (100/1000 person-years), a constant relative effect of treatment (e.g. relative risk of 0.90) would lead to very different impacts of the intervention.

Estimates of absolute effects are valuable for weighing certain outcomes against others. For example, a large relative increase in the risk of a rare outcome (e.g. anaphylaxis) may be of less concern than a modest relative increase in the risk of a common outcome (e.g. myocardial infarction). Communicating the magnitude of relative effects improves if absolute effects (such as risk difference and number needed to treat) are included. Providing both absolute and relative measures of effect provides with more complete information on the potential benefits and harms of a given treatment.

The other elements of the study design and analysis will need to be informed by the choice of effect measures. For instance, some relative effect measures are unbiased when the outcome is assessed with perfect specificity (no false positives) and there are no differences by treatment group in the sensitivity. In contrast, the absolute effect measure (risk difference) is unbiased when the sensitivity is maximised, without differences by treatment group in the specificity. Thus, the choice of effect measure has implications for selecting an outcome definition that maximises specificity or sensitivity.

- Competing risk events

A competing risk is an event that precludes the outcome of interest from occurring for that individual. It is not merely the inability to observe the outcome of interest, but also eliminating the outcome from ever occurring, observed or unobserved. The most common competing risk is death. In any study in which mortality is not the outcome of interest, death before the event of interest will serve as a competing risk. Other competing risks are perhaps less obvious but equally important to address including, for example, hysterectomy in studies

of uterine cancer, hospital discharge in studies of in-hospital mortality, complete mastectomy in studies of breast cancer recurrence.

Appropriate handling of competing risks is critical for the analysis plan. Many analyses erroneously treat competing risks like all other censoring events. This approach leads to the imputation of events for these individuals based on the observed event rate among those who remain uncensored in the analysis at later follow-up times. In doing so, the resulting estimates of the risk of the outcome of interest from the complement of the Kaplan-Meier curve will be inflated and therefore overestimate the risk.

If the competing risk is also of interest as an outcome relevant to the estimation of treatment effects, one simple approach is to create a composite outcome in which the occurrence of either outcome is used to estimate the time to event.

- Data correction methods

Population-adjusted methods for indirect comparisons are useful in situations in which an NMA is performed but there is some doubt regarding whether the similarity assumption is valid for some effect modifiers. This doubt can be resolved by applying a population-adjusted method that contains the corresponding effect modifiers to confirm the results of the NMA.

Two early approaches for population-adjusted methods for indirect comparisons were developed for situations involving two trials, one comparing treatment A versus treatment B (AB trial) and one comparing A versus C (AC trial), with IPD only available for the AB trial. A third approach extended the standard NMA framework. It is important to note that the target population for the population adjustment may differ from that of the pivotal clinical trial(s) for the intervention under assessment. The population for which the relative treatment effect is estimated must be clearly stated, while bearing in mind that this can often differ from the population of interest.

The simulated treatment comparison (STC) method fits an outcome regression model using IPD from the AB trial to predict the average effect of A versus B in the AC population dependent on the covariates, and finally a population-adjusted average effect of B versus C in the AC population. The method also relies on the assumption of “conditional constancy of relative effects”, which means that the model contains all relevant effect modifiers. Furthermore, the validity of STC depends on the correct specification of the outcome regression model.

The matching-adjusted indirect comparison (MAIC) method uses reweighting methods similar to inverse propensity score weighting to predict a population-adjusted average (marginal) effect of B versus C in the AC population. It is important to note that this method is only valid where there is sufficient overlap between the patient populations in both trials. Furthermore, MAICs are limited to providing a comparison that is adjusted to the population of the study for which only aggregate data are available, which may not match the target population for the decision. The method also requires the assumption of "conditional constancy of relative effects" to hold, which means that the model contains all relevant effect modifiers.

The multilevel network meta-regression (ML-NMR) approach for population-adjusted indirect comparisons was proposed by extending the standard framework for NMA. As the other population-adjusted methods, ML-NMR depends on the assumption of “conditional

constancy of relative effects” and on correct specification of the outcome regression model. This approach has some conceptual advantages in facilitating inferences from larger networks with any number of treatments. ML-NMR targets conditional treatment effects, and is compatible with non-linear link functions. The population-adjusted treatment effects can be estimated for any target population with given covariate values, and not just the population of the trial for which only aggregated data are available.

- Key moments:

- For cases in which the property of similarity does not hold, the usual methods for indirect comparisons are invalid. In this scenario, population-adjustment methods might be considered as an alternative approach, provided the network is connected and there is good evidence a priori that such an adjustment is likely to reduce bias. To this end, model and covariate selection strategies should be prespecified and based on transparent criteria.
- Access to IPD from at least one treatment arm in some of the studies included is required in order to adjust for imbalances between trials.
- Treatment effects estimated from population-adjusted indirect comparisons are associated with additional uncertainty arising from several sources.
- The target population for which the treatment effect is estimated via a population-adjusted method has to be described in detail.

