www.advance-vaccines.eu

Accelerated Development of VAccine beNefit-risk Collaboration in Europe

Grant Agreement nº115557

D4.3 Report on appraisal of vaccine benefit-risk methodology

WP4 Methods for burden of disease, vaccination coverage, vaccine safety and effectiveness, impact and benefit-risk monitoring

V2.0
[Final]

Lead beneficiary: P95
Date: 04/11/2014
Nature: Report
Dissemination level: CO

TABLE OF CONTENTS

DOCUMENT INFORMATION

Grant Agreement Number	115557	**Acronym**	ADVANCE
Full title	Accelerated Development of VAccine beNefit-risk Collaboration in Europe		
Project URL	http://www.advance-vaccines.eu		
IMI Project officer	Angela Wittelsberger (angela.wittelsberger@imi.europa.eu)		

Deliverable	**Number**	4.3	**Title**	Report on appraisal on the vaccine benefit-risk methods
Work package	**Number**	4	**Title**	Methods for burden of disease, vaccination coverage, vaccine safety and effectiveness, impact and benefit-risk monitoring

Delivery date	**Contractual**	Month 12	**Actual**	
Status	Current version / v2.0		Draft ☐ Final x	
Nature	Report x Prototype ☐ Other ☐			
Dissemination Level	Public ☐ Confidential x			

Authors (Partner)	Vincent Bauchau (GSK), Eduoard Ledent (GSK), Bennett Levitan (J&J), Thomas Verstraeten (P95), Yuan Zhong (J&J)			
Responsible Author	Kaatje Bollaerts (P95)		**Email**	Kaatje.bollaerts@p-95.com
	Partner	P95	**Phone**	+32 485 78 96 57

Description of the deliverable	Report
Key words	Benefit-risk, Methodology, Vaccine

DOCUMENT HISTORY

NAME	DATE	VERSION	DESCRIPTION
Kaat Bollaerts (P95), Bennett Levitan (J&J),Eduoard Ledent (GSK), Zhong Yuan (J&J), Vincent Bauchau (GSK), Thomas Verstraeten (P95)	20/07/2014	1.0	First draft
Vincent Bauchau(GSK), Thomas Verstraeten (P95)	10/08/2014	1.0	Reviewed
Kaat Bollaerts (P95)	25/08/2014	1.1	Second draft
Shahrul Mt-Isa (ICL), Bennett Levitan (J&J), Zhong Yuan (J&J), Marianne Van Der Sande (RIVM), Matti Lehtinen (UTA),Tyra Krause (SSI)	07/10/2014	1.1	Reviewed
Kaat Bollaerts (P95), Bennett Levitan (J&J),Eduoard Ledent (GSK)	22/10/2014	1.2	Third draft
Zhong Yuan (J&J), Thomas Verstraeten (P95)	25/10/2014	1.2	Reviewed
Kaat Bollaerts (P95)	3/11/2014	2.0	Final version

DEFINITIONS

- Participants of the ADVANCE Consortium are referred to herein according to the following codes:
  - **EMC**. Erasmus Universitair Medisch Centrum Rotterdam (Netherlands) - **Coordinator**
  - **UNIBAS**. Universitaet Basel (Switzerland) - **Managing entity of the IMI JU funding**
  - **AEMPS.** Agencia Española de Medicamentos y Productos Sanitarios (Spain)
  - **ASLCR.** Azienda Sanitaria Locale della Provincia di Cremona (Italy)
  - **AUH.** Aarhus Universitetshospital (Denmark)
  - **CRX**. Crucell Holland BV (Netherlands)
  - **ECDC.** European Centre for Disease Prevention and Control (Sweden)
  - **EMA.** European Medicines Agency (United Kingdom)
  - **GSK.** GlaxoSmithKline Biologicals, S.A. (Belgium) – EFPIA Coordinator
  - **KI.** Karolinska Institutet (Sweden)
  - **LSHTM.** London School of Hygiene and Tropical Medicine (United Kingdom)
  - **MHRA.** Medicines and Healthcare products Regulatory Agency (United Kingdom)
  - **NOVARTIS.** Novartis Pharma AG (Switzerland)
  - **OU.** The Open University (United Kingdom)
  - **P95.** P95 (Belgium)
  - **PEDIANET.** Società Servizi Telematici SRL (Italy)
  - **PFIZER**. Pfizer Limited (United Kingdom)
  - **RCGP.** Royal College of General Practitioners (United Kingdom)
  - **RIVM.** Rijksinstituut voor Volksgezondheid en Milieu * National Institute for Public Health and the Environment (Netherlands)
  - **SP MSD.** Sanofi Pasteur MSD (France)
  - **SP.** Sanofi Pasteur (France)
  - **SSI.** Statens Serum Institut (Denmark)
  - **SURREY.** The University of Surrey (United Kingdom)
  - **SYNAPSE.** Synapse Research Management Partners, S.L. (Spain)
  - **TAKEDA**. Takeda Pharmaceuticals International GmbH (Switzerland)
  - **UTA.** Tampereen Yliopisto (Finland)
  - **WIV-ISP.** Institut Scientifique de Santé Publique (Belgium)

- **Grant Agreement.** The agreement signed between the beneficiaries and the IMI JU for the undertaking of the ADVANCE project (115557).
- **Project.** The sum of all activities carried out in the framework of the Grant Agreement.
- **Work plan.** Schedule of tasks, deliverables, efforts, dates and responsibilities corresponding to the work to be carried out, as specified in Annex I to the Grant Agreement.
- **Consortium.** The ADVANCE Consortium, comprising the above-mentioned legal entities.
- **Project Agreement.** Agreement concluded amongst ADVANCE participants for the implementation of the Grant Agreement. Such an agreement shall not affect the parties' obligations to the Community and/or to one another arising from the Grant Agreement.

- **List of abbreviations:**

ACBC	Adaptive choice based conjoint
ADVANCE	Accelerated development of vaccine benefit-risk collaboration in Europe
AF	Aetiological fraction
B-R	Benefit-Risk
BCoDE	Burden of Communicable Diseases in Europe
BLRA	Benefit-less risk analysis
BRAT	Benefit-Risk Action Team
BRR	Benefit-Risk ratio
CBC	Choice based conjoint
CIN	Case impact number
CIRS	Centre for Innovation in Regulatory Science
COBRA	Consortium on Benefit-Risk assessment
CPM	Confidence profile method
CUI	Clinical utility index
DALY	Disability adjusted life years
DCE	Discrete choice experiment
DES	Discrete event simulation
DIN	Disease impact number
ECIN	Exposed case impact number
EIN	Exposure impact number
GBR	Global benefit-risk
GRADE	Grading of Recommendations Assessment, Development and Evaluation
HALE	Healthy life expectancy
HTA	Health Technology Assessment
IMI	Innovative Medicines Initiative
INHB	Incremental net health benefit
IRBR	Incremental benefit-risk ratio
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
MCDA	Multi-criteria decision analysis
MCE	Minimum clinical efficacy
MPES	Multi-parameter evidence synthesis
MTC	Mixed treatment comparison
NEAR	Net Efficacy Adjusted for Risk
NEPP	Numbers of events prevented in your population
NHB	Net health benefit
NITAG	National Immunization Technical Advisory Group
NNH	Number needed to harm
NNT	Number needed to treat
NNTB	Number needed to treat for one person to benefit
NNTH	Number needed to treat for one person to harm
NNV	Number needed to vaccinate
PAF	Population attributable fraction

PAR	Population attributable risk
PBRER	Periodic benefit-risk evaluation reports
PhRMA	Pharmaceutical Research Manufacturers of America
PICO	Population, Intervention, Comparison, Outcome
PIN	Population impact number
PIN-ER-t	Population impact number of eliminating a risk factor
PrOACT-URL	Problem, Objectives, Alternatives, Consequences, Trade-offs, Uncertainty, Risk and Linked decisions framework
PROTECT	Pharmacoepidemiological Research on Outcomes of Therapeutics
PSA	Probabilistic sensitivity analysis
PSM	Probablistic simulation method
QALY	Quality-adjusted life years
QoL	Quality of life
QTWIST	Quality adjusted time without symptoms and toxicity
RBAT	Risk-benefit acceptability threshold
RRM	Random regret minimization
RUM	Random utility maximization
RV	Relative utility value
SABRE	Southeast Asia Benefit Risk Evaluation initiative
SMAA	Stochastic multi-criteria acceptability analysis
STIKO	Standing Committee on Vaccination
STM	State transition model
TURBO	Transparent Uniform Risk/Benefit Overview
UMBRA	Unified Methodologies for Benefit-Risk Assessment

EXECUTIVE SUMMARY

Most benefit-risk methodologies have been developed to assess the benefit-risk balance of (therapeutic) drugs or devices. Several recent reviews of existing benefit-risk methodologies exist. Therefore, we did not perform a formal systematic review of the literature on benefit-risk methodology. Instead, we revisited all methods described in the systematic reviews by the ISPOR Risk-Benefit Management Working Group and by PROTECT and appraised their suitability for the benefit-risk assessment of vaccines, with the exception of the estimation techniques described in the PROTECT review. Additionally we evaluated the modelling techniques and evidence-synthesis techniques most commonly used in Health Technology Assessment.

In this report, we first describe the vaccine specificities. Then, we describe a literature review of benefit-risk methods, evaluating (1) qualitative or semi-quantitative frameworks, (2) benefit-risk measures, (3) composite health measures, (4) quantitative benefit-risk frameworks, (5) modelling approaches commonly used in Health Technology Assessment, (6) parameter estimation and uncertainty and (7) preference elicitation techniques.

In the final section, we recommend the use (or development of a vaccine-specific) of qualitative or semi-quantitative frameworks while exploring the use of evidence grading methodology, the use of a toolbox of selected quantitative methodologies, the quantification of various sources of uncertainty and the investigation of how to adapt common preference elicitation techniques to the field of vaccination.

1. INTRODUCTION

1.1 ADVANCE project

Benefit-risk (BR) analysis are indispensable to govern decision-making regarding pharmaceutical products, whether it is the manufacturer's decision for further pharmaceutical research and development1, the regulator's decision for approval, restriction or withdrawal of the product2, 3 or the recipient's decision to take the product4. Evaluating benefits and risks is however, a complex exercise. It involves the integration of data and information from different sources, typically associated with different levels of uncertainty. In addition, measuring the benefit-risk balance is to a certain extent subjective as it involves value judgements for gaining certain benefits versus avoiding certain risks.

To allow more transparent, consistent, reproducible and communicable benefit-risk assessments of pharmaceutical products, the need for more structured approaches is generally acknowledged5. Following this need, the Benefit-Risk Methodology Project by the European Medicines Agency (EMA)2 and the "Pharmacoepidemiological Research on Outcomes of Therapeutics" (PROTECT)6 project funded by the Innovative Medicines Initiative (IMI) were both launched in 2009. Recognizing that vaccines are different from therapeutics for monitoring benefits and risks, the IMI funded the "Accelerated development of vaccine benefit-risk collaboration in Europe" (ADVANCE) project. The ADVANCE project was launched in October 2013 and brings together 200 researchers from more than 30 institutions, including the European Centre for Disease Prevention and Control (ECDC), the EMA, vaccine manufacturers, academics, regulators, public health institutes and authorities and small and medium enterprises (SMEs). The overall objective of the ADVANCE consortium is to review, develop and test methods, data sources and procedures that should feed into a blueprint of an efficient and sustainable European framework that can rapidly deliver quantitative data to support manufacturers, regulators, public health authorities, health professionals and the general public to make informed decisions regarding vaccines. The activities of the project have been grouped in seven work packages, amongst which work package 4 (WP4) on the appraisal, development and testing of methods for burden of disease, vaccination coverage, vaccine safety, vaccine effectiveness/impact and benefit-risk methodology. This report is on the appraisal of benefit-risk methodology.

1.2 Objectives and scope of the report

Most of the benefit-risk methodology has been developed to assess the benefits and risks of (therapeutic) drugs or devices. In this report, we describe the existing benefit-risk methodologies and evaluate whether they are suited (or extendible) to conduct benefit-risk analysis of vaccines, with an emphasis on the post-licensure setting and specific ADVANCE objectives (such as timeliness and integration). The ultimate objective of this review is to identify suitable methods as well as knowledge gaps and recommend further methodological developments to support the overall ADVANCE objective of developing a framework for rapid delivering of quantitative benefit-risk data to support vaccine decisions.

1.3 Structure of the report

In this report, we evaluate methods for their suitability for benefit-risk assessments of vaccines. Therefore, the features that are specific to vaccines (vaccine specificities) are described first (Section 2). Then, a literature review of benefit-risk methodologies was performed, evaluating (1) qualitative or semi-quantitative frameworks, (2) benefit-risk measures, (3) composite health measures, (4) quantitative benefit-risk frameworks, (5) modelling approaches commonly used in Health Technology Assessment, (6) parameter estimation and uncertainty and (7) preference elicitation techniques (Section 3). Finally, some concluding remarks and recommendations are formulated (Section 4).

2. SPECIFICITIES of BENEFIT-RISK ASSESSMENTS of VACCINES

A benefit-risk framework aims to provide transparency to the process of assessing benefits and risks by structuring this process and making a clear distinction between evidence and preferences. Most of the previous work on methodologies for benefit-risk assessments is related to drugs and devices (e.g. see reviews7-10). In this section, we identify the vaccine specificities that warrant special consideration when assessing the benefits and risks of a vaccine or vaccination programme. The identified vaccine specificities will guide the appraisal of the benefit-risk methodology originally developed for drugs for their use (or extendibility) to vaccines. The identified vaccine specificities do not apply to all vaccines and all vaccination usage.

A first feature that sets vaccines apart is the potential population-level impact by reducing disease transmission within the (partially) vaccinated population. Second, there are different stakeholders involved in decision-making about vaccines, with a special role to play for public health authorities. Further vaccine specificities include the limited tolerance for adverse effects, the large exposure numbers, the potential immediate risks contrasted to the long-term benefits, high levels of uncertainty, challenging preference elicitation and the importance of post-licensure observational studies. These vaccine specificities will be discussed in turn below.

Population-level impact of vaccination

The population-level impact of vaccination refers to the effect of vaccination on the entire population, including vaccinated and unvaccinated persons, and is a combination of direct and indirect effects11, 12. Direct effects refer to the effect vaccination is having on the targeted health outcome among vaccinated individuals, resulting from immunological protection rather than from altered disease transmission in the population. Indirect effects refer to the effect of vaccination resulting from altered disease transmission and are most easily assessed in unvaccinated individuals. Indirect effects are typically beneficial (herd immunity), but may in some cases be detrimental. Examples of indirect effects that are detrimental are: increased risk of varicella complications through an increased average age at infection as a result of herd immunity and risks to immuno-compromised people that come into close contact with a person

vaccinated with some live vaccines. The indirect effects of vaccination further depend on the vaccine uptake within that population, the assortative behaviour of the population and other population characteristics (such as the prevalence of the vaccine-preventable disease and population density) as well as on the infectious disease dynamics. Therefore, to assess the benefits and risks of vaccination at population-level, additional information is needed regarding the vaccine uptake, potentially other population characteristics and the infectious disease dynamics.

Different stakeholders and different perspectives

Vaccine manufacturers, regulatory authorities, public health authorities, health care providers and vaccine recipients are all stakeholders involved in decision-making concerning vaccines. Public health authorities play a prominent role in decision-making for vaccines, particularly for diseases that are highly contagious. Indeed, public health authorities recommend or even mandate vaccination in order to protect the entire population, evoking ethical issues regarding individual autonomy versus population protection. The different stakeholders might have different value judgements of the various vaccine effects (e.g. public health authorities tending to attach more importance to the indirect effects than some vaccine recipients might do).

Regarding perspectives, the distinction is often made between the individual perspective (including only benefits and risks to the vaccine candidate) and the population perspective (including all benefits and risks to the population). For vaccines, both perspectives are relevant as the potential benefits and risks are not always borne by the same individual. Indeed, unvaccinated individuals might benefit from vaccination as a result of reduced likelihood of disease exposure while not being exposed to the risks induced by vaccination. Examples of vaccination strategies that exploit indirect protection are maternal immunization (i.e. protection through transplacental transfer of maternal vaccine-induced antibodies) and cocooning (i.e. protection through exposure reduction by vaccinating close contacts of the individual to be protected), which are typically used to protect newborns. Depending on the vaccine and how it is used, the individual perspective (e.g. travellers vaccines) or the population perspective (e.g. measles, influenza) might be more or less important. In addition, there might be instances where the population perspective renders a different benefit-risk decision than the individual perspective. The extreme situation is that of disease eradication

where the direct protective effect of vaccination is completely absent, yet vaccination may be required to keep the disease from being reintroduced (e.g. polio, measles).

Limited tolerance for adverse effects

Vaccines are usually administered to otherwise healthy individuals, often very young or vulnerable. They may be administered to a large fraction of the population and vaccination is mandatory in some countries. Therefore, a high level of safety is expected from vaccines[13]

Large exposure numbers and likely detection of very rare adverse events

Many vaccines are recommended for large population groups (to reach the critical vaccination coverage) and therefore administered to very large number of persons, frequently entire age cohorts. This large number of vaccine recipients increases the power to detect very rare adverse events. Oftentimes, a single rare serious adverse event, detected in the post-authorisation era, dominates the risk profile of a vaccine.

Immediate risks, long-term benefits and high levels of uncertainty

As opposed to the risks that are often immediate or relatively short term, the benefits of some vaccines may be long-term (e.g. cancers related to human papillomavirus infection (HPV)). Furthermore, because vaccines are often licenced based on surrogate endpoints like immunogenicity due to the lack of immediate endpoints and/or low disease incidence, substantial uncertainty exists regarding the expected benefits at the time of authorisation, even for the direct effects. The uncertainty around the population-level impact of vaccination may be even larger, because impact depends on the actual implementation of the vaccination programme and is influenced by changes in the transmission dynamics of the disease targeted by vaccination. Even long after authorisation, establishing the vaccine benefits remains challenging due to factors such as heterogeneity in immune responses, waning of protection and lack of a comparable unvaccinated population. This uncertainty will challenge the feasibility of benefit-risk assessments needed in the face of more immediate or short term risks.

Challenging preference elicitation

The elicitation of preferences or value judgements is challenging for various reasons. The first difficulty is knowing who to ask. For some vaccines (e.g. travellers vaccines), both the potential benefits and risks are borne by the same individual. Hence, preferences from candidate vaccine recipients are informative. For vaccination programs aiming to reduce disease transmission within the general population, one might argue that the general population and/or public health experts play an important role in generating these preferences, which may further depend on whether the vaccine is recommended or mandated. In addition, preference elicitation for vaccination might involve surrogate decision making, which is typically invoked when the care-receiver lacks decision making capacity[14]. For vaccines given to young children, the authority to make the vaccination decision on behalf of a child usually falls to the child's parents and hence, parent's preferences are informative. For vaccines given to adolescents (e.g. HPV vaccination), one might argue that the preferences of the adolescents, parents or both are informative[15]. Finally, although patients suffering (or having suffered) from the vaccine preventable disease are typically not the ones being vaccinated, their preferences are informative as well because they are well placed to value the benefits of vaccination.

The second difficulty is knowing how to ask. Vaccines are primary preventive measures, implying that the candidate vaccine recipients are only very rarely confronted with the disease they are protected against. This distorts the perceived benefits of vaccination. On the other hand, a very low risk tolerance exists because vaccines are generally given to healthy people, typically to young children, often as part of a vaccination recommendation or mandate. The risk perception is further influenced by the public concerns about vaccines and the enhanced media attention for vaccine-related issues[16].

Importance of observational studies

The rare adverse events and long-term benefits typical to vaccines cannot be fully investigated using pre-authorisation studies, which are relatively small compared to the ultimately exposed

population and limited in follow-up time. Therefore, there is a strong emphasis on post-authorisation studies for vaccines. However, post-authorisation studies are mainly observational, making them much more vulnerable to bias and confounding (such as confounding by indication and by health-care seeking behaviour) compared to well-controlled pre-licensure studies. Vaccine post-authorisation studies might be additionally sensitive to confounding by age- and seasonal effects because vaccines are often administered following recommended age-dependent immunization schedules (e.g. childhood vaccinations) or during specific seasons (e.g. influenza). Nevertheless, post-authorisation information collected in 'real life' often complements the already available pre-authorisation information obtained through well-controlled clinical trials. Identifying appropriate exposure (vaccinated) and control (non-vaccinated) cohorts is a critical step in post-authorisation observational studies for establishing benefit-risk assessments.

Concluding remarks

The benefit-risk assessment of vaccines is very important, albeit challenging for reasons outlined above. Generally, for the benefit-risk assessment of vaccines, both the individual and the population perspective are relevant. When the population perspective is adopted, the benefit-risk methodology should also account for the potential indirect effects of vaccination. The limited tolerance for adverse effects needs to be taken into account, while dealing with the likelihood of very rare events being associated to vaccination. Furthermore, the uncertainty in the benefit-risk assessments of vaccines (being stochastic uncertainty, parameter uncertainty, heterogeneity and/or structural uncertainty) is substantial and therefore, appropriate benefit-risk methodology should be able to account for this. Because post-authorisation studies often provide important information regarding the benefits and risks of vaccines, an ideal benefit-risk methodology should allow integrating pre- and post-authorisation information. Finally, preference elicitation has particular challenges with uncertainties on who to elicit these from as well as how to do so most appropriately.

