David Hughes, E. Waddingham, S. Mt-Isa, Alesia Goginsky, Edmond Chan, Gerald Downey, C. E. Hallgreen, K.S. Hockley, J. Juhaeri, Alfons Lieftucht, M.A. Metcalf, R.A. Noel, L.D. Phillips, Deborah Ashby and Alain Micaleff

Recommendations for benefit–risk assessment methodologies and visual representations

Article (Accepted version)
(Refereed)

Recommendations for Benefit-Risk Assessment Methodologies and Visual Representations

Diana Hughes*, Vice President, Worldwide Safety Strategy, Pfizer, NY Office 219/9/5, 219 East 42nd Street, New York, NY 10017

Ed Waddingham*~, Research Assistant, Imperial College London, Faculty of Medicine, St Mary's Campus, Norfolk Place, Paddington, London W2 1PG

Shahrul Mt-Isa, Research Fellow, Imperial College London, Faculty of Medicine, St Mary's Campus, Norfolk Place, Paddington, London W2 1PG

Alesia Goginsky, Business Process Analyst, Pfizer, 235 East 42nd Street, New York, NY 10017 USA

Edmond Chan, Director, Oncology Signal Detection and Analytics Physician Team Lead, Global Medical Organisation, Division of Janssen-Cilag Ltd, Johnson & Johnson , 50 - 100 Holmers Farm Way, High Wycombe, Buckinghamshire, HP12 4DP

Gerald F. Downey, Biostatistics Senior Manager, Amgen Limited, 240 Cambridge Science Park, Milton Road, Cambridge, CB4 0WD

Christine E. Hallgreen, Honorary research assistant, Imperial College London, Faculty of Medicine, St Mary's Campus, Norfolk Place, Paddington, London W2 1PG; Assistant professor, University of Copenhagen, Faculty of Health and Medical Science, Department of Pharmacy, CORS, Universitetsparken 2, 2100 Copenhagen Ø, Denmark

Kimberley S. Hockley, Research Associate, Imperial College London, Faculty of Medicine, St Mary's Campus, Norfolk Place, Paddington, London W2 1PG

Juhaeri Juhaeri, Head, Pharmacoepidemiology, Global Pharmacovigilance and Epidemiology, Sanofi, 55 Corporate Drive, Bridgewater, NJ 08807

Alfons Lieftucht, Director, Benefit-Risk Evaluation, RD Chief Medical Office, GlaxoSmithKline UK Ltd, 1-3 Ironbridge Road, Uxbridge, UB11 1BT

Marilyn A. Metcalf, Sr Dir, Benefit Risk Evaluation, US Safety Mgmt, RD Chief Medical Office, GSK, 5 Moore Drive, PO Box 13398, RTP, NC 27709-3398, United States

Rebecca A. Noel, Director, Benefit-Risk Assessment, Eli Lilly and Company, Lilly Corporate Center, Indianapolis IN 46285

Lawrence D. Phillips Emeritus Professor, Department of Management, London School of Economics, Houghton Street, London NW3 1AH

Deborah Ashby†, Co-Director, Imperial Clinical Trials Unit, Imperial College London, Faculty of Medicine, St Mary's Campus, Norfolk Place, Paddington, London W2 1PG

Alain Micaleff†, Former employee, MerckSerono SA, Terre Bonne Business Park, route de Crassier 7

CH- 1262 Eysins, Switzerland

for the PROTECT Benefit-Risk Group

*indicates joint lead authors

†indicates joint senior authors

~indicates corresponding author – email e.waddingham@imperial.ac.uk

Word count: 3,668 without Acknowledgements or 4,089 with Acknowledgements

Key messages

- Formal and transparent discussion of multiple viewpoints, interests and priorities facilitates mutual understanding of complex decision problems
- Benefit-risk assessments of treatments should be undertaken in a structured way so that it is clear how a decision on the overall balance of a treatment's effects has been reached
- Various structured approaches and singular methodologies/visual representations are available to support benefit-risk assessment of medicines, but so far universal agreement as to the most suitable method for structured benefit-risk assessment has been lacking
- A team combining expertise from public and private institutions carried out a review of benefit-risk methods and visual representations, including application of the tools to case studies based on real regulatory scenarios
- The project produced a clear set of practical recommendations for undertaking benefit-risk assessments, organised around a generic, five stage benefit-risk assessment roadmap

Keywords: Benefit-risk, decision making, drug development, regulation

Acknowledgements

Abstract

Purpose

To draw on the practical experience from the PROTECT BR case studies and make recommendations regarding the application of a number of methodologies and visual representations for benefit-risk assessment.