The first step in the assessment of the statistical analysis is to consider whether the method used is correct for the network of evidence. The Bucher indirect treatment comparison is appropriate for a network comprising two treatments indirectly compared through a common comparator. The Bucher method can be applied in star-shaped networks to obtain indirect comparisons of each pair of treatments via a shared comparator and can also be applied iteratively in ladder networks to indirectly compare treatments connected by paths of length greater than two. Multi-arm trials can only be included as pairwise comparisons, but the generated effect estimates are correlated, and the corresponding standard errors are inappropriate. This correlation will be problematic if the aim is to use the estimates in a decision model because the method assumes independence between pairwise comparisons.

In cases with several different pairwise comparisons, or when a random-effects approach is deemed appropriate, a network meta-analysis encompassing all this evidence should be considered. Frequentist and Bayesian methods are equally applicable. Naive comparisons (i.e., comparisons of absolute outcomes without any adjustment for confounding) should not be used because they do not preserve randomisation. Disconnected evidence networks cannot be analysed with these methods.

If the method for analysis is deemed appropriate (assumptions met), the appropriateness of the model used must be validated. This includes, but is not limited to, justification for the use of a fixed-effect model over a random-effects model, the choice of informed, uninformed, or vague priors (Bayesian), and baseline risk adjustment models. Additionally, any subgroup or meta-regression analysis for different levels of identified effect modifiers must be described and justified. Further considerations must be given to the number and heterogeneity of studies informing each contrast, number of events (rare versus common events), scale (OR, RR, HR, RD, or MD), quality of evidence, and so on, when assessing the appropriateness of the method and model choices.

Population-adjusted methods are used in the context of an ITC or more general NMA, in which there is concern that the similarity assumption might not hold. These methods aim to adjust for this imbalance to obtain an unbiased estimate of the relative treatment effect in the scenario in which IPD is available for one or more trials in the network, and only aggregate data (AgD) for others. MAIC and STC should not be used when full IPD is available for all studies; IPD network meta-regression is generally the appropriate method to adjust for covariate imbalances in this case.

The validity of all population-adjusted methods depends on the inclusion of all effect modifiers as covariates in the relevant model. In the case of both MAIC and STC, only effect modifiers of the relative effect being estimated in the IPD trial are needed to carry out the adjustment. However, interpretation of the results also requires knowledge of effect modifiers for the AgD trial. In the case of more-complex networks of evidence (e.g., using ML-NMR), knowledge of effect modifiers for all pairwise comparisons is typically needed.

In the case of anchored STCs, the inclusion of additional prognostic variables (that are not also effect modifiers) in the outcome model will not reduce bias, but could improve precision of the estimated treatment effect and, therefore, can be considered. If this is done, then the additional variables should be specified a-priori and justified.

The STC and ML-NMR methods can generate estimates of the treatment effect in any target population by substituting the relevant mean covariate values into the outcome regression model. This can be useful if the population of interest differs from the trial populations. However, the validity of these estimates is unknown outside the range of covariate values included in the IPD study; extrapolation beyond this region might not generate meaningful estimates of the treatment effect. The usual approach to STC involves substituting mean covariate values from the AgD population into the outcome regression model, which estimates the conditional treatment effect at this level of the covariates (i.e., the predicted individual-level response). When MAIC is used to carry out population adjustment, the principal concern is whether the weighted pseudo-population has the same distribution of effect modifiers (anchored and unanchored comparisons) and prognostic variables (unanchored only) as the target population. These distributions should be reported, and their similarity assessed; if nontrivial differences exist for one or more variables after matching, then the results of the MAIC will likely be biased.

Population-adjusted methods for indirect comparisons are also used when considering disconnected networks. The validity of the results depends on all relevant prognostic variables (as well as effect modifiers) being included as covariates in the relevant model, which is unlikely to be satisfied in practice. In general, this will substantially increase the amount of adjustment required. The process used to identify prognostic variables is analogous to that described previously for effect modifiers and should be reported transparently in the submission.

Differences in patient characteristics are typically more likely to affect the absolute values of outcomes than the relative effects, which means that more covariates must be included in the adjustment model to obtain an unbiased estimate of the treatment effect. For example, if two hypothetical treatments, A and B, aimed at lowering blood pressure were to be compared in an unanchored comparison, then adjustment would need to be carried out for all covariates potentially affecting blood pressure, such as age, sex, smoking status, race, geographical location, body mass index, diabetes status, and many others that might not have been recorded. By contrast, an anchored comparison of an A and B via a common (e.g.,

placebo) comparator would only require adjustment for covariates affecting response to treatment.