3. BENEFIT-RISK METHODOLOGY

Several recent reviews of existing benefit-risk methodology exist, all focussing on the use of benefit-risk methodology for pharmaceutical drugs and devices 7-10; of which the reviews by the ISPOR Risk-Benefit Management Working Group (2010)8 and by PROTECT (2014)9 were systematic. Therefore, we did not perform a formal systematic review of the literature on benefit-risk methodology. Instead, we revisited all methodologies described in the two systematic reviews8, 9 and appraised their suitability for the benefit-risk assessment of vaccines, with the exception of the estimation techniques described in the PROTECT review9. These estimation techniques include generic statistical techniques that are not unique to benefit-risk assessments, but were used in combination with other benefit-risk measures. Instead, we evaluated the modelling techniques and evidence-synthesis techniques most commonly used in Health Technology Assessment (HTA), some of which were also described by PROTECT. We did this because it has been recognised earlier that cost-effectiveness analyses and benefit-risk assessments share a lot of commonalities17 and notwithstanding this, the applications of (adjusted) cost-effectiveness techniques to the field of benefit-risk assessment are sparse18.

In this section, we subsequently describe and appraise (1) descriptive or semi-quantitative frameworks, (2) benefit-risk measures, (3) composite health measures, (4) quantitative benefit-risk frameworks, (5) modelling approaches commonly used in HTA, and (6) preference elicitation techniques.

3.1 Descriptive or semi-quantitative frameworks

It has been recognised that structured qualitative processes must precede quantification19. Such processes or frameworks ensure that all elements of the benefit-risk balance have been considered and rendered explicit and this to improve transparency and communication in decision-making9. Descriptive or semi-quantitative frameworks are structured stepwise processes that might include graphical and/or tabular summaries of the metrics associated with the key benefits and risks. The descriptive and semi-quantitative frameworks are to be distinguished from the quantitative frameworks, in which an overall benefit-risk score is calculated20 (Section 3.4). In this section, we will first discuss common summary tools (Section

3.1.1) and the most commonly used descriptive/semi-quantitative frameworks in benefit-risk assessment, i.e. the BRAT and PrOACT-URL frameworks (Section 3.1.2-3.1.3). Other frameworks are touched upon as well (Section 3.1.4).

3.1.1 Summary tools

Attribute trees, tabular summaries and forest plots are primary benefit-risk summary and visualisation tools. An **attribute tree** (or value tree) is a visual, hierarchic display of the key attributes or criteria relevant to the decision. A generic example of an attribute tree for vaccines is given in Figure 3.1. Attribute trees are useful to clarify the different benefits and risks, to facilitate communication and to enhance common understanding[21].

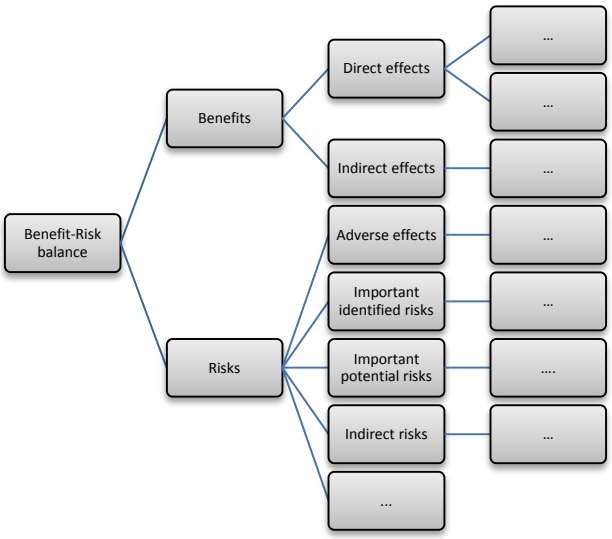


Figure 3.1. Generic example of an attribute tree for vaccines.

The **key benefit-risk summary table** (Table 3.1) and **effects table** (Table 3.2) are tabular presentations of all key benefits and risks relevant to the benefit-risk decision. The table rows match the terminal branches of the attribute tree and minimally include the titles of the benefits/risks and their reported values (and possibly units of measurements, ranges, uncertainties, treatment differences or comments). The term "key benefit-risk summary table" is used within the BRAT framework (Section 3.1.2) whereas the term "effects table" is used within the PrOACT-URL framework (Section 3.1.3), though they are conceptually very similar and are flexible as to what columns of information are included. The effects table is one of the

four benefit-risk tools recommended within the scope of the EMA benefit-risk project (in addition to the PrOACT-URL framework, MCDA and graphical displays)[2].

Table 3.1. Example of key benefit-risk summary table for triptans in migraine (from [22]).

		Outcome	Study Drug Risk (/1,000 pts)	Comparator Risk (/1,000 pts)	Odds Ratio (95% CI)
Benefits	↓ Pain	Rapid onset	271	248	1.13 (1.00,1.27)
		Headache relief	643	633	1.04 (0.94,1.15)
		Pain free response	383	349	1.16 (1.03,1.30)
		Sustained response	285	295	0.95 (0.80,1.14)
	↓ Sensitivity	Reduced sensitivity to sound and light	530	505	1.10 (0.94,1.30)
	↓ Other	Reduction in functional disability	540	480	1.28 (1.09,1.49)
		Reduction in nausea or vomiting	604	517	1.43 (1.22,1.67)
Risks	↑ Individual Risks	Transient triptans sensations	43	52	0.83 (0.61,1.14)
		CNS AEs	53	45	1.18 (0.92,1.51)
		"Chest-related" AEs	58	21	2.93 (2.04,4.20)

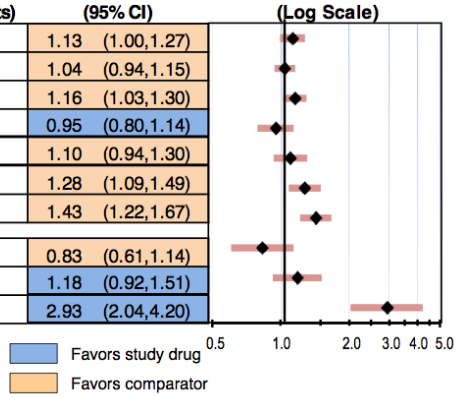


Table 3.2. Example of effects table, for Caprelsa, a drug for treatment of inoperable thyroid cancer (from [23]).

		Effect	Description	Best	Worst	Units	Placebo	300 mg
Favourable effects	Primary endpoint	Progression-free survival Hazard Ratio	Date of randomisation to the date of objective progression or death (blinded independent review)	0	1	unitless	1	0.46
	Secondary endpoints	Progression-free survival (median)	Date of randomisation to the date of objective progression or death (Weibull model)	60	0	months	19.3	30.5
		Objective Response (RECIST)	Proportion of complete or partial responders (at least a 30% decrease in the sum of the longest diameter of target lesions compared to baseline)	100	0	%	13	45
Unfavourable effects		Diarrhoea, Grade 3-4	Increase of ≥7 stools per day over baseline; incontinence; IV fluids ≥24 hrs; hospitalisation; severe increase in ostomy output compared to baseline; interfering with activities of daily living; Life-threatening consequences	0	100	%	2.0	10.8
		QTc related events, Grade 3-4	QTc >0.50 second; life-threatening signs or symptoms (eg, arrhythmia, CHF, hypotension, shock syncope); Torsade de pointes	0	100	%	1.0	13.4
		Infections, Grade 3-4	IV antibiotic, antifungal, or antiviral intervention indicated; interventional radiology or operative intervention indicated; life-threatening consequences.	0	100	%	36.4	49.8

A **forest plot** (Figure 3.2) is a simple graphical representation that complements the effects table[24]. It provides a graphical representation of the risk differences (say between a treatment and a comparator/baseline) for multiple dichotomous endpoints and their associated uncertainty, potentially using different colours for the efficacy and safety endpoints.

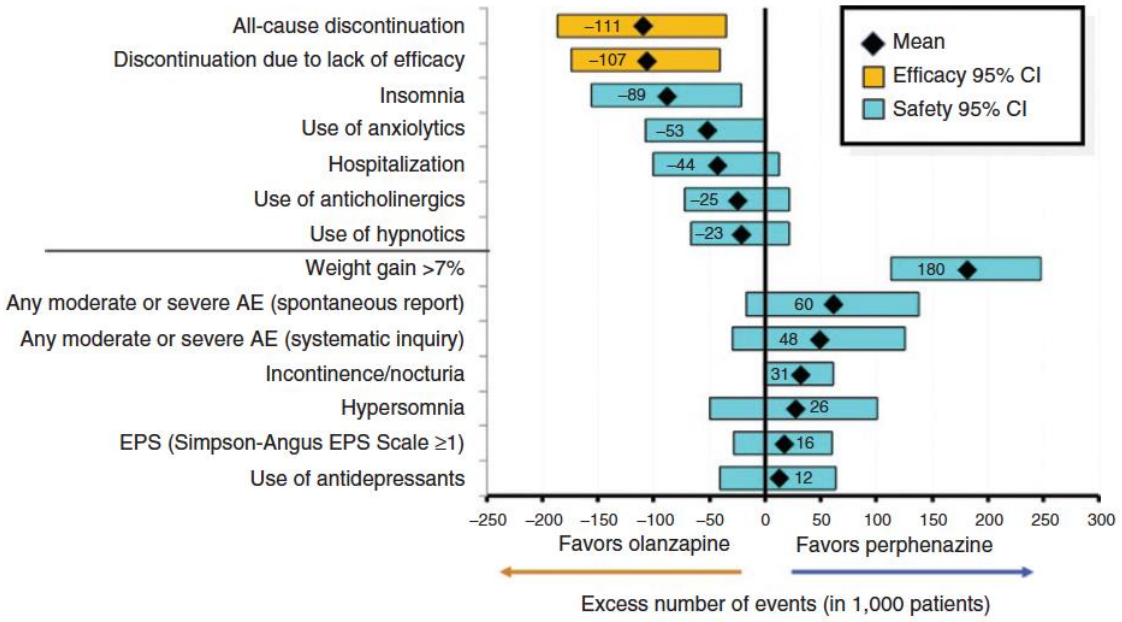


Figure 3.2. Forest plot representing risk differences for a variety of endpoints, comparing two treatments of schizophrenia. The x-axis represents the excess number of cases for a hypothetical population of 1000 patients (from [24]).

3.1.2 BRAT framework[1]

The Benefit-Risk Action Team (BRAT) framework is a general platform for benefit-risk assessment that facilitates the selection, organization, summarization, and interpretation of evidence relevant to benefit-risk decisions. The BRAT framework originated in 2005, when the Pharmaceutical Research Manufacturers of America (PhRMA) implemented a 5-year project to develop a transparent, systematic approach for pharmaceutical benefit-risk assessment. It has been iteratively developed and tested, first using hypothetical scenarios that incorporated the complexities found in real-work benefit-risk assessment[22, 25] and subsequently in a pilot program with PhRMA companies[26]. Since the pilot, the BRAT framework has been used by

[1] This section is written by Bennett Levitan (J&J) and Rebecca Noel (Eli Lilly)

various companies and has appeared in FDA Advisory Committee Meetings and periodic benefit-risk evaluation reports (PBRERs)[27-29]. It has also been tested in five case studies in the IMI PROTECT project where it was considered valuable in facilitating benefit-risk assessments. The BRAT framework formally consists of six steps (see Figure 3.3), though it has been modified and extended by individual companies that have implemented custom versions of BRAT. One highlight of the BRAT framework is its use of tabular and graphic displays to clearly depict difference between treatments in all benefits and harms included, such as key benefit-risk summary tables (Table 3.1) and risk difference forest plots (Figure 3.2).

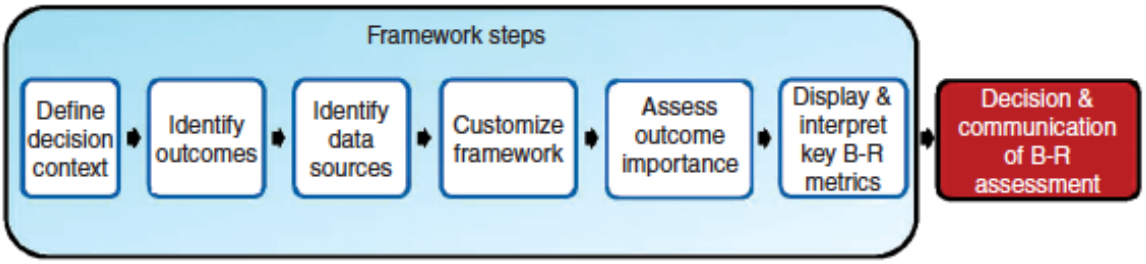


Figure 3.3. Steps in using the Benefit Risk Action Team (BRAT) benefit-risk assessment framework (from [25]).

For assessment based on the individual perspective, the BRAT framework is an appropriate tool for the structuring the benefit-risk assessment. The rationale is well described elsewhere in[22, 25, 26, 30-33] and will not be discussed here. For assessment based on the societal perspective, we will consider each step of the framework:

- Decision context: With some modifications, the BRAT decision context can easily apply to vaccines. The context typically includes drug, formulation/dose, comparator, indication, population, time horizon and perspective of decision-makers. For vaccines, the time horizon should include the duration of the exposure to the product (i.e. time frame) and the time period over which the benefit-risk events are measured (i.e. analytic horizon). Therefore, the time horizon is based on the nature of the illness (e.g. seasonal for influenza, lifetime for rubella), and the policies by which the vaccine is administered (e.g., age for vaccination).

- Identifying outcomes: Identifying outcomes can differ considerably for vaccines than for drugs. Other than rare averse events, the risks for a drug or device are generally known (either observed throughout the development program or predicted through the pharmacological mechanism) at the time of approval, while those for vaccines are often not clear. Vaccine benefit-risk assessment, at least for novel vaccines, will more often rely on post-marketing data to identify the risks, and potentially effectiveness when administered in a real-world setting. This adds an additional layer of complexity in the use of any framework at the time of vaccine authorisation. For vaccines with characteristics similar to existing vaccines, it may be possible to identify the key risks to candidate vaccine recipients at approval. If risks such as mutation to more virulent, infectious or resistant forms of the disease are a consideration, such endpoints can be added as risks to the framework.

- Identify and extract source data: Registry studies and observational data may play a critical role in the application of the BRAT and other frameworks to vaccines. However, unlike drug or device benefit-risk assessments, vaccines assessments will generally require an analytical model. These models, particularly dynamic transmission models, require a large number of parameters and assumptions regarding transmission dynamics, heterogeneity in contact patterns, policy parameters for the vaccination program, etc. Accordingly, reviewers of such assessments will have many more questions about the data, the underlying assumptions, and their implications for the validity of the final results than in a typical drug benefit-risk assessment.

- The incorporation of an analytical model has several other consequences on the use of BRAT:
    - The majority of applications of benefit-risk assessments for drugs and devices consider a small number of alternative treatments. In contrast, vaccine programs can be implemented in many ways, and therefore just comparing a vaccine to alternative vaccines may be insufficient. The assessment may also need to consider alternative policies for the implementation of the vaccine programs.

- The BRAT framework currently lends itself to comparing only a small number of alternatives and may need modification to accommodate a wide variety of vaccine programs.
- There are various sources of uncertainty to be considered in drug benefit-risk assessment – statistical uncertainty, whether the right endpoints were used, applicability of clinical trial data to real-world use, conduct and quality of the trials, implications of missing data, etc. Vaccines require considering additional uncertainties such as the disease transmission factor and the uncertainties related to vaccine policy and acceptance by individuals. These considerations introduce a new dimension of complexity for vaccine B-R assessment. Therefore, BRAT analyses and displays may need to be extended to accommodate some additional factors.
- The displays for BRAT have not been developed for cases where there is dependence between endpoints – they are not designed to show any implications of such dependency. Vaccine benefit-risk may need advancement of BRAT to account for such dependencies.

- Customize the framework: This step is important when accounting for differences between the ideal set of endpoints and those for which data is available. Framework customization applies to vaccine benefit-risk with no complications.
- Assess outcome importance: This step in BRAT is the assessment of the relative clinical impact, or weight, for the outcomes included in the assessment. It is not always required, and in many cases, qualitative clinical judgment is sufficient to render a benefit-risk decision. When needed, the assessment of weights for vaccine benefit-risk problems can be done with the same methodologies as for drug benefit-risk.

  The question of whose weights to use for vaccines is challenging. Since the decision is intended to affect public policy, the preferences of health authority decision-makers may be informative. Government policies may dictate some of the weights or at least specify some preference trade-offs amongst the benefits and harms. For a non-communicable disease, it is ultimately the patient who is taking the risks, so conceivably patient preferences are informative. However, for a transmittable disease with a major implication on community, regional, and global public health, the preferences of the general, non-patient public may be most informative. There may be cases where the

preferences of these different groups may render different decisions. In particular, a public health perspective based on health authority weights may yield different decisions than a decision based on a individual-level perspective.

- Display and interpret key benefit-risk metrics: The key benefit-risk summary table in BRAT lends itself well to vaccine benefit-risk assessment (Table 3.1); though the display may need modification to account for the potentially larger number of vaccine alternative scenarios that may need to be considered. The forest plot often used in BRAT (Figure 3.2) is limited to dichotomous endpoints, though continuous and categorical endpoints can be used if clinically meaningful threshold change can be used to dichotomize the endpoint. If a mixture of different types of endpoints are used in a vaccine assessment, some modifications will be necessary to ensure that the endpoints are comparable. There may also be some legibility issues when the magnitudes of the endpoints are not similar in size.

In summary, the BRAT framework is generally appropriate for vaccine benefit-risk assessment, but may need extensions to reliably account for the larger number of alternatives (e.g. different vaccines and different implementations of the vaccination program), the dynamic nature of the infectious disease model (e.g. the indirect effects that depend on the vaccine uptake and other population characteristics), and correlation in endpoints.

3.1.3 PrOACT-URL framework

The PrOACT-URL framework is conceptually similar to the BRAT framework. Multi-criteria decision analysis (MCDA) is a quantitative instantiation of the PrOACT-URL framework and will be discussed in Section 3.4.3. PrOACT-URL has only recently been applied to pharmaceutical B-R assessment after having been used in other fields, especially operations reseach[34]. PrOACT-URL structures the process of B-R assessment and contains the following eight steps: (1) '**Pr**oblem', (2) '**O**bjectives', (3) '**A**lternative(s)', (4) '**C**onsequences' (5) '**T**rade-off' between benefits and risks, (6) '**U**ncertainty', (7) '**R**isk attitude' of the decision maker and (8) '**L**inked decisions'. A more detailed description of the 8-step PrOACT-URL framework is given in Table 3.3 along with a description of its suitability for vaccines. Step 4 of the framework suggests creating an effects table, similar to the Key Benefit-risk Table in BRAT. The PrOACT-URL framework is one of the four methodologies recommended within the scope of the EMA

benefit-risk project[2] and was evaluated positively by the PROTECT consortium[9]. Compared to the BRAT framework, the PrOACT-URL framework contains the additional step 'Uncertainty' and 'Linked decisions', but does not specifically require customizing the framework as the BRAT framework does. The Uncertainty step in PrOACT-URL may refer to uncertainty in the trade-offs (weights) and the data (consequences). BRAT includes uncertainty in the data but does not explicitly incorporate weighting.

Similar to the BRAT framework, the PrOACT-URL framework is generally appropriate for vaccine benefit-risk assessments given that special consideration is being paid to vaccine specificities such as the time horizon, the analytic horizon, low risk tolerance and the high levels of uncertainty. Multiple effects table (commonly used within PrOACT-URL) might be needed to summarize the evidence for vaccines with a substantial public health impact (e.g one for vaccine uptake of 30%, one for an uptake of 50%, etc).

Table 3.3. Description of the PrOACT-URL framework (adjusted from [19]).