Methods

Eight case studies based on the benefit-risk balance of real medicines were used to test various methodologies that had been identified from the literature as having potential applications in benefit-risk assessment. Recommendations were drawn up based on the results of the case studies.

Results

A general pathway through the case studies was evident, with various classes of methodologies having roles to play at different stages. Descriptive and quantitative frameworks were widely used throughout to structure problems, with other methods such as metrics, estimation techniques and elicitation techniques providing ways to incorporate technical or numerical data from various sources. Similarly, tree diagrams and effects tables were universally adopted, with other visualisations available to suit specific methodologies or tasks as required. Every assessment was found to follow five broad stages: 1) Planning, 2)Evidence gathering and data preparation, 3) Analysis, 4) Exploration, and 5) Conclusion and dissemination.

Conclusions

Adopting formal, structured approaches to benefit-risk assessment was feasible in real-world problems and facilitated clear, transparent decision making. Prior to this work, no extensive practical application and appraisal of methodologies had been conducted using real world case examples, leaving users with limited knowledge of their usefulness in the real world. The practical guidance provided here takes us one step closer to a harmonised approach to benefit-risk assessment from multiple perspectives.

Introduction

Benefit-risk assessments play a critical role in bringing treatments to market, providing crucial information for decisions regarding (among others) drug development, licensing and reimbursement. In such situations, judgements by individuals or committees have traditionally been the main approach. However, without an explicit, systematic framework to capture the logic around these assessments, there has been increasing concern among companies and regulators about non-standardised, implicit and often qualitative approaches, with the Council for International Organizations of Medical Sciences (CIOMS) IV suggesting that explicit, quantitative statements would improve the transparency and consistency of decisions.[1] It was this concern that led the European Medicines Agency to establish the three-year Benefit-Risk Methodology Project in 2009,[2] and the ongoing testing of tools and processes for balancing the key benefits and risks of a new medicinal product.[3] In the US since 2010, the Food and Drug Administration and industry worked together to introduce a formal framework for benefit-risk assessment into the reauthorization of the Prescription Drug User Fee Act.[4]

Various structured approaches to decision-making have been developed and widely employed in other fields to address similar problems, and many could theoretically be applied in benefit-risk assessment to address concerns about the decision-making process. However, these have not traditionally been used in this field, and no single agreed method exists for integrating benefit-risk data or to determine the overall balance, and hence arrive at a treatment decision. Importantly, while several methodologies had been proposed prior to this project, a thorough appraisal of methodologies and practical applications in a large number of different real-life case studies had been missing, leaving practitioners with limited guidance.

The Innovative Medicines Initiative's PROTECT project (Pharmacoepidemiological Research on Outcomes of Therapeutics by a European ConsorTium) was established with the aim of strengthening the monitoring of the benefit-risk balance of medicines in Europe. This paper reports the findings of PROTECT's Benefit-Risk Group (PROTECT BR) since inception in September 2009. The group's objectives were to: "1) identify, characterise and test methods of collating data on benefits and risks from various data sources, parameters and strengths of evidence, and of integrating them with decision-criteria and formal assessment of values of patients, healthcare providers, regulators, the pharmaceutical industry and in benefit-risk assessment; 2) identify, test and compare modelling approaches that would allow continuous benefit-risk risk-modelling along the lifecycle of the product, and support decision-making; and 3) develop methods of graphical expression of the benefits and risks of the medicinal products for use by patients, healthcare providers, the pharmaceutical industry and regulators along the lifecycle of the product"[5]. A variety of organisations from the public, private and academic sectors participated in the group.

