www.advance-vaccines.eu

Accelerated Development of VAccine beNefit-risk Collaboration in Europe

IMI JU Grant Agreement nº115557

Proof-of-Concept Benefit-Risk Analyses

Testing a system for new approaches to monitoring benefit/risk with pertussis vaccines as test case.

Part IV: benefit-risk analysis of pertussis vaccines in pre-school children comparing whole-cell and acellular formulations in the post-marketing setting

WP5 – Proof-of-concept of a framework
to perform vaccine benefit-risk monitoring

Disclaimer: *The analyses described in this protocol are conducted as part of the IMI ADVANCE project with the aim to test methodological aspects of the design, conduct and reporting of studies for vaccine benefit-risk monitoring activities.*

The protocol presented herein relates solely to the testing of these methodologies and is not intended to inform regulatory or clinical decisions on the benefits and risks of the exposures under investigation. Therefore any use of information from these studies should carry over this warning and be used accordingly.

V 1.1

Title	Testing new approaches to monitoring benefit/risk with pertussis vaccines as test case. Part IV: benefit-risk analysis of pertussis vaccines in pre-school children comparing whole-cell and acellular formulations in the post-marketing setting
Medicinal product	All available whole-cell pertussis- and acellular pertussis-containing vaccines
Product reference	Any whole-cell pertussis- and acellular pertussis-containing vaccines
Research question and objectives	The overall ADVANCE proof-of-concept (POC) question is to test the system for benefit-risk monitoring of vaccines in Europe. This will first be done by using test cases. For this POC, the following research question is used: "Has the initial benefit-risk profile in children prior to school-entry booster been maintained after the switch from whole-cell pertussis vaccines to acellular pertussis vaccines?" The objectives of this specific analysis, which focuses on testing methods for benefit-risk analysis with pertussis vaccines as test case, are the following: 1. To analyze the benefit-risk balance of pertussis-containing vaccines in children comparing wP and aP formulations at the time of the switch from wP to aP adopting a public health perspective (historical benefit-risk) 2. To investigate the impact of (1) statistical uncertainty in benefit and risk estimates as obtained from the literature, clinical trials, observational databases (uncertainty analyses), (2) differences in preferences and (3) subjective model choices (scenario analyses). 3. To identify the benefit and risk criteria that would most likely modify the benefit-risk balance in case they would change over time (i.e. the pivotal parameters). 4. To assess the feasibility of (retrospectively) monitoring the benefit-risk balance of pertussis-containing vaccines over time (this to mimic prospective monitoring) 5. To re-analyze the benefit-risk balance of pertussis-containing vaccines in children comparing wP and aP formulations from a public health perspective using all currently available evidence (current assessment).
Countries	Participating electronic health care databases from ADVANCE partners and associated partners in Denmark (Aarhus and national), UK (RCGP, THIN), Spain (IDIAP, FISABIO, BIFAP), Netherlands (IPCI/Osiris) and Italy (Pedianet, ASL Cremona), based on quality assessment.
Authors	**B/R POC Team**: Kaatje Bollaerts, Edouard Ledent (preference elicitation), Vincent Bauchau (benefit-risk monitoring) and Jorgen Bauwens, Kartini Gadroen, Alena Khromava, Scott McDonald, Raphaële Roten, Charlotte Switzer, Nicoline van der Maas

	Authors from POC Outline Document (Parts of this protocol have been copied from the POC Outline Document) **Main Authors**: Nicoline van der Maas, Kaat Bollaerts, Denis Macina, Miriam Sturkenboom, Vincent Bauchau **Reviewers**: Simon de Lusignan, Hanne Dorthe Emborg, Mendel Haag, Michael Greenberg, Ulrich Heininger, Alena Khromava, Piotr Kramarz, Xavier Kurz, Harshana Liyanage, Patrick Mahy, Laurence Pagnon, Tin Tin Htar Myint, Marianne van der Sande, John Weil, Eddy Ziani

Study protocol version	Read and approved by (name)	Role	Signature	Date
		Principal investigator		

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ADVANCE	Accelerated Development of VAccine beNefit-risk Collaboration in Europe
aP	acellular (pertussis vaccine)
DALY	Disability-Adjusted Life Years
ELS	extensive limb swelling
HHE	hypotonic-hyporesponsive episode
IMI	Innovative Medicines Initiative
MC	Monte Carlo
MCDA	multi-criteria decision analysis
POC	proof-of-concept
ROC	rank-order centroid
SMAA	stochastic multi-criteria acceptability analysis
wP	whole-cell (pertussis vaccine)

RESPONSIBLE PARTIES

MAIN AUTHOR(S) OF THE PROTOCOL

Name	*Role*	*Contribution*
Kaatje Bollaerts (P95) Edouard Ledent (GSK)	PI	BR modelling preference elicitation
Kaat Bollaerts (P95) Edouard Ledent (GSK) Vincent Bauchau (GSK) by Nicoline van der Maas (RIVM) Scott McDonald (RIVM) Jorgen Bauwens (UNIBAS) Alena Khromava (SP) Miriam Sturkenboom (EMC)	Main authors	 Section 7.1.2 (step5-step6-step7), 7.1.3 Section 8 Section 7.1.1 (step1) Section 7.2 rewriting - editing rewriting Section 5
Miriam Sturkenboom (EMC), John Weil (Takeda), Cedric Mahé (SP), Patrick Mahy (WIV), Pjotr Kramarz (ECDC), Marianne van der Sande (RIVM), Mendel Haag (Novartis), Xavier Kurz (EMA), Laurence Pagnon (SPMSD), Susanne Hartwig (SPMSD), Sandrine Gilhet-Mailfait (SPMSD), Eddy Ziani (SPMSD), Lydie Marcelon (SP), Stéphane Morisset (SP), Denis Macina (SP), Shahrul Mt-Isa (ICL)	Steering Committee members Reviewers BR expert	Review
Miriam Sturkenbooom (EMC), Vincent Bauchau (GSK)	WP5 lead	Review/comments and alignment of different WP5 Pertussis protocols

PRINCIPAL INVESTIGATORS

- Kaatje Bollaerts, P95 Pharmacovigilance and Epidemiology Services
  Address: Vlierbeeklaan 18, 3001 Leuven, Belgium
  Tel: + 32 485 789657
  Email: kaatje.bollaerts@o-95.com

- Edouard Ledent, GSK Vaccines
  Address: Avenue Fleming 20, B-1300 Wavre, Belgium
  Tel: + 32 10 85 85 00
  Email: EDOUARD.Y.LEDENT@GSK.COM

ABSTRACT

Date of Protocol Abstract:
19 November 2015

Title: Testing new approaches to monitoring benefit/risk with pertussis vaccines as test case.

Part IV: benefit-risk analysis of pertussis vaccines in pre-school children comparing whole-cell and acellular formulations in the post-marketing setting.

Observation Period: 01 January 1990 – 31 December 2015

Rationale and Background: This protocol aims to propose and test an approach for synthesizing and near real-time monitoring of coverage, benefits and risks on pertussis vaccination. The key decision question addressed is: Has the initial benefit-risk profile in children prior to school-entry booster been maintained after the switch from whole-cell pertussis (wP) vaccines to acellular pertussis (aP) vaccines?

Research Question and Objectives:

The objectives of this specific analysis, which focuses on testing methods for benefit-risk analysis with pertussis vaccines as test case, are the following:

1. To analyze the benefit-risk balance of pertussis-containing vaccines in children comparing wP and aP formulations at the time of the switch from wP to aP adopting a public health perspective (historical benefit-risk)
2. To investigate the impact of (1) statistical uncertainty in benefit and risk estimates as obtained from the literature, clinical trials, observational databases (uncertainty analyses), (2) differences in preferences and (3) subjective model choices (scenario analyses).
3. To identify the benefit and risk criteria that would most likely modify the benefit-risk balance over time (i.e. the pivotal parameters).
4. To assess the feasibility of (retrospectively) monitoring the benefit-risk balance of pertussis-containing vaccines over time (this to mimic prospective monitoring)
5. To re-analyze the benefit-risk balance of pertussis-containing vaccines in children comparing wP and aP formulations from a public health perspective using all currently available evidence (current assessment).

Methods:

The multi-criteria decision analysis (MCDA) methodology will be used to analyze the benefit-risk balance of wP and aP vaccines at the time of the switch, using all currently available evidence. To construct the MCDA effects table, hypothetical wP and aP cohorts of 1,000,000 children receiving at least one dose of pertussis vaccine followed since the start of pertussis vaccination until pre-school booster will be built. A state transition model will be developed to simulate the number of events for the various benefit and risk outcomes within the hypothetical cohorts. Preference weights will be obtained using MCDA swing-weighting. Uncertainty in outcome measurements and preferences will be explored using stochastic multi-criteria acceptability analysis (SMAA). The incremental net health benefit (INHB) using disability weights approach will be explored for prospective monitoring.

Data Sources:

- Electronic health care databases (record linkage, surveillance and GP-based databases) currently available in the ADVANCE consortium and eligible are located in Denmark, Spain, Italy, The Netherlands and UK

• Informative data sources: ECDC pertussis schedules in Europe and switch points of national ministries of health
Population: Children from birth until their school-entry pertussis booster if any (4th or 5th dose) within all eligible ADVANCE databases
Informed Consent and Ethical Approval: This analysis will be conducted on the basis of secondary use of electronic healthcare records. Each database will apply local governance and privacy rules prior to aggregating and sharing anonymized data.
Milestones: Draft protocol: July 31 2015 Submission to SC: August 6, 2015 Comments from SC: August 31, 2015 Submission for stakeholder consortium review: September 2015 Finalized and cleared protocol: November 20 September 30 2015 Submission to Ethics Committee/Institutional Review Board: January 2016 Updated protocol after review: April 15, 2016 Final data extraction to CDM: June 15, 2016 Running scripts and submission to RRE: June 30, 2016 Data analysis: July 2016 Data interpretation and reporting: August 2016 Final report of study results: September 2016

AMENDMENTS AND UPDATES

Protocol amendments following IRB approval:

Table 1: Overview of Protocol Amendments and Updates

Number	*Date (DDMMMYY)*	*Section of the study protocol*	*Amendment or update*	*Reason*
1				
2				
….				

MILESTONES

Table 2: Overview of Milestones

Milestones:
Draft protocol: July 31 2015
Submission to SC: August 6, 2015
Comments from SC: August 31, 2015
Submission for stakeholder consortium review: September 2015
Finalized and cleared protocol: November 20 September 30 2015
Submission to Ethics Committee/Institutional Review Board: January 2016
Updated protocol after review: April 15, 2016
Final data extraction to CDM: June 15, 2016
Running scripts and submission to RRE: June 30, 2016
Data analysis: July 2016
Data interpretation and reporting: August 2016
Final report of study results: September 2016

1. RATIONALE AND BACKGROUND

The ADVANCE vision is to deliver "best evidence at the right time to support decision-making on vaccination in Europe". The mission is to establish a prototype of a sustainable and compelling system that rapidly provides best available scientific evidence on vaccination benefits and risks post-marketing for well-informed decisions. In light of this goal, the ADVANCE platform aims to provide evidence on the benefits and risks of vaccines to support decision-making by all stakeholders in a wide range of contexts. Examples of scenarios are the inclusion of a new vaccine in a vaccination program, and the occurrence of a new safety issue, e.g. when the benefits of the vaccine are questioned or when a new population is targeted (see Pertussis POC Outline).

The concept this POC analysis aims to demonstrate is as follows: in the event that an important decision regarding a health intervention is to be made, a benefit-risk assessment will be carried out. Upon a favorable benefit-risk assessment, the health intervention is implemented and the benefits and risks are monitored to investigate whether the benefit-risk balance is changing over time. The benefit-risk monitoring may focus primarily on the benefits and risks that could potentially modify the benefit-risk balance. If there is a strong indication that the benefit-risk has changed over time, a full re-assessment of the benefit-risk balance of the health intervention may be triggered using all accumulated evidence available at that point in time. To inform the benefit-risk assessment and monitoring, electronic health care databases available within Europe will be used.

To be able to prove this concept of benefit-risk monitoring in ADVANCE without waiting for the evidence to accumulate prospectively, we will start from a historical decision and simulate monitoring through a retrospective analysis. Pertussis vaccination, particularly comparing wP and aP vaccination, was chosen by the ADVANCE Steering Committee as the subject of the first POC analysis. Therefore, the starting point of the current POC analysis is the historical decision to switch from wP to aP vaccination in children in the pioneering countries.