1. Problem	Determine the nature of the problem, its context and frame the problem. This includes a description of the vaccine preventable disease epidemiology, the product, the indication(s) for use and the unmet medical need. *Examples of vaccine related benefit-risk decision problems are: approval of a new vaccine, restriction of vaccinations, update of an existing benefit-risk assessment after safety signal, the decision by public health authorities to offer routine vaccination, to recommend or mandate vaccination, to change the vaccination schedule and to launch a vaccination catch-up programme.* *For vaccines, it will be important to mention the target population (e.g. neonates, infants, pregnant women, high-risk groups, elderly), the vaccination schedule (recommended age at vaccination and number of dosages), the time horizon (i.e. the duration of vaccine exposure), the analytic horizon (the time period over which the benefits and risks will be measured), the perspective (individual or societal) and the decision maker (e.g. public health authority, regulators, candidate vaccine recipient).*
2. Objective	Establish the objectives that indicate the overall purposes to be achieved (e.g. approval, restriction, update after safety signal) and identify criteria of benefits (favourable effects) and risks (unfavourable effects), that is build the attribute tree. *Criteria that might be relevant for vaccines are direct benefits, indirect benefits, (serious) adverse events, important identified risks, important potential risks and indirect risks. For all criteria, it is important to mention the relevant time window for observation.*
3. Alter-natives	Identify the options to be evaluated against the criteria. *For vaccines, there are many alternatives to consider. These alternatives include no vaccination, the use of an alternative vaccine, the use of other preventive measures and other vaccination implementations.*
4. Conse-quences	Describe how the options perform on the different criteria (i.e. the magnitudes of all effects, their desirability or severity, their incidence). Create the effects table. *The information to be included in the effects table might come from various sources, including a.o. clinical trials, epidemiological studies, observational database analyses and infectious disease models.*
5. Trade-offs	Assess the balance between favourable (benefits) and unfavourable (risks) effects (i.e. clinical judgement and rationale). *Depending on the benefit-risk decision to be taken, the preferences from candidate vaccine recipients, the general population, public health experts and/or patients are informative (see also Section 2).*
6. Uncertainty	Assess the uncertainty associated with the effects (e.g. statistical uncertainty, bias and representativeness of the studies, correlates of protection). Consider how uncertainty affects the balance by conducting sensitivity analyses and scenario analyses on the model.
7. Risk tolerance	Judge the relative importance of the decision maker's risk attitude for this product and indicate how this affects the balance reported in 5. *It is important to consider whether vaccination is recommended or mandated.*

8. Linked decisions	Consider the consistency of this decision with similar past decisions, and assess whether making this decision could have an impact on future decisions.

3.1.4 Other frameworks

Currently, PrOACT-URL and BRAT are the frameworks most commonly used in pharmaceutical and regulatory science. Other pharmaceutical frameworks have been developed, which we will only briefly refer to. For a more in depth discussion of the frameworks developed by regulatory and industry, we refer to Noel[19]. The Centre for Innovation in Regulatory Science (CIRS) engaged a consortium of regulators from four countries (i.e. Health Canada, Australia's Therapeutic Goods Administration, Swissmedic and the Singapore Health Science authority – the **CASS** group) to develop a common framework, referred to as the **COBRA** framework (Consortium on Benefit-Risk assessment). A similar initiative to harmonize regulatory activity was taken by the regulatory agencies of China, Indonesia, Malaysia, Philippines, Singapore, South Korea & Taiwan (the Southeast Asia Benefit Risk Evaluation initiative - **SABRE**). Also the **US FDA** developed a framework of its own, including the following steps: analysis of condition, current treatment options, benefits, risks and risk management. In addition, CIRS established the Unified Methodologies for Benefit-Risk Assessment (**UMBRA**) to provide a platform for the coordinated development of benefit-risk methodologies that can be used internationally during drug development and regulatory review[35].

Apart from the pharmaceutical research and regulatory oriented frameworks, the German National Immunization Technical Advisory Group (NITAG), called the Standing Committee on Vaccination (**STIKO**), developed a decision framework or Standard Operation Procedure (SOP) to guide decisions related to vaccine recommendations[36]. The key questions addressed in this SOP fall within five categories: (1) pathogen characteristics, (2) characteristics of the target disease, (3) vaccination characteristics, (4) vaccination strategy and (5) implementation of the recommendation. Based on these questions, a benefit-risk assessment is conducted. The frameworks further utilizes existing tools: formulation of questions following the PICO method (Population, Intervention, Comparison and Outcome); assessment of the quality of individual studies following the Cochrane risk of bias tool[37]; assessment of the quality of body of evidence following the GRADE system (Grading of Recommendations Assessment, Development and Evaluation)[38] and assessment of the quality of systematic reviews following the AMSTAR methodology[39]. The **framework of population impact analysis** has been proposed to assess the impact of an intervention or risk factor on a local population based on systematic

reviews of public health literature[40]. The framework involves the following steps: (1) ask the right question (using PICO-T with T for type of evidence required), (2) collect the evidence, (3) understand, synthesise and appraise the information and (4) use the information in policy making. The authors suggest using Population Impact Measures (Section 3.2.4) to collect the evidence. The latter two frameworks are developed to support vaccine implementation decisions and therefore, contain cost-effectiveness related elements. Ashby & Smith argue that a Bayesian approach offers a natural framework for medical decision-making[41]. They described the following general structure for medical decision-making: (1) who is the decision-maker, (2) possible actions, (3) uncertain consequences, (4) sources of evidence and (5) utility assessments. Finally, two other frameworks developed for non-pharmaceutical benefit-risk decision making were mentioned by PROTECT[9], being the BRAFO framework for benefit-risk analysis for foods[42] and the OMERACT 3 x 3 framework for assessing outcome measures in rheumatology[43].

3.1.5 Concluding remarks

Currently, BRAT and PrOACT-URL are the frameworks most commonly used in pharmaceutical and regulatory science, with the FDA referring to the BRAT framework[27-29] within its documentation and the EMA to (the effects table of) the PrOACT-URL framework[23]. Both frameworks are generally appropriate for vaccine benefit-risk assessments given that special consideration is paid to the vaccine specificities (Sections 3.1.2-3.1.3). The use of tabular and graphic displays to summarize the benefits and risks (e.g. forest plot, effects table, key benefit-risk summary table) might need to be adjusted to account for the dynamic nature of vaccine-preventable infectious diseases (e.g. increasing indirect benefits with increasing vaccine uptake). Finally, although cost-effectiveness analyses are out of the scope of the ADVANCE project, frameworks to support the vaccine implementation decisions (in particular, the STIKO framework[36]) contain elements that are worth considering for benefit-risk assessments of vaccines. Particularly, because (post-marketing) benefit-risk assessment of vaccines relies on heterogeneous sources of data (including registries and observational data), it might be interesting to take into account the strength of the evidence in a structured way (e.g. using the GRADE system[38] as done within the STIKO framework[36]).

3.2 Benefit-risk measures

Benefit-risk measures are metrics that are used in a benefit-risk assessment. They encompass a broad range of metrics, some of them explicitly trading off benefits and risks. In total, we discuss 21 benefit-risk measures. We subsequently describe (1) numbers needed to treat and variants thereof, (2) benefit-risk measures based on differences in benefits and risks and (3) benefit-risk measures based on ratios of benefits and risks.

3.2.1 Numbers needed to treat and variants thereof

A measure with an intuitive clinical interpretation is the '**number-needed-to-treat**' (NNT), which is to be interpreted as 'the average number of patients who need to be treated to prevent one additional unfavourable event compared to a control treatment' [44]. The NNT applies to dichotomous outcomes and is defined as the reciprocal of the absolute risk reduction (ARR) or

$NNT = 1/ARR = 1/(P_C-P_T) = 1/\Delta P$,

where P_C is rate of an unfavourable outcome event rate (%) in the control group and P_T is the rate (%) in the treated group.

When the health outcome is related to safety, this measure is often called '**number-needed-to-harm**' (NNH), or

$NNH = 1/ARR = 1/(Q_T-Q_C) = 1/\Delta Q$,

where Q_C is the rate of adverse events (%) in the control group and Q_T is the rate of adverse events (%) in the treatment group. The reason to take the reciprocal of the absolute risk reduction lies in its interpretation. Indeed, the NNT –NNH is rooted in statistical theory, with NNT-NNH being characterized as the expected value of the geometric distribution, which is a discrete waiting time distribution[45]. The geometric distribution describes the total number of trials that must be undertaken before the first 'success' (with success probability p) is reported and its expected value (average) is 1/p, the reciprocal of the success probability. The NNT is likewise a reciprocal, but of a difference in probabilities. However, taking the reciprocal of a difference in probabilities results in undesirable statistical and mathematical properties[46,47]. In addition, the interpretation of the confidence interval for the NNT is not straightforward in case the ARR is not statistically significant[48]. Then the confidence interval (CI) of the ARR includes zero, implying that the corresponding CI for the NNT must include infinity (∞). In this case,

Altman[48] suggest a (NNH: # to ∞; NNT: # to ∞) representation of the CI. To illustrate this, assume that the 95% CI for the ARR equals [-5%, 25%]. Then, the CI of the corresponding NNT-NNH is represented as [NNH: ∞ to 20; NNT: 4 to ∞], which emphasizes that ∞ is included in the CI. To evaluate the benefit-risk balance of a product, NNT and NNH are compared with NNT < NNH indicating a positive benefit-risk balance and NNT > NNH a negative benefit-risk balance. However, this comparison implies that equal importance is attached to benefit and harm. Furthermore, the NNT-NNH comparison lacks the ability to account for multiple benefits and risks and only applies to dichotomous outcomes.

Several extensions to NNT have been proposed to address these limitations. To allow differential weighting of benefit and risk, Guyatt[49, 50] proposed to add the **relative utility value** (RV) to the NNH calculation, with RV given by

RV = (1-utility of AE)/(1-utility of disease of interest).

Utility u is defined as the numeric representation of the patients' preference for a specific outcome, with $u = 0$ representing 'death' and $u = 1$ representing 'perfect health'. RV is then interpreted as the value of avoiding an adverse event relative to avoiding the disease of interest. NNH adjusted for the relative utility values is

RV-NNH = 1/(RV* ΔQ)= (1/RV) * NNH.

Then, to evaluate the benefit-risk balance of product accounting for the preferences of avoiding an adverse event relative to avoiding the disease of interest, NNT and RV-NNH are compared. To account for multiple (k) adverse events, the RV-NNH approach is readily extended as

RV-NNH = $1/\left[\sum_{i=1}^{k} RV_i(Q_{ti} - Q_{ci})\right] = 1/\left[\sum_{i=1}^{k} RV_i(\Delta Q_i)\right]$

where Q_{Ci} is the rate of adverse event i (i = 1,2,…k) in the control group (%) and Q_{Ti} is the corresponding rate (%) in the treatment group[50, 51]. Again, the benefit-risk balance is assessed through comparing NNT with RV-NNH, where RV-NNH is now the reciprocal of a weighted sum of absolute risk reductions. The (bad) statistical properties associated with NNT carry forward to the RV-NNH. Furthermore, although RV-NNH is an extension of NNH, the 'waiting time' interpretation associated with NNH-NNT is lost and as such, also the reason to take the reciprocal.

Further modifications to NNT have been proposed. Riegelman[52] introduced the **utility and time adjusted NNT** (UT-NNT) to adjust for differences in timing of the health outcomes. The modified NNT uses life expectancy and time-discounting and measures the number needed to

treat in order to produce one additional year of quality-adjusted life at present value. Similar as RV-NNH, the UT-NNT has bad statistical properties and lacks a clear (waiting time) interpretation.

Schulzer and Mancini[45] introduced the concept of **NNT for an 'unqualified success'** (treatment success without treatment-induced adverse event, NNT_{us}) and **NNH for an 'unmitigated failure'** (treatment failure with treatment-induced side effects, NNH_{uf}) to use for therapies that are associated with severe treatment-related adverse events. These numbers are typically calculated under the assumption of independence between treatment success and induction of an adverse event, although they can be adjusted for use when treatment success and induction of an adverse event are correlated, which is particularly relevant when the therapeutic window of the treatment is narrow[45, 53]. Similar to NNT, these modified measures can only accommodate one favourable and one adverse endpoint (or combined endpoints) and they have bad statistical properties. However, and unlike RV-NNH and UT-NNH, these modified measures do have a clear 'waiting time' interpretation.

The **number needed to vaccinate**[54, 55] is defined as 'the number of people needed to vaccinate in order to prevent one event of disease per year' and is obtained as

NNV = 1/(inc x VE),

with *inc* being the annual incidence among the unvaccinated and VE being the vaccine effectiveness. This is equivalent to the reciprocal of the annual absolute risk reduction, since VE measures the relative risk reduction. Additional measures have been proposed to support the estimation of the economical costs associated with the implementation of the intervention; 'numbers needed to target for vaccination', 'the number of vaccine doses needed' and 'vaccine cost' [54]. The NNV has been criticized because it only takes into account the direct protective effects of vaccination and ignore the indirect effects generated through herd immunity[55].

Impact numbers allow the assessment of the wider impact of a treatment[56, 57] or risk factor[58] on the population (and not only on those actually treated or exposed).

The **disease impact number** (DIN) and the **population impact number** (PIN) are NNT-like measures that can be used to assess the impact of an intervention in a population. The DIN is defined as 'the number of those with the disease in question among whom one event will be prevented by the intervention' and is given by

$DIN = 1/((I_u - I_e) \times P_{e|d})$,

where I_e and I_u are the event rates (%) among the exposed and unexposed and where $P_{e|d}$ is the prevalence of exposure to intervention among the diseased[57]. The PIN is defined as 'the number of those in the whole population among whom one case is prevented by the intervention'. It is calculated as

$$PIN_{interv} = 1/((I_u - I_e) \times P_{e|d} \times P_d) = 1/((I_u - I_e) \times P_e)$$

with P_d being the disease prevalence (%) and P_e the exposure prevalence (%) within the whole population. It is clear that for every intervention, disease and population impact numbers are higher than NNT. The DIN will be much higher than the NNT (and hence unfavourable) in case only a small proportion of the diseased population is having access to the intervention (e.g. because it is cost intensive, technically demanding, often contra-indicated). The PIN combines the probability of intervention success, the accessibility of the intervention and disease occurrence. Highly accessible and effective interventions for a common disease will have a favourable PIN. The DIN and PIN have a sound population impact interpretation, provided that the absolute risk difference, exposure prevalence and disease prevalence apply to the same population[56, 57]. The DIN and PIN suffer from the same statistical issues as the NNT.

The **population impact number** (PIN), the **case impact number** (CIN), the **exposure impact number** (EIN) and the **exposed case impact number** (ECIN) are proposed to measure the impact of a risk factor on the population[58]. These impact numbers are interpreted as 'the number of people within a given population among whom one case is attributable to the risk factor'. The PIN, CIN, EIN and ECIN differ with respect to the population they are referring to; the PIN refers to whole population; the CIN to the population of cases; the EIN to the population of exposed and the ECIN to the population of exposed cases. The impact numbers are the reciprocals of epidemiological measures; the PIN is the reciprocal of the population attributable risk (PAR); the CIN of the population attributable fraction (PAF); the EIN of the absolute risk reduction (ARR) and the ECIN of the aetiological fraction (AF). The PIN, CIN, EIN and ECIN are all NNT-like measures and hence, suffer from the same statistical issues as the NNT.

An impact measure related to the PIN is the **Numbers of events prevented in your population** (NEPP)[59]. Starting from the number of people in that population who are eligible for treatment (NTP), or NTP = population size x $P_{e|d} \times P_d$.

The NEPP is then obtained as

NEPP = NTP * ARR = population size/PIN,

which is straightforwardly interpreted as 'the numbers of events prevented in the population by the intervention'. Another related measure is the **population impact number of eliminating a risk factor** (PIN-ER-t), which intends to measure 'the potential number of disease events prevented in your population over the next t years by eliminating a risk factor' [60].The PIN-ER-t is calculated as

PIN-ER-t = population size x I_t x PAR = population size x I_t/PIN,

where I_t is the incidence of the outcome in the whole population over *t* years. Unlike the other impact measures, the NEPP and the PIN-ER-t do not take the reciprocal of differences in probabilities, avoiding the associated undesirable statistical and mathematical properties[46,47].

3.2.2 Differences in benefits and risks

The **minimum clinical efficacy** (MCE) determines the minimal therapeutic benefit for a treatment at which the treatment is worth administering[61]. The minimal clinical efficacy of a new treatment compared to a standard treatment (control) is

$E_t >= E_c + (Q_t - Q_c)/P_0$,

where Q_t and Q_c represent the adverse event rate (%) in the new treatment and control group, P_0 represents the event rates (%) in the untreated population and E_t refers to the efficacy of the treatment *i* (*i*=1,2) relative to no treatment or

$E_i = (P_0 - P_i)/P_0$,

where P_i represents the event rate in the population receiving treatment *i* (either new treatment *t* or control treatment *c*). It is very easy to show that the MCE comes down to comparing absolute risk differences in benefits and risks. Indeed, from above two equations it follows that

$(P_0 - P_t)/P_0 >= (P_0 - P_c)/P_0 + (Q_t - Q_c)/P_0$,

which readily simplifies as

$P_c - P_t >= Q_t - Q_c$.

As such, MCE makes an analogous comparison as NNT-NNH without taking the reciprocal of the absolute risk differences. Analogous to RV-NNH, the MCE has been extended with **relative utility values** (RV) to account for multiple adverse events and the relative importance of the adverse events compared to the disease of interest or

$E_t >= E_c + ((Q_{C1}-Q_{T1}) \times RV_1 + \ldots + (Q_{Ci}-Q_{Ti}) \times RV_i + \ldots (Q_{Ck}-Q_{Tk}) \times RV_k))/P_0$.

Net Health Benefit (NHB) and **Incremental Net Health Benefits** (**INHB**) are commonly used in Health Technology Assessment[62]. NHB is the difference between the sum of the benefits and the sum of the risks of a treatment, with all outcomes expressed using the same metric. This is done by quantifying benefits and risks using available clinical or post-marketing surveillance data and attaching preferences to each outcome. **Incremental Net Health Benefit** (**INHB**) is then the difference between the NHB of the treatment of interest and the control treatment, or $NHB_T - NHB_c$. The INHB can also be calculated as the difference in benefits from the new treatment compared with a standard, adjusted for the differences in risks or

$$\text{INHB} = \left[\sum_{i=1}^{k} E_{ti} - \sum_{i=1}^{k'} E_{ci}\right] - \left[\sum_{j=1}^{l} R_{tj} - \sum_{j=1}^{l'} R_{cj}\right],$$

where E refers to the expected benefits and R to the expected risks expressed using the same metric. A positive INHB indicates that the net benefits of the new treatment are positive compared to the control treatment and favours the new treatment. The use of INHB has been advocated for quantitative benefit-risk assessment of drugs[63]. Typically, the NHBs are expressed using QALY's (e.g.[64] [65]), but other metrics can be used as well, such as Life Years (e.g. [66]).

3.2.3 Ratios in benefits and risks

The **Benefit-risk ratio** (BRR) is simply the ratio of a given benefit (e.g. with rate P) and a given risk (e.g. with rate Q), or

$BRR = P/Q$.

The BRR has an easy interpretation and is suited in case the benefit-risk profile is dominated by a single benefit and a single risk. Similar to the simple NNT-NNH comparison, the BRR lacks the ability to account for multiple benefits and risks, and implies that equal importance is attached to benefit and risk.

The **Net Efficacy Adjusted for Risk (NEAR),** which is the relative risk (RR NEAR) or odds ratio (OR NEAR) of treatment success without adverse event of a new treatment compared to a control treatment[67]**.** It can be considered the 'ratio-variant' of the NNT for an 'unqualified success'[45]. The NEAR is derived from a 2x2 table of treatment success and treatment-induced side effects, typically assuming independence between treatment success and induction of

adverse events. Just like NNT for 'unqualified success', the NEAR is mainly relevant for therapeutic drugs with more common and more serious adverse events. The NEAR compares a single risk with a single benefit and assumes equal weighting of benefit and risk.

Similar in spirit as the NEAR, Chuang-stein developed the **global benefit-risk** (GBR) methodology to describe the benefit-risk of an intervention by comparing differences in GBR scores using asymptotic statistical distributions. The method consists of creating benefit-risk outcome categories (e.g. individual experiencing (a) benefit without adverse events, (b) benefit with adverse events, (c) no benefit and no adverse events, (d) no benefit and no adverse event and (e) serious adverse event leading to withdrawal/death) and making statistical comparisons of treatment groups. Using probabilities of belonging to a certain category and weights assigned to each category, three measures were proposed; linear score, ratio score and composite ratio score. The basic principle is to discount benefits by the presence of untoward safety experiences according to pre-specified levels at the individual patient level. The method is mainly useful for clinical data. The method incorporates weights, but is still limited to comparing a single benefit with a single risk (or composite measures).

The **incremental risk-benefit ratio** (IRBR) is analogous to the incremental cost-effectiveness ratio (ICER) used in Health Technology Assessment. The IRBR is simply the ratio of the difference in risk to the difference in benefit comparing the new treatment with a standard treatment. In the case of comparing a single benefit with a single risk, the IRBR is equal to the ratio of NNT to NNH or

$IRBR = (Q_T - Q_c)/ (Pc - P_T) = NNT/NNH,$

where Q_T and Q_c are adverse event rates (%) and P_T and P_c are the unfavourable outcome event rate (%) in the treatment and control group. The interpretation of the IRBR is the number of serious adverse event incurred for every efficacy event. Similar to the INHB, the IRBR can be extended to account for multiple events. Similar as the NNT, the IRBR has undesirable statistical and mathematical properties. The ratio will approach infinity when the denominator approaches zero. These properties have been extensively discussed for the ICER, the cost-effectiveness analogue of the IRBR[68].

3.2.4 Concluding remarks

The benefit-risk measures were divided into (1) numbers-needed-to-treat (NNT) like measures, (2) differences in benefits and risk and (3) ratios of benefits and risks. The NNT has been valued for its intuitive interpretation by clinicians and policy makers (i.e. 'the effort that must be spent in order to accomplish a treatment target'). However, the variance estimation of the NNT (and all its variants taking the reciprocal of differences) is problematic, implying that NNT is not a good measure to reflect uncertainty. The NNT-measures have their virtue in easy communication.