A total of thirteen methodologies with features representative of their categories were selected for investigation in the case studies. This list was not intended to be exhaustive or restrictive, but it was believed that, taken together, the selected methodologies would be a sufficiently powerful toolbox for most benefit-risk assessments.[6] The thirteen methodologies, their classification and the abbreviations used to refer to them are shown in Table 1.

Due to the limited space available for the main text of this paper, a degree of familiarity with the names and basic features of the methods and visualisations is assumed on the part of the reader. For those with no knowledge of the methods, a description of each method and visual type is provided in the Appendices. The purpose of this paper is to augment the existing descriptions and theoretical appraisals of these benefit-risk assessment methods and visual representations with practical experience and recommendations from the PROTECT BR case studies.

<<Table 1 here>>

Methods

The case studies were selected based on real world scenarios involving medicines where the publicly available data suggested a marginal benefit-risk balance and which presented various practical challenges to stress-test candidate methods. The drugs and indications that formed the focus of the case studies are shown in Table 2. The methods were tested in two 'waves' comprising four case studies each. The first wave established the feasibility of many candidate methods by applying them to a straightforward benefit-risk assessment problem, while the second explored more complex scenarios or methods, and applied selected visual representation techniques. Each case study was highly collaborative in nature, with participants drawn from the range of public and private organisations within PROTECT BR. The case study teams worked independently of each other, but some individuals worked on more than one case study.

<<Table 2 here>>

The case study teams used only publicly available evidence on treatment effects. The qualitative and quantitative methods applied throughout the case studies drew heavily on the principles of decision analysis, which provided the foundations for both the BRAT and PrOACT-URL frameworks, and the theory underlying the most frequently used quantitative approach, MCDA, was used in some form in all of our case studies.

Following the case studies, the key findings and lessons learnt from the reviews and case studies were organised according to five stages representing an approximate chronological order of activities to be undertaken for a benefit-risk assessment (the Recommendations Roadmap), which aims to provide an overarching view of benefit-risk assessment for those who are new to the process, while also allowing more experienced readers to hone in on technical aspects or methodologies of interest.[41]

The efalizumab study has been chosen to provide the illustrations in this paper as it has features that were shared by many of the case studies, is comprehensive enough to illustrate a range of approaches and provides visual examples that are relatively simple.

Results

Case studies

Each of the eight case studies applied several methodologies in combination or in parallel, as shown in Table 1. In total, ten of the thirteen methodologies recommended by Mt-Isa et al[6] were tested; no suitable data could be found for the QALY, Q-TWiST or INHB methodologies in any case study. Two additional quantitative frameworks were also tested: weighted Net Clinical Benefit[42] (wNCB) in the natalizumab case study and Sarac's Benefit-Risk Assessment Method[43] (SBRAM) in the telithromycin case study. wNCB is a utility-weighted extension of NNT/NNH, and can be seen as a special case of MCDA that is simpler to apply in some situations. SBRAM has a similar structure to the other frameworks but with a unique way of scoring the data on each treatment effect.

Methods were chosen according to their suitability for the underlying decision problem and compatibility with the available data (for example, some methods such as NNT/NNH only work with binary outcome data). Methods were selected by group consensus within the case study teams; it is acknowledged that individuals may have proposed methods based on their own experience or research interests. In the first wave of case studies, teams were required to use as many methods as possible, giving reasons for any excluded methods, in order to encourage a comprehensive evaluation. The reasons given for selecting or excluding each method at this stage are set out in Appendix 2 and summarised in Table 1. Each case study in the second wave, by contrast, was aimed at testing a particular method or subset of methods as indicated in Appendix 3.

A common structure to the case studies emerged, with each team using a *descriptive* (i.e. qualitative) *framework* to document key contextual aspects of the benefit-risk assessment and to act as a foundation and a guide for the application of formal methods for specific subtasks, namely: measuring/expressing outcomes (*metrics*), extracting data (*estimation techniques*), eliciting preferences (*utility survey techniques)* and integrating effects data with preferences (*quantitative frameworks).* This common structure is shown in Figure 1.