Within this POC analysis , three main phases are distinguished:

- **Historical benefit-risk analysis**: First, the benefit-risk analysis around the time of the switch from wP to aP was made will be simulated, under a public health perspective. Only information available at that time will be used, which is information from clinical trials, literature and possibly observational databases for wP vaccination, and information from clinical trials for aP vaccination. A sensitivity analysis will identify pivotal parameters, i.e. parameters that might modify the benefit-risk balance.
- **Benefit-risk monitoring using retrospective analyses**: Second, the monitoring of the pivotal benefits and risks parameters of aP vaccination over time will be simulated, as compared to the benefits and risks of wP vaccination based using retrospective analyses in observational databases.
- **Current benefit-risk analysis**: Third, a benefit-risk analysis will be performed using all currently accumulated evidence as if re-assessment was triggered today. We will use information from clinical trials, literature and observational databases. The decision to opt to act as if a trigger occurred today is taken pragmatic reasons. This allows the performance of

the multi-country observational databases within the ADVANCE consortium to be tested without having to wait for a trigger from the benefit-risk monitoring.

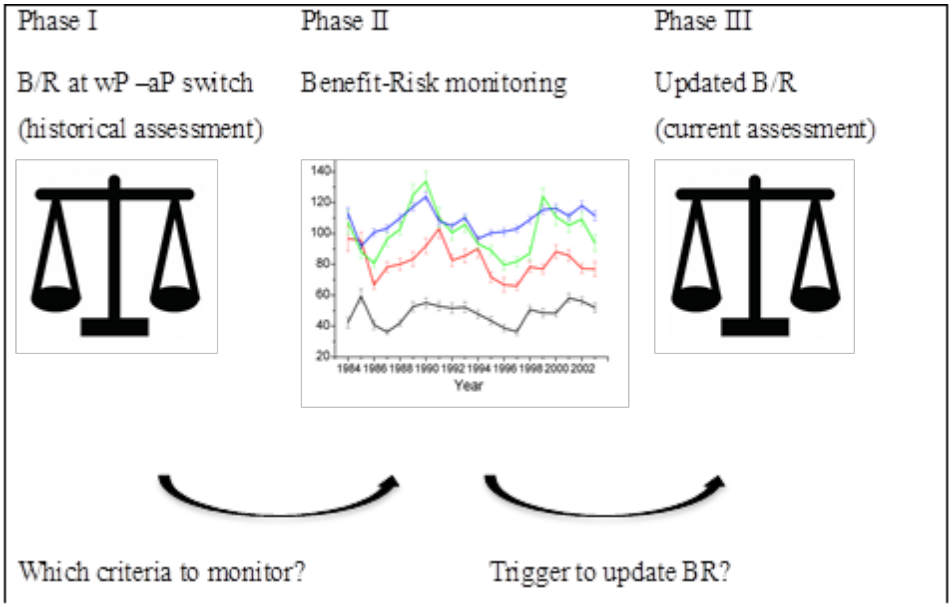


Figure 1: Schematic Overview of the Concept of Monitoring the Benefit-Risk of Vaccination Post-licensure

This protocol is about analyzing information on coverage, benefits and risks that will be collected and made available by the different POC pillars. Since data will be derived from different sources and with different designs, these pillars have separate protocols. The different protocols are aligned to ensure that the data are collected as required for the benefit-risk analyses proposed in this protocol.

2. RESEARCH QUESTION AND OBJECTIVES

The overall ADVANCE POC question is to test the currently available system for benefit-risk monitoring of vaccines in Europe. The system will be tested around the following B/R question: Has the initial benefit-risk profile in children prior to school-entry booster been maintained after the switch from wP to aP pertussis vaccines?

2.1. Process and methodology for system-testing

Although the system testing will occur largely outside of this study, it is summarised here.

The system testing follows several steps which are visualized in the figure 2 and described in the following chronological order:

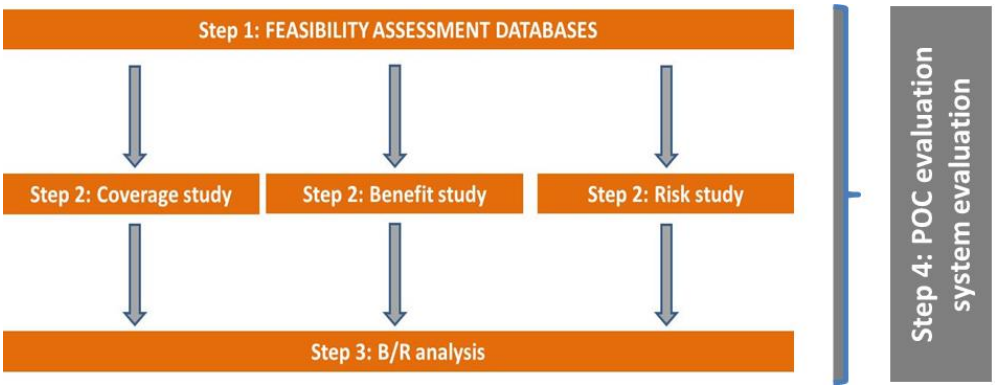


Figure 2 Visualization of the stepwise approach to the system testing

Step 1: Feasibility assessment of the databases: this step will assess whether the quality of the candidate database is sufficient inclusion in the study. The focus will be on what type of data are available in the databases and whether population, events, and exposure may be misclassified. This step is largely based on the so-called fingerprinting which has been described in deliverable D5.2. A summary of the components and methods is provided in appendix 1. A quality assessment summary will be created per database, with decisions whether the data-sources can or cannot participate in the different pillar studies (see below under 7.2). Responsibility of fingerprinting lies with the workpackage leaders.

Step 2: Estimation and delivery of the rates for coverage, benefits and risks, this is described in the different 'pillar' protocols in the databases that may generate adequate results according to the feasibility assessment. Responsibilities are with the study teams that have generated the protocols

Step 3: B/R analysis: integration of the incidence rates (generated from step 2) with the utilities to generate a B/R model, as described in the B/R analysis protocol, responsibility with the B/R study team

Step 4: Evaluation of the studies and the systems used. This is conducted by a POC evaluation team which is separated on purposes from the POC study teams. A description of the framework for the POC evaluation is attached in appendix 2.

The purpose of this protocol is to describe in detail the methods for the B/R analysis (step 3). In particular, the objectives of the B/R analysis, which focuses on testing methods, are the following: To investigate the impact of

- statistical uncertainty in benefit and risk estimates as obtained from the literature, clinical trials, observational databases (uncertainty analyses),
- differences in preferences
- subjective model choices, such as different case definitions (scenario analyses).
- To identify the benefit and risk criteria that would most likely modify the benefit-risk balance over time (i.e. the pivotal parameters).

- To assess the feasibility of (retrospectively) monitoring the benefit-risk balance of pertussis-containing vaccines over time (this to mimic prospective monitoring). This will be further developed within WP4 on benefit-risk monitoring.
- To re-analyze the benefit-risk balance of pertussis-containing vaccines in children comparing wP and aP formulations from a public health perspective using all currently available evidence (current assessment).

3. RESEARCH METHODS

To demonstrate the concept of benefit-risk monitoring of vaccinations, different methodologies will be used (see Figure 2 for a schematic overview). For the historical and current benefit-risk analysis, the MCDA methodology with MCDA swing-weighting will be used to elicit preferences from representatives of public health and regulatory authorities, and from family physicians/clinicians (candidate vaccine recipients or their parents are not included for this POC for simplification reasons). The public health authorities are the main decision makers. The preferences from the regulatory authorities and family physicians/clinicians are elicited as well to inform the decision making by the public health authorities and to test the methodology of eliciting preferences from different stakeholders. Monte Carlo (MC) simulation will be performed to investigate the impact on the benefit-risk balance of: (1) statistical uncertainty in the benefit and risk estimates (uncertainty analyses), (2) differences in preference, and (3) subjective model choices (e.g. different case definitions). Additional sensitivity analyses will be performed to identify the pivotal benefit and risk outcomes.

Disability weights for the various benefit and risk criteria will also be extracted to be used for monitoring the benefit-risk and for validating the preferences obtained using MCDA swing-weighting (section 7.2). For the benefit-risk monitoring, roll-back analyses will be used that monitor the 'pivotal' outcomes periodically.

Details on the methods and the rationale for their choice are given in the subsequent sections.

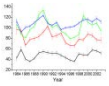
Phase I	Phase II	Phase III
Historical Benefit-Risk assessment (wP – aP switch) 	Benefit-Risk monitoring 	Current Benefit-Risk assessment 
• MCDA • MCDA swing-weighting • Disability weights • MC: uncertainty, preferences, subjective model choices, pivotal outcomes	Periodic monitoring of critical outcome rates	• MCDA • MCDA swing-weighting • Disability weights • MC: uncertainty, preferences, subjective model choices, pivotal outcomes

Figure 2: Schematic Overview of Methodology Tested for Monitoring the Benefit-Risk of Vaccination Post-licensure

3.1. Multi-criteria Decision Analysis

3.1.1. Rationale

The benefit-risk assessments will be carried out following MCDA methodology. Among the benefit-risk methodologies for medicines as appraised by PROTECT and for vaccinations appraised by the ADVANCE consortium, MCDA is the only quantitative framework sufficiently flexible to use for post-marketing benefit-risk evaluation. Its use for pharmaceutical benefit-risk assessment has been recommended by the CMPH working group on benefit-risk methods [1] and the PROTECT consortium [2]. MCDA has also been used when subjective preferences are involved in decision-making; good practice guidelines are currently being developed (International Society for Pharmacoeconomics and Outcomes Research [ISPOR] Good Practices Task Force: MCDA in health care decision making).

MCDA has been developed to support decision makers in structured evaluation and explicit selection of alternatives. It is a quantitative methodology for appraising alternatives on individual, often conflicting criteria and combining them into one overall appraisal, through incorporating elicited preferences (weights). Its use is particularly recommended to assess and compare multiple alternatives on multiple criteria. MCDA methods are valuable to systematically identify where the differences in preferences between persons are and to achieve consensus necessary for decision-making. In addition, MCDA (in combination with its stochastic extensions, such as SMAA and probabilistic sensitivity analyses [PSA]) allows sensitivity analyses to be conducted to investigate the impact of uncertainty in benefit, risk and preference estimations on the overall benefit-risk balance.

Within this POC, MCDA will be used for the historical and current benefit-risk analyses because:

- Within this POC, multiple alternatives (i.e. wP and aP vaccination) will be compared on multiple benefit and risk criteria.
- Different stakeholder groups (e.g. public health authorities, regulatory authorities, health care providers) who may have different preferences are involved in vaccine-related decision-making. MCDA makes any differences in preference explicit and requires consensus to be reached in order to establish the benefit-risk balance.
- There may be substantial uncertainty about the available estimates on the benefits and risks from multiple countries and data sources.
- The sensitivity analyses can be tailored to identify the benefit and risk parameters that might modify the benefit-risk balance (i.e. the pivotal parameters), which will be subsequently prioritized for benefit-risk monitoring.

The use of MCDA in the context of periodic monitoring is an open question, which will be addressed by this analysis.

Preferences will be elicited during a decision conference from the following stakeholder groups:

1. Public health authorities (main decision makers),
2. Regulatory authorities, and
3. Family physicians/clinicians.

As the Steering Committee had decided to adopt a public health perspective for the current POC, the public health authorities are the primary stakeholder group. Preferences from regulatory authorities and family physicians/clinicians are elicited as well, and these can be taken into account by public health authorities in their decision-making. The family physicians/clinicians are considered to represent the (candidate) vaccine recipients.

Although the importance of patient and public involvement in healthcare/medical decision-making is recognized, preferences from candidate vaccine recipients or their parents will not be elicited in the current POC. Direct preference elicitation from candidate vaccine recipients is believed to be more difficult as this requires a proper understanding of the clinical relevance of both the benefit and risk outcomes. Indirect preference elicitation (such as discrete choice experiments - which are more resource and time consuming) seems more appropriate to use with healthy subjects.

In particular, preferences will be elicited for MCDA using a value measurement model, involving swing-weighting. Thokala et al. [3] describe the merits of three different approaches to MCDA (Table 3). As compared to outranking methods and goal programming, the value measurement model is the easiest approach to propagate uncertainty in outcome measurements and is therefore the preferred choice.