The impact measures DIN, PIN, CIN, EIN, ECIN are all NNT-like concepts, allowing the assessment of the wider impact of a treatment or risk factor on the population. Similar to the NNT, their variance estimation is problematic. Impact measures are developed to support resource-allocation[40]. As such impact measures are generally more suited for the purpose of evaluating cost-effectiveness rather than for benefit-risk assessment. It is also highly questionable whether e.g. the incidence of disease or the prevalence of exposure to the intervention should be taken into account for the benefit-risk assessment of a given medicine[69].

For preventive measures, population impact matters. Indeed, the expected benefit of a preventive measure depends on the effectiveness of the preventive measure and on the incidence of the preventable disease, as properly reflected by number-needed-to-vaccinate[54]. For some vaccines, it is also important to measure impact defined as the overall effect of the vaccination programme on the (partially) vaccinated population (including direct and indirect effects)[12,55].

Measures such as (relative-value adjusted) minimum clinical efficacy (MCE and RV-MCE) and (incremental) net health benefit (NHB and INHB) are based on differences in benefits and risks. These measures do not suffer from the undesirable statistical properties of the NNT-like measures. Particularly, the measures of (I)NHB may be suitable for benefit-risk assessment, because they allow integrating multiple benefits and risks, as well as value judgements.

Ratios of benefits and risks have been proposed as well for benefit-risk assessment to provide a relative measure between two outcomes or set of outcomes. However, and similar to the NNT-like measures, complex or undesirable (or sometimes undefined) statistical properties result when comparing treatment options using ratios (e.g. IRBR).

3.3 Composite Health Measures

Composite health measures combine different health outcomes (morbidity and mortality) into a single commensurable score. By far, the most common composite health measures are QALYs (typically used in health economics) and DALYs (typically used in public health)[70, 71]. They use validated methods to determine preferences and are used as metrics in some benefit-risk measures described earlier (i.e. IRBR, INHB).

3.3.1 QALY

Quality-adjusted life years (QALYs) represent the number of life years spent in discrete health states (i= 1,…k)[72]. QALYs, in their most simple form, are calculated as

$$\text{QALY} = \sum_{i=1}^{k} q_i \, t_i,$$

where q_i represent the Quality of Life (QoL) associated with health state i and t_i the time spent in this state. The q_i 's are also referred to as 'health utilities' and represent the quality of life (encompassing both physical and mental dimensions) enjoyed by individuals in health state i. They are elicited from the general population or from groups of patients, typically using elicitation techniques such as time-trade-off and standard gamble. The utilities produced represent the valuations attached to each health state with zero being equivalent to death and one representing perfect health state. Negative QALYs are possible as well. However, the utilities for short-term illnesses in young children are virtually non-existent and the appropriate methodology for obtaining them among children is subject to debate[73]. QALYs can be calculated with time discounting (the further the events in the future, the less heavily they are weighted) and age weighting (to recognise the added social value of people in their middle years) and the choice of the time discount and age weights strongly influences the results. For benefit-risk assessments, time discounting might be considered (3% annual discount rate is often used), whereas age weighting seems inappropriate.

QTWIST is an extension of QALY developed for the application to cancer treatments[74]. QTWIST is obtained by dividing the survival time of a patient into discrete health states (time with toxicity effects, time without toxicity and disease and time from relapse to death), to which different utilities are attached.

3.3.2 DALY

Disability-adjusted life years (DALYs) represent the number of healthy life years lost due to a certain disease or condition[75, 76]. DALYs are calculated by adding the adjusted number of years lived with disability (YLDs) to the number of years of life lost due to premature death (YLLs). Basically, DALYs are calculated as

$$\text{DALY} = \sum_{i=1}^{k} d_i \, t_i + E_{(age\ at\ death),}$$

where d_i represents the disability associated with health state *i* and t_i the time spent in this state and where E is the residual life expectancy at age of death. The disability weights d_i indicate to which extent the health state *i* reduces the patient's physical capacity, ranging from zero (perfect health) to one (worst possible health state). The disability weights tend to be based on a universal set of standard weights based on expert opinion, with the Global Burden of Disease (GBD) disability weights being the most commonly used ones. DALYs are also used within the BCoDE (Burden of Communicable Diseases in Europe) study to calculate the burden of infectious diseases in the European member states[77]. There is a debate on the validity of the DALYs. It is argued that the DALY method is problematic because it uses the residual life expectancy at age of death[78]. This feature might cause a life extending intervention to increase the disease burden. Indeed, if the years of life gained (of low quality) due to the intervention are lower than the additional years of life lost due to the patient living longer (and hence having a higher residual life expectancy). This will only happen if older patients are treated with an intervention that extends life with a limited amount of time and of poor quality. Similar as QALYs, the DALYs can be calculated with age weighting and time discounting.

3.3.2 HALE

Healthy life expectancy (HALE) is a metric developed by the WHO that connects life expectancy and good health[79]. HALE is the average number of years that a person can expect to live in full health by taking into account years lived in less than full health. This measure is less commonly used.

3.3.4 Concluding remarks

Vaccination programmes can be evaluated by estimating the QALYs gained or DALYs averted. However, these evaluations typically do not take into account adverse events[80]. Both DALYs and QALYs are reasonable choices for benefit-risk assessment of vaccines, although not without issues. QALYs have already been used for benefit-risk assessment as the common metric for Incremental Net Health Benefits (e.g.[64] [65]). The composite health measures could be further tailored to the specificities of benefit-risk assessment. A systematic overview of QALYs and DALYs is given in Table 3.4.

Table 3.4. Systematic overview of QALYs and DALYs.

QALYs	DALYs
Combines quantity of life with quality of life.	Combines morbidity and mortality
Requires time spent in the different health states and the associated health utility.	Requires time spent in the different disabled health states, the associated disability weight and the residual life expectancy at age of death.
Utilities are derived from the general population or patient population using common elicitation techniques.	Disability weights are standard sets based on expert opinion.
Obtaining utilities from young children is subject to debate.	The validity of DALYs is questioned. DALYs are mainly problematic to evaluate life-extending interventions.
	DALYs are global and are used within the BCoDE project.
Cross-vaccine comparisons are possible.	Cross-vaccine comparisons are possible.

3.4 Quantitative benefit-risk frameworks

Quantitative frameworks are structured stepwise processes (like the descriptive and semi-quantitative frameworks), as part of which an overall benefit-risk score is calculated.

3.4.1 'Principles of Threes', TURBO and Beckmann

The 'Principle of Threes'[81], the TURBO model[82] and Beckmann model[83] can be considered as first attempts to develop a quantitative benefit-risk framework. A description of these models can be found in Mussen et al[69]. For all models, various limitations have been identified[69]. One of the most important limitations of the models is that they essentially cover only one benefit and one risk criteria. Furthermore, these models (e.g. criteria, grading systems) have not been

validated and their actual use seems to be very limited[69]. The models can be considered as predecessors of more structured benefit-risk assessment. A strong point of the Beckmann model is the importance attached to the strength of the evidence.

3.4.2 Benefit-less-risk analysis

In line with the basic principle of GBR to discount benefits by the presence of untoward safety experiences, Chuang-Stein introduced **Benefit-less-risk analysis** (BLRA)[84]. The same benefit-risk outcome categories are created as in GBR. Then, for each individual *i*, the risk-adjusted benefit measure is obtained by discounting the benefit by a multiple of the aggregated risk score (RS) or

$$E_i^* = E_i \ - f \times RS_i,$$

where *f* controls the amount of discounting. Then, if applied in a clinical trial comparing different treatments, statistical significance tests can be performed. The model provides a detailed methodology for assessing the safety data (organised by body functions), proposes to use sensitivity analysis and provides a structure for combining benefits and risks into a single measure. The model has been criticized because it requires weights that reflect the relative seriousness of groups of adverse events organised by body function[69]. Furthermore, the model is developed for individual-level data on benefits and risks of an intervention, and therefore suitable data may not be available in the post-authorization benefit-risk surveillance data sources typically used for vaccines.

3.4.3 Multi-criteria decision analysis and extensions

Multi-criteria decision analysis (MCDA) is a methodology for integrating various benefits and risks, and consequently includes value judgements. The use of MCDA in the context of drug benefit-risk analysis was first proposed by Mussen et al[69] and has been extensively evaluated by the PROTECT consortium[6]. Both EMA[2] and PROTECT[9] recommend further investigation into the use of MCDA for pharmaceutical benefit-risk assessment. Developing a MCDA model involves different steps, which are summarized in Table 3.5. For worked examples we refer to [6, 85].

MCDA provides a highly structured approach based on the PrOACT-URL framework[9]. It allows assessing and integrating multiple benefits and risks criteria and comparing multiple options. MCDA can be applied to benefit-risk assessment of vaccines given that special consideration is paid to the vaccine specificities as discussed when appraising the BRAT and PrOACT-URL frameworks (see Sections 3.1.2 – 3.1.3). A particularly valuable aspect of MCDA for vaccines is that it can accommodate many types of inputs or attributes. The ability to include continuous endpoints, dichotomous endpoints, categorical attributes and potentially more complex inputs could be potentially very important when combining information from heterogeneous sources, such as clinical trials, epidemiological studies, observational data analyses and infectious disease models.

MCDA assumes fixed values for the criteria measurements as well as for the weights reflecting the clinical relevance of the different criteria. Simple deterministic sensitivity analyses (i.e. change the input parameters and re-run the model)[86] are then used to evaluate the uncertainty in model inputs. However, this might be cumbersome if many model parameters (both criteria measurements and weights) are to be evaluated.

Stochastic multi-criteria acceptability analysis (SMAA) is a family of probabilistic extensions of MCDA methods that allow defining preference information and criteria measurements with uncertain or missing values[87]. SMAA uses Monte Carlo simulation to evaluate the uncertain information. Missing preference information is represented using uniformly distributed normalized weights (e.g. the feasible weight space in the 3-criterion case is a triangle with corners (1,0,0), (0,1,0) and (0,0,1)). Uncertain preference information can also be presented by weight intervals or importance ranking of the criteria that constrain the uniform weight space. Uncertain criteria measurements are represented by suitable probability distributions (e.g. beta distributions for probabilities, Poisson distributions for counts). SMAA is very similar to **probabilistic sensitivity analysis** (PSA) (see also Section 3.5). In a PSA, the uncertain model inputs are represented by probability distributions and Monte Carlo simulation is used to evaluate the model. Recently, Wen et al. (2014)[88] applied this Monte-Carlo approach to account for data uncertainty in MCDA models. In addition, they proposed using the delta-method as an alternative way to account for data uncertainty in MCDA models.

Table 3.5. Steps in creating and exploring an MCDA model (from [85]).

1. Context		
	Establish the decision context	• Identify the medicinal product • Identify the therapeutic area and the indication for use • Recognise the unmet medical need, severity and morbidity of condition, affected population, patients' and physicians' concerns, time frame for health outcomes • Define the decision problem (what is to be decided and by whom)
2. Alternatives		
	Identify the options	• Describe the medicinal products • Describe the comparators
3. Criteria		•
	Identify and define the criteria for assessing the effects of each alternative. Represent these in an effects tree	• Select the favourable effects • Select the unfavourable effects
4. Weighting		•
	Establish a measurement scale for each criterion and assess the relative importance of the scales	• Define each effect's measurement scale and its units (e.g., mean, median scores, proportions) and determine upper and lower limits that encompass a plausible range for the data • Assess swing weights to represent the clinical relevance of the swing from the lower to the upper limit of each scale
5. Scoring		•
	Describe how the alternatives perform for each of the criteria and show how to convert input data into preference values (i.e. assess value functions).	• Gather available data, pooling or performing meta-analysis of multiple data sources, to give data summaries and confidence intervals • Provide data summaries in effects table with alternatives in columns and criteria in rows • Assess linear or nonlinear value functions using direct (more means better) for favourable effects, and inverse (more means worse) for unfavourable effects
6. Results		
	Calculate results and provide graphical displays	• Multiply preference values and criterion weights and sum the products to obtain overall value (usually carried out by appropriate software) • Construct preference value bar graphs for favourable and unfavourable effects, and for individual effects • Calculate difference displays for pairs of alternatives
7. Sensitivity analyses		
	Explore effects of uncertainty on the benefit-risk balance	• Vary individual weights over their entire range from 0 to 1; display the overall results graphically • Change input data over ranges of uncertainty (e.g. pessimistic values for favourable effects and optimistic ones for unfavourable effects) • Examine the overall BR-balance under possible future scenarios (e.g. adverse events) by changing input data and criteria weights • Revise any of the above numbered steps and tasks as insights emerge
8. Recommendation		•
	Formulate recommendations	• Judge the relative importance and effect of the decision maker's risk tolerance for this product • Consider how this decision is consistent with similar past decisions in the future easier or more difficult • Metabolize the results before making any decisions (newly constructed preferences can change with reflection and new insights surface)

3.4.4 Data-driven benefit-risk assessment (Sarac et al.)

Sarac et al. proposed a data-driven benefit-risk assessment method, where two or more drugs can be compared[89, 90]. Their approach is similar to MCDA, but is tailored directly to drug development and approval. They proposed a structured 8-step process, involving: (1) defining the decision context, (2) defining the decision profile, (3) weighting the criteria, (4) scoring the performance of the drugs (and the comparator) for each of the selected criteria, (5) evaluating the uncertainty, (6) calculating the weighted scores, (7) presenting the results and (8) obtaining the overall conclusion. Their method is data-driven, based on the analysis of clinical data, and simple and transparent rules for weighting and scoring have been proposed. Each benefit and risk criterion is assigned a weight of 1 (low), 2 (medium) and 3 (high) and weighting is done independently of the data. Then, for each criterion, the drug is scored relative to the comparator using a simple and transparent scale: -1 (inferior), 0 (non-inferior or equivalent) and 1 (superior). The method of scoring may be different for different types of data (e.g. difference distribution scoring for continuous variables, confidence interval scoring for rare events). They further proposed to account for uncertainty through the use of bootstrapping and to visualize the results using tornado-like diagrams.

The method proposed by Sarac is specifically developed to provide structure and support to the benefit-risk interpretation of clinical trial data. Particular features of the analysis of clinical trial data (e.g. inferiority, non-inferiority, superiority) play a special role within Sarac's approach, making it less suited for post-marketing surveillance. However, the simplicity and full transparency of this approach are extremely valuable.

3.4.5 Clinical utility index

The **clinical utility index** (CUI) quantifies the tradeoffs between different product attributes by providing a single metric for the multiple attributes (i.e. criteria) of the product profile[91]. The additive multivariate utility function is

$$CUI = \sum_{i=1}^{n} w_i \, U_i(x_i),$$

with i indexed over n drug attributes, with weight w_i and utility function $U_i(x_i)$ that transforms the attribute from its original scale into the (0,1) scale of utility. The additive multivariate utility

function is also called net clinical benefit (NCB). The attributes are typically limited to clinical attributes. Key steps in clinical utility analyses are: (1) identification of the key attributes, (2) normalization of the attributes, (3) assigning importance weights to the attributes and (4) sensitivity analysis and measurement of uncertainty. The CUI tool shows a lot of similarities with MCDA (e.g. multiple criteria, scaling, weighting and sensitivity analysis) and has been mainly used in early drug development.

3.4.6 Concluding remarks

All quantitative frameworks described are tailored towards drug development, with the exception of MCDA. MCDA has been valued because it provides a highly structured approach in line with the PrOACT-URL framework, allowing to assess and integrate multiple benefits and risks for multiple treatment or intervention options[9]. Important choices are to be made when building and trimming the value tree, constructing the effects tables, defining the (scales of the) value functions and eliciting the weights. MCDA and its stochastic variants (SMAA) seem generally applicable to assess the benefit-risk balance of vaccines given that special consideration is paid to vaccine specificities such as time window, indirect effects and high levels of uncertainty.

3.5 Modelling approaches (HTA)

(Decision) Analytical modelling techniques facilitate the estimation of the consequences of health care decisions and are commonly used in Health Technology Assessment (HTA)[92]. As also recognised by Lynd & O'Brien[17], it is natural to apply techniques used for cost-effectiveness analysis to benefit-risk assessment as well. Indeed, the risks of a medicine can be considered a nonmonetary cost and benefits are synonymous with effectiveness. Therefore, we describe the most commonly used modelling techniques in HTA[93]: decision trees, state-transition models, discrete event simulation and dynamic transition models, as well as common meta-analytic approaches. We give examples of their application for benefit-risk assessment. Finally, we will discuss the different sources of model uncertainty and how they can be dealt with.

3.5.1 Decision trees

One way to analytical modeling is the **decision tree** (Figure 3.5). There are three types of nodes on a decision tree: decision nodes, chance nodes and terminal nodes. The first branch in the tree is the decision node and represents the decision question. The pathways that follow each decision option are the logical consequences, possibly emanating from chance nodes. The options at a chance node should be exhaustive and mutually exclusive and their probabilities should sum to 1. The end points at each pathway are the terminal nodes, to which values (e.g. QALY's, utilities) are assigned. Then, the decision tree is averaged out or 'rolled back' to calculate the expected value of each option. Decision tree models may be more realistic if the branching probabilities (and possibly the values or utilities) are represented by distributions rather than point estimates. Monte Carlo simulation techniques[94] and Bayesian analyses[95] can then be used for modeling this uncertainty.

Decision trees offer a valuable graphical representation of the decision problem. They aid structuring the decision problem and clarifying the options and their consequences). However, decision trees lack a time variable and do not allow for time-dependent variables (such as time-to-event) or recurrent events and interactions. As such, they are not suited for modeling complex, time-dependent and dynamic decision problems. Decision trees have already been used for benefit-risk analyses[66], even for vaccines [96, 97]. Decision tree models have also been used for vaccine-related economic analysis[97, 98]

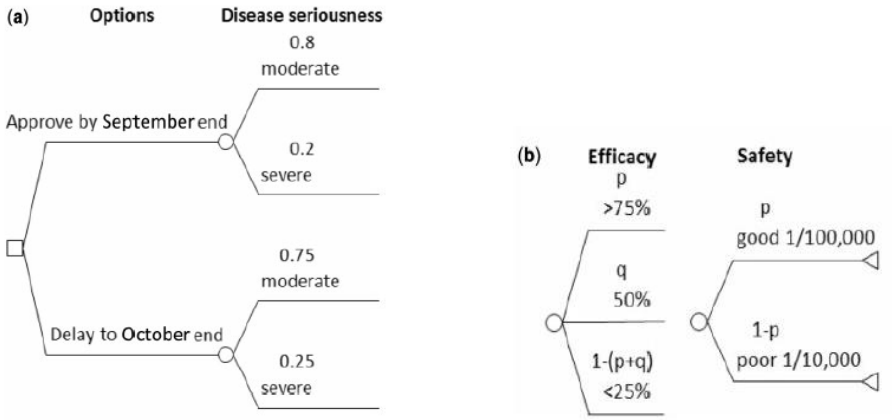


Figure 3.5. Decision tree on modelling the risk–benefit impact of H1N1 influenza vaccines (figure from [96]). (a) The initial part of the decision tree: a decision node (square) with two decisions followed by Disease seriousness nodes (circles) with two possible outcomes and their probabilities, which are conditional on the decision. (b) The subsequent events' efficacy and safety, and their outcomes. The Safety node attaches at the end of each branch of the Efficacy node, which in turn attaches at the end of each Disease Seriousness node's outcome branch The triangles at the end of each path receive the number of DSDs (deaths and serious disabilities) appropriate for the outcome of the uncertain events in that path.

3.5.2 State transition models

State transition models (STMs) assume that a patient is in one of a finite number of discrete health states (also called Markov states) at any point in time and make transitions between the health states. In the area of infectious diseases, frequently used states are: Susceptible, Exposed, Infected and Recovered (SIR or SEIR models). The probability of staying in a health state or moving towards another health state is determined by a set of transition probabilities. A STM is typically evaluated as a cohort simulation or Monte Carlo simulation. The constituent elements of a state-transition model are: the initial state vector, states, transitions with certain transition probabilities, cycle length and state values (e.g. life years, QALY's).

When cohort simulation is used, a hypothetical (closed) cohort of patients transitions to the model simultaneously at specified time intervals (Figure 3.7a). These models are also called **Markov models** or **cohort models**. These models are relatively simple to develop and communicate. Unlike decision trees, Markov models permit a flexible sequencing of outcomes, including time-dependent parameters such as recurring outcomes and time-to-event outcomes. An important limitation of the Markov model is the assumption that the transition probability only depends on the current health state and not on the previous ones (also called

the Markovian property). Solutions have been proposed to this problem. However, they result in overly complex models (i.e. state explosion)[99].

Individual-based state transition models are a special case of micro-simulation models and are evaluated using first order Monte-Carlo simulation. In these models, patients are randomly selected from the hypothetical cohort and they transition to the model one at a time (Figure 3.7b). Unlike the Markov model, the individual-based STM is not characterized by the Markovian property because the state history of an individual can be traced and the state transition probabilities can be adjusted accordingly. Compared to Markov models, the individual-based STM is computationally much more intensive.