<<Figure 1 here>>

A common pathway through each case study also emerged, with each team proceeding in broadly the following order:

1. select a *descriptive framework*
2. consider and document basic aspects of the decision context, beginning with a statement of the decision question itself in fairly general terms and moving towards more specific practical aspects of the problem such as establishing what alternative treatments exist and what data are available
3. examine the data and establish what *metrics* could be used to express the favourable and unfavourable effects
4. extract the data in the desired form using *estimation techniques*
5. optionally, elicit preference information using *utility survey techniques* and integrate this with the effects data using a *quantitative framework*
6. bring the results of the previous step back into the *descriptive framework* and proceed to conduct sensitivity analyses and communicate the findings.

This critical path is shown on Figure 1 by the curved arrow. It is important to note that, although this indicates the general order in which to proceed, the teams often found that it was necessary to look a few steps ahead during the process and/or to revisit earlier tasks in the light of what was uncovered later (for example, the choice of quantitative framework could in some cases limit the range of metrics that could be adopted). The next few paragraphs deal with each step of the path in turn.

The *descriptive frameworks* PrOACT-URL and BRAT were found to be useful guides for planning and executing a benefit-risk assessment and are considered suitable for use at any stage of a medicinal product's life cycle. Each was used in six out of the eight case studies, with the first wave of four case studies employing both frameworks in parallel. The frameworks provide a structure for breaking down a benefit-risk problem into a stepwise thought process. The list of steps is somewhat similar in both frameworks, although they do not perfectly map onto one another. At the time of the initial methodology review, PrOACT-URL and BRAT were the two most promising descriptive frameworks; other frameworks may have since emerged, but we have not reviewed these and are unable to comment on their suitability.

The *descriptive frameworks* encourage clear delineation of the decision problem, which in the benefit-risk context means setting out the treatment under investigation, the indication and target population of interest, and any specific efficacy or safety concerns that have prompted the assessment. Clinical expertise can then guide the choice of appropriate comparators, i.e. existing alternative treatments that act as benchmarks for the benefit-risk balance.

The favourable and unfavourable effects in the assessment should include, at a minimum, key efficacy measures, any adverse events that may have prompted the assessment, and the key side effects of the treatment under investigation and of all the comparators. In practice, our case study teams found selecting a complete set of relevant effects to be a surprisingly difficult task. The assessor typically starts with an exhaustive list of clinical outcomes for a given treatment and indication, and must attempt to narrow this down to those benefits and risks that have a substantial impact on the benefit-risk balance, aiming to represent the range of treatment effects as fully as possible while avoiding problems such as double-counting of endpoints[44]. The effects are often displayed in a tree diagram or value tree such as that shown in Figure 2, which is a simplified version of the final tree from the efalizumab case study. Such tree diagrams can be used with any methodology that handles multiple outcomes in order to aid understanding of the structure of the problem, and are typically drawn up in the early stages of the PrOACT-URL or BRAT frameworks and/or used to guide the weighting process in quantitative methods such as MCDA. To avoid bias, decision-makers should establish a common understanding and state the assumptions underlying the selection of benefits and risks. This allows for a transparent and auditable selection process. Ideally, both clinical expertise and patients' views inform selection of the most relevant benefits and risks, and sufficient time should be allowed to resolve any disagreements through group discussion.

<<Figure 2 here>>

Our case studies were retrospective in nature and used publicly available data, principally from published trial reports and public registration documents (such as European public assessment reports and periodic safety update reports in the European Union, or periodic adverse drug experience reports in the United States), which are a convenient summary of the data from pivotal studies allowing replication and further exploration by others. However, these documents have clear limitations for benefit-risk assessment. They are low on detail, and reporting standards and outcome definitions frequently vary. In some cases, to facilitate testing of the more complex methods, the teams made strong modelling assumptions on an ad hoc basis to align the data (for example, to convert between related outcome measures where the same measure was not reported for all comparators). We recognise that this may have introduced bias in some instances and that real-world assessors would need to proceed more carefully with any data manipulations, or to use more qualitative methods and avoid the problem altogether. Nevertheless, we feel our approach was justified for the purpose of this project, which was to test methodologies rather than comment on the benefit-risk balance of medicines.