Table 3: Comparison of Different Multi-Criteria Decision Analysis Approaches

	Value measurement models	*Outranking approach*	*Goal programming*
Weights	Swing weights are used to capture both the effect of measurement scales and the importance of the criteria. Weights need to satisfy the preferential independence of criteria and the trade-off requirements.	Weights are uninfluenced by the scale of the value functions. They convey the relative importance of criteria in the assertion that one alternative is better than the other. Weights do not have to satisfy any conditions	Weights are attached to the deviations and represent the relative importance of criteria by specifying an overall measure of deviations from the goals. Weights do not have to satisfy any conditions.
Measuring the performance of the criteria	Performance scores $v_i(a)$, monotonic functions of the attribute values $z_i(a)$, need to be developed for all criterion i. Significant effort is needed to develop these performance scores.	Outranking approach can use either performance value scores $v_i(a)$ or the attribute values $z_i(a)$, saving on the effort needed to develop performance scores.	Goal programming method operates directly on the attribute values, $z_i(a)$. No need to develop performance scores.
Complexity of the MCDA model	Weighted sum approach is easy to understand and use by the decision makers. The parameters can be changed in real time to observe their effect.	Intuitive and easy to follow. With right software, assumptions can be changed and results can be observed almost instantaneously.	Easy to understand but requires significant computational time to provide results. Real-time updating is not possible.
Presentation of the results	Easy to follow and enables further deliberation, well suited for good visual presentation of the results.	Moderately easy to follow, can be presented visually but difficult with multiple alternatives.	Results easy to follow, but they cannot be represented visually.
Incorporating uncertainty	Probabilistic sensitivity analysis can be used to propagate parameter uncertainty quite easily.	Moderately difficult to include uncertainty, needs specialist software.	Quite difficult to include uncertainty, complex stochastic programming techniques are needed.

Source: Thokala et al [3]

3.1.2. Multi-criteria Decision Analysis: Different Steps

There are eight steps involved in developing and exploring an MCDA model (see Table 4) [4]. Some of these steps were already described in the Pertussis POC Outline (see Appendix 1), which was approved by the Steering Committee and is the basis for this protocol.

Table 4: Steps in Conducting a Multi-Criteria Decision Analysis, Linkage to the Different Pertussis POC Analyses and Data Sources

MCDA STEP	*Historical Benefit-Risk (switch wP to aP)*	*Current BR*
1. **Context**: establish the decision context and describe the perspective	Public health decision: switching from wP to aP?	Public health: is the benefit-risk balance of aP still more favorable compared to wP?
2. **Alternatives**: identify the alternatives to be appraised	wP (all marketed formulations) versus aP (all marketed formulations) pertussis vaccines	
3. **Criteria**: identify and define the benefit and risk criteria and organize in a value tree	Benefits: pertussis disease, pneumonia, seizures, death. Risks: injection site reactions, fever, somnolence, persistent crying, convulsions, hypotonic-hyporesponsive episodes (HHE) and extensive limb swelling (ELS).	
4. **Scoring**: criteria measurements, assess the performance of each alternative against the criteria (effects table)	Primary sources: 1. Clinical development (as public) 2. Literature	Primary sources: 1. Clinical development (as public), 2. Literature 3. Observational databases (POC benefit protocol, POC risk protocol)
5. **Value functions**: transform the scores to preferences on the 0-1 scale	The type of value functions will be decided upon during the decision conference.	
6. **Weighting**: assign a weight to each criterion	Types of weights: 1. MCDA swing weights elicited during the decision conference 2. Disability weights (Disability-Adjusted Life Years [DALYs])	
7. **Results**: calculate results and provide graphs	Incremental Net Health Benefit (INHB)	
8. **Sensitivity analysis:** explore the effects of uncertainty on the benefit-risk balance	- MC simulation to investigate uncertainty in criteria measurements. - SMAA to investigate the uncertainty/heterogeneity in preference measures. - Scenario analyses if required. - 'Pivotal' analyses to identify the criteria that might potentially modify the benefit-risk balance.	

Step 1: Establish the Decision Context

Whole-cell (wP) pertussis vaccines were used first and are typically available in combination with diphtheria (D) and tetanus (T) vaccines. Concerns regarding the reactogenicity and neurological effects of the wP component in DTwP vaccines have been expressed but the neurological effects could never be proven (and therefore will not be addressed in this POC). To address adverse reactions observed with wP vaccines, aP vaccines were developed. The aP vaccines have gradually supplanted the use of wP vaccines in industrialized countries. As evidence emerged that the duration of vaccine protection was less than expected, based on the results of clinical trials, several countries introduced additional booster doses later in life. Those booster doses further influenced/changed the pertussis epidemiology.

In this POC, we will investigate whether the initial benefit-risk profile associated with wP vaccination was maintained after the switch from wP to aP, or changed over time during the use of aP vaccines. For the sake of simplicity, and to focus on proving the concept of monitoring benefit-risk post-licensure, the population of interest is children from first dose until their school-entry pertussis booster, for the reason that the introduction of booster doses hampers the comparison of benefits during the wP and aP vaccine era beyond childhood.

In this POC, public health authorities are selected as the decision-making body, i.e. deciding on whether to use wP or aP in their vaccination programs. However, there may be differences in perspectives from other stakeholder groups that need to be taken into account when making their decision. Therefore, preferences from other stakeholder groups (i.e. regulatory authorities and family physicians/clinicians) will be elicited in this POC as well.

Step 2: Identify Alternatives to be Appraised

Two alternatives will be considered:

1. Any marketed vaccines containing aP (any brand), any series of doses recommended by the National Immunization Programs in the ADVANCE participating countries
2. Any marketed vaccines containing wP (any brand), any series of doses recommended by the National Immunization Programs in the ADVANCE participating countries

Because vaccine types and not vaccine brands are compared, this comparison rather reflects a public health perspective than a regulatory perspective.

The schedule of the primary series depends on the countries and has been changed over time. Details on changes in type of pertussis-containing vaccine and coverage will be provided in the protocol developed by the POC team on coverage.

Step 3: Identify and Define the Benefit and Risk Criteria and Organize in a Value Tree (Outcome Tree)

The full list of benefit and risk criteria that will be used to compare aP and wP vaccines are graphically organized in the outcome tree shown in Figure 3. Within the current Pertussis POC, the benefits and risks of pertussis vaccination in children from first dose until pre-school booster will be compared, with a focus on direct effects only for simplicity reasons. The benefits include

prevention of pertussis disease, and complications (pneumonia and seizures) and deaths due to pertussis. In this POC, the risks are limited to the known adverse reactions and include injection site reactions, fever, somnolence, persistent crying, convulsions, HHE and ELS. The case definitions, time windows and codes are described in the risk and benefit protocols that will generate the data for the benefit-risk analysis. The case definitions are also included Appendix 2.

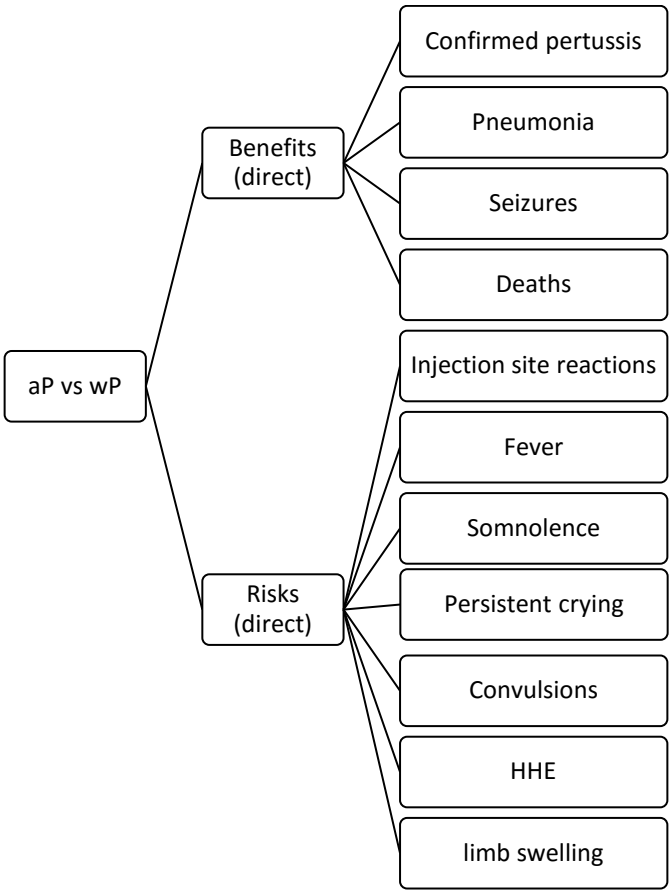


Figure 3: Pertussis-containing Vaccine Benefit-Risk Outcome Tree

Step 4: Scoring: Assess the Performance of Each Alternative Against the Criteria (Health Outcomes)

The benefits and risks of both aP- and wP-containing childhood vaccinations will be estimated through the incidence rates for the various health outcomes listed in the outcome tree.

- 'Historical' benefit-risk analysis: information on incidence rates for each of the outcomes in the outcome tree will be retrieved from the literature and electronic health care databases when these cover the time period of interest.
- 'Current' benefit-risk analysis: information on incidence rates for each of the outcomes in the outcome tree will be retrieved from the literature and electronic health care databases.

To facilitate the comparison of different health outcomes (very rare versus common outcomes) hypothetical cohorts of 1,000,000 children receiving at least one dose followed up since the start

of pertussis vaccination until pre-school booster will be built. The hypothetical population is assumed to be static, meaning that no new subjects will enter the hypothetical cohort at a later stage than first dose, and that no subjects will leave the hypothetical cohort before the age of pre-school booster, with the sole exception that deaths due to pertussis might occur. Two hypothetical cohorts will be built; one with aP-vaccinated children and one with wP-vaccinated children. The two hypothetical cohorts are assumed to be identical, except for the pertussis vaccine.

For each hypothetical cohort, the number of cases per health outcome will be estimated during the period from first dose until pre-school booster. To this end, a state transition model will be built: every child within the hypothetical cohorts will move through various periods (i.e. 'states') one at a time. The states are defined by age at vaccination and time (Figure 4). Each child within each state has a certain probability of developing the event. This probability depends on the duration of the state and on the age-specific incidence rates.

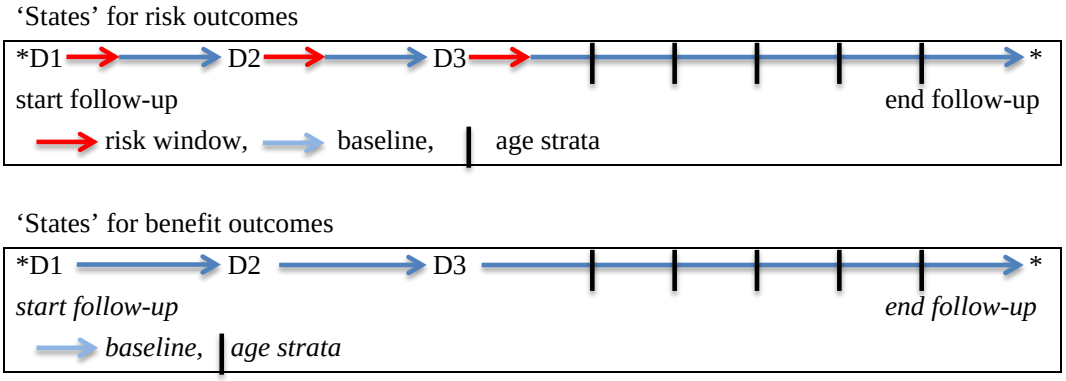


Figure 4: Pertussis Vaccination-related States from First Dose until End of Follow-up

Note: for the various benefit and safety outcomes, the relevant benefit-risk windows might differ

For every child and for every outcome, the number of events within each state will be simulated, and the sum of the events over the states/children taken to obtain the number of events within the hypothetical cohorts. The number of estimated cases of safety and benefit outcomes within each hypothetical cohort will be used to build the effects table. An outline of the effects table is given in Table 5. In addition, age at disease onset and duration of disease will be obtained to describe the criteria (outcomes) to the experts participating in the decision conference.

The information needed to inform the parameters of the hypothetical cohorts will be retrieved from the literature and from the databases. Information on duration of disease will be obtained from the literature only.