Consensus-based guidelines for the application of STMs in the context of health care (ISPOR) exist[99]. They are recommended when the decision problem can be framed in states, interactions between individuals are not relevant and the population is a closed cohort. STMs (Markov models) have already been successfully used for benefit-risk analyses[100], also for vaccines[101-104].

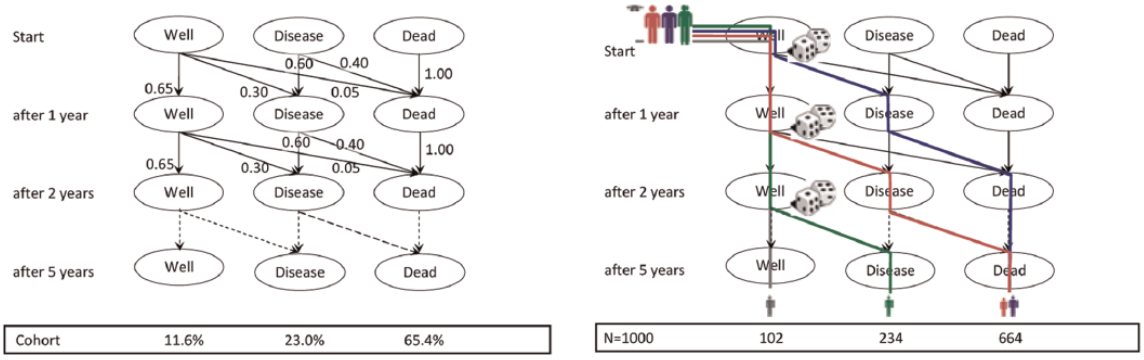


Figure 3.7. Example of a Markov model (left) and individual-based state transition model (right). In a Markov model, the entire cohort is redistributed across states after each cycle. In an individual-based model, first order Monte-Carlo simulation is used to move individuals across states (figure from [99]).

3.5.3 Discrete event simulation

Discrete event simulations (DES) describe the progress of individuals through health care processes or systems as a discrete sequence of events in time. The system is assumed to not change between consecutive events and therefor, the simulation model can jump in time from one event to the next. DES is an operational research model, being originally developed for industrial planning. The constituent elements of a DES are: entities (e.g. patients), events (e.g.

adverse events, hospitalization), attributes (e.g. age, sex, past events) and resources (e.g. number of beds). In addition, time is a fundamental component of a DES as well. An example of a graphical representation of a DES is given in Figure 3.8. Unlike decision trees and state-transition models, a DES allows entities (e.g. patients) within a system to interact or compete with each other. Unlike state-transition models, the timing of the events is not fixed by the cycle lengths, but can be stochastic. Compared to decision trees and state-transition models, a DES is more complex to build, to understand and to communicate.

Consensus-based guidelines for the application of DES in the context of health care (ISPOR) exist[105]. A DES is best used when the modeled system involves competition for resources, the timing of the event is stochastic and when there are interactions between events or entities. DES has been used for benefit-risk analysis as well[64, 106].

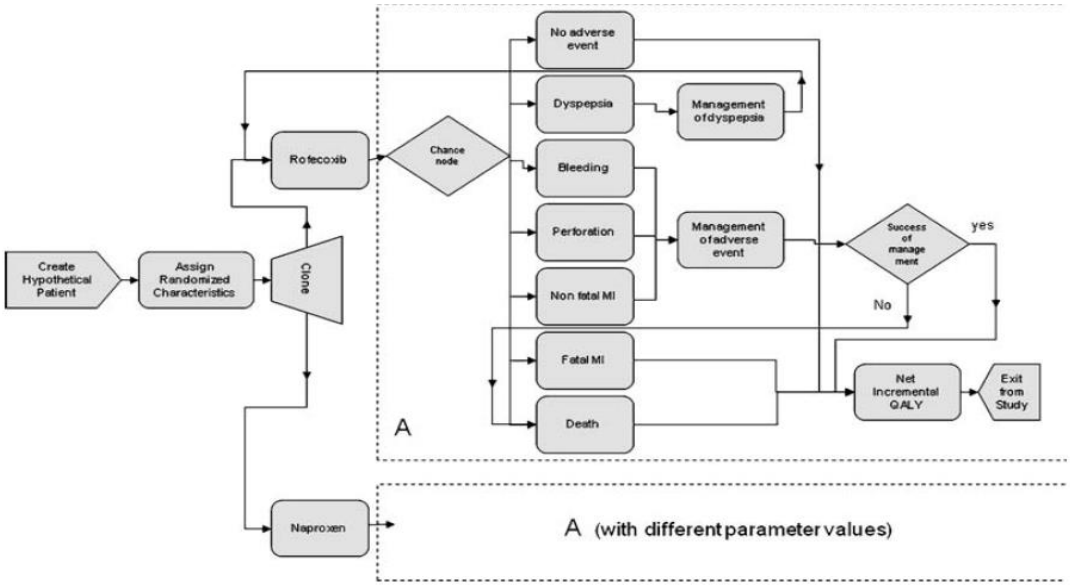


Figure 3.8. Example of a discrete event simulation model, used to calculate the incremental net benefit of rofecoxib relative to naproxen in arthritis patients over a 1-year time horizon (figure from [64]).

3.5.4 Dynamic transmission models

Dynamic transmission models are mathematical models used to model infectious diseases, explicitly modeling disease transmission. Static models (e.g. Markov models, discrete event simulation) assume a constant risk of infection (or force of infection) whereas dynamic models allow the force of infection to depend on the number of infectious agents within the population

at a given point in time. There are two main groups of dynamic transmission models: compartmental models and individual-based models. Compartmental models divide the population according to infection status (e.g. SIR-model: susceptible-infectious-recovered) and assume homogeneous mixing of the population. Individual-based models are a collection of individuals (agents) and rules specifying how they behave within a specific environment. Compared to compartmental models, individual-based models are more flexible and more computational-intensive. The basic infectious disease models can be extended to account for heterogeneities in the host population structure, different modes of transmission, waning immunity, mutations and more. Consensus-based guidelines for the application of dynamic models in the context of health care (ISPOR) exist[107]. They are recommended when evaluating an intervention against an infectious agent, when the intervention affects disease transmission and when the intervention affects a pathogen's ecology (i.e. strain replacement).

3.5.5 Meta-analytic approaches

Meta-analysis is a well-established technique to combine multiple sources of quantitative evidence. It is common practice to use pairwise meta-analysis methods to estimate the effectiveness of two specific interventions (A versus B comparisons). **Mixed treatment comparisons** (MTCs) (also called Multiple Treatment Comparisons or network meta-analysis) are a generalisation of pairwise meta-analysis. MTC allows the simultaneous estimation of the effectiveness of multiple treatments using a network of evidence that individually do not compare all treatment options, but each has a common option to another (e.g. using A versus B, B versus C and A versus C comparisons)[108]. The basic assumption of MTC (as for pairwise meta-analysis) is that the different studies are sufficiently homogeneous to be quantitatively combined. Pairwise meta-analysis and mixed treatment comparisons can both be formulated within a common Generalised Linear Model (GLM) framework, which can be applied in either frequentist or Bayesian contexts[109].

The **confidence profile method** (CPM) was developed in the late eighties as a Bayesian method to evaluate evidence from different types of empirical studies, adjust individual pieces of evidence for biases, combine evidence from different studies and incorporate subjective judgements to derive a probability distribution for the intervention effects[110]. Multiple pieces of evidence are incorporated by successive applications of the Bayes' theorem. The **multi-parameter evidence synthesis** (MPES) approach builds on and extends the confidence

profile method[111]. Within a MPES model, multiple sources of information on either model parameters and/or functions of several parameters are incorporated. An important feature of MPES is the ability to incorporate information on more functions of parameters than there are parameters in the model, allowing to validate the model and to check the consistency of the different sources of evidence. MPES has been used in epidemiology (to estimate disease impact [112, 113]) and health technology assessment. The MPES approach is a powerful technique to combine evidence from different sources and is gaining popularity. However, developing a MPES model is time-consuming and technically complex.

3.5.6 Concluding remarks

Decision analytical modelling techniques are at the core of health technology assessment (HTA), but are sparsely used for benefit-risk assessment, although their use has been advocated[17, 18]. Their merits lie in their capability of synthesising evidence from different sources while being able to account for different sources of uncertainty (e.g. through using first and second-order Monte Carlo simulation or by adopting a Bayesian approach). The modelling techniques as they are commonly applied in HTA do not reflect stakeholders' preferences, though they can be modified to do so. Lynd et al.[106] gave a good example of the modifiability of the techniques used in HTA for the purpose of benefit-risk assessment. The authors used discrete event simulation in combination with QALY's with preference weights derived using conjoint analysis (see Section 3.6) to quantify the incremental net health benefit (INHB) of alosetron[106].

When assessing the benefit-risk of a vaccine adopting the individual perspective, decision trees, state-transition models and discrete event simulations might be appropriate. When assessing the benefit-risk balance of vaccination adopting the societal perspective, dynamic transition models may be more suitable, at least insofar as the indirect effects are judged to be important. Adhering to the fundamental modelling principle of keeping a model as simple as possible (but not too simple), static models might be acceptable to model benefit-risks from a societal perspective when (1) the target group for intervention is not epidemiologically influential for transmitting the disease, (2) when the effects of vaccination are expected to be entirely direct (e.g. when herd immunity does not play an important role) or (3) when the static model suggests that the intervention has a positive benefit-risk profile, and that the indirect effects would enhance this (adjusted from [107]).

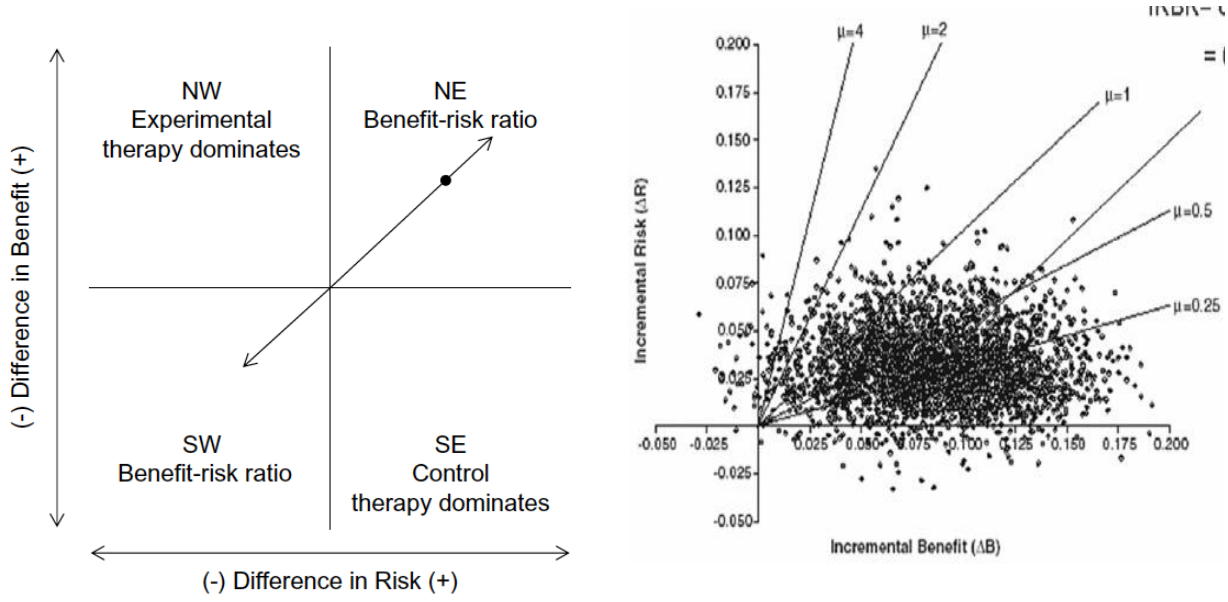
From the meta-analytic approaches, multi-parameter evidence synthesis methods are particularly suited to model benefits and risks of vaccines because they are powerful methods to combine different sources of evidence. Multi-parameter evidence synthesis is particularly relevant if direct evidence to inform the model parameters is unavailable, but the model parameters can be 'indirectly' informed based on evidence of functions of parameters. The applicability of pairwise meta-analyses and MTC might be more limited for vaccines because studies on infectious disease interventions are often very heterogeneous as a result of the rapidly changing infectious disease dynamics.

3.6 Parameter estimation and uncertainty

Different types of uncertainty in analytical modelling techniques exist, i.e. stochastic uncertainty, parameter uncertainty, heterogeneity and structural uncertainty. Stochastic uncertainty (or first order uncertainty) relates to the fact that 'identical' patients (i.e. patients having the same genetic predisposition and sharing the same environmental factors) will respond differently to disease or intervention due to chance. 'Identical' patients have the same probability of developing a specific outcome, but the realisations of these outcomes might still be different. Some of the models described earlier do account for stochastic uncertainty (i.e. individual-based STM, DES, individual based dynamic models) whereas others not (i.e. decision tree, Markov models, compartmental models). Parameter uncertainty (or second order uncertainty) relates to the fact that the model parameters themselves (e.g. outcome probabilities) are subject to uncertainty because they have been estimated. Parameter uncertainty results from the finite sample size of the study used to inform the parameter, 'conflicting' multiple studies and might be further enhanced by bias and confounding in these studies. In a frequentist context, parameter uncertainty can be represented by deterministic sensitivity analysis or probabilistic sensitivity analysis. In a **deterministic sensitivity analysis**, parameter values are varied according to a predefined set of values and the impact on the model results is assessed. Such a deterministic sensitivity analysis can be easily improved upon by using (probabilistic) Monte Carlo simulation, randomly sampling parameter values from predefined probability distributions (i.e. second order Monte Carlo simulation). This is commonly called **probabilistic sensitivity analysis** (PSA) by health economic modellers and is used by Lynd & O'Brien, although they used the broad term **probabilistic**

simulation method[17]. In a Bayesian context, parameter uncertainty is naturally incorporated as prior information[95]. Both types of sensitivity analyses (deterministic and probabilistic) as well as Bayesian analysis can be used to account for potential sources of bias and confounding that might have affected the studies used to inform the model parameters[114]. A third type of variability is heterogeneity, which refers to differences in parameters across patients, patient populations. In contrast to parameter uncertainty, heterogeneity cannot be reduced through performing additional or better studies, but should be acknowledged using e.g. stratified analyses. Finally structural uncertainty (or model uncertainty) relates to the assumptions inherent to the model. Structural uncertainty can be addressed by specifying several (plausible) assumptions about the model structure. For guidelines on accounting for several sources of uncertainty in analytical modeling techniques, we refer to Bilcke et al[115] and Briggs et al[116].

Uncertainty can be graphically represented. An example of such a graphical representation is the **risk-benefit plane**[17, 117], which is a two-dimensional plot with the differences in risk on the x-axis and the differences in benefit on the y-axis (Figure 3.9). The plane is divided in four quadrants (NE, SE, NW and SW). In the SE quadrant, the new therapy dominates the old therapy (more benefit and less risk), and vice versa in the NW quadrant (less benefit and more risk). In the NE and SW quadrants, the decision to prefer the new treatment over the old one, depends on the **risk-benefit acceptability threshold** (RBAT), which is the maximum number of additional adverse events the decision maker is willing to accept to realize one additional beneficial outcome. The risk-benefit acceptability threshold is represented as the slope of the line passing through the origin and crossing the NE and SW quadrants (Figure 3.9). A risk-benefit plan with RBAT can be used to graphically display both statistical uncertainty in risk-benefit ratio (scatterplot of risk-benefit pairs generated trough Monte-Carlo simulation) and uncertainty in preferences (varying RBAT slopes).



(a) risk-benefit plane (b) joint distribution of risk and benefit

Figure 3.9. (a) Risk-benefit plane with the line through the origin representing the risk-benefit acceptability threshold (figure from[117]). (b) Results of a Monte Carlo simulation presented as a scatter plot of benefit-risk pairs within a risk-benefit plane (figure from [17]).

The **risk-benefit contour plot** is an alternative way to graphically represent the probability of benefit and risk and associated uncertainty[118]. The probabilities can be derived from the reported confidence intervals[118] or from simulation studies[17] and plotted as contour lines. The various contours provide degrees of probability of both benefit and risk. For example, a clinician might recommend the new treatment if there is at least 10% survival benefit compared with another treatment and if the probability of severe harm is not increased by more than 30% compared with the other treatment. The contour plot given in Figure 3.10 shows a 70% probability that these two conditions will be met (clinician A).

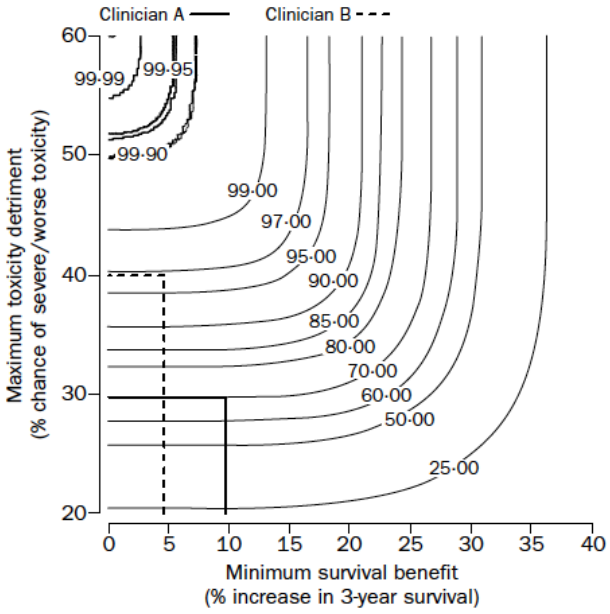


Figure 3.10. Example of a benefit-risk contour plot (figure from [118]).

3.7 Preference elicitation[2]

The section below assumes the reader is acquainted with Appendix 11 of the IMI-PROTECT WP5 report[119].

1. What is aimed at with utility stakeholder surveys

Surveys were used in marketing and transportation domain for many years to inform decision-makers on what matters really to customers. Likewise, utility surveys applied to benefit-risk assessment aim at improving the decision quality by decision makers. Opinion and preferences of "users" or "consumers" are collected (or "elicited") through some kind of methodology (qualitative or quantitative) and accounted in a quantitative model for decision-making.

2. Qualitative and Quantitative approaches

Thoroughly quantitative survey methods like conjoint analysis often initiates with qualitative preference assessment that narrows down the problem considered and focus on the most important information to clarify for the decision-making. Among others, focus groups, individual and group interviews, or open questionnaires are those sorts of qualitative approaches that may be very informative for decision-making although not immediately covered in PROTECT. We suggest covering some aspects on the quantitative analysis of those qualitative assessments of preferences and opinion.

3. Revealed and stated preference survey

Retrospective survey may include database search reflecting on subject's actual behaviour when purchasing or selecting a health care alternative. The conclusions driven from these analyses are identified as "Revealed Preferences". That strategy for preference assessment and benefit-risk evaluation is uncommon and out of the scope of this section.

[2] This section is written by Edouard Ledent (GSK)

The alternative pathway involves asking the same individuals to state their preferences in hypothetical (or virtual) settings (or markets). The conclusions that follow that strategy are collectively known as "Stated Preferences" techniques and are the focus of the following paragraphs[120].

4. Groups of respondents to consider for the survey

Collecting preferences from various group representatives presents technical challenges. It also involves a broader issue on how those opinions from different groups should be balanced and how trade-offs should be made. Although preferences from regulatory authorities, public health authorities, health care professionals, vaccine recipients from general populations and minorities would be collected, no clear framework exists on how those preferences from different groups interact to form recommendations and how the decision is communicated and articulated. Prior to collecting those feedbacks, we would benefit from more clarity on how to use that information for decision-making.

5. Communication of utility survey results

PROTECT WP5 report[119] mentioned the absence of standard visualization from utility survey techniques which seems a limitation in communicating the results to a broader audience, and enhancing trust and transparency to the public. A proposal for visualizations was made by Sur D & al [121] in the context of policy makers.

Statistical technology allows that survey results are communicated as predictions (i.e. prospectively rather than retrospectively) for a specific subject profile, accounting for his/her individual parameters. That sort of communication channel may improve the impact of such survey results to the public. Utility survey methodologies allowing such improved communications to a large audience might present additional advantages as compared to other methodologies.

6. Innovative perspectives

Since the PROTECT systematic review[119], a few improvements or innovative proposals were made that may ease quantitative benefit-risk using a mixed level of qualitative and/or

quantitative information. No recommendations identifying the most critical innovations for utility survey in the vaccine field is available. Instead, retrospective analysis of what might be pertinent for those surveys is performed through publication reviews and a couple of those are described below.

(a) Proposals were made to adapt the questionnaires of ongoing surveys based on the results of information already collected aiming at focusing on attributes that requires more data to support better the decision-making process. General software (Sawtooth Software ACBC[122]) and academic proposals[123, 124], for adaptive survey are available but have rarely been used in health care.