Metrics are measures used to numerically express the absolute or relative value of treatment effects. This includes everyday outcome measures such as incidence rates, which were not specifically reviewed by PROTECT-BR owing to their familiarity and ubiquity in medical reporting, but were nevertheless indispensable in the case studies. The

group of metrics known as Impact Numbers go further in describing the numbers of people affected by (binary) events in a population, but were not found to be particularly useful for making benefit-risk decisions. The NNT/NNH and BRR metrics were developed to compare multiple effects for benefit-risk assessment, but in their standard form they suffer from being able to compare only one favourable and one unfavourable effect, and from not explicitly acknowledging any difference in importance between the effects (even, arguably, implying they are of equal importance). To address these shortcomings, extensions to NNT/NNH have been developed, effectively resulting in the wNCB quantitative framework which is discussed below.

Estimation techniques are designed to facilitate the extraction and/or synthesis of appropriate values for the chosen metrics. Again, widely known epidemiological/biostatistical methods such as 2x2 tables or meta-analyses can be seen to fall within this class but have not been evaluated by PROTECT BR owing to their familiarity. The two specific estimation techniques evaluated were ITC/MTC and PSM, and both were found particularly useful in the case studies as they lend themselves naturally to benefit-risk assessments. ITC/MTC is designed to bring together evidence on several treatments where each source study does not compare all treatments simultaneously, a situation that arose in both the natalizumab and rimonabant case studies. PSM is used to propagate uncertainty through complex multi-variable models, and as such was the only tool capable of quantifying the uncertainty of the overall benefit-risk balance in many of the case studies.

Quantitative frameworks integrate objective treatment effects data with subjective preference data, i.e. utilities and/or weights that give information regarding the relative importance of the treatment effects. MCDA was the most commonly used quantitative framework in the case studies, owing partly to its comprehensiveness, flexibility (unlike some other methods it is not restricted in the types of outcome metrics it can handle) and its natural link with the PrOACT-URL descriptive framework. SMAA was also used widely in the case studies; this method is an extension of MCDA that explores all possible weightings in the event that utility/preference information is missing or limited, essentially applying PSM to the weights in an MCDA model. This is certainly a useful tool, but we would not recommend it as the default approach; it would be a shame if most benefit-risk assessments employed quantitative decision models and yet made no attempt to understand the underlying trade-offs. Also evaluated were wNCB, a weighted extension of the NNT/NNH metric and which is similar to MCDA but limited to binary outcomes, and SBRAM, which is again similar to MCDA but uses a simplified scoring system that does not always discriminate well between different options.

Utility survey techniques are used to obtain utility/preference information for use in a quantitative framework. MCDA is usually implemented via a simple pairwise weighting process whereby stakeholders directly quantify the importance of outcomes ("swing weighting"), but the rimonabant case study also combined MCDA with DCE, a utility survey technique that presents participants with a set of binary choice scenarios in which the outcomes take different values. For simple problems, a well-designed DCE with a sufficiently large number of responses can arguably provide the most comprehensive preference information, but DCEs require significant resources to design, and it has been argued that they do not work well when more than seven outcomes are to be considered simultaneously (a limit that was exceeded in most of our case studies). Swing weighting is more easily adaptable to different problems and was therefore favoured in the case studies.

The reports available at http://www.imi-protect.eu/benefitsRep provide detailed accounts of the application of the methods to the case study problems. This paper includes several examples from the efalizumab case study which asked, hypothetically, whether the overall benefit-risk balance of the drug as a treatment for chronic plaque psoriasis was favourable based on publicly available data at the time of regulatory review. However, this work was intended to test methodologies and representations for the evaluation of benefit and risk of medicines. It neither replaces nor is intended to replace or comment on any regulatory decisions made by national regulatory agencies, nor the European Medicines Agency.