VI. Methods of data processing applied by the NCPR

- Methods – software (versions and analytical tools) – a platform for analyzing large amounts of structured and unstructured real-world data from various sources. The system handles regularly changing data as information is processed in real time.

- Statistical code used – The following endpoints were used to monitor the effect of drug therapies with real-world data:

1. Overall Survival (OS) - overall survival, defined from the start date of treatment to death from any cause;

2. Progression-Free Survival (PFS) – progression-free survival, defined from the start date of treatment to the recorded response of "progression" or death from any cause;

3. Clinical Benefit Rate (CBR) - the proportion of patients who have a complete response, partial response, or stable disease.

The above endpoints are standard in clinical trials as well as in studies based on real-world data.

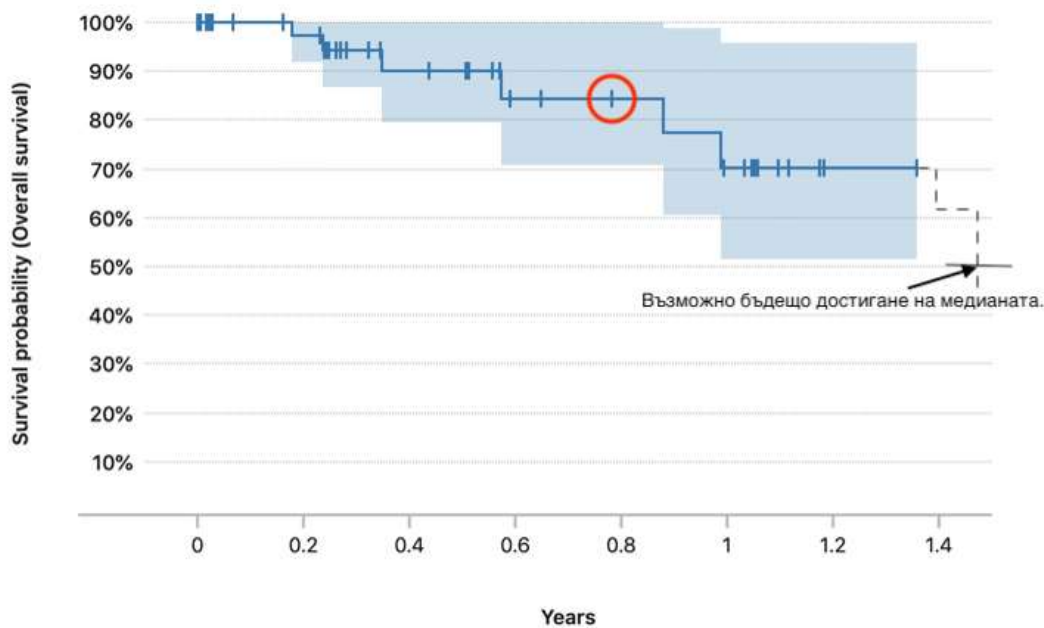
- Statistical methods – a description of all variations, including the models; the method of variable selection for the model; methods used to control error; methods used for missing data; how follow-up time and changes in exposures for processed data are tracked; subgroup analyses; data validation.

- Quantitative determination of the accuracy of all estimates using confidence intervals - Survival results are examined using time-to-event analysis and forecasted using the Kaplan-Meier method. The curve describes the percentage of patients for whom an event has not occurred (i.e., they are still alive and/or no progression has occurred) at a given point in time. The confidence interval shows the error in measurements. As the number of patients and events increases, the probability of error decreases, resulting in a more accurate probability value. Confidence intervals are most commonly constructed with confidence levels of 95% or 99%, with 95% used for comparisons. By establishing a 95% confidence interval using the sample mean and standard deviation, assuming a normal distribution represented by the bell curve, researchers determine an upper and lower boundary that contains the true mean value 95% of the time.

For accurate and statistically reliable reporting of the 'overall survival' indicator, which is defined from the initial date of treatment until death from any cause, when analyzing real-world data, data on overall mortality in Bulgaria are recorded. These data are obtained from the National Statistical Institute (NSI) / National Health Information System (NHIS).

An example graph is shown below, illustrating how the survival curve is interpreted.

Figure 2. Survival curve



The ordinate axis (y-axis) describes the probability of survival, while the abscissa axis (x-axis) describes time. The curve represents the percentage of patients for whom an event has not occurred (i.e., they are still alive and/or no progression has occurred) at a given point in time. In the specific example of overall survival, at the end of the first year, the overall survival is ~70%, indicating that 70% of all patients who have been on therapy for at least 1 year are still alive. The median corresponds to the time at which half of the patients have died. On the graph, this is the point where the y-axis intersects on the 50% threshold with the survival curve in dark blue. In the example, the curve does not intersect the 50% threshold, indicating that the median has not been reached. A dashed line indicates a possible future moment when the median is reached (the curve intersects the 50% threshold).

