Table S1. Overview of the available quantitative methodologies to assess benefit-risk balance, as given by the CHMP [14].

Method	Advantages	Limitations	References
Discrete event simulation	Detailed simulation based on differential equations and continuous variables. Ability to handle multiple assumed characteristics and simultaneously assess impact of multiple effects on health economics.	Complexity, complicate adaptability, lack of transparency and validation. Risk of underestimation in case of prediction limited to short term effects. No clear assessment of unfavourable effects.	[14,54,55]
System dynamics	Account for non-linearity using feedback and time-delays, both short and long term. Possibility of input data from different sources.	No recorded use in drug development. Focus on pharmacoeconomics. No consideration of (weighted) unfavourable effects, such as ADEs.	[14,56]
Bayesian beliefs networks	Network of nodes representing risks, benefits, observations and assessments, connected by conditional arrows, which input probabilities result in probability distribution for all nodes. Inclusion of both objective data and subjective expert opinion. Visualisation of effect of factors on each other.	Requires structural similarity across cases, which in BR might only be appropriate for similar indications. Probability input as a subjective element remains unsupported. Uncertainty of indirect effects introduces bias in their impact on the outcome.	[14,57]
Evidence-based BR model	Model visualised as a set of scales, including the benefit 'box' with efficacy, including responder rate and evidence and the risk 'boxes', for each ADE, with seriousness, frequency and evidence. The method correlate to EMA's definition, as the first two criteria of either box are (un)favourable effects and the third includes uncertainty of effects.	Simplified multi-criteria model with limited (three) criteria each. There is no application supporting the translation of effects into one unit. Preference weights are not accounted for.	[14]
Incremental net health benefit	Incremental net health benefit is the difference between unfavourable effects and favourable effects derived from the treatment options, where all effects are normalised into one unit. The method is transparent and theoretically sound, including uncertainty and extrapolation in time.	Although this is a version of a multi-criteria model, such as MCDA, translation to a single unit requires another methodology (e.g., value-adjusted life years, or QALY). QALY can only be transferred to health benefit when costs are not considered, in other words the willingness to pay is infinite. Weighting of effects is also dependent on another methodology, like conjoint analysis. This methodology on itself is incomplete. Subject to bias by confounders.	[1,14,58–60]

Principle of threes	Simplified method in which only three criteria per risk/benefit are scaled with three possible outcomes (e.g., low, medium, high), benefit and risk are summed up.	Very limited in number of criteria. No weighting of the criteria.	[14]
TURBO	Simplified method in which only two criteria per risk/benefit are scaled up with five possible outcomes. Pairs of outcomes are weighted and assessed. Frequency, probability, severity and extent are included into the choices of criteria.	Very limited in terms of the number of criteria. There is no way of knowing prior to assessment which criteria to choose. Choice might be arbitrary. No theoretical basis.	[14,60]
NNT	Easy understandable measure used in the clinic, stating the number of patients required to treat one occurrence of the disease (or to have one more ADE in NNH). Patient preferences can be included using Relative Value Adjusted NNT (RV-NNT).	Limited statistical power and because of lack of preference data, misinterpretation by different risk perceptions, as well as by using the same scale without proper weighting effects. Ratio of NNT/NNH assumes independence and similar timescale.	[1,4,14,32, 33,60,61]
Contingent valuation	Benefits are translated to financial values by enquiring the prize patients are willing to pay for it.	Not focused on BR assessment	[14]
Stated preferences	Collection of methods using preference values to determine utility functions of different stakeholders. Measures e.g., the extent patients are willing to experience unfavourable effects to achieve favourable effects.	Empirical method that does not account for uncertainty or weighting. Overlaps with conjoint analysis. Gathering of individual patient data is time consuming.	[1,14,60]
Probabilistic simulation	Complementary to point estimate statistics, as it states the impact of risk and benefit as a probability distributions based on simulated random draws from study data. More precise, accounts for uncertainty in trade-offs. Can account for correlation, if suitable data is available.	Limited if using non-validated or non-representative probability distributions for simulation. Benefits or risks are not weighted, shown by the fatal adverse event in the adalimumab-study, which did not seem to affect the simulation analysis.	[1,14,60,62,63]
Bayesian statistics	Prior and posterior probabilities based on available evidence. Together describe the likelihood of an effect and its uncertainty, combined with utility function in the Bayesian approach. Methodology improves as more data are gathered, as it involves iterative learning.	Significance levels state something about the data, not about the hypotheses, so it cannot directly be included into a formal BR assessment. The model itself doesn't include multiple criteria. Mathematical models can get complex.	[14,64]

Markov model	Describes time-dependent dynamic processes, using transitions between health states and their probability distributions.	Probability data might be sparse before approval. Complex health states might be oversimplified.	[14,65,66]
QALY	Multiple dimensions are scored and weighted for preference, outcome measured in life years on population level.	Limited in uncertainty and unique (disease/patient) data representation, more focussed on health- and pharmacoeconomics. Threshold is debatable.	[14,67]
Kaplan-Meier estimation	Function of survival over time, impact measured in ratio of differences, useful in Markov models.	Limited representation of (un)favourable effects, for example in non-fatal indications. It does not account for uncertainty and cumulative probabilities can be misleading due to lack of correlation structure (e.g., competing events).	[14,68]
Conjoint analysis	Covers preferences of different stakeholders, utility weight is based on preferred trade-offs. Realistic method helpful in weighting.	Labour intensive if all stakeholders are included. Weight might not be independent from methodological decisions. Does not account for uncertainty.	[14,69]
Clinical Utility Index	Multi-attribute utility analysis with weighted trade-offs. Utility function introduces clinical meaning to the assessment. CUI is flexible over different indications and endpoints. Transparent method with possibility of sensitivity analysis.	In case of limited applicable data, complex modelling with high variability and uncertainty is required. Subjective discussion on clinically relevant factors remains unsupported. More useful for a no-go than for a go-decision	[70–72]
Decision tree	Overview of all possible outcomes with their probabilities, calculated using the branches and nodes leading to said outcome. The decision tree is a useful framework.	Too simple for complex cases. Uncertainties are only limited covered, as probabilities are often empirically determined.	[14,46]
MCDA	Multi-criteria method breaking up the problem, followed by scoring and weighted assessment of benefits and risks as most representative presentation of data. Sensitivity analysis prevents unwanted impact. Incorporates uncertainty.	Might be too comprehensive for a simple analysis. Does not account for possible correlations between endpoints. Preference value determination is accounted for in the weighting step.	[1,14,24,25,49–52,60,61]