(b) Theoretical considerations that underneath better the collection and analysis of preferences for benefit-risk in vaccines might not follow the utility-maximization approach that is mostly used for conjoint analysis methods (see definition below) [125]

(c) Improved probabilistic simulation methods can make better use of qualitative or semi-quantitative information from respondents to proceed with quantitative benefit-risk assessment[126]. Qualitative or semi-quantitative survey information is faster and easier to collect and probably less prone to large or uncertain bias due to study design issues or poor understanding of the background decision problem. The outcome of benefit-risk evaluation might account for larger uncertainty but maybe not to an extent that a decision cannot be made. If such situation would apply, further (fully) quantitative survey would then focus on the value of information regarding to what is most critical for the decision.

Defining an ideal strategy for an efficient and effective collection of information pertaining to benefit-risk assessment would provide better guidance to researchers on what innovations would have the highest impact.

4.7.1 Focus groups

Definition and context

Focus groups are defined as "carefully planned series of discussions designed to obtain perceptions on a defined area of interest in a permissive, non-threatening environment."[127]. They consist of small groups of people who have been gathered together for a group discussion in order to gain insight on a particular topic[127]. Sullivan[128] provides a thorough review of focus

groups for MCDA, involving swing-weighting, Analytic Hierarchy Process (AHP) or MACBETH as simplified forms of preference elicitation for decision-making as compared to conjoint analysis.

Those structured group interviews are conducted by a trained facilitator and can provide useful qualitative information to feed initial selection of attributes and levels for surveys or key information for quantitative benefit-risk frameworks relying on qualitative data only[126]. Comparison between focus groups and other types of surveys or discussions were summarized by Grudens-Schuck & al[129] and reproduced in Table 3.6.

Table 3.6. Comparing and contrasting focus groups and other types of discussion groups (from [129]).

	Focus Groups	**Other Small Discussion Groups[1]**	**Large Discussion Groups[2]**
Application			
Identify problems	Recommended	Recommended	Limited use
Design programs	Limited use	Limited use	Not recommended
Evaluate programs	Limited use	Not recommended	Not recommended
Educate or inform participants	Not recommended	Recommended	Recommended
Build consensus	Not recommended	Recommended	Recommended
Purpose	Designed to encourage divergent thinking and disclosure of personal perceptions and behaviors	Designed to study and/or generate ideas and solutions	Designed to build consensus, educate, or persuade
Participant selection	Participants are selectively invited, based on similar characteristics	Participants invited or required to participate because of their organizational affiliation. Similarity between participants is not a qualifier and may be a limitation in some situations.	Open to everyone in an organization or community
Group size	Group size from 6 to 12 individuals	Group size from 6 to 20 individuals	Group size from 6 to 100 or more individuals, depending on the issue
Event environment	Open, trusting environment	Open, trusting environment	Open, trusting environment

1 For example, Study circle, Delphi Technique, Search Conference

2 For example, town meeting

Set-up and conduct

Focus groups scenarios are set up prior to the conduct of the group interviews by facilitators that will make sure that each group member can have the opportunity to express him/herself or acknowledge his/her agreement or disagreement with what was said by other attendees. Focus groups are usually taped or video-recorded after personal identifiable information is taking care of and transcripts are made available for further analysis. Each intervention should be related to a unique participant in a way to allow between-subjects analysis but also to follow how the opinion of each individual evolves along the interview depending on what ideas are mentioned.

Quantitative Analysis

Content analysis of focus group data[130] provides vital and robust information that will be used to build the design of a conjoint experiment. The survey quality and biases may be driven by poor preliminary review of expert opinions, interviews or focus group feedback.

No clear standard in analysing focus groups for health care purposes is available and the recommend process[130] is followed in various ways, possibly leading to poor surveys. Computer Aided Qualitative Data Analysis Software (CAQDAS) products are available to ease the process (e.g. ATLAS.ti, MAXQDA, NVivo, Hyperresearch and QDA Miner) although the structure of the group interview may allow formal identification of ideas, attributes or concepts by the participants themselves .

Publication search results

A publication review through Scopus on abstract, titles & keywords, searching for focus groups and conjoint analysis in the vaccine field provided one reference only (i.e. [131]), although the use of focus groups together with conjoint analysis in any fields provided 112 hits mostly in medicine, agriculture, nursing social sciences and engineering. Focus groups may have been used to set-up surveys in the 48 references related to conjoint analysis in the vaccine field although not mentioned in the abstract (e.g. [132]).

4.7.2 Conjoint analysis overview

Introduction and historical perspectives

Louviere (2011)[133] presented to the Isaac Newton Institute an introduction to DCE and the several important challenges that researchers are facing when implementing such widely spread survey approach. Bryan Orme[134] provides a short summary and practical knowledge required to implement these techniques:

- Conjoint methods were based on work in the sixties by mathematical psychologists Luce and Tukey. Discrete choice methods come from econometrics, building upon the work of McFadden (1974)[135], 2000 Nobel prize winner in economics.

Marketers have thought that the word "conjoint" refers to respondents evaluating features of products or services "CONsidered JOINTly". In reality, the adjective "conjoint" derives from the verb "to conjoin" meaning joined together. The key characteristic of conjoint analysis is that respondents evaluate product profiles composed of multiple conjoined elements (attributes or features). Based on how respondents evaluate the combined elements (the product concepts), we deduce the preference scores that they might have assigned to the individual components of the product that would have resulted in those overall evaluations. Essentially, it is a back-door, decompositional approach to estimating people's preferences for features rather than an explicit, compositional approach of simply asking respondents to rate the various features.

The estimation procedure relies on a statistical model that is expected to provide (partial) insight into the respondent's answers. What cannot be explained by the model, either due to variables not captured or simply respondent's mistakes, is absorbed into "residuals" that are given a specific statistical distribution and potentially some correlation features.

PROTECT mentioned utility measures used in MCDA like Swing-weighting, Analytic Hierarchy Process, MACBETH, ELECTRE or PROMETHEE as "conjoint" approaches. For clarity, we will not refer those methods as "conjoint analysis" and reserve that wording to the methods that share the features described by Orme above.

Types of Conjoint Analysis

Several types of conjoint analysis are available, although the Choice-Based-Conjoint (CBC), also called Discrete Choice Experiment (DCE), now represents ~80% of the conjoint experiments implemented[134]. A classification of conjoint experiments is provided Table 3.7, although the criteria interact heavily and consulting Orme's guidance using consult Sawtooth Software's *"Interactive Advisor"* for selecting the most appropriate conjoint method is recommended.

Table 3.7. Summary classification of conjoint analysis

	CVA Conjoint Value Analysis	**ACA** Adaptive Conjoint Analysis	**CBC** Choice-based-conjoint	**PPCBC** Partial-Profile Choice-based-conjoint	**ACBC** Adaptive Choice-Based-Conjoint	**Menu-based** (a)	**Best-worst scaling** (b)
# attributes	≤6	(≥ 8)	≤4, 5-7	((5-7)), (≤12), ≥ 12	(5-7), ≥ 8		≥ 8
# levels/attribute			≤15	≤15, ≥ 15	≤15, ≥ 15		15 - 40
Interview method	Paper, (PC)	PC, (Phone)	Paper, PC	Paper, PC	PC only	PC	Paper, PC
Sample size	Small, (≤ 100), (high)	(≤ 100)	≥ 100	≥ 100	(≤ 100)		
interview time			≤8 min, average, (longer)	≥ 8 min, longer			
Monetary research			Yes	Yes			

(..)The preferred approach or feature for a method is presented without parenthesis that indicates that the methodology can be selected for the feature but another methodology may be more appropriate.

(a) More appropriate when the respondent may select the product's attributes using a menu-based approach.

(b) Where the goal is to estimate the relative importance or preferences for each of the items separately but NOT being able to estimate how multiple items taken together affect overall preference.

4.7.3 Choice Based Conjoint methods

The scope of current document will be limited to the choice-based conjoint methods (CBC, PPCBC and ACBC) as they represent the most prominent methodologies currently used for utility surveys although other approaches might reveal valuable to some specific vaccine applications.

Introduction and general guidance

On the contrary to standard interview, CBC methods rely on a statistical model to get insights in respondent's preferences and choice motivations. The alternatives described to the respondents in the questionnaire rarely correspond to an existing alternative. Conversely, the researcher is focusing on how the characteristics that build those alternatives influence the chance for the respondent to select the proposed alternatives. Guidance on how to build efficient and effective CBC experiments is provided by ISPOR[136, 137]. A convenient summary of the process by which to conduct CBC is illustrated in Figure 3.11. Examples of CBC experiments for benefit-risk[138] and for health care decision-making[139, 140] exist.

Those methods rely heavily on behavioural and cognitive theories explaining how humans are making choices. The challenges are therefore not only of statistical optimization nature. The methods should also account for that part of irrational behind each of us. The researcher is facing both theoretical and practical issues when selecting a survey strategy that will provide the most effective understanding on respondent's motivations. The sections below are aiming at providing some aspects of those challenges.

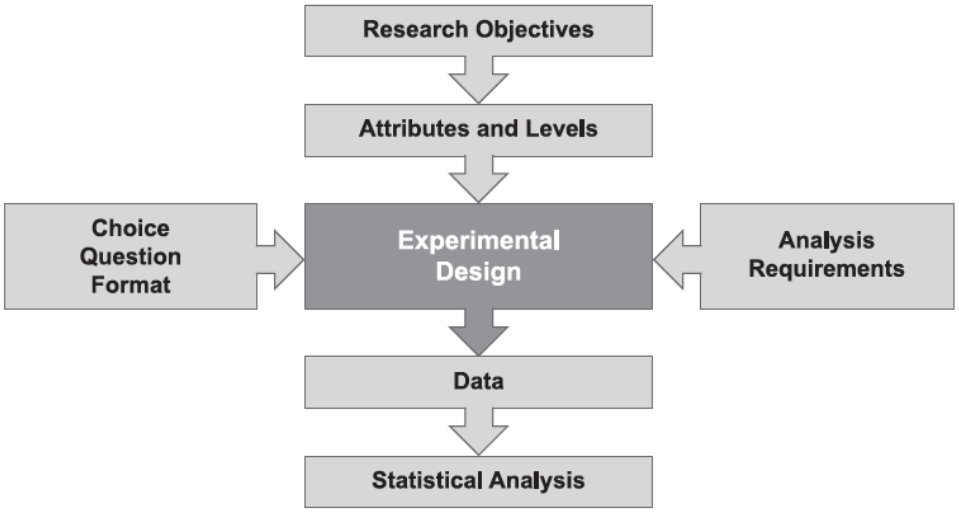


Figure 3.11. Key stage for developing a discrete-choice-experiment (from[136])

Compensatory and non-compensatory behaviours

The random utility theory that dominated the field assumes that respondents are prone to select an alternative presenting a less-then-desirable characteristic provided that alternative presents at least one characteristic that has more value to his/her preferences. The respondent

is ready to compensate the lower value of the first characteristics with the higher value of the second.

Respondents are not always ready to compensate for all attributes. Some alternatives might be selected due to the presence of that characteristic that dominates all other characteristics. Conversely, an alternative may always be rejected due to the presence, or absence, of that characteristic, irrespective of the other constituents of that alternative. The respondent presents therefore a non-compensatory behaviour that may bias the data collection, corrupt the model parameters and obscure the real respondent's motivations to the researcher.

Deviations from expected behaviour are usually absorbed by the model's residuals and pose no major problems to the researcher when such behaviour does not dominate. When the alternative covers a critical domain of the respondent's life, the researcher may anticipate that some questions may trigger an emotion aiming at some non-compensatory behaviour. The focus group's objectives may include identifying those situations and avoiding them in the questionnaire. When such mitigation approach cannot be implemented, the researcher has other recent options; of which two, adaptive CBC[134, 141] and random-regret theory[139, 142] are presented below.

Asking healthy respondents about choices to be made with regards of their health or the health of their children may trigger some emotional reactions, different from what might happen with respondent suffering from a non-severe disease or when asking a customer about purchase preferences. Vaccination involves a broader scope of personal values, including e.g. social orientation and altruism[143, 144], when selecting a prevention alternative. Exploring the extent to which the answers to CBC questions may be altered by deviations to random utility theory could help the researcher to anticipate such situations and select the appropriate strategy when building the questionnaire.

Experimental design

In comparison to "Revealed Preferences" analyses, "Stated Preferences" methodologies control as much as possible the parameters (or attributes of alternatives) that determines stimulus administered to the respondents. The experimental design describes how many and which combinations of alternative characteristics (or attribute's levels) must be evaluated by respondents to provide sufficient data for unambiguous analysis and decision-making.

No gold standard exists in selecting an efficient design and options vary depending on the analysis technique and the research objectives. Precise guidance on the questionnaire features and experimental design is beyond the scope of this document and can be found in a series of articles or books (i.e. [120, 124, 145-148]), from which some paragraphs below have been summarized.

Experimental design will account for the following topics:

1) Model identification that refers to the ability to obtain unbiased estimates from the data for every parameter in the conjoint model. Statistical criteria (e.g. D-optimality, D-efficient, S-optimality, Kullback–Leibler divergence) for optimal design also determine how alternative characteristics should be combined together to form the questionnaires. Iterative algorithms are available that searches for such design. It is recommended to first focus on identification and then on efficiency of the design since the latter can be improved by increasing the sample size although the former cannot be changed once the design is constructed[149]

2) Parameter interaction applies when the respondent's preference towards characteristic may change depending on the presence of other characteristics also present in the same alternative. The features are interacting and the experimental design and analysis model should account for those possible interactions when the researcher may anticipate their relevance. Including all interactions is not practical and would lead to implausible combinations. Focus groups and expert input may inform of the need to consider interaction parameters. The researcher can therefore select a design efficient to estimate the relevant fraction (i.e. fractional factorial design) of the design involving all possible combinations (full factorial design).

Statistical efficiency refers to minimizing the confidence intervals around parameter estimates for a given sample size. Perfectly efficient design presents each level equally often within an attribute (balanced design) and each pair of levels appears equally often across all pairs of attributes within the design (orthogonal design). Constraints on the combinations of parameters alter the efficiency but moderate reduction of orthogonality is preferred over generating implausible combinations of attribute levels. Eventually, balance and orthogonality in design are limited by the actual respondent's choices and the resulting dataset. Statistical efficiency can be improved by asking a large number of difficult trade-off questions, which however affects the response efficiency. The overall precision of the design results in the combination of both the response and the statistical efficiencies.

3) Response efficiency refers to measurement error resulting from respondent's inattention to the choice questions or other unobserved, contextual influences. Improvements can be made by asking a smaller number of easier trade-off questions. An attribute may present the same level for all alternatives in a choice set. Such overlaps improve the response efficiency but potentially limit the trade-off information collected. Sources of reduction in response efficiency include, but are not limited to:

   i) Short-cut by the respondent in making choices that are inconsistent with utility maximization or other error-model considered.

   ii) Respondent fatigue resulting from a large number of choice questions or respondent inattention resulting from a scenario much too different from respondent's reality.

   iii) Confusion, misunderstanding, assumptions made or heterogeneous interpretation by respondents, poorly constructed attributes or levels

   iv) Unobserved prognostic variable influencing respondent's choices

   Some researchers implement logic tests in the questionnaire, identify those respondents that fail the test(s) and exclude them from the analysis. Ryan et al. (Chapter 9)[120] provides a thorough discussion on that topic. Respondents often have "reasonable" arguments to explain their "irrational" responses. A large fraction of those respondents may lack a consistent and coherent choice criterion across choice sets; some irrational responses were due to strict preferences and some others appeared to have reformulated the experiment in some way in their mental process. Practical considerations are given to reduce the proportion of irrational responses.

4) Labelled alternatives involve an external reference (e.g: a label, a name, a brand…) that provides meaning to the respondent, in addition to the list of characteristics for that alternative. The attributes may involve different levels for each of the labels involved and therefore all levels for that attribute cannot exist for all alternatives; which would alter the efficiency of the design and for which recommendations were provided[150], including the use of "L^{MA} design" implemented by Lancsar[133].

5) Constant alternative refers to the presence of the same alternative, with unchanging attribute levels in all choice sets, and describe a reference condition, a status-quo or an option to not participate. Specific considerations are made when such alteration to optimal design are necessary[146, 147].

6) Block-assignment of respondents to the questionnaire refers to the need of several subjects to answer all questions required by design. The response efficiency may imply a lower number of questions per subjects as compared to the optimal number of questions according to statistical efficiency. The questionnaire is therefore split into several blocks of sub-questionnaires and the total number of subjects will be (ideally) a multiple of the number of blocks.

Sample Size

Rose & Bliemer[151] provide detailed methods for sample size estimation based on asymptotic t-statistics using prior information. The focus is based on "statistical significance" (i.e. type-1 error), considering a maximum of 5% "false-positive risk" to identify a parameter as being "relevant" when that parameter is actually not relevant to the decision.

A valid decision however can be made with a "false-positive risk" higher than 5%; or that the researcher may be more concern about the "false-negative risk" associated with the failure of identifying an opportunity. A Bayesian statistician would potentially consider the distributions of gains (or losses) of the alternatives under considerations based on predictions made using the available information; and would recommend decisions based on maximizing the mean

gain, minimizing the percentile of the distribution of gains below a gain-threshold (i.e. the value-at-risk), minimizing the mean gains below that threshold (i.e. expected value-at-risk); or even consider a combination of approaches.

Sample size considerations are therefore far from being only driven by a formula and depend very much on the questions under consideration and the resources being available. Studies aiming at estimating utilities over the whole population may be limited to a few hundreds of subjects, depending on the prevalence of preferences among the population. Studies aiming at differentiating utilities between different sub-populations may require thousands of respondents. Models for which parameter-interactions are the topic of interest require more subjects than studies for which effects are constant whatever the level of other attributes.

Considering an adaptive approach to the sample size and the experimental design may be valuable to the decision-maker. Respondent recruitment can be stopped based on interim analysis of the data accumulated so far if the decision can be made. The experimental design can be altered to focus on critical attributes for which the precision should be improved. The researcher may use the predictive distributions of the model parameters at interim analysis to anticipate on the consequences of accumulating additional data on the decision to be made and compared them with the costs of increasing the sample size. A maximum sample size would probably be identified for operational reasons but theoretical considerations do not prevent from moving beyond that limit.

The sequential clinical trial design[152] and many other adaptive approaches[153] suggest various adjustments to researchers concerned by inflating the "false-positive risk" as a consequence of making several analyses, or by altering the design of the current study based on preliminary data. Interestingly, none of the ~6200 publications on conjoint analysis in a Scopus search had mentioned "interim analysis" in title, abstract or keywords.

Selection of choice-based conjoint models

A good understanding of the various models that can be used for the analysis of conjoint experiments allows the scientists to use efficient designs when generating the questionnaire. The model classification is made difficult by the absence of unique denomination across the whole conjoint literature[3].

Considerations should be given to the fixed effect model, the random effect model and/or the variance-covariance structure (i.e. parameters used to model the residual errors) when selecting the model that, prior to data collection, seems more relevant for an experiment (see Figure 3.12 below). Providing recommendations on which model to select for a specific application is beyond the scope of this document. More references on conjoint analysis models can be found in [120, 145, 154-156]

Decisions on selecting the fixed effect model should account for the need of interactions between simple effects, but also the number of scenarios to be considered by the respondent, which will lead to a different set of parameters for each scenario. Alternatives for which the label has a specific meaning that may influence choices (like a brand name) should lead also to a specific set of parameters.

The construction of variance-covariance structure should account, for instance, for the likelihood that the probability of selecting the best option from a set of alternatives does not change if a subset of those alternatives is considered instead (e.g. independence of irrelevant alternatives). Also, heterogeneity in preferences among the respondent may lead to considering random parameters and, therefore, mixed effects models (e.g. mixed logits) or hierarchical Bayesian models would be more appropriate.

Logit models used for choice-based conjoint experiments include therefore a large number of parameters and require enough data for an appropriate estimation. In addition, the scientist may be more interested in estimating the probability of selecting each alternative and compare them. For those reasons, sample size calculations are made difficult and usually require simulations[120, 155, 156].

[3] For instance, the wording "hybrid conjoint model" may refer to mixed logit models that address the covariance structure, but may also refer to fixed effect model that presents characteristics of both the conditional logit model, allowing for a same set of parameters for each alternative, and the standard multinomial logit model allowing for a different set of parameters for each alternatives.