The light blue shaded area represents the 95% confidence interval. At the 1-year mark, the interval is between 52% and 96%, meaning that if the experiment were theoretically repeated 100 times, the probability of survival would fall between 52% and 96% in 95% of the cases. The confidence interval indicates the error in measurements. As the number of patients and events increases, the probability of error decreases, resulting in a more accurate probability value.

The vertical lines (dashes) on the curve mark censored patients (excluded from the statistics). These patients have not experienced an event (death and/or progression, depending on what the curve shows – OS/PFS) within the respective time period marked by the vertical line (dash). The patient is excluded from the statistics because there is no information about their therapy beyond the duration marked by the vertical line/dash.

For example (highlighted in red above), a patient is shown whose therapy duration is approximately 0.8 years. The graph depicts overall survival. Therefore, we conclude that the patient underwent therapy for about 0.8 months and is alive at this point (no negative event has occurred - no death). We have no data on the patient beyond this time interval, so we exclude them from the statistics.

The drops in the curve indicate patient(s) for whom a negative event (death and/or progression) occurred at the corresponding point where the curve declines.

When applying the described methodology, a comparison is made between a drug designated for therapy monitoring and its comparative alternative using collected and analyzed real-world data. The results are graphically presented as follows:

Figure 3. Comparison between PFS of a medicinal product, designated for monitoring of the effect, and of a comparative alternative

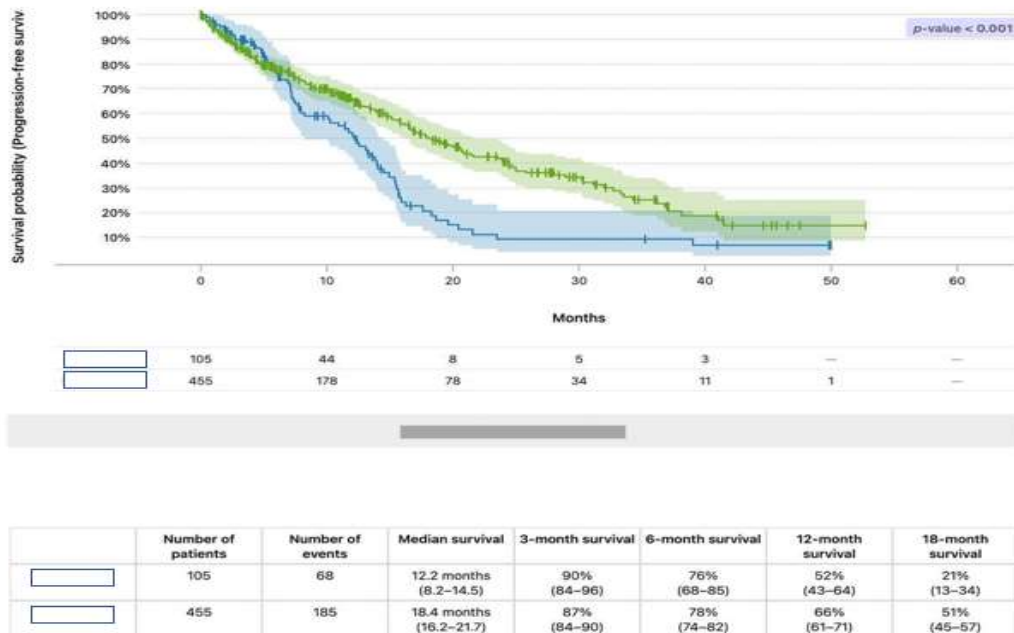
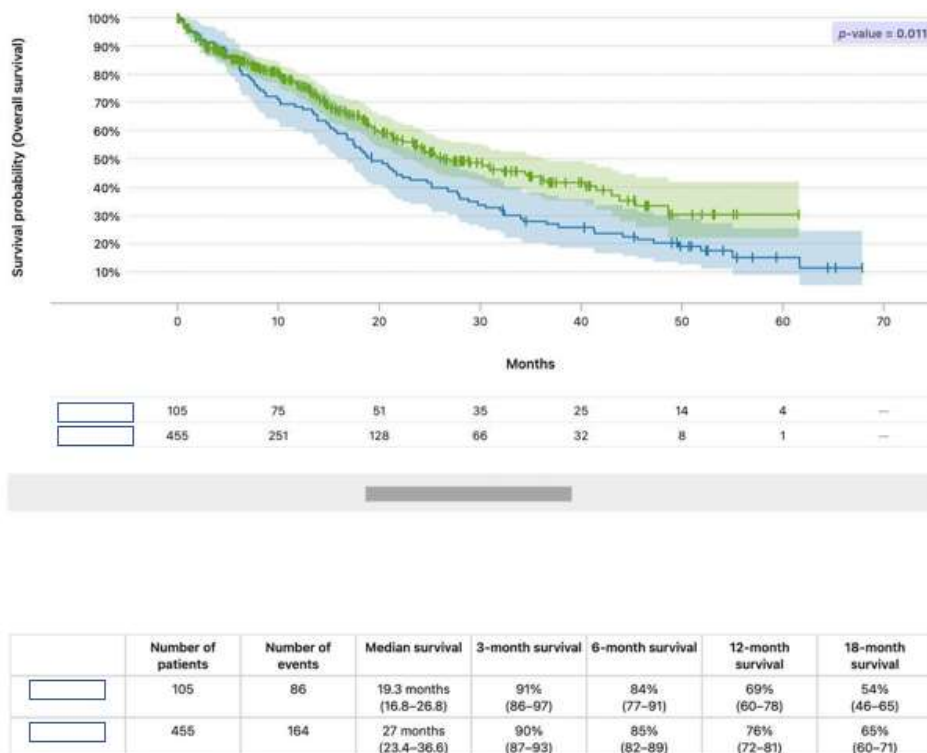