Appraisal of visual representations

Among the variety of visual representations applied throughout the case studies, two visual types stood out as playing a fundamental role in every assessment. These are the *value tree* (or *tree diagram*) (Figure 2) and *effects table* (or *data table*) (Table 3) (examples taken from the efalizumab case study). The value tree is a simple visual hierarchy that displays the favourable and unfavourable effects (with clear definitions provided) and organises them into smaller groups to aid understanding (by adopting a logical structure based on, say, body systems) and/or facilitate elicitation of preferences (by grouping together outcomes according to seriousness, duration or any other factor that aids comparison).

<<Table 3 here>>

The effects table also lists the favourable and unfavourable effects, not necessarily grouped this time but with the numerical values of the outcome metrics for each comparator and other key data clearly displayed in each row. The data shown in the effects table is purely descriptive, i.e. it does not incorporate any in-depth analysis or preference information. The effects table is an important milestone in a benefit-risk balance because it summarises all the objective data and prompts the assessor to consider what more is required. The data may clearly show that one treatment has a superior benefit-risk balance than all its comparators. If, alternatively, the data in the effects table shows no clear advantage for any one treatment then quantitative modelling of the benefit-risk balance (i.e. assigning preferences to the treatment effects) may be considered, either on a fully quantitative basis (with explicit utilities) or a partially quantitative one (a complete or partial ranking of the treatment effects that eliminates any doubt as to the overall balance). Figure 3 is a flowchart showing the possible ways in which a benefit-risk assessment may proceed at this point.

<<Figure 3 here>>

Besides the value tree and effects table, a variety of other visual types were used in the case studies. Figures 4 and 5 are a *difference display* and a *line graph* from the efalizumab case study, showing, respectively, the results of a quantitative analysis using MCDA and a sensitivity analysis on a key preference parameter.

PROTECT BR's visual review workstream investigated visuals in greater depth, identifying 14 visual types. The team made recommendations on how to create these visuals by considering four audience-visual compatibility criteria, and how to determine appropriate visuals for benefit-risk information through a series of key benefit-risk questions. Further recommendations on visuals, including their style and design, were published in a full report[45] and in another article[46]. More details, including interactive visual displays created within PROTECT BR's case studies, are available at http://protectbenefitrisk.eu/visualisations.html.

<<Figure 4 here>>

<<Figure 5 here>>

Recommendations and Conclusions

The critical path through a benefit-risk assessment shown in Figure 1 can be organised according to five broad stages common to all benefit-risk assessments:[41] 1) Planning, 2) Evidence Gathering and Data Preparation, 3) Analysis, 4) Exploration and 5) Conclusion and Dissemination, as shown in Figure 6 with key recommendations for each stage.

<<Figure 6 here>>

Regulators, such as the European Medicines Agency, and pharmaceutical companies have begun implementing structured approaches to benefit-risk assessments, but each organisation has adopted its own set of frameworks and tools, leading some commenters to propose a harmonisation initiative.[1, 47, 48] It is clear that many of the frameworks and methods have common elements; identifying these and finding a shared, transparent language to describe them is arguably more realistic than finding a one-size-fits-all approach to benefit-risk assessment. PROTECT BR has stopped short of recommending outright any particular framework or methodology, emphasising instead the importance of a structured approach with careful planning and execution. The resulting recommendations provide guidance on the tools available for benefit-risk assessment as the discipline evolves and a more harmonised approach begins to emerge. We suggest that techniques reviewed by this project lend themselves to the inclusion of stakeholders who bring many perspectives and whose input can be included systematically, in qualitative and quantitative ways, to enhance the overall benefit-risk assessment. We strongly encourage the use of structured approaches to provide added clarity around the assessment of favourable and unfavourable effects of medicines.

Unanswered questions and ongoing research
The scope of this project was limited in that all case studies refer to pharmaceutical prescription medicines. Vaccines and over-the-counter medicines were not evaluated and may require different techniques. The usefulness of methodologies recommended by PROTECT BR for the benefit-risk evaluation of vaccines is being evaluated within IMI ADVANCE[49].

One theme whose importance has