Table 5: Outline of the Effects Table: Estimated Number of Cases for the Various Benefits and Risks in Hypothetical Cohorts of 1,000,000 wP- and aP-vaccinated Children Followed from First Dose Until Pre-School Booster; by Geographical Area, and Calendar Time

	wP			*aP*		
	Number of cases	*Duration of disease**	*Age at disease onset*	*Number of cases*	*Duration of disease**	*Age at disease onset*
	est. (95% CI)	*Mean (IQR)*	*Mean (IQR)*	*est. (95% CI)*	*Mean (IQR)*	*Mean (IQR)*
Benefits						
Pertussis disease						
Pneumonia						
Seizure						
Death						
Risks						
Injection size reaction						
Fever						
Somnolence						
Persistent crying						
Convulsion						
Hypotonic-hyporesponsive episode						
Extensive limb swelling						

*Obtained from the literature only

Step 4a: Source – Literature

Information on incidence rates for each of the outcomes in the outcome tree will be retrieved from the clinical trials and observational studies. Priority will be given to studies on children under the age of six conducted in Western countries. If these cannot be found, we will look at other age groups and studies conducted in African and Asian countries as well. The following systematic reviews will be the primary sources:

1. Zhang L, Prietsch SO, Axelsson I, Halperin SA. Acellular vaccines for preventing whooping cough in children. Cochrane Database Syst Rev 2014;9:CD001478.
2. Pertussis vaccines: WHO position paper - September 2015. Wkly Epidemiol Rec 2015;90:433-58.
3. Lambert LC. Pertussis vaccine trials in the 1990s. J Infect Dis 2014;209 Suppl 1:S4-9.
4. Herzog C. Changing from whole-cell to acellular pertussis vaccines would trade superior tolerability for inferior protection. Expert Rev Vaccines 2015;14:1065-72.
5. Jefferson T, Rudin M, DiPietrantonj C. Systematic review of the effects of pertussis vaccines in children. Vaccine 2003;21:2003-14.

If appropriate, information from different studies will be pooled using a random effects meta-analysis. For the current benefit-risk model, an additional literature/clinical trial search for studies/trials published after 2014 will be conducted.

For the historical benefit-risk model, only information prior to ~1995 will be used (to be exactly determined based the average year of switch for the European countries). The rationale for this choice is that several European countries switched from wP to aP around this time; earlier switches were based on limited data whereas for later switches post-marketing information was available.

Step 4b Source – Observational Databases

Information on incidence rates for each of the outcomes in the outcome tree will also be retrieved from the observational data. The information that will be provided by the coverage pillar is given in Table 6. This information will be used to build cohort models reflecting real-life vaccine usages (i.e. age at vaccination and number of doses).

Table 6: Frequency Table of the Number of Vaccinated Subjects by Age at First Dose, Time Since First/Second Dose at the Subsequent Doses, by Year of Birth and Country

Year of Birth	*Dose 1*	*Dose 2*	*Dose 3*	*Frequency*
	Age (months)	*Time since first dose (wks)*	*Time since second dose (wks)*	
199X				
199X				
199X				
199X				
199X				
…	…	…	…	…
200X				

The specific evidence required from the observational databases to inform the cohort model and to subsequently build the effects table is given in Table 7 for the safety outcomes and in Table 8 for the benefit outcomes. The incidence rates (or baseline rates multiplied with a measure of relative risk) will be used to simulate the number of events within the hypothetical cohorts using state-transition modelling. For every health outcome, estimates and confidence intervals are provided by country and by birth year. Time trends will be investigated by the benefit and risk POC teams. It is still an outstanding issue whether data from different databases/countries can be pooled or not. If pooling of data is appropriate, a 'European' benefit-risk model will be developed. If this cannot be done, country-specific benefit-risk models will be developed.

Table 7: Evidence Required from the Databases for Safety Outcomes to Build the Effects Table by Outcome, Country, and Year of Birth

Incidence in risk period after aP or wP vaccination						
	*Dose 1**		*Dose 2*		*Dose 3*	
*Time since dose***	*Inc*	*95% CI*	*Inc*	*95% CI*	*Inc*	*95% CI*
Day 0						
Day 1-3						
Day 4-7						
Baseline incidence						
2-3 months						
3-4 months						
4-5 months						

* number of doses might depend on the country and birth year

** risk windows might be different depending on the safety outcome

Table 8: Evidence Required from the Databases for Benefit Outcomes to Build the Effects Table by Outcome, Country, and Year of Birth

	Unprotected [a]		*Dose 1* [b]		*Dose 2* [c]		*Dose 3* [d]	
	Inc	*95% CI*	*Inc*	*95% CI*	*Inc*	*95% CI*	*Inc*	*95% CI*
Whole cell								
Pertussis disease (not resulting in complications/death)								
Complications [e] (not resulting in death)								
Death [f]								
Acellular								
Pertussis disease (not resulting in complications/death)								
Complications [e] (not resulting in death)								
Death[f]								

a Pertussis from start follow-up until 4 weeks after first dose

b Pertussis from 2 weeks after Dose 1 until 2 weeks after Dose 2

c Pertussis from 3 weeks after Dose 2 until 2 weeks after Dose 3

d Pertussis from 3 weeks after Dose 3 until school-enter booster/age 6

e Complications until one month after each dose

f Death until three months after each dose

Step 5: Create Value Functions

This step it will decided how to convert the scores (in this case the number of cases within the hypothetical cohort) to preference values on a 0 to 1 scale.

First, the range of the scores needs to be defined. This range will be based on the mean or median vaccine evaluations and should not be influenced by the level of precision of the criteria measurements. Large confidence intervals may influence the expert in the assessment of the importance (weights) to be given to the criterion during the swing-weighting process.

Then, the value function, which maps the scores (*x*-axis) to the preferences (*y*-axis), needs to be defined. The simplest method is the direct linear conversion, by which differences in scores are preserved. A concave or convex function shortens the distance on the preference scale (*y*-axis) for the range of scores (*x*-axis) where the function is flat: a change on the *x*-axis translates into a very small change on the *y*-axis. For the range of scores where the curve increases or decreases sharply, a small change on the *x*-axis translates into a substantial change on the *y*-axis.

During the decision conference, participants will decide on the shape of the value functions using a software tool by which they can explore the shape of the value function (i.e. linear, convex, concave, s-shaped, piecewise linear) by altering two sliders.

Step 6: Assign a Weight to Each Criterion Using Swing-Weighting

In MCDA swing-weighting, the value of a change on the preference scale of one (benefit or risk) criterion is compared with the value of a change of another criterion. The changes or 'swings' over these scales are considered by the decision-makers to assign weights. The theoretical background of the consistency between preferences and weights is provided in Appendix 4.

The following hypotheses are made when applying MCDA swing-weighting:

1. **Compensatory behavior of the expert**: The expert agrees that a loss on one endpoint can be compensated by an increase in another endpoint. On the contrary, if a range of evaluations for a criterion renders the alternative completely unacceptable, the criterion is non-compensatory.
2. **Independence between criteria**: The utility for a criterion can be calculated independently from the evaluations on another criterion. MCDA accounts for main effects only and does not account specifically for interactions between main effects. Such interactions, if critical for decision-making, should be handled outside of the MCDA workshop in an ad-hoc manner.
3. **Independence between evaluations on different endpoints**: Correlation would apply if the occurrence of an endpoint increases the probability for another (independent) endpoint to occur. For example, it is likely that crying is correlated with pain. Not accounting for the correlation between the events will introduce biased results.

   Double counting of the same event often occurs and data sets should be revised prior to being used for MCDA input.

The weighting is performed after all alternatives are scored on all criteria and all value functions are decided on. The software facilitates the swing-weighting through visual interfaces by guiding the participants through the following steps (see Appendix 4 for more details):

1. Select a subgroup of criteria to be weighted together, starting from the lowest level of the hierarchy of the outcome tree hierarchy to the highest (i.e. bottom-up approach).
2. Order the sub-set of criteria according to the increase in overall preference generated by improving a criterion from its worst to its best performance (called swing). The first-order criterion presents the most valuable swing and is assigned a swing-weight of 100.

3. Elicit the relative weights of lower-order criteria in comparison to the reference criterion assigned the swing-weight of 100.
4. Verify the weighting consistency.
5. Move up the value tree, assessing the relative weights of different reference criteria.

In the following, distinction will be made between:

- **Criteria swing-ratings**: for which the most important criterion is the reference and is always assigned a weight of 100.
- **Relative weights**: these weights are normalized criteria swing-ratings for which the sum equals 1 among a family (or subgroup) of criteria weighted together.
- **Cumulative weights**: these weights are equal to the product of the normalized weights from the top of the hierarchy down to each branch of the final criterion and for which the sum equals 1 among the whole outcome tree.

Ordering the criteria and rank-based weighting

Among a subgroup of criteria, the participants are asked to identify the criterion that is the most important to bring from its lowest (worst) level to its highest (best) level and to proceed similarly across all remaining criteria. This way the order of the criteria is established. The degree of preferences among the ordered criteria is then elicited to the expert and described in the next section.

When the number of criteria is very large, the expert may use a rank-based criteria weighting. The rank-order centroid (ROC) has been shown as one of the best alternatives among other rank-based methods. More details on ROC calculations are provided in Appendix4.

Weighting the criteria

The first-order criterion is given a swing-rating of 100 (%) by default and is considered the reference to which the other criteria will be compared. The worst level for the first criterion will be noted as Eval1Worst and its best level as Eval1Best, identifying therefore the swing as (Eval1Worst, Eval1Best).

A lower-order endpoint, with swing (Eval2Worst, Eval2Best), is then considered. The expert is asked to determine the value Eval1Mid (Eval1Mid ≤ Eval1Best) for which the utility of the swing (Eval1Worst, Eval1Mid) on the first-order criterion would be similar to the utility of the swing (Eval2Worst, Eval2Best) on the second criterion. Appendix 4 provides further insight on the rational for this procedure. Figure 5 provides a visual explanation of the process.

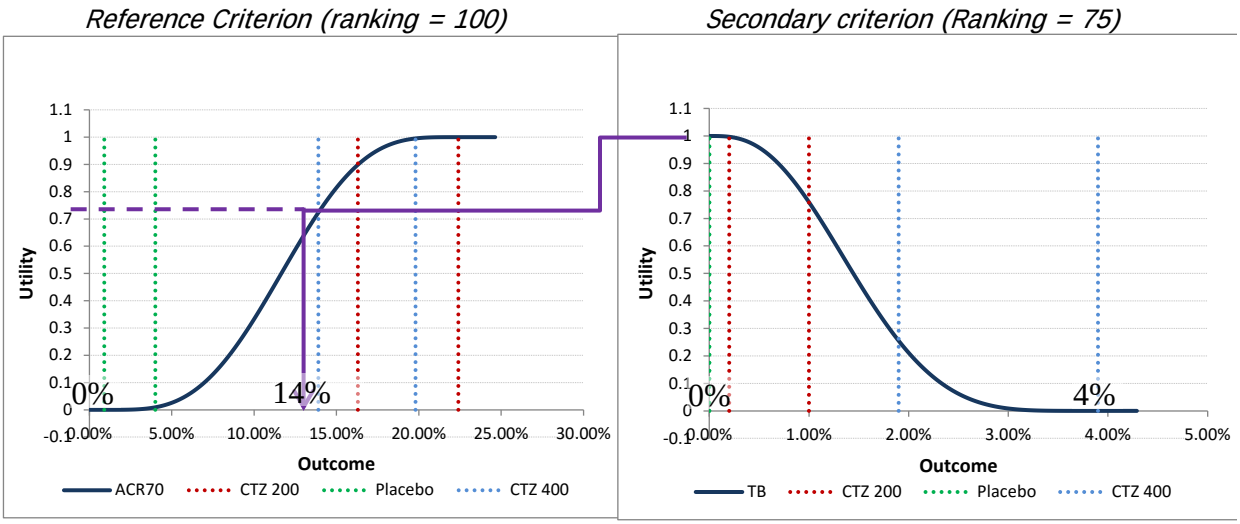


Figure 5: Criteria Weighting – Swing Rating

In this example, the swing range [0%-4%] of the secondary criterion is given a weight of 75 in comparison to the 100 rating given, by default, to the primary swing range [0%-25%]. The expert then confirms his indifference between a change of 0% to 14% on the reference criterion and a change of 0% to 4% on the secondary criterion. If the expert expresses a preference for either of the two proposals, it suggests that the weights need to be changed.

Relative weights are calculated through normalization to the sum of swing-ratings. The relative weights between two criteria can be translated into a coefficient of 'stretch' or 'shrinkage' applied to the two ranges of values on the separate scales in order to present an equal utility.

When moving up the outcome tree, the criterion that was considered as the reference among the sub-set of criteria is now the representative of that sub-group and may be considered for ordering with the reference criteria of other criteria sub-sets.