Figure 4. Comparison between OS of a medicinal product, designated for monitoring of the effect, and of a comparative alternative



- Reporting the threshold of statistical significance. The p-value is a statistical measure used to validate a hypothesis against observed data, measuring the probability of obtaining the observed results assuming the null hypothesis is true. Generally, a p-value less than 0.05 is considered statistically significant, in which case the null hypothesis should be rejected, with the p-value = 0.05 defining the threshold of statistical significance. This somewhat corresponds to the probability that the value of the null hypothesis (often zero) is contained within the 95% confidence interval.

- When comparing real-world data with results from clinical trials for a given therapy, typically the patient cohort in real-world practice has a different distribution of characteristics compared to the cohort in the clinical trial (such as gender, age, ECOG, and others). Without further adjustment of the characteristics of the real population to those in the clinical trial, the obtained results may be influenced by these differences (for example, if the real population is older or has a higher ECOG, it is highly likely to have lower progression-free survival (PFS) and overall survival (OS) outcomes than those in the clinical trial). When comparing real-world data with data from clinical trials, the Iterative Proportional Fitting (IPF) algorithm is used to align characteristics, correcting real-world data to match those of the clinical trial. This adjustment allows progression-free survival (PFS) and overall survival (OS) curves to be more comparable to those observed in the study. The algorithm functions as follows:

1. A table is compiled with the characteristics of the population from the clinical trial, represented as a percentage of the total cohort adjusted to 100% for each characteristic. For example (column PERC):

RCT		
CHAR	VAL	PERC
ECOG	0	61,4
ECOG	1	38,6
ECOG	else	0
Age	<23	0
Age	23-62	50
Age	63-91	50
Age	>91	0
ERpos	Yes	99,4
ERpos	No	0,6
PRpos	Yes	81,1
PRpos	No	18,9
DeNovo	Yes	34,1
DeNovo	No	65,9
PrevCTx	Yes	43,7
PrevCTx	No	56,3
PrevET	Yes	52,4
PrevET	No	47,6

2. The percentages for the same characteristics of the population are dynamically calculated based on real-world practice data. For example (column PERC):

RWD		
CHAR	VAL	PERC
ECOG	0	44,9
ECOG	1	55,1
ECOG	else	0
Age	<23	0
Age	23-62	47
Age	63-91	53
Age	>91	0
ERpos	Yes	97,4
ERpos	No	2,6
PRpos	Yes	84,7
PRpos	No	15,3
DeNovo	Yes	47,1
DeNovo	No	52,9
PrevCTx	Yes	34,9
PrevCTx	No	65,1
PrevET	Yes	33,3
PrevET	No	66,7

3. The IPF algorithm computes coefficients based on real-world practice data, aligning the percentages from real-world practice to those from the clinical trial. For example (column ADJ. COEF):

CHAR	RWD			RCT
	VAL	PERC	ADJ. COEF	PERC. ADJ
ECOG	0	44,9	1,3674833	61,4
ECOG	1	55,1	0,70054446	38,6
ECOG	else	0	0	0
Age	<23	0	0	0
Age	23-62	47	1,06382979	50
Age	63-91	53	0,94339623	50
Age	>91	0	0	0
ERpos	Yes	97,4	1,02053388	99,4
ERpos	No	2,6	0,23076923	0,6
PRpos	Yes	84,7	0,95749705	81,1
PRpos	No	15,3	1,23529412	18,9
DeNovo	Yes	47,1	0,72399151	34,1
DeNovo	No	52,9	1,24574669	65,9
PrevCTx	Yes	34,9	1,252149	43,7
PrevCTx	No	65,1	0,86482335	56,3
PrevET	Yes	33,3	1,57357357	52,4
PrevET	No	66,7	0,71364318	47,6

The obtained coefficients are applied as weights to individual patient data when calculating progression-free survival (PFS) and overall survival (OS).