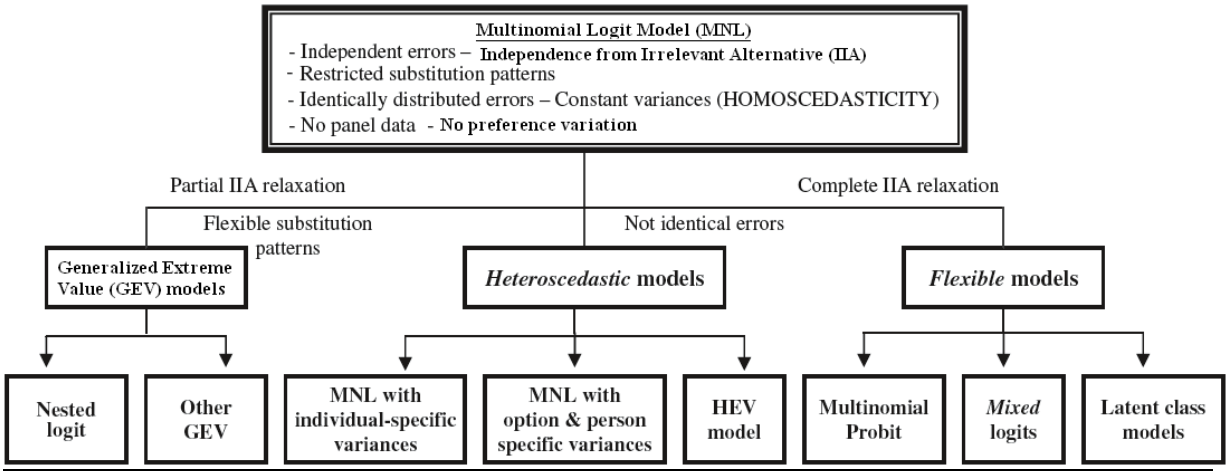


Figure 3.12. Families of choice-based-models and relaxations of assumptions (from[136])

Adaptive choice-based conjoint approaches

The use of web-interface allows the researcher to accrue interim data and make decision on potential (pre-planned) modifications of the questionnaire features. Various approaches to handling adaptive CBC[4] questionnaire have been proposed[134, 141] In those applications, the heterogeneity in the respondent preferences is taken into account. The model assumes that the individual characteristic-values (i.e. path-worths) follow a statistical distribution that fits the between-subjects variability. The researcher is therefore not only interested in the average preferences among the population but also in the preferences of individual respondents. Bayesian estimation procedures are usually applied to such models, which allow the design of future questionnaire in the ongoing study to account for the data already collected.

Other ACBC features include the ability to concentrate the questionnaires on those parameters that present lower precisions and adapt better to the respondent's actual profile. That flexibility may alter the efficiency of the design on other aspects. The time needed to identify the most efficient design may take several minutes to a multi-core multi-threads computer, which cannot be achieved during an interview. Therefore, the researcher will adopt a pragmatic approach to questionnaire building rather than using the most efficient design.

In its commercial Adaptive-Choice-Based-Conjoint software, Johnson & Orme[157] first assess the respondent's optimal choice. The questionnaire is then optimized around that optimal choice in order to improve the respondent's experience, allowing for more attributes and levels

[4] ACBC should not be confused with ACA, an older and much less efficient approach to conjoint experiment.

per attribute to be included in the survey, as compared to standard CBC approach. The respondent will also be screened for any non-compensatory behaviour and the ACBC software will exclude those characteristics (i.e. attribute's levels) from further questions. A near-to-optimal approach to questionnaire design is used to limit the time required between 2 choice sets. Health care applications were limited so far[158-160].

Crabbe et al[123] presents major advances in identifying such design in switching the optimization criteria from (Bayesian) D-optimal to Kullback–Leibler divergence (also called the Kullback–Leibler information or the Kullback–Leibler distance between 2 statistical distributions), much easier to estimate but as efficient to identify effective designs. More specifically, and applied to the discrete choice setting: in order to select the next best choice set for a specific respondent, one maximizes the divergence between the current posterior of the coefficients (obtained with the choice data at hand) and the updated posterior one will obtain with the additional response information from the next choice set.

Finally, using the answers to early questions in a conjoint interview to select later questions may induce (endogeneity) biases in the estimated parameters (i.e. parth-worths) when the researcher overlook the need to account for those answers to early questions in the final analysis. Liu & al. (2006)[161] has shown that including all data collected (i.e. early and later answers) for the analysis of such questionnaires ensures the validity of the results as it adheres to the likelihood principle. Such adaptive procedure however requires advance software that handles properly the variance components of earlier or later sections of the questionnaire. Such approach is implemented for instance by Otter[162] in Sawtooth's ACBC based on his original Matlab implementation, but renders very difficult the proper analysis of the same dataset with another software without appropriate coding of Otter's algorithm.

Random Regret Theories

Chorus (2010) introduced recently a modelling approach based on the notion of regret minimization-driven choice behaviour for analysing data from conjoint experiments. The minimization of anticipated regret may be an important factor when choices are perceived by the individual as difficult and important to them or their relatives[163].

The Random Regret Minimization (RRM) assumes that regret associated with an alternative depends on the performance of each attribute relative to other alternatives in the set of choices proposed to the respondent. In contrast, most of the standard Random-Utility-Maximization (RUM) theories postulate that the utility of an alternative only depends on its own performance[142].

In standard RUM analysis of CBC experiments, effects are coded to represent the levels of categorical variables and (often) numerical are coded to represent the linear or non-linear effects of quantitative variables. In RRM, however, the variables code the differences between an attribute level in a given product profile and the total regret of not being able to choose the more attractive levels for that attribute in the choice set[164]. RRM will give equivalent results to RUM when categorical effects only are included. RRM's interest increases if the model includes quantitative attributes to model as (non-) linear functions.

RRM also has a compromise effect that can allow alternatives with attributes at intermediate levels of utility to perform better in some RRM simulations than in standard RUM CBC simulations. RRM also departs from RUM model in how the choice probability ratios can be greatly impacted by the introduction of new alternatives to the choice sets, although standard RUM exhibit independence towards irrelevant alternatives property.

The RRM approach is implemented in the NLOGIT software and can be coded into software like SAS, Matlab or Gauss. Sawtooth Software users can estimate MNL, Latent Class MNL or hierarchical Bayesian (HB) MNL models for RRM by employing user-specified coding of the variables.

Comparisons made between RRM and RUM[165] do not show definite superiority of RRM versus RUM. RRM seems more appropriate to predict choices between alternatives that are comparable in terms of their attributes. That scenario does not apply, for instance, when a no-choice option (i.e. opting-out the choice-set proposed for trade-off) is available to the respondent. Future applications using RRM may bring clarity on when RUM theory is outperformed. The foundations of regret theory may however be very useful to explain deviations to choice predictions and provide alternatives to RUM models.

4. DISCUSSION and RECOMMENDATIONS

It is essential to understand that pharmaceutical benefit-risk assessment involves both the accurate measurement of the benefits and risks and the value judgments about the relative importance of the various benefits and risks. Based on the current review, the following points for attention and areas for further research were identified.

Qualitative or semi-quantitative framework

It is recognised that the application of any quantitative method must be based on a qualitative framework[1]. The use of such a framework enhances communication and adds transparency and consistency to the process of benefit-risk assessment. Such a framework should preferably support decisions throughout the life cycle of a vaccine and should by suited for use by all stakeholders. The pre-authorization benefit-risk assessment could then be used to inform which health outcomes to monitor during the post-authorization benefit-risk monitoring. The BRAT[25] and the PrOACT-URL framework[34] are currently the most commonly used ones. The UMBRA framework, aiming to unify methodologies for benefit-risk assessment, is still in testing phase. We recommend using (and potentially modifying) these frameworks for the benefit-risk assessment of vaccines. We further recommend investigating the use of evidence grading methodology (e.g. GRADE[38]) for post-authorization benefit-risk assessment because it typically involves the integration of various sources of information of different quality (e.g. clinical trials, observational database analyses, epidemiological studies and infectious disease modelling).

Toolbox of quantitative methodologies

For some benefit-risk assessments, a qualitative or semi-quantitative approach may not be sufficient and quantitative methodologies may be needed. We believe that the various quantitative methodologies described in this report are complementary and that no single approach can cover all issues related to the benefit-risk assessment of vaccines. We advocate the use of a toolbox containing methodologies that can be applied depending on the perspective taken (individual or societal) and on the complexity of the benefit-risk profile (profile dominated by one benefit and one risk, profile with multiple benefits and multiple risks, profile with indirect effects). Furthermore, there is a merit in combining several methodologies. For example, one might think of developing a cohort model and, given weights derived from

utility values, preference studies or other sources, summarizing the model results using incremental net health benefits (INHB) and using (a particular version of) number-needed-to-vaccinate (NNV) for communicating the results of the benefit-risk assessment to vaccine candidates or public health decision makers.

Based on the current appraisal, we initially recommend to further investigate the following methodologies for post-authorisation benefit-risk assessments of vaccines: numbers-needed-to-vaccinate (NNV[54]) including the extension proposed by Tuite & Fisman[55], benefit-risk ratio (BRR), QALYs[72] and DALYs, multi-criteria decision analysis (MCDA[69]) and stochastic extensions[87,88], incremental net health benefit (INHB[68]) and modelling techniques, particularly cohort models[99], dynamic transmission models[107] and multi-parameter evidence synthesis (MPES)[111]. See sections 'concluding remarks' of the current report for argumentation. The majority of these methodologies were also recommended by PROTECT for the benefit-risk assessment of medicines; NNT (NNV in case of vaccines), BRR, QALYs, MCDA including the stochastic extensions and INHB. We additionally recommend the use of cohort models, dynamic transmission models and MPES because these models are specifically developed or already successfully used to model (the impact of interventions on) infectious diseases in a given population.

PROTECT also recommended the use of Q-TWIST[74], impact numbers (PIN, NEPP, PIN-ER-t and related measures[40, 56-60]) and mixed treatment comparison (MTC[108]) for the benefit-risk assessment of medicines. We do not recommend these methodologies for vaccines because Q-TWIST is developed for cancer treatment[74], impact numbers are developed to support resource allocation[40] and do not measure the impact of a vaccination programme[12] and MTC assumes homogeneous study populations[108], which is difficult to achieve for infectious disease studies due to the dynamic nature of infectious diseases. Finally, and contrary to the recommendation by PROTECT to not use DALYs for benefit-risk assessment of medicines, we do recommend the use of DALYs for benefit-risk assessment of vaccines and vaccination programmes. We recommend their use because DALYs are commonly and successfully used to estimate the Burden of Disease of infectious diseases in Europe (BCoDE project funded by ECDC[77]) and to estimate the cost-effectiveness of vaccination programmes (guidelines World Health Organisation[80]). The validity of DALYs is questioned but these concerns are related to the use of DALYs to evaluate life-extending interventions[78] and are not related to vaccination.

Uncertainty

Benefit-risk assessments of vaccines or vaccination programmes might be heavily subject to uncertainty. Potential sources of uncertainty are: stochastic uncertainty, parameter uncertainty, heterogeneity and/or structural uncertainty[115, 116].

We recommend quantifying uncertainty rather than ignoring. Two commonly used techniques to quantify uncertainty are Monte Carlo simulation and Bayesian statistical modelling. We also recommend investigating the use of sensitivity analyses[166] to assess how the uncertainty in the output of benefit-risk assessment can be apportioned to the uncertainty in the different sources of evidence. One might then link the uncertainty apportioned to a specific source of evidence to the strength of the evidence (see recommendation 'qualitative and semi-quantitative framework). This might be useful to identify the sources of evidence having a low evidence grade, though causing substantial uncertainty in the benefit-risk assessment. If this would happen, the benefit-risk assessment is problematic and better evidence should be obtained.

Preference elicitation

Asking healthy subjects about their preferences on favorable and unfavorable aspects of vaccination is much different than asking a patient about aspects of a disease or side effects of a drug. Implementing preference elicitation implies that the groups of respondents are clearly identified together with the topics of the decision they can inform and the procedures that will be followed to prevent bias. Providing guidance on how to set up preference elicitation experiments for vaccines will be a first necessary step. The most appropriate populations to be surveyed and the means of obtaining the preference values should be the focus of such research. Visual tools like influence diagram mapping the various sources of uncertainties can help in identifying the decision parameters for which elicitation may seem relevant.

Some cautions should be taken in how implementations of conjoint experiments, primarily applied to transportation and marketing fields, are extended to health care area. Accepting vaccinations with the promise of also protecting your (more vulnerable) peers at the expense of some burden might involve more subtleties than when collecting preferences for a purchase.

Behavioral attitude, aspects of communication about vaccination to the public and the type of societal values conveyed in the population may influence on how respondents make their choices, and would also influence the design of conjoint experiments to capture them.

Sample sizes and model selections are particularly difficult with conjoint experiment as they determine the design of the questionnaires and, therefore, the limitations of the study itself. Some recommendations or tools would be welcome to guide the practitioner. Adaptive features involving for instance e.g. the list of attributes to consider and the number of respondents, may improve the usefulness of the technique for policy makers and vaccine manufacturers. The evaluations of new decision-making paradigms to the health care area, as compared to random utility theory, are ongoing and their potential usefulness should be clarified further.

The use of conjoint experiments to inform on preferences and utilities of health care interventions has increased over the last 5 to 10 years. The implementation to the vaccine field is challenging and still poses open questions. The technique brings however some unique advantages that clarify the choices made by a population with regard to an intervention that matters to a large number of individuals. We suggest therefore that several types of preference elicitations are tested during proof-of-concept studies in order to identify which approaches fit better for specific scenarios. Among others, MCDA decision-conference, focus groups or the recently proposed clinically-informed simulation approach[126] are probably easier to implement when quick turn-around time is required. Discrete-choice experiment, due to the experimental design and modelling framework to consider, seems to fit better to a more in-depth analysis of preference setting across a population, after preliminary information was collected using the first set of methods.

Limitations

This is a theoretical appraisal of a large set of very heterogeneous methodologies potentially useful for benefit-risk assessment of vaccines. We limited our review to a literature review and review of a number of ongoing projects in the area of pharmaceuticals and devices. We have not explored applications developed in other fields (outside of health care) and may have missed some ongoing initiatives. In addition, a number of statements are made on theoretical grounds only as we are in the early (appraisal) stages of the ADVANCE project. Some of the opinions expressed in this appraisal can be expected to change as more experience is gained in this rapidly evolving field.

References

1. Sashegyi A, Felli, J., Noel, R. Benefit-Risk Assessment in Pharmaceutical Research and Development. Chapman & Hall, 2014.
2. European Medicines Agency (EMA) Benefit-Risk methodology Project, EMA/213482/2010. . 2010.
3. Food and Drug Administration (FDA), Draft PDUFA V Implementation Plan. 2013.
4. Gagnon MP, Desmartis M, Lepage-Savary D, et al. Introducing patients' and the public's perspectives to health technology assessment: A systematic review of international experiences. International journal of technology assessment in health care. 2011; 27: 31-42.
5. European Medicines Agency. Report of the CHMP working group on benefit-risk assessment models and methods. 2007.
6. Innovative Medicines Initiative PROTECT project. 2010.
7. European Medicines Agency. Work package 2 report: applicability of current tools and processes for regulatory benefit-risk assessment. London2010.
8. Guo JJ, Pandey S, Doyle J, Bian B, Lis Y and Raisch DW. A review of quantitative risk-benefit methodologies for assessing drug safety and efficacy-report of the ISPOR risk-benefit management working group. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2010; 13: 657-66.
9. Mt-Isa S, Hallgreen CE, Wang N, et al. Balancing benefit and risk of medicines: a systematic review and classification of available methodologies. Pharmacoepidemiology and drug safety. 2014.
10. Puhan MA, Singh S, Weiss CO, Varadhan R and Boyd CM. A framework for organizing and selecting quantitative approaches for benefit-harm assessment. BMC medical research methodology. 2012; 12: 173.
11. Halloran ME, Haber M, Longini IM, Jr. and Struchiner CJ. Direct and indirect effects in vaccine efficacy and effectiveness. American journal of epidemiology. 1991; 133: 323-31.
12. Hanquet G, Valenciano M, Simondon F and Moren A. Vaccine effects and impact of vaccination programmes in post-licensure studies. Vaccine. 2013; 31: 5634-42.
13. Guidenline on good pharmacovigilance practices (GVP): Product- or population-specific considerations I: Vaccines for prophylaxis against infectious diseases. EMA/488220/2012. European Medicines Agency 2013.
14. Buchanan AE, Brock, D.W. Deciding for others: the ethics of surrogate decision making. Cambridge: Cambridge University Press, 1990.
15. Vietri JT, Chapman GB, Li M and Galvani AP. Preferences for HPV vaccination in parent-child dyads: Similarities and acknowledged differences. Preventive medicine. 2011; 52: 405-6.
16. Larson HJ, Cooper LZ, Eskola J, Katz SL and Ratzan S. Addressing the vaccine confidence gap. Lancet. 2011; 378: 526-35.
17. Lynd LD and O'Brien B J. Advances in risk-benefit evaluation using probabilistic simulation methods: an application to the prophylaxis of deep vein thrombosis. Journal of clinical epidemiology. 2004; 57: 795-803.
18. Towse A. Net clinical benefit: the art and science of jointly estimating benefits and risks of medical treatment. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2010; 13 Suppl 1: S30-2.

19. Noel RA. Systematic Approaches to Benefit-Risk Assessment. In: Sashegyi A, Felli, J., Noel, R., (ed.). Benefit-risk assessment in pharmaceutical research and development. Chapman & Hall, 2014.
20. Levitan B, Mussen, F. Evaluating benefit-risk during and beyond drug development: an industry view. Regulatory Rapporteur. 2012; 9: 5.
21. Levitan B, Cross, J. Pharmaceutical benefit-risk assessment in early development. In: Sashegyi A, Felli, J., Noel, R., (ed.). Benefit-risk assessment in pharmaceutical research and development. Chapman & Hall, 2014.
22. Levitan BS, Andrews EB, Gilsenan A, et al. Application of the BRAT framework to case studies: observations and insights. Clinical pharmacology and therapeutics. 2011; 89: 217-24.
23. Zafiropoulos N, Phillips, L. D., Pignatti, F., Luria, X. Evaluating benefit-risk: an Agency perspective. Regulatory Rapporteur. 2012; 62: 5-8.
24. Levitan B. A concise display of multiple end points for benefit-risk assessment. Clinical pharmacology and therapeutics. 2011; 89: 56-9.
25. Coplan PM, Noel RA, Levitan BS, Ferguson J and Mussen F. Development of a framework for enhancing the transparency, reproducibility and communication of the benefit-risk balance of medicines. Clinical pharmacology and therapeutics. 2011; 89: 312-5.
26. Noel R, Herman R, Levitan B, Watson DJ and Van Goor K. Application of the Benefit-Risk Action Team (BRAT) Framework in Pharmaceutical R&D: Results From a Pilot Program. Drug Information Journal. 2012; 46: 736-43.
27. Rivaroxaban Cardiovascular and Renal Drugs Advisory Committee Briefing Document September 8, 2011. 2011.
28. Rivaroxaban Cardiovascular and Renal Drugs Advisory Committee Briefing Document March 19, 2009 2009.
29. Belatacept Cardiovascular and Renal Drugs Advisory Committee Briefing Document March 1, 2010 2010.
30. Nixon R, Stoeckert I, Hodgson G, Pears J, Tzoulaki I and Montero D. IMI WP5 Report 1:b:iv Benefit-Risk Wave 1 case study report: NATALIZUMAB. 2013.
31. Nixon R, Waddingham E, Mt-Isa S, et al. IMI PROTECT WP5 IMI Report 2:b:iv Natalizumab Wave 2 Case Study Report. 2013.
32. Innovative Medicines Initiative PROTECT Project. 2012.
33. Juhaeri J, Mt-Isa S, Chan E, Genov G, Hirsch I and Bring J. IMI Work Package 5: Report 1:b:i Benefit - Risk Wave 1 Case Study Report: Rimonabant. 2011.
34. Hammond J, Keeney, R., Raiffa, H. Smart choices: a practical guide to making better decisions. Boston: Harvard University Press, 1999.
35. Science CfIiR.
36. Standard Operating Procedure of the German Standing Committee on Vaccinations for the systematic development of vaccination recommendations.
37. Higgins JP, Altman DG, Gotzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. Bmj. 2011; 343: d5928.
38. Guyatt GH, Oxman AD, Schunemann HJ, Tugwell P and Knottnerus A. GRADE guidelines: a new series of articles in the Journal of Clinical Epidemiology. Journal of clinical epidemiology. 2011; 64: 380-2.
39. Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC medical research methodology. 2007; 7: 10.