The notion of swing weights captures both the concept of 'importance' of the criterion and the extent to which the evaluation scale used discriminates between the alternatives. For instance, fatalities are amongst the worst outcomes, but the swing weight may be very small (or even null) if the maximum death swing between the alternatives equals 1 death over 109 subjects. One of the most common errors in naïve scoring models is to assume that weights are independent of the evaluation scales used. From the equations in Appendix 5 it is clear that the effect of the weight parameter w_i is directly connected to the scale used for calculating the utility for the corresponding criterion.

Step 7: Calculate the Results

The overall preference value for an alternative will be calculated as the weighted average of the preference values of the benefit and risk criteria.

Step 8: Conduct an Uncertainty Analysis

The obtained overall benefit-risk measure reflects only the average scenario without quantifying the variability of this measure. Because many uncertainties may be involved in the benefit-risk analysis, it is important to understand how strongly the benefit-risk balance depends on these uncertainties. The uncertainty might arise from uncertain criteria measurements (i.e. incidences are measured with a certain precision), from uncertain or heterogeneous preferences (i.e. stakeholders are uncertain regarding their preferences or different stakeholders have different preferences) and from subjective choices made when building the MCDA effects table. The uncertainty analyses will be explained in detail in Section 7.3.

3.1.3. Selection of Stakeholder Groups

The primary stakeholder group is the public health authorities, as the decision-making authority. The preferences of representatives of regulatory authorities and family physicians/clinicians or other health care professionals will be elicited as well. Participants should:

- have a solid clinical training to understand the clinical relevance and consequences of the benefit and risk criteria;
- make quantitative decisions on priorities between two sorts of endpoints / pathologies associated to two different cohorts. This cognitive exercise is difficult for people who have no real-life experience in dealing with such decisions.

Consistency Assessment

Criteria weights in compensatory approaches express the substitution rates of one criterion in comparison to another. The participants to the decision conference may use this property to assess the consistency of their judgments. The D-Sight Swing-Weighting tool is using that concept to elicit weights from experts.

After the criteria swing-rating have been provided by experts, the relative weights (i.e. the criteria swing-ratings normalized to unity) are calculated and the importance of each criterion may be compared pairwise to the reference criterion. Experts may then decide to adapt the weight of the (non-reference) criterion if the comparison seems unfair.

After all criteria have been rated, a global consistency check consists in presenting all criteria according to their cumulative weights (weights that sum to unity across the whole endpoint tree). The participants will review the discriminative power of the criteria ranges between alternatives and decide on the need to revise the weights.

Consensus Discussions

Only few participants are selected to participate in the decision conference and their opinions do not necessarily reflect the whole diversity of opinions. Their recommendations therefore pertain to their own opinions, but are nonetheless informative to decision-makers. The calculation of the average scores among participants is meaningful provided a consensus has been reached during the decision conference. In absence of consensus, calculating average scores is less meaningful. Instead, it is more meaningful to identify divergences and initiate a discussion on the reasons for this among the participants. The participants are then invited to update their opinions after the discussion and consensus is attempted again.

Training

The MCDA training will focus on understanding the swing-weighting process. An Excel tool similar to the one below (which was set up for neuralgia and reactogenicity pain) will be used to train the participants and to make them familiar with the IT tools, swing-weighting and the cognitive processes involved in stating their preferences.

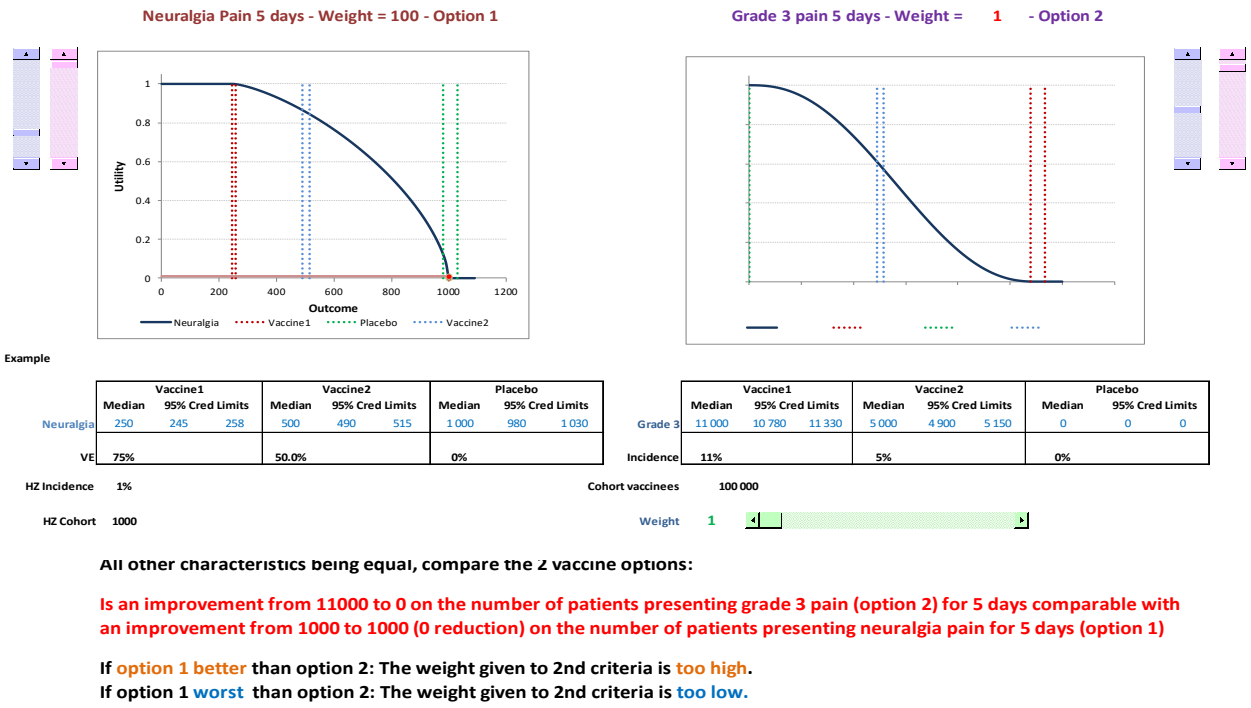


Figure 6: Swing-weighting Example for Neuralgia and Reactogenicity Pain

Evaluating the Process of Involvement

The participants will be asked to evaluate the decision conference by rating how strongly they (dis)agree with a series of statements on a five point Likert scale. Examples of such statements are:

1. "It was easy to make comparisons between the outcomes."

2. “The questions adequately reflect the aspects of pertussis vaccination that I feel are important.”
3. “Sufficient information was provided, in a clear and understandable format, to enable me to answer the questions.”
4. “I would be happy to take part in similar decision conferences in the future.”

Participants will also be invited to provide additional free text comments or suggestions to improve the preference elicitation methodology.

Number of Participants

There is no theoretical minimum number of participants at a decision conference. If different stakeholder groups are present, that minimum number of participants per stakeholder group can be reduced, maintaining the number per group well above 4. Based on experience, the maximum number of participants, including the facilitator, is about 20. More participants would make it difficult to manage the decision conference. A minimum number of participants (about 8) is needed as well to reflect the diversity in opinions. Hence, the optimal number of participants ranges between 8 and 15.

Trimming the Outcome Tree

A large number of pertussis complications has been identified and may lead to a large outcome tree. A subset of pertussis complications and sequels may be identified and endorsed by the experts prior the decision conference. The outcome tree will first be trimmed down by omitting those criteria that are less clinically relevant, having a small population impact or providing no discrimination between alternatives.

Software: D-Sight

The preference elicitation will be carried out using the web-based software D-Sight. D-Sight is preferred to HiView3 since D-Sight offers the option to organize a virtual decision conference (a video conference) using a web-based MCDA interface for each participant individually and a central repository. A server will handle all consensus calculations, data presentations and visualizations.

Prior to the MCDA workshop, the facilitator will set up the web-server and create the alternatives and criteria scores for each alternative. A linear value function will be defined as default for all criteria, and equal swing-weights will be defined for all criteria among the all sub-groups. Ideally, the facilitator will investigate whether preliminary consensus can be reached on the grouping of benefit and risk criteria and, possibly, the order of importance of criteria within these groups.

The MCDA interface used for this POC is based on a previous version of the standard software of D-Sight and has been optimized for benefit risk balance applications about two years ago. The software (https://brb.d-sight.com/) is freely available for testing purposes by individuals for a period of 30 days.

D-Sight can upload Excel files from a PC to the web-server to initiate the outcome tree, to define the alternatives and to define the criteria scores for each alternative. The format of the Excel file is provided in Appendix 2.

The XML files generated by D-Sight web-interface can be retrieved and converted into SAS datasets. An SQL script is currently being created to retrieve the relational database aspects of the MCDA database. All critical information will therefore be available in SAS datasets. Additional graphs, tables and analysis that are not available from D-Sight might be implemented using the D-Sight – SAS bridge.

Resources

Resources that need to be considered:

1. A meeting place for participants to attend the MCDA session: if possible, participants working close-by should meet face-to-face, possibly in the presence of a local facilitator who is knowledgeable of the MCDA process and IT tools.
2. Web-meeting platform to share interim and final results of MCDA across all experts, possibly with video enabled. The session may become long and non-verbal feedback might be important to catch. The software selected presently is D-Sight.
3. A group of three to four experts available for each of the stakeholder groups.
4. A main MCDA facilitator and possibly a local facilitator at each location to guide participants on how to use the interface.
5. A secretary to capture main issues occurring and decisions made during the decision conference. A statistician may also be useful to support bridge between the D-Sight web-platform and the SAS system.

A training session will be organized for the experts to get used to the MCDA interface and to understand the Swing-Weighting process. We anticipate a three hour session. It will also be an opportunity to reach a consensus on the list of criteria considered, alternative evaluations, and possibly the clinical relevance ordering of the criteria. Several training sessions can be organized if all experts cannot attend simultaneously. About three to four weeks after the training session, the actual MCDA decision conference will be organized. The presence of identified participants is mandatory during the four to five hours planned for the decision conference.

3.2. Disability Weights

For the benefit-risk analyses, disability weights as used for DALY calculations will be used. Disability weights will be obtained from the literature if possible. This would allow an intermediate benefit-risk analysis to be carried out. The intermediate benefit-risk analyses can be used for validation and, in case of the historical benefit-risk analysis, will be the starting point for the benefit-risk monitoring.

For this intermediate benefit-risk assess analysis, the disability weights developed previously for the estimation of pertussis burden within the BCoDE project [5], or within the GBD-2010 project [6] will be used. In the former, a set of disability weights for specific health outcomes were

assembled through literature review/expert opinion, and were attached to a natural history model for pertussis, beginning with symptomatic pertussis infection. In the GBD project, new disability weights were determined for a set of 220 more generic health states [7]; these were elicited using a survey-based procedure, with the general public as respondents. For both projects, if disability weights for a specific health state/sequela were unavailable, a proxy weight (for another health outcome deemed similar in severity, according to expert opinion) could be applied.

3.3. Uncertainty and Sensitivity Analyses

3.3.1. Rationale

There are many sources of uncertainty that might affect the benefit-risk assessment. Uncertainty analyses will be conducted to assess the impact on the overall benefit-risk result of:

- the uncertainty in benefit and risk criteria measurements as obtained from the literature, clinical trials or observational databases;
- differences in preferences;
- subjective modelling choices if applicable (e.g. different case definitions).

Additionally, a sensitivity analyses will be conducted to identify the benefit and risk criteria that might modify the benefit-risk balance (i.e. the pivotal criteria, to be used for monitoring).

3.3.2. Probabilistic Uncertainty Analysis through Monte-Carlo Simulation

In benefit-risk assessment, the uncertainty of the criteria measurements (or statistical uncertainty) propagates or 'carries forward' to the uncertainty on the final benefit-risk result. MC simulation is perhaps the most common technique for propagating uncertainty and is especially recommended when having multiple inputs. Wen et al. applied the Monte-Carlo approach to account for uncertainty in MCDA models [8]. MC simulation applied to MCDA involves:

1. Specifying suitable probability distributions for the criteria measurements (e.g. beta distributions for probabilities, Poisson distributions for counts). Parameters for these probability distributions will be derived mainly from the literature/clinical trials for the 'historical' benefit-risk assessment and from the literature/clinical trials/observational databases for the 'current' benefit-risk assessment. Correlations between the criteria measurements will be estimated if appropriate.
2. Randomly sampling (correlated values) from these probability distributions.
3. Calculating the overall benefit-risk result.
4. Applying steps 1-3 multiple times, say 10,000 times.

This way, a distribution of benefit-risk results is obtained. Based on this distribution, the median, the standard deviation and coverage intervals are calculated reflecting uncertainty in criteria measurements. Correlated values will be generated using Gaussian's copula's.