Results

Figure 5. IPF-adjusted Progression-free survival (PFS)

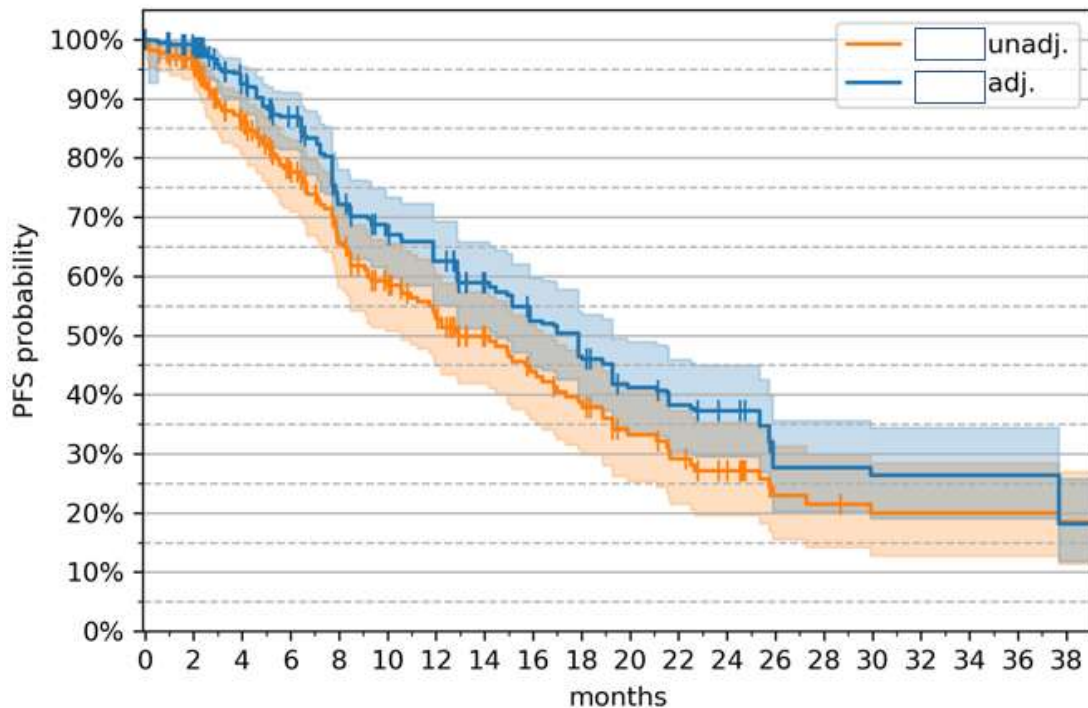
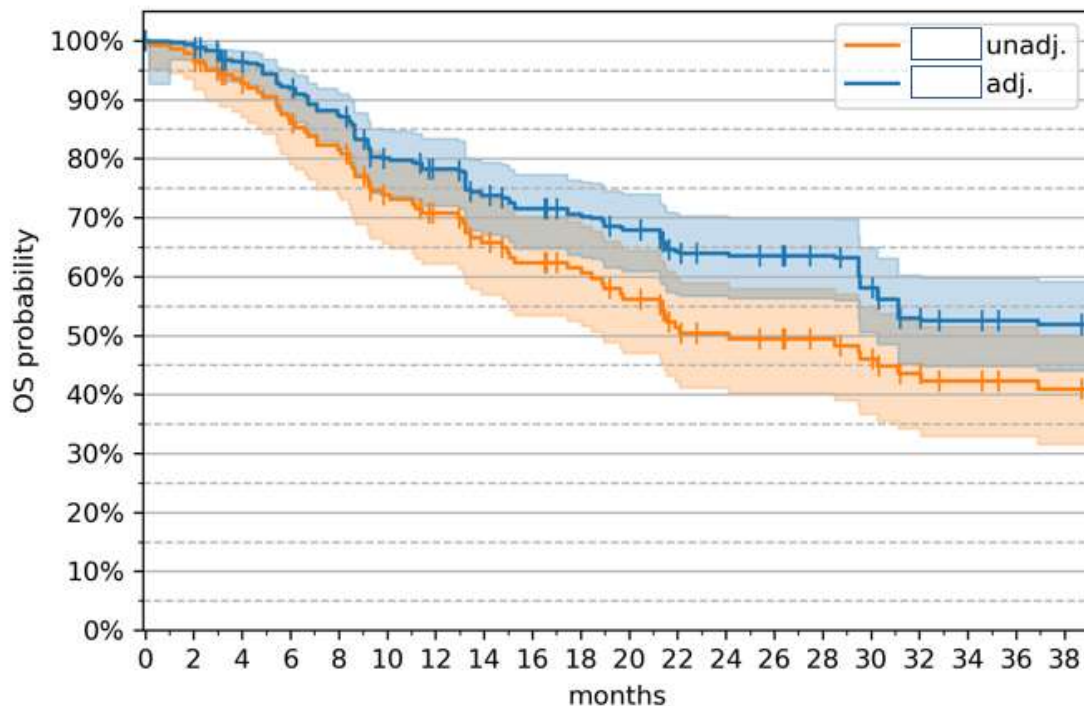