40. Verma A, Torun P, Harris E, et al. Population Impact Analysis: a framework for assessing the population impact of a risk or intervention. Journal of public health. 2012; 34: 83-9.
41. Ashby D and Smith AF. Evidence-based medicine as Bayesian decision-making. Statistics in medicine. 2000; 19: 3291-305.
42. Hoekstra J, Hart A, Boobis A, et al. BRAFO tiered approach for Benefit-Risk Assessment of Foods. Food and chemical toxicology : an international journal published for the British Industrial Biological Research Association. 2012; 50 Suppl 4: S684-98.
43. Boers M, Brooks P, Fries JF, Simon LS, Strand V and Tugwell P. A first step to assess harm and benefit in clinical trials in one scale. Journal of clinical epidemiology. 2010; 63: 627-32.
44. Laupacis A, Sackett DL and Roberts RS. An assessment of clinically useful measures of the consequences of treatment. The New England journal of medicine. 1988; 318: 1728-33.
45. Schulzer M and Mancini GB. 'Unqualified success' and 'unmitigated failure': number-needed-to-treat-related concepts for assessing treatment efficacy in the presence of treatment-induced adverse events. International journal of epidemiology. 1996; 25: 704-12.
46. Lesaffre E and Pledger G. A note on the number needed to treat. Controlled clinical trials. 1999; 20: 439-47.
47. Hutton J. Number needed to treat: properties and problems. Journal of the Royal Statistical Society Series A. 2000; 163: 403-19.
48. Altman DG. Confidence intervals for the number needed to treat. Bmj. 1998; 317: 1309-12.
49. Guyatt GH, Sinclair J, Cook DJ and Glasziou P. Users' guides to the medical literature: XVI. How to use a treatment recommendation. Evidence-Based Medicine Working Group and the Cochrane Applicability Methods Working Group. JAMA : the journal of the American Medical Association. 1999; 281: 1836-43.
50. Holden WL, Juhaeri J and Dai W. Benefit-risk analysis: examples using quantitative methods. Pharmacoepidemiology and drug safety. 2003; 12: 693-7.
51. Holden WL. Benefit-risk analysis : a brief review and proposed quantitative approaches. Drug safety : an international journal of medical toxicology and drug experience. 2003; 26: 853-62.
52. Riegelman R and Schroth WS. Adjusting the number needed to treat: incorporating adjustments for the utility and timing of benefits and harms. Medical decision making : an international journal of the Society for Medical Decision Making. 1993; 13: 247-52.
53. Mancini GB and Schulzer M. Reporting risks and benefits of therapy by use of the concepts of unqualified success and unmitigated failure: applications to highly cited trials in cardiovascular medicine. Circulation. 1999; 99: 377-83.
54. Kelly H, Attia J, Andrews R and Heller RF. The number needed to vaccinate (NNV) and population extensions of the NNV: comparison of influenza and pneumococcal vaccine programmes for people aged 65 years and over. Vaccine. 2004; 22: 2192-8.
55. Tuite AR and Fisman DN. Number-needed-to-vaccinate calculations: fallacies associated with exclusion of transmission. Vaccine. 2013; 31: 973-8.
56. Attia J, Page J, Heller RF and Dobson AJ. Impact numbers in health policy decisions. Journal of epidemiology and community health. 2002; 56: 600-5.
57. Heller RF and Dobson AJ. Disease impact number and population impact number: population perspectives to measures of risk and benefit. Bmj. 2000; 321: 950-3.

58. Heller RF, Dobson AJ, Attia J and Page J. Impact numbers: measures of risk factor impact on the whole population from case-control and cohort studies. Journal of epidemiology and community health. 2002; 56: 606-10.
59. Heller RF, Edwards R and McElduff P. Implementing guidelines in primary care: can population impact measures help? BMC public health. 2003; 3: 7.
60. Heller RF, Buchan I, Edwards R, Lyratzopoulos G, McElduff P and St Leger S. Communicating risks at the population level: application of population impact numbers. Bmj. 2003; 327: 1162-5.
61. Djulbegovic B, Hozo II, Fields KK and Sullivan D. High-Dose Chemotherapy in the Adjuvant Treatment of Breast Cancer: Benefit/Risk Analysis. Cancer control : journal of the Moffitt Cancer Center. 1998; 5: 394-405.
62. Stinnett AA and Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. Medical decision making : an international journal of the Society for Medical Decision Making. 1998; 18: S68-80.
63. Garrison LP, Jr., Towse A and Bresnahan BW. Assessing a structured, quantitative health outcomes approach to drug risk-benefit analysis. Health affairs. 2007; 26: 684-95.
64. Lynd LD, Marra CA, Najafzadeh M and Sadatsafavi M. A quantitative evaluation of the regulatory assessment of the benefits and risks of rofecoxib relative to naproxen: an application of the incremental net-benefit framework. Pharmacoepidemiology and drug safety. 2010; 19: 1172-80.
65. Minelli C, Abrams KR, Sutton AJ and Cooper NJ. Benefits and harms associated with hormone replacement therapy: clinical decision analysis. Bmj. 2004; 328: 371.
66. Chawla AJ, Mytelka DS, McBride SD, et al. Estimating the incremental net health benefit of requirements for cardiovascular risk evaluation for diabetes therapies. Pharmacoepidemiology and drug safety. 2014; 23: 268-77.
67. Boada JN, Boada C, Garcia-Saiz M, Garcia M, Fernandez E and Gomez E. Net efficacy adjusted for risk (NEAR): a simple procedure for measuring risk:benefit balance. PloS one. 2008; 3: e3580.
68. Craig BA and Black MA. Incremental cost-effectiveness ratio and incremental net-health benefit: two sides of the same coin. Expert review of pharmacoeconomics & outcomes research. 2001; 1: 37-46.
69. Mussen F, Salek,S., Walker, S. Benefit-risk appraisal of medicines. John Wiley & Sons, 2009.
70. Sassi F. Calculating QALYs, comparing QALY and DALY calculations. Health policy and planning. 2006; 21: 402-8.
71. Gold MR, Stevenson D and Fryback DG. HALYS and QALYS and DALYS, Oh My: similarities and differences in summary measures of population Health. Annual review of public health. 2002; 23: 115-34.
72. Weinstein MC, Torrance G and McGuire A. QALYs: the basics. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2009; 12 Suppl 1: S5-9.
73. Ungar WJ. Challenges in health state valuation in paediatric economic evaluation: are QALYs contraindicated? PharmacoEconomics. 2011; 29: 641-52.
74. Gelber RD, Cole, B.F, Gelber, S., Aron., G. Comparing treatments using quality-adjusted survival: The Q-Twist method. The American Statistician. 1995; 49: 161-9.
75. Murray CJ. Quantifying the burden of disease: the technical basis for disability-adjusted life years. Bulletin of the World Health Organization. 1994; 72: 429-45.

76. Murray CJ, Lopez AD and Jamison DT. The global burden of disease in 1990: summary results, sensitivity analysis and future directions. Bulletin of the World Health Organization. 1994; 72: 495-509.
77. Kretzschmar M, Mangen MJ, Pinheiro P, et al. New methodology for estimating the burden of infectious diseases in Europe. PLoS medicine. 2012; 9: e1001205.
78. Airoldi M and Morton A. Adjusting life for quality or disability: stylistic difference or substantial dispute? Health economics. 2009; 18: 1237-47.
79. Mathers CD, Murray CJ, Salomon JA, et al. Healthy life expectancy: comparison of OECD countries in 2001. Australian and New Zealand journal of public health. 2003; 27: 5-11.
80. WHO guide for standardization of economic evaluations of immunization programmes. Immunization, Vaccines and Biologicals. World Health Organization2008.
81. Edwards R, Wiholm BE and Martinez C. Concepts in risk-benefit assessment. A simple merit analysis of a medicine? Drug safety : an international journal of medical toxicology and drug experience. 1996; 15: 1-7.
82. IV CWG. Benefit-risk balance for marketed drugs. Evaluation safety signals. Geneva: Council for International Organizations of Medical Sciences, 1998.
83. Beckmann J. Basic aspects of risk-benefit analysis. Seminars in thrombosis and hemostasis. 1999; 25: 89-95.
84. Chuang-Stein C. A new proposal for benefit-less-risk analysis in clinical trials. Controlled clinical trials. 1994; 15: 30-43.
85. Phillips LD. Benefit-Risk modeling of medicinal products: methods and applications. In: Sashegyi A, Felli, J., Noel, R., (ed.). Benefit-Risk assessment in pharmaceutical research and development. CRC Press, 2014.
86. Mussen F, Salek S and Walker S. A quantitative approach to benefit-risk assessment of medicines--part 1: the development of a new model using multi-criteria decision analysis; part 2: the practical application of a new model. Pharmacoepidemiology and drug safety. 2007; 16 Suppl 1: S42-6.
87. Tervonen T, van Valkenhoef G, Buskens E, Hillege HL and Postmus D. A stochastic multicriteria model for evidence-based decision making in drug benefit-risk analysis. Statistics in medicine. 2011; 30: 1419-28.
88. Wen S, Zhang, L, Yang, B. Two approaches to incorporated clinical data uncertainty into multiple criteria decision analysis for benefit-risk assessment of medicinal products. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2014; Published online.
89. Sarac SB, Rasmussen CH, Rasmussen MA, et al. A comprehensive approach to benefit-risk assessment in drug development. Basic & clinical pharmacology & toxicology. 2012; 111: 65-72.
90. Sarac SB, Rasmussen CH, Afzal S, et al. Data-driven assessment of the association of polymorphisms in 5-Fluorouracil metabolism genes with outcome in adjuvant treatment of colorectal cancer. Basic & clinical pharmacology & toxicology. 2012; 111: 189-97.
91. Poland B, Hodge FL, Khan A, et al. The clinical utility index as a practical multiattribute approach to drug development decisions. Clinical pharmacology and therapeutics. 2009; 86: 105-8.
92. Caro JJ, Briggs AH, Siebert U, Kuntz KM and Force I-SMGRPT. Modeling good research practices--overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--1. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2012; 15: 796-803.

93. Petrou S and Gray A. Economic evaluation using decision analytical modelling: design, conduct, analysis, and reporting. Bmj. 2011; 342: d1766.
94. Critchfield GC and Willard KE. Probabilistic analysis of decision trees using Monte Carlo simulation. Medical decision making : an international journal of the Society for Medical Decision Making. 1986; 6: 85-92.
95. Spiegelhalter D. Incorporating Bayesian Ideas into health-care evaluation. Statistical Science. 2004; 19: 8.
96. Phillips LD, Fasolo B, Zafiropoulous N, et al. Modelling the risk-benefit impact of H1N1 influenza vaccines. European journal of public health. 2013; 23: 674-8.
97. Che D, Zhou H, He J and Wu B. Modeling the impact of the 7-valent pneumococcal conjugate vaccine in Chinese infants: an economic analysis of a compulsory vaccination. BMC health services research. 2014; 14: 56.
98. Rozenbaum MH, Sanders EA, van Hoek AJ, et al. Cost effectiveness of pneumococcal vaccination among Dutch infants: economic analysis of the seven valent pneumococcal conjugated vaccine and forecast for the 10 valent and 13 valent vaccines. Bmj. 2010; 340: c2509.
99. Siebert U, Alagoz O, Bayoumi AM, et al. State-transition modeling: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--3. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2012; 15: 812-20.
100. Thompson JP, Noyes K, Dorsey ER, Schwid SR and Holloway RG. Quantitative risk-benefit analysis of natalizumab. Neurology. 2008; 71: 357-64.
101. Clark M and Cameron DW. The benefits and risks of bacille Calmette-Guerin vaccination among infants at high risk for both tuberculosis and severe combined immunodeficiency: assessment by Markov model. BMC pediatrics. 2006; 6: 5.
102. Cho BH, Clark TA, Messonnier NE, Ortega-Sanchez IR, Weintraub E and Messonnier ML. MCV vaccination in the presence of vaccine-associated Guillain-Barre Syndrome risk: a decision analysis approach. Vaccine. 2010; 28: 817-22.
103. Desai R, Cortese MM, Meltzer MI, et al. Potential intussusception risk versus benefits of rotavirus vaccination in the United States. The Pediatric infectious disease journal. 2013; 32: 1-7.
104. Clark A, Jit M, Andrews N, Atchison C, Edmunds WJ and Sanderson C. Evaluating the potential risks and benefits of infant rotavirus vaccination in England. Vaccine. 2014; 32: 3604-10.
105. Karnon J, Stahl J, Brennan A, et al. Modeling using discrete event simulation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--4. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2012; 15: 821-7.
106. Lynd LD, Najafzadeh M, Colley L, et al. Using the incremental net benefit framework for quantitative benefit-risk analysis in regulatory decision-making--a case study of alosetron in irritable bowel syndrome. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2010; 13: 411-7.
107. Pitman R, Fisman D, Zaric GS, et al. Dynamic transmission modeling: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--5. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2012; 15: 828-34.
108. Lu G and Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Statistics in medicine. 2004; 23: 3105-24.

109. Dias S, Sutton AJ, Ades AE and Welton NJ. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. Medical decision making : an international journal of the Society for Medical Decision Making. 2013; 33: 607-17.
110. Eddy DM. The confidence profile method: a Bayesian method for assessing health technologies. Operations research. 1989; 37: 210-28.
111. Ades AE, Welton NJ, Caldwell D, Price M, Goubar A and Lu G. Multiparameter evidence synthesis in epidemiology and medical decision-making. Journal of health services research & policy. 2008; 13 Suppl 3: 12-22.
112. Goubar A, Ades AE, De Angelis D, et al. Estimates of human immunodeficiency virus prevalence and proportion diagnosed based on Bayesian multiparameter synthesis of surveillance data. Journal of the Royal Statistical Society: Series A (Statistics in Society). 2008; 171: 541-80.
113. Harris RJ, Ramsay M, Hope VD, et al. Hepatitis C prevalence in England remains low and varies by ethnicity: an updated evidence synthesis. European journal of public health. 2012; 22: 187-92.
114. Steenland K and Greenland S. Monte Carlo sensitivity analysis and Bayesian analysis of smoking as an unmeasured confounder in a study of silica and lung cancer. American journal of epidemiology. 2004; 160: 384-92.
115. Bilcke J, Beutels P, Brisson M and Jit M. Accounting for methodological, structural, and parameter uncertainty in decision-analytic models: a practical guide. Medical decision making : an international journal of the Society for Medical Decision Making. 2011; 31: 675-92.
116. Briggs AH, Weinstein MC, Fenwick EA, et al. Model parameter estimation and uncertainty: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--6. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2012; 15: 835-42.
117. Shaffer ML and Watterberg KL. Joint distribution approaches to simultaneously quantifying benefit and risk. BMC medical research methodology. 2006; 6: 48.
118. Shakespeare TP, Gebski VJ, Veness MJ and Simes J. Improving interpretation of clinical studies by use of confidence levels, clinical significance curves, and risk-benefit contours. Lancet. 2001; 357: 1349-53.
119. Mt Isa S, Wang, N., Hallgreen, C, Callréus, T., Genov, G., Hirsch, I., Hobbiger, S., Hockley, K., Luciani, D., Philips, L., Quartey, G. Sarac, S., Stoeckers, I?, Micaleff, A., Ashby, D. Tzoulaki, I. . Review of methodologies for benefit and risk assessment of medication. IMI-PROTECT. London2013.
120. Ryan M, Gerard, K., Amaya-Amaya, M. . using discrete choice experiments to value health and health care. Springer, 2008.
121. Sur D, Cook, J., Chatterjee, S., Deen, J., Whittington, D. . Increasing the transparancy of stated choice studies for policy analysis: Designing experiments to produce raw response graphs. journal of Policy analysis and management. 2007; 26: 10.
122. Sawtooth Software interactive advisor.
123. Crabbe M, Akinc, D, Vandebroek, M. Fast algorithms to generate individualized designs for the mixed logit choice model. Transportation Research Part B: methodological. 2014; 60: 15.
124. Kessels R, Jones, B, Goos, P. Bayesian optimal designs for discrete choice experiments with partial profiles. Journal of choice modelling. 2011; 4: 22.

125. de Bekker-Grob EW and Chorus CG. Random regret-based discrete-choice modelling: an application to healthcare. PharmacoEconomics. 2013; 31: 623-34.
126. Caster O. Quantitative methods to support drug benefit-risk assessment. Department of computer and system sciences. Stockholm university, 2014.
127. Krueger RA, Casey, M.A. Focus groups; a practical guide for applied research (4th edition). Thousand Oaks, CA: Sage Publications, 2009.
128. Sullivan T. Using MCDA (Multi-Criteria Decision Analysis) to prioritise publicly funded health care. New Zealand: University of Otaga, 2012.
129. Grudens-Schuck N, Lundy Allen, B., Larson, K. Focus group fundamentals - methodology: brief.
130. Stewart WD, Shamdasani, P., Rook, D. Focus Groups: Theory and Practice. Thousand Oaks; SAGE Publications, 2007.
131. Lee SJ, Brooks RA, Newman PA, Seiden D, Sangthong R and Duan N. HIV vaccine acceptability among immigrant Thai residents in Los Angeles: a mixed-method approach. AIDS care. 2008; 20: 1161-8.
132. de Bekker-Grob EW, Hofman R, Donkers B, et al. Girls' preferences for HPV vaccination: a discrete choice experiment. Vaccine. 2010; 28: 6692-7.
133. Louviere. A brief history of DCEs and several important challenges. Isaac Newton Institute for Mathematical Sciences. 2011.
134. Orme B. Which conjoint method should I use? 2013.
135. McFadden D. Conditional logit analysis of qualtiative choice behavior. In: Zarembka p, (ed.). Frontiers in econmetrics. New York: Academic Press, 1978.
136. Reed Johnson F, Lancsar E, Marshall D, et al. Constructing experimental designs for discrete-choice experiments: report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2013; 16: 3-13.
137. Bridges JFP, Hauber AB, Marshall D, et al. Conjoint Analysis Applications in Health—a Checklist: A Report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2011; 14: 403-13.
138. Van Houtven G, Johnson FR, Kilambi V and Hauber AB. Eliciting benefit-risk preferences and probability-weighted utility using choice-format conjoint analysis. Medical decision making : an international journal of the Society for Medical Decision Making. 2011; 31: 469-80.
139. Bekker-Grob d. Discrete Choice Experiments in Health Care. Erasmus University Rotterdam, 2009.
140. Lancsar E and Louviere J. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. PharmacoEconomics. 2008; 26: 661-77.
141. Yu J, Goos, P, Vandebroek, M. Individually adapted sequential Bayesian designs for conjoint choice experiments. Journal of choice modelling. 2011; 4: 22.
142. Chorus CG. A new model of random regret minimization. European Journal of Transport Infrastructure Research. 2010; 10: 16.
143. Luyten J, Desmet P, Dorgali V, Hens N and Beutels P. Kicking against the pricks: vaccine sceptics have a different social orientation. European journal of public health. 2014; 24: 310-4.
144. Hershey JC, Asch, D.A. Thumassathit, T, Meszaros, J, Waters, V.V. The roles of altruism, free riding and bandwagoning in vaccination decisions. Organisation behavior and Human decision processes. 1994; 59: 11.

145. Louviere J, Hensher, D, Swait, J. Stated Choice methods: analysis & applications. Cambridge University, 2000.
146. Vermeulen B, Goos, P, Vandenbroek, M. Models and optimal designs for conjoint choice experiments including a no-choice option. International Journal of Research in Marketing 2008; 25: 9.
147. Rose J, Bliemer, M, Hensher, D, Collins, A. Designing efficient stated choice experiments in the presence of reference alternatives. Transportation Research Part B: methodological. 2008; 42: 11.
148. Bliemer M, Rose, J, Hensher, D. Efficient stated choice experiments for estimating nested logit models. Transportation Research Part B: methodological. 2009; 43: 16.
149. Louviere J, Lancsar, E. Choice experiments in health: the good, the bad, the ugly and toward a brighter future. Health Economics, Policy and Law. 2009; 4: 19.
150. Blettner M, Rose, J. Efficient Designs for Alternative Specific Choice Experiments. Institute of Transport and Logistics studies, 2005.
151. Rose J, Bliemer, M. Sample Size requirements for stated choice experiments. Transportation Research Part B: methodological. 2013; 40: 20.
152. Jennison C, Turnbull, B. Group sequential methods: applications to clinical trials. Chapman & Hall, 2000.
153. Chang M. Modern issues and methods in Biostatistics. Springer, 2011.
154. Hesse S. Advanced discrete choice models with applications to transport demand. London: Imperial College London, 2005.
155. Kjaer T. A review of the discrete choice experiment with emphasis on its application in health care. Health Economics Papers. University of Southern Denmark, 2005.
156. Train K. Discrete choice methods with simulation. Cambridge University Press, 2003.
157. Johnson FR, Orme, B. A new approach to adaptive CBC Sawtooth Software Inc in Sequim. Washington2007.
158. Cunningham CE, Deal K and Chen Y. Adaptive choice-based conjoint analysis: a new patient-centered approach to the assessment of health service preferences. The patient. 2010; 3: 257-73.
159. de Groot IB, Otten W, Smeets HJ, Marang-van de Mheen PJ and group C-s. Is the impact of hospital performance data greater in patients who have compared hospitals? BMC health services research. 2011; 11: 214.
160. Whitman CB, Shreay S, Gitlin M, van Oijen MG and Spiegel BM. Clinical factors and the decision to transfuse chronic dialysis patients. Clinical journal of the American Society of Nephrology : CJASN. 2013; 8: 1942-51.
161. Liu Q, Otter, T, Allenby, G. Investigating endogeneity bias in marketing. marketing science. 2006; 26: 8.
162. Otter T. HB-analysis for multi-format adaptive CBC. Sawtooth Software Inc in Sequim. Santa Rosa2007.
163. Zeelenberg M, Pieters, R. A theory of regret regulation. Consumer Psychology. 2007; 17: 15.
164. Chrzan K, Forkner, J. The random regret minimization choice modelling paradigm: an introduction with empirical tests. Sawtooth Software Inc in Sequim. Orem, Utah2014.
165. Chorus CG. Random regret minimization: an overview of model properties and empirical evidence. Transport reviews. 2012; 32: 17.
166. Saltelli A, Chan, K, Scott, E.M. Sensitivity analysis. Wiley Series in Probability and Statistics. New York: John Wiley and Sons, 2000.