For the current POC, the following approaches to include the preference weights might be considered:

1. The average preferences across all stakeholder groups if appropriate (overall model).
2. The average preferences within each stakeholder group if appropriate (stakeholder group-specific models).
3. The individual member preferences (heterogeneity models, overall and stakeholder group-specific).

3.3.3. Stochastic Multi-criteria Acceptability Analysis: Exploring Preferences

Stochastic Multi-criteria Acceptability Analysis (SMAA), one of the methods recommended by the PROTECT consortium, is a stochastic extension of MCDA [9]. SMAA allows investigation of the impact of uncertainty and heterogeneity in preferences on the benefit-risk assessment. It is based on exploring the weight (preferences) space in order to describe the preferences that would make each alternative the most preferred one. SMAA can be computed efficiently through MC simulation using probability distributions for the preferences. The main results of SMAA are:

1. **Rank acceptability indices**: This index measures the variety of different preferences that grant a certain alternative the first (or second) rank and is most conveniently expressed as percentages.
2. **Central weight vectors**: These are calculated as the average of the preferences that grant a certain alternative the first rank. The central weights represent the 'typical' preferences supporting a certain alternative.
3. **Confidence factors**: These are calculated as the probability for an alternative to obtain the first rank when the central weight vector is chosen.

In this POC, SMAA analysis will be performed by extending the MC analyses described above. In particular, rank acceptability indices, central weight vectors and confidence factors will be calculated using the following preference distributions:

- Stochastic distributions derived from the elicited preferences, for instance using Dirichlet distribution to describe uncertainties on the cumulative weights
- Uniform weight distributions (all values between 0 and 1 are equally likely), reflecting a lack of preferences (so called missing-weight analysis).

3.3.4. Scenario-analysis: Subjective Choices

It is anticipated that subjective choices will have to be made to build the MCDA effects table and that these choices may impact the benefit-risk assessment. For every subjective choice made that is expected to have some impact on the benefit-risk assessment, a scenario-analysis will be performed. There may be also interest to vary parameters outside of their observed range. A scenario-analysis implies that the benefit-risk model is adjusted for every choice option. Possible scenarios include:

- Different case definitions (e.g. laboratory confirmed pertussis or clinical case definition)
- Different vaccination schedules

3.3.5. Sensitivity Analyses: Identifying the Pivotal Criteria

Sensitivity analysis has different meanings depending on the context in which it has been used. Sensitivity analyses will be performed to provide an answer to the question: "How much can the criteria measurements and associated weights differ in comparison to the baseline without tipping the benefit-risk balance?" and "Which parameters are the most likely to impact the overall benefit-risk balance?" These most 'pivotal' criteria will be prioritized for sequential benefit-risk monitoring.

In this POC, the following analyses will be performed:

1. **Deterministic sensitivity analyses**: Starting from the MCDA with average values for the criteria measurements and preferences, criteria measurements and preference weights will be varied until the benefit-risk balance is tipped. The measurements/preference weights for which the least percentage-wise change is needed before tipping the benefit-risk balance are the most 'pivotal'. Sensitivity analyses will be conducted by varying each criteria measurement/preference weight separately. This is a one-factor-at-the-time approach, since when one factor is varied all others are held constant. In addition, sensitivity analyses will be conducted by jointly varying each criteria measurement and corresponding preference weight.
2. **Stochastic sensitivity analyses**: The deterministic analyses described above do not account for the uncertainty in the benefit-risk model. The analyses will be repeated starting from the MC model developed in Section 7.3.2. For every change in percentage points in criteria measurements and/or preference weights, the probability of tipping the benefit-risk balance will be obtained.

3.4. Benefit-risk monitoring

As described above, this POC analysis will apply a three phase approach to simulate what could have been performed with the wP to aP switch for pertussis vaccines. The phases are:

1. Simulation of the historical benefit-risk analysis to support the decision for the switch, based on data available at that time.
2. Simulation of prospective monitoring of critical benefit-risk parameters (identified in phase I), using input from retrospective analyses based on the electronic health care databases in ADVANCE. In order to monitor the pivotal parameters (which may be beyond the parameters that are currently collected) which will only be identified after phase 1 the benefit and risk protocols may need to be amended.
3. Current benefit-risk analysis using all currently accumulated evidence as if re-assessment were triggered today.

This section focuses on phase II, which is described in Table 9, where the activities in this POC analysis are listed (in the second column) in regards with what would be (have been) done in an ideal world where ADVANCE would be implemented.

Table 9: Stepwise Approach to Monitoring of Benefit-Risk of Vaccines

Step-by-step process in real life (i.e. post-ADVANCE), according to ADVANCE vision	*How will this be mimicked in the Pertussis POC*
Benefit-risk analysis to support the decision to launch a new vaccine on the market	Benefit-risk analysis (MCDA) to illustrate the historical decision to switch from wP to aP
Sensitivity analysis of the benefit-risk analysis used to identify critical parameters (large statistical uncertainty, large potential impact on benefit-risk ratio or difference)	Idem
Critical parameters identified from benefit-risk sensitivity analysis translated into outcomes and measurements from post-marketing surveillance data	Idem
Agreement on threshold, periodicity, total duration of monitoring	Idem, but can be simplified (see below)
Risk and benefit analyses designed to produce periodical values of critical benefit-risk parameters	Idem
Periodical output from the safety and benefit analyses, near real-time, prospective	Idem; 3 month periodicity. 'Real-time' situation mimicked by roll-back extraction in the retrospective databases
Periodical analysis with cumulative number (rates) of events to assess if still within acceptable limits, using a control chart statistical method (such as CUSUM, SPRT) [10-14]	For the sake of simplicity and feasibility, this will be reduced here to monitoring by simple visualization (rates and confidence intervals of 'critical' parameters) of periodic output from the analyses (As a consequence, there is no need for predefined threshold nor total duration) This POC will be limited to only few parameters, but there should be at least one risk parameter and one benefit parameter.
End of monitoring and conclusion if no alert, or new actions if threshold for alarm is reached	Not applicable (In this POC, phase III will be performed regardless of phase II outcome)

4. LIMITATIONS

- This benefit-risk analysis has been designed mostly to test new approaches to benefit-risk monitoring, not to provide a definite answer to the comparison between the two alternatives
- The benefit-risk model is truncated: the risks are from all vaccine components (not only pertussis), and the benefits are limited to those before booster age.
- In the benefit-risk analysis, evidence on the benefits and risks of pertussis vaccination obtained from various sources will be combined. The quality of the sources will be different. However, these differences in quality will not be taken into account in the current benefit-risk analyses.
- For reasons of simplification, indirect effects will not be considered, although a public health perspective is adopted.
- It was decided to elicit preferences using MCDA swing-weighting from representatives of public health and regulatory authorities as well as from family physicians/clinicians. The public health authorities are the main decision makers. In this analysis, preferences from the general public will not be elicited.
- The preferences that will be elicited for the historical benefit-risk analysis as part of this benefit-risk analysis do not necessarily reflect the preferences at that time and are to a certain extent artificial.
- Within this POC, we will not monitor changes in preferences over time because this cannot be retrospectively done.
- The quality of the benefit-risk assessment is as good as the quality of the evidence that feeds into the benefit-risk assessment. Issues of bias, confounding, missing data and missing information that often characterize epidemiological studies, will affect the benefit-risk analysis they inform.
- The primary objective of benefit-risk methodology is to support transparent and consistent decision making. Complex methodology hampers transparency.
- To conduct the benefit-risk analyses, evidence on benefit and risk health outcomes is required. It might not be possible to find accurate evidence for some of the required health outcomes and the benefit-risk analyses might be subject to huge uncertainty, leading to inconclusive results.
- It is still questionable whether the MCDA swing weights obtained for the historical benefit-risk analyses can be used for the subsequent benefit-risk monitoring.
- It is still unclear whether there will be sufficient evidence per country to conduct a country-specific benefit-risk analysis. On the other hand, it is also still unclear how generalizable the results are across countries.
- The comparability of the evidence on the benefits and risks of wP and aP vaccination is questionable as a results of the sudden switch from wP to aP and hence, the huge potential for time-varying confounding.

- Pivotal parameters may not yet be collected in the benefits and risk protocols which would mean they need to be amended.

5. REGULATORY STATUS

This analysis is based on secondary use of data.

This analysis is not considered as a post-authorization safety surveillance (PASS) because the aim is not to identify, characterize or quantify a safety hazard, confirm the safety profile of the medicinal product, nor of measure the effectiveness of risk management measures.

While the analysis is being conducted, the marketing authorization holder (MAH) shall monitor the data generated and consider its implications for the risk-benefit balance of the medicinal product concerned. Any new information that might influence the evaluation of this risk-benefit balance shall be communicated to the competent authorities of Member States in which the medicinal product has been authorized. The channel for communicating this information is the notification of an Emerging Safety Issue.

6. REFERENCES

1. European Medicines Agency. Report of the CHMP working group on benefit-risk assessment models and methods. Available at: http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2010/01/WC500069668.pdf (last accessed 15 October 2015).
2. Mt-Isa S, Hallgreen CE, Wang N, et al. Balancing benefit and risk of medicines: a systematic review and classification of available methodologies. Pharmacoepidemiol Drug Saf 2014;23:667-78.
3. Thokala P, Duenas A. Multiple criteria decision analysis for health technology assessment. Value Health 2012;15:1172-81.
4. Mussen F, Salek S, 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. Pharmacoepidemiol Drug Saf 2007;16 Suppl 1:S42-6.
5. Kretzschmar M, Mangen MJ, Pinheiro P, et al. New methodology for estimating the burden of infectious diseases in Europe. PLoS Med 2012;9:e1001205.
6. Murray CJ, Vos T, Lozano R, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012;380:2197-223.
7. Vos T, Flaxman AD, Naghavi M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012;380:2163-96.
8. Wen S, Zhang L, Yang B. Two approaches to incorporate clinical data uncertainty into multiple criteria decision analysis for benefit-risk assessment of medicinal products. Value Health 2014;17:619-28.
9. Tervonen T, van Valkenhoef G, Buskens E, Hillege HL, Postmus D. A stochastic multicriteria model for evidence-based decision making in drug benefit-risk analysis. Stat Med 2011;30:1419-28.
10. Gagne J, Bykov K, Najafzadeh M, et al. Prospective benefit-risk monitoring of new drugs for rapid assessment of net favorability in electronic healthcare data. ISPOR Annual International Meeting , Philadelphia, USA: May, 2015. Available at: http://www.ispor.org/research_pdfs/49/pdffiles/PRM7.pdf (last accessed 15 October 2015).
11. Greene SK, Kulldorff M, Lewis EM, et al. Near real-time surveillance for influenza vaccine safety: proof-of-concept in the Vaccine Safety Datalink Project. Am J Epidemiol 2010;171:177-88.
12. Donegan K, Beau-Lejdstrom R, King B, et al. Bivalent human papillomavirus vaccine and the risk of fatigue syndromes in girls in the UK. Vaccine 2013;31:4961-7.
13. Nelson JC, Yu O, Dominguez-Islas CP, et al. Adapting group sequential methods to observational postlicensure vaccine safety surveillance: results of a pentavalent combination DTaP-IPV-Hib vaccine safety study. Am J Epidemiol 2013;177:131-41.
14. McDonald L, Yiannakoulias N, Svenson L. Abstracts from the 2007 Conference of the International Society for Disease Surveillance: A novel application of surveillance algorithms in childhood immunization program monitoring. Available at: http://faculty.washington.edu/lober/www.isdsjournal.org/htdocs/articles/2081.pdf (last accessed 15 October 2015).

APPENDIX 1: POC OUTLINE

See ADVANCE SharePoint:

https://publication.wiv-isp.be/workspaces/advance/wp5/Shared%20Documents/POC%20Study%20Teams/COVERAGE/Background%20docs/2_POC%20pertussis%20outline_version%201.5_tobedistributed.docx

APPENDIX 2: CASE DEFINITIONS

- Injection site reactions: Brighton Collaboration case definition: Gidudu J, Kohl KS, Halperin S, et al. A local reaction at or near injection site: case definition and guidelines for collection, analysis, and presentation of immunization safety data. Vaccine 2008;26:6800-13.

- Fever: Brighton Collaboration case definition: Marcy SM, Kohl KS, Dagan R, et al. Fever as an adverse event following immunization: case definition and guidelines of data collection, analysis, and presentation. Vaccine 2004;22:551-6.