Figure 6. IPF-adjusted Overall survival (OS)



4. Limitations - Missing Data and Deviations

For certain inclusion and exclusion criteria in clinical trials (such as types and values of biomarkers, genetic markers, etc.), values are not entered for 100% of the patients in the primary documents (discharge summaries, oncology protocols, outpatient procedures) as structured or unstructured data. In these cases, patients with missing data are excluded from the comparison.

When comparing data from clinical trials with real-world data, we differentiate between internal bias and external bias. To identify internal bias, we compare the details of the completed study against the protocol initially set. To identify external bias, we compare data from the clinical trial with results from real-world practice. Ultimately, the difference in individual external biases is distributed proportionally across the groups in the IPF alignment of the two populations (from the clinical trial and real-world practice).

For internal biases (internal bias), we use the following categorization: selection bias; performance bias; attrition bias; detection bias. Selection bias refers to systematic differences between the comparative groups at the baseline level. Presentation biases represent systematic differences, different from the comparison in the study; this can be caused by inadequate blinding of trial participants and incorrect classification of exposure in observational studies. Attrition biases reflect systematic differences between the comparative groups in exclusions

and dropouts. Detection bias describes systematic differences between groups in the assessment of the outcome, which may result from lack of standardization and blinding of those assessing the outcome.

For external biases, the categories represent the differences between clinical trials and the four components of the target study: population bias (between clinical trials and real-world data); intervention bias; control bias; outcome bias. For example, population bias may reflect differences in age, gender, or health status of study participants compared to the real-world population. Intervention bias relates to differences in dosage, timing, or method of treatment delivery in clinical trials compared to real-world practice, and similarly, control bias relates to differences in the control strategy. Outcome bias may represent differences in the definition, timing, or method of outcome assessment in clinical trials compared to real-world practice.

The reporting of therapeutic outcomes, as well as the presence/absence of progression, varies in detail and frequency between clinical trials and real-world data, leading to differences in months of follow-up and distribution among patient groups with complete or partial response, or stable disease. These differences do not impact overall comparison results in relatively large patient cohorts and long-term follow-ups, but results in small patient numbers and relatively short follow-up periods should be treated with caution. In comparisons, results are tested for statistical significance ($p\text{-value} \leq 0.05$) for rejecting the null hypothesis.

5. Personal Data - Patient data is fully anonymized. For the purpose of aligning basic characteristics of populations in clinical trials and real-world practice, the patient's gender and age are known. For comparisons of Progression-Free Survival (PFS) and Overall Survival (OS), data on deceased patients are also obtained. Anonymous primary documents (discharge summaries, oncology protocols, outpatient procedures) are collected for patients, from which data on disease history and types of treatments applied are aggregated.

- Data security

The National Council continuously monitors and takes measures to ensure information security. The database has been migrated to the State Hybrid Private Cloud, optimizing resources for building virtual storage, enhancing information security, and accelerating the recovery of critical information resources in case of sudden failure. Access to the information system is strictly governed based on the job responsibilities of respective staff members. The implemented module for therapy effect monitoring anonymizes information at the hospital level before sending it to the server of NCPR.