- Somnolence: Reduced interest in surrounding or increased sleeping. Grade1: Sleepier than usual or less interested in surroundings; Grade 2: Not interested in surroundings or did not wake up for a feed/meal; Grade 3: Sleeping most of the time or difficult to wake up

- Persistent crying, irritability: Brighton Collaboration case definition: Bonhoeffer J, Vermeer P, Halperin S, et al. Persistent crying in infants and children as an adverse event following immunization: case definition and guidelines for data collection, analysis, and presentation. Vaccine 2004;22:586-91.

- Generalized convulsions: Brighton Collaboration case definition: Bonhoeffer J, Menkes J, Gold MS, et al. Generalized convulsive seizure as an adverse event following immunization: case definition and guidelines for data collection, analysis, and presentation. Vaccine 2004;22:557-62.

- HHE: Brighton Collaboration case definition: Buettcher M, Heininger U, Braun M, et al. Hypotonic-hyporesponsive episode (HHE) as an adverse event following immunization in early childhood: case definition and guidelines for data collection, analysis, and presentation. Vaccine 2007;25:5875-81.

- Extensive limb swelling: Brighton Collaboration case definition: Kohl KS, Walop W, Gidudu J, et al. Swelling at or near injection site: case definition and guidelines for collection, analysis and presentation of immunization safety data. Vaccine 2007;25:5858-74.

- Pertussis is defined clinically as well as by laboratory assessement. *Clinical criteria* for the diagnosis of pertussis are (1)Any person with a cough lasting at least two weeks and at least one of the following three: paroxysms of coughing, inspiratory 'whooping', post-tussive vomiting, or (2) Any person diagnosed as pertussis by a physician, or (3) Otherwise

unexplained apnoeic episodes in infants. For *laboratory criteria* at least one of the following is needed for diagnosis: (1) Isolation of *B. pertussis* from a clinical specimen, (2) Detection of *B. pertussis* nucleic acid in a clinical specimen and (3)*B. pertussis*-specific antibody response.

- Pneumonia: A new infiltrate on a chest radiograph together with two or more clinical symptoms (dyspnoea, cough, sputum production, chest pain, and/or body temperature > 38°C or < 36.1°C) and/or WBC > 109 cells/L) or plasma CRP > 30 mg/L). Severe pneumonia was defined as pneumonia with a CURB-65 score ≥ 2. Community acquired pneumonia was defined as: (1) the presence of symptoms consistent with acute lower respiratory tract infection (at least one of increasing breathlessness, cough, sputum orfever); and (2) the presence of a new infiltrate on the chest radiograph.

APPENDIX 3: EXCEL FILE FORMAT FOR UPLOAD INTO D-SIGHT

An example of structure of an excel file used to upload into D-Sight.

	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
		Endpoint 1			Endpoint 2			Endpoint 3	
Alternative 1	Low11	Mid11	High11	Low12	Mid12	High12	Low13	Mid13	High13
Alternative 2	Low21	Mid21	High21	Low22	Mid22	High22	Low23	Mid23	High23
Alternative 3	Low31	Mid31	High31	Low32	Mid32	High32	Low33	Mid33	High33
Alternative 4	Low41	Mid41	High41	Low42	Mid42	High42	Low43	Mid43	High43

APPENDIX 4: SWING-WEIGHTING

The following sections contain extracts from publications providing details on swing-weighting and implementation.

HYDE KM. (2006) UNCERTAINTY ANALYSIS FOR MULTI-CRITERIA DECISION ANALYSIS

Page	*Extracts*
	Compensatory approaches In compensatory methods, the CWs amount to being substitution rates, allowing differences in preferences, as they relate to different criteria, to be expressed on the same scale i.e. that the weights be proportional to the relative value of unit changes in their attribute value functions (Hobbs, 1980; Poyhonen and Hamalainen, 2000). These parameters are in fact scaling constants needed for the cardinal criteria-functions to be commensurate in some way. In other words, if $CW_1 = 2$ and $CW_2 = 4$, actors must be indifferent between the change in $V_1(X_1)$ of 1 and a change in $V_2(X_2)$ of 0.5. This condition also implies that weights are on a "ratio level of measurement". That is, $CW_1 = 2$ and $CW_2 = 4$ means that a unit change in $V_1(X_1)$ must be half as valuable as a unit change in $V_2(X_2)$.
50	Compensatory approaches In compensatory methods, the CWs amount to being substitution rates, allowing differences in preferences, as they relate to different criteria, to be expressed on the same scale i.e. that the weights be proportional to the relative value of unit changes in their attribute value functions (Hobbs, 1980; Poyhonen and Hamalainen, 2000). These parameters are in fact scaling constants needed for the cardinal criteria-functions to be commensurate in some way. In other words, if $CW_1 = 2$ and $CW_2 = 4$, actors must be indifferent between the change in $V_1(X_1)$ of 1 and a change in $V_2(X_2)$ of 0.5. This condition also implies that weights are on a "ratio level of measurement". That is, $CW_1 = 2$ and $CW_2 = 4$ means that a unit change in $V_1(X_1)$ must be half as valuable as a unit change in $V_2(X_2)$.

Thus, in these approaches, CWs have no absolute or intrinsic meaning and there is no sense in attempting to derive them without knowledge of the criterion or its value function. If the value trade-offs are done properly and address the question of how much of one specific criterion is worth how much of another specific criterion, the insights from the analysis are greatly increased (Bana e Costa *et al.*, 1997).

50

Thus, in these approaches, CWs have no absolute or intrinsic meaning and there is no sense in attempting to derive them without knowledge of the criterion or its value function. If the value trade-offs are done properly and address the question of how much of one specific criterion is worth how much of another specific criterion, the insights from the analysis are greatly increased (Bana e Costa *et al.*, 1997).

The AHP assumes that actors take the set of alternatives explicitly into account when they assess the CWs. Value theory based methods assume that actors give preference statements about the CWs so that they reflect the criteria ranges (Poyhonen and Hamalainen, 2000). Value theory based weighting methods include SMART (von Winterfeldt, 1986), SWING (von Winterfeldt, 1986) and SMARTER (Barron and Barrett, 1996).

50

The AHP assumes that actors take the set of alternatives explicitly into account when they assess the CWs. Value theory based methods assume that actors give preference statements about the CWs so that they reflect the criteria ranges (Poyhonen and Hamalainen, 2000). Value theory based weighting methods include SMART (von Winterfeldt, 1986), SWING (von Winterfeldt, 1986) and SMARTER (Barron and Barrett, 1996).

D1.9 Swing weights

App D 1.9

The Swing method is similar to the SMART procedure, but the procedure starts from the most important criterion, keeping it as the reference. The DM begins by rank ordering criteria in terms of their associated value ranges. Assuming that each criterion is at its worst possible level, the DM is asked which criterion they would most prefer to change from its worst level to its best level. The criterion chosen has the most important value range. Next, the DM is asked which criterion they would next most prefer to change from its worst to its best level. To quantify the relative value ranges, the DM next assigns a relative importance weight between 0 and 100. The criterion with the most preferred swing is most important and is assigned 100 points. Proceeding in this fashion, the DM rank orders the criteria and assigns relative importance weights to their value ranges. The final step in the Swing weight procedure is to normalise the relative importance of the weights (Jia et al., 1998).

Examples of the use of this method can be found in: Schoemaker and Waid (1982), Poyhonen and Hamalainen (2001) and Bell *et al.* (2001).

THOKALA P. (2011) MULTIPLE CRITERIA DECISION ANALYSIS FOR HEALTH TECHNOLOGY ASSESSMENT

Page	*Extracts*
	This is achieved by swing weights which represent the gain in overall value by going from tl worst value to best value in each criterion i.e. for any two criteria i and k, the ratio w_i/w_k is tl change in $v_k(a)$ that should compensate for a unit loss on $v_i(a)$. There are a number of ways which these swing weights can be elicited, these techniques are not discussed here as they a explained in detail in section VII and in literature.[7]

BRUHN BARFOD M. (2012) OPTIMISING TRANSPORT DECISION MAKING USING CUSTOMISED DECISION MODELS AND DECISION CONFERENCES

Page	*Extracts*
41 (a)	The weight assigned to a criterion is essentially a scaling factor which relates scores on that criterion to scores on all other criteria. Thus if criterion A has a weight which is twice that of criterion B this should be interpreted as the decision-maker values 10 points on criterion A the same as 20 points on criterion B and would be willing to trade one for the other. These weights are often referred to as swing weights to distinguish them from the less well defined concept of importance weights. Thus the notion of swing weights captures both the psychological concept of "importance" and the extent to which the measurement scale adopted in practice discriminates between alternatives. One of the most common errors in naive scoring models is to assume that weights are independent of the measurement scales used. It can be seen from the algebraic structure of (3.1), however, that the effect of the weight parameter w_i is directly connected to the scaling used for $v_i(a)$, so that the two are intimately connected.
	The weight assigned to a criterion is essentially a scaling factor which relates scores on that criterion to scores on all other criteria. Thus if criterion A has a weight which is twice that of criterion B this should be interpreted as the decision-maker values 10 points on criterion A the same as 20 points on criterion B and would be willing to trade one for the other. These weights are often referred to as swing weights to distinguish them from the less well defined concept of importance weights. Thus the notion of swing weights captures both the psychological concept of "importance" and the extent to which the measurement scale adopted in practice discriminates between alternatives. One of the most common errors in naive scoring models is to assume that weights are independent of the measurement scales used. It can be seen from the algebraic structure of (3.1), however, that the effect of the weight parameter w_i is directly connected to the scaling used for $v_i(a)$, so that the two are intimately connected.

Swing weights

A method for eliciting weights for criteria is available with the swing weight technique (von WInterfeldt and Edwards, 1986). The technique is usually considered to be the theoretical most correct method for eliciting criteria weights, but also difficult to use in practice.

The technique presupposes that the decision-makers consider the swing from worst to best performance within each criterion and rank the criteria based on which swing gives the highest increase in overall value. Afterwards the swings within each of the criteria are assigned a numerical value reflecting its importance compared to the swing within the most important criterion.

41 (b) It can be useful to work with graphically supported scales as decision-makers generally seem to be comfortable with this and may be willing to assess the relative magnitude of the swing weights directly using this means. An example of such graphical support is illustrated in Figure 3.3 where a small example concerning criteria for the selection of a new by-pass road is presented. First the criteria are ranked in order of importance by considering which swing from worst to best performance within each criterion that gives the greatest increase in overall value, the next greatest increase in overall value, and so on, until a ranking is established. The swing from worst to best within the highest ranked criterion is then assigned a value of 1 (see the column for "Urban development"). The swing from worst to best within the second highest ranked criterion ("Landscape") is then using the visual scale compared with the swing within the highest ranked criterion. The process is repeated with the remaining criteria.

42 (a)

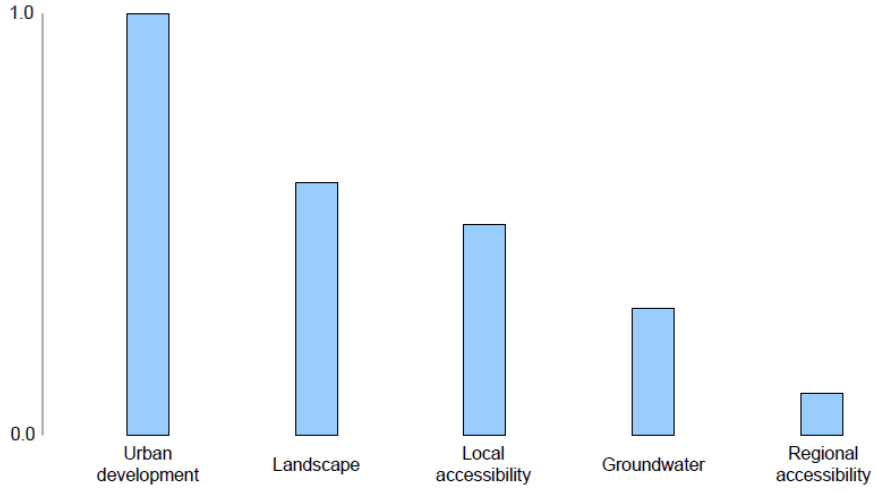


Figure 3.3. Swing weights: a visual analogue scale has been used for determining the magnitude between the five criteria in the example

This visual scaling provides a means for communicating a good sense of the magnitude of judgments whilst removing the need for numerical precision. However, it is important that this degree of imprecision is not forgotten when information is aggregated.

This visual scaling provides a means for communicating a good sense of the magnitude of judgments whilst removing the need for numerical precision. However, it is important that this degree of imprecision is not forgotten when information is aggregated.

The weights implied by the visual representation in Figure 3.3 may be translated into numerical values, as shown in Table 3.3 below. The second column of the table lists the weights standardised with the largest weight set to 1. It is usual, although not essential, to normalise weights to sum to 1 or 100, as shown in the third column of Table 3.3. Such normalisation allows the decision-makers to interpret for example the weight of landscape in Table 3.3 as constituting 24% of the total importance weight. This often seems to be a useful interpretation. However, in specific cases decision-makers may find it more intuitive to specify a reference criterion whose units are weighted 1 and against which all other criteria are compared, as shown to be the original weights with urban development as the reference criterion.

42 (b)

The weights implied by the visual representation in Figure 3.3 may be translated into numerical values, as shown in Table 3.3 below. The second column of the table lists the weights standardised with the largest weight set to 1. It is usual, although not essential, to normalise weights to sum to 1 or 100, as shown in the third column of Table 3.3. Such normalisation allows the decision-makers to interpret for example the weight of landscape in Table 3.3 as constituting 24% of the total importance weight. This often seems to be a useful interpretation. However, in specific cases decision-makers may find it more intuitive to specify a reference criterion whose units are weighted 1 and against which all other criteria are compared, as shown to be the original weights with urban development as the reference criterion.

Table 3.3. Swing weights – original and normalised values for the example

Criterion	Original weights	Normalised weights
Landscape	0.6	0.24
Groundwater	0.3	0.12
Urban development	1.0	0.40
Local accessibility	0.5	0.20
Regional accessibility	0.1	0.04

Weights in value trees

When the problem is structured as a multi-level value tree consideration has to be given to weights at different levels of the tree. It is useful to define relative weights and cumulative weights. Relative weights are assessed within families of criteria – i.e. criteria sharing the same parent – the weights within each family being normalised to sum to 1. The cumulative weight of a criterion is the product of its relative weight in comparison with its siblings and the relative weights of its parent, parent's parent, and so on to the top of the tree.

For illustration the example from above has been divided into different levels of criteria in a value tree, see Figure 3.4. By definition, the cumulative weights of all bottom-level criteria (leaves on the tree) sum to 1 – thus the normalised weights shown in Figure 3.4 are cumulative weights. The cumulative weight of a parent criterion is the total of the cumulative weights of its descendants.

43 As illustrated for the example problem, if the value tree does not have too many leaves, then weights can be assessed by directly comparing all bottom-level criteria to give the cumulative weights. Weights at higher levels of the tree are then to be determined by adding the cumulative weights of all members of a family to give the cumulative weight of the parent. Relative weights are determined by normalising the cumulative weights of family members to sum to 1. Relative and cumulative weights for the example problem are illustrated in Figure 3.4.

For larger models it is easier to begin by assessing relative weights within families of criteria. Weights at higher levels of the value tree can be assessed top-down or bottom-up. The top-down approach would assess relative weights within families of criteria by working from the top of the tree downwards. However, the analyst must be aware of the difficulty of interpreting weights at higher levels of a value tree – the weight of a higher level criterion is the sum of the cumulative weights of all its sub-criteria. Thus, in comparing two higher level criteria the decision-maker should be thinking in terms of a swing from 0 to 100 on all sub-criteria of the two higher level criteria. If the top-down approach is used it is important to carry out cross family checks on the cumulative weights of bottom level criteria.

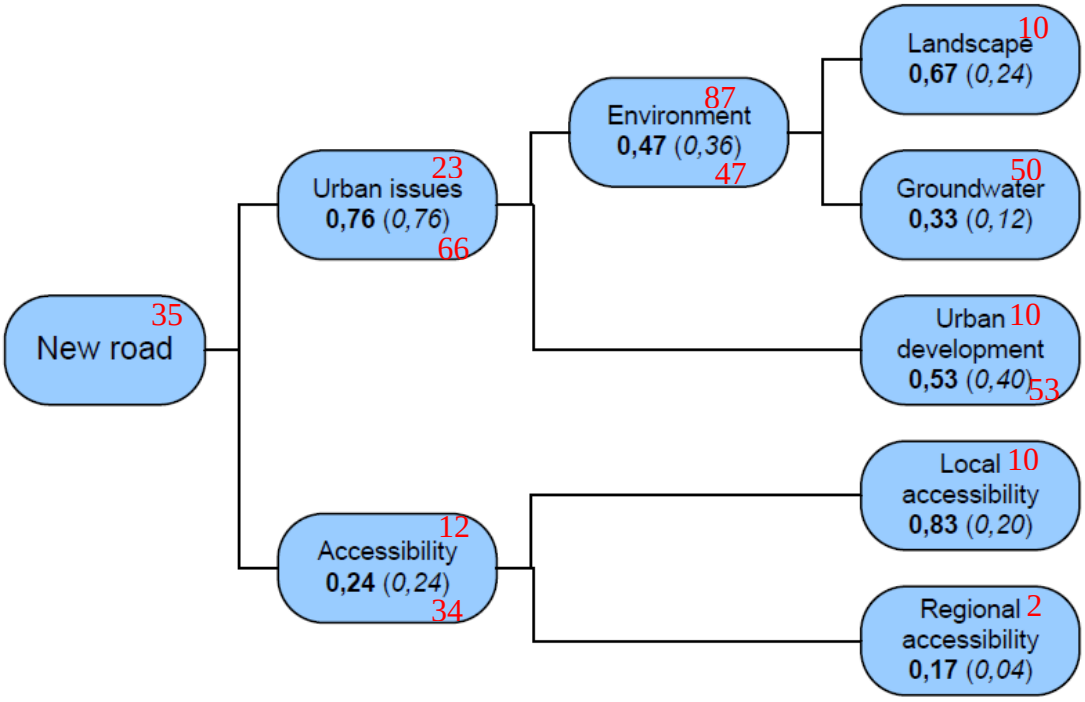


Figure 3.4. Relative weights (in bold) and cumulative weights (in italics) for the example. The bottom-up approach begins by assessing relative weights within families which contains only bottom level criteria and then carrying out cross family comparisons using one criterion from each family (perhaps the most highly weighted criterion in each family) and comparisons with any unitary bottom level criteria. This process would eventually give the cumulative weights of the bottom level criteria which can be aggregated to higher levels as described before.

Fig 3.4: *Urban issues* cumulative weight of 76% is the sum of cumulative weights below that node since all criteria were evaluated simultaneously. Red figures were added to represent what might be the respondent's ratings if a stepwise process had been used. *Environment* node is given a rating of 87 since the references *Landscape* and *Urban Development* references were compared in favor of the later, leading to a relative weight of 47% and 53% respectively. In other words, the *Landscape* Swing Weight is worth 87% of the *Urban Development* Swing Weight. The ratings sum for the *Urban Issues* node equals 231 and includes the rating for *Urban Development* and 87% of the rating sum for the Environment node (= 100 + 0.87*150). The rating sum for the *Accessibility* node is simply the sum of the criteria ratings below the node. A comparison of rating sums of 231 and 120 would lead to apriori cumulative weights of 66% and 34% **if both reference criteria were given the same weights**. The respondent can now modify those weights to reach 76% and 24% respectively.

Page ***Extracts***

Determining what weights ($w_1, w_2, \ldots, w_n$) to use is a separate problem from that of choosing the functional form (e.g., additive or multiplicative). Edwards and Barron (1994) proposed a method to approximate
35 the w_js using the decision maker's rank order of the relative importance of the attributes. In particular, they proposed using the rank order centroid method to derive weights for a set of attributes, a method that was later extensively evaluated by Barron and Barrett (1996).

The rank order centroid approximation

The decision maker's major input is to produce a rank order of the relative importance of the attributes in order to differentiate the priority of the vaccine candidates. This induces a rank order on the weights in the additive model. Suppose that the rank order is $w_1 \geq w_2 \geq \ldots \geq w_n$ for n attributes. The rank order centroid approximation for the constants in an additive model would then be as follows:

$$w_1 = \left(\frac{1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n}}{n} \right)$$

$$w_2 = \left(\frac{0 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n}}{n} \right)$$

$$w_3 = \left(\frac{0 + 0 + \frac{1}{3} + \ldots + \frac{1}{n}}{n} \right)$$

$$w_n = \left(\frac{0 + 0 + 0 + \ldots + \frac{1}{n}}{n} \right)$$

More compactly the weights can be expressed by

$$w_i = \sum_{j=i}^{n} \frac{1}{j} \quad i = 1 \cdots n$$

Barron and Barrett showed this rank order centroid approximation for weights to be superior to other often-proposed methods, such as the normalized sum of ranks. It is important to realize that rank order centroid weights are not essential to the multi-attribute utility models; rather they are an approximation used to reduce the workload of the potential user.

APPENDIX 5: THEORETICAL RELATIONSHIP BETWEEN UTILITY AND WEIGHTS

For each attribute (1, .. i ..,n) of a set of alternative, a range of performance $\left[x_i^{(0)}, x_i^{(1)}\right]$ is defined. The additive utility function is scaled in order to equal zero (0) for the worst alternative $x^{(0)}$ and to equal one (1) for the best conceivable alternative $x^{(1)}$. Using its additive property, the total utility of an alternative can be expressed using the part-utilities relative to each criterion.

Equation 1:

$$\min_x U(x) = U\left(x^{(0)}\right) = U\left(x_1^{(0)}, x_2^{(0)}, x_3^{(0)}, \dots, x_n^{(0)}\right) = \sum_{i=1}^{n} U_i\left(x_i^{(0)}\right) = 0$$

$$\max_x U(x) = U\left(x^{(1)}\right) = U\left(x_1^{(1)}, x_2^{(1)}, x_3^{(1)}, \dots, x_n^{(1)}\right) = \sum_{i=1}^{n} U_i\left(x_i^{(1)}\right) = 1$$

A scaling constant w_i is introduced in order to also normalize each part-utility $U_i(x_i)$ on the [0, 1] interval. That scaling constant w_i is the difference in utility between an alternative presenting the best level for the attribute (i) minus an alternative presenting the worst level for the same attribute, *all other attributes being equal*.

Equation 2:

$$\begin{aligned} U(x) &= \sum_{i=1}^{n} U_i(x_i) = \sum_{i=1}^{n} \left[U_i(x_i) - U_i\left(x_i^{(0)}\right)\right] \\ &= \sum_{i=1}^{n} \left\{ \left[U_i\left(x_i^{(1)}\right) - U_i\left(x_i^{(0)}\right)\right] \frac{\left[U_i(x_i) - U_i\left(x_i^{(0)}\right)\right]}{\left[U_i\left(x_i^{(1)}\right) - U_i\left(x_i^{(0)}\right)\right]} \right\} \\ &= \sum_{i=1}^{n} \{w_i * U_i'(x_i)\} \end{aligned}$$

$$where\ w_i = U_i\left(x_i^{(1)}\right) - U_i\left(x_i^{(0)}\right)\ and\ \sum_{i=1}^{n} w_i = 1$$

$$U_i'(x_i) = \frac{\left[U_i(x_i) - U_i\left(x_i^{(0)}\right)\right]}{\left[U_i\left(x_i^{(1)}\right) - U_i\left(x_i^{(0)}\right)\right]}$$

The normalized part-utility $U_i'(x_i)$ is given by the utility function elicited to the expert for each attribute (i). Equation 2 clearly establishes that the weight w_i elicited to the experts for each attribute (i) depend on the range $\left[x_i^{(0)}, x_i^{(1)}\right]$ and express the substitution rate of one criterion in comparison to another.

Assuming weights w_i and w_j ($w_i \geq w_j$) were given by an expert to attributes (i) and (j), the facilitator may then present $\left[x_i^{(0)}, x_i'\right]$ and $\left[x_j^{(0)}, x_j^{(1)}\right]$ to the expert and ask whether the expert is indifferent to the part-utilities $U_i(x_i')$ and $U_i\left(x_j^{(1)}\right)$, where x_i' is calculated as below.

Equation 3:

$$U_i(x_i') = U_j\left(x_j^{(1)}\right)$$
$$\{w_i * U_i'(x_i')\} = \left\{w_j * U_j'\left(x_j^{(1)}\right)\right\}$$
$$U_i'(x_i') = \frac{w_j}{w_i} * U_j'\left(x_j^{(1)}\right)$$
$$x_i' = U_i'^{-1}\left(\frac{w_j}{w_i} * U_j'\left(x_j^{(1)}\right)\right)$$
$$x_i' = U_i'^{-1}\left(\frac{w_j}{w_i}\right)$$
$$since\ U_j'\left(x_j^{(1)}\right) = 1\ following\ normalization$$