VII. Report on the monitoring of the effect of the therapy

In accordance with the requirements of Article 31b, paragraph 6, and Article 35a, paragraph 11 of the Regulation on the Conditions, Rules, and Procedures for the Regulation and Registration of Prices of Medicinal Products, for monitoring of the effect of the therapy of medicinal products included in the Positive Drug List, the National Council for Pricing and Reimbursement of Medicinal Products conducts an analysis of the collected and processed

statistical information (accumulated data). This analysis is presented to the Ministry of Health and the National Health Insurance Fund:

- Annual Report on the monitoring of the effect of therapy of medicinal products included in Appendix No. 1 of the Positive Drug List;
- Annual Report on the monitoring of the effect of therapy of medicinal products for the treatment of solid tumors and hematological diseases included in Appendix No. 2 of the Positive Drug List, for conducting an analysis of the effective and efficient expenditure of public funds for medicinal products designated for monitoring.

The annual reports consist of three main parts:

- Reporting on the effect of the therapy of medicinal products designated for monitoring by the National Council on Pricing and Reimbursement of Medicinal Products.
- Comparison of therapy effects between different medicinal products (direct comparative alternatives) for the same ICD code.
- Comparison of therapy effects between real-world practice and clinical trials.

Endpoints:

For monitoring the effect of the therapy of medicinal products using real-world data, the following endpoints have been used:

- Overall Survival (OS) - defined as the time from the start of treatment until death from any cause.
- Progression-Free Survival (PFS) - defined as the time from the start of treatment until documented disease progression or death from any cause.
- Clinical Benefit Rate (CBR) - the proportion of patients who achieve a complete response, partial response, or stable disease.

These endpoints are standard in both clinical trials and studies based on real-world data.

Comparison of Real-World Data with Clinical Trial Results

When comparing real-world data with results from clinical trials for a given therapy, the patient cohort in real-world practice generally exhibits different distributions of characteristics compared to the clinical trial cohort (such as gender, age, ECOG status, among others). Without additional adjustment of the characteristics of the real-world population to match those of the clinical trial population, the obtained results may be influenced by these differences. For example, if the real-world population is older or has higher ECOG scores, it is highly likely to show lower progression-free survival (PFS) and overall survival (OS) outcomes compared to the clinical trial population.

Matching of characteristics between the real-world population and those of clinical trials is conducted using the Iterative Proportional Fitting (IPF) algorithm to adjust real-world data, aligning them with clinical trial data. This adjustment aims to make the progression-free survival (PFS) and overall survival (OS) curves more comparable to those observed in the

clinical trials. This method is applied to therapeutic products with sufficiently large patient cohorts in real-world data.

Based on the collected and analyzed real-world data using the IPF algorithm, results are obtained that can be included in scientific publications.

Preparation of an annual report submitted to the National Health Insurance Fund (NHIF) or Ministry of Health (MoH) (data consumption) and publication of real-world data in scientific articles in Bulgaria (publication) represent secondary use of treatment data from Bulgarian patients.

References:

NICE real-world evidence framework, Corporate document [ECD9], Published: 23 June 2022, Updated March 2024

Real-world data and real-world evidence in regulatory decision making, Report of the CIOMS Working Group XIII, The Council for International Organizations of Medical Sciences (CIOMS), Geneva 2024

Good Practice Guide for the use of the Metadata Catalogue of Real-World Data Sources, v 1.0, European Medicines Agency and Heads of Medicines Agencies, 1 September 2022

Recommendations on a Data Quality Framework for the European Health Data Space for secondary use, Deliverable 6.3, Towards European Health Data Space (TEHDAS), 26 September 2023

Real-world evidence framework to support EU regulatory decision-making, Report on the experience gained with regulator-led studies from September 2021 to February 2023, EMA-HMA

Guidance for Reporting Real-World Evidence, CADTH Methods and Guidelines, May 2023

The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). Guide on Methodological Standards in Pharmacoepidemiology (Revision 11). EMA/95098/2010.

Dang A. Real-World Evidence: A Primer. *Pharmaceut Med.* 2023 Jan;37(1):25-36. doi: 10.1007/s40290-022-00456-6. Epub 2023 Jan 5. PMID: 36604368; PMCID: PMC9815890.

Graili P, Guertin JR, Chan KKW, Tadrous M. Integration of real-world evidence from different data sources in health technology assessment. *J Pharm Pharm Sci.* 2023 Jul 17;26:11460. doi: 10.3389/jpps.2023.11460. PMID: 37529633; PMCID: PMC10387532.

"Guiding Principles of Real World Data Used to Generate Real World Evidence (Trial)" China, April 2021

Methodological Guideline for Quantitative Evidence Synthesis: Direct and Indirect Comparisons, Adopted on 8 March 2024 by the HTA CG pursuant to Article 3(7), point (d), of Regulation (EU) 2021/2282 on Health Technology Assessment

Practical Guideline for Quantitative Evidence Synthesis: Direct and Indirect Comparisons, Adopted on 8 March 2024 by the HTA CG pursuant to Article 3(7), point (d), of Regulation (EU) 2021/2282 on Health Technology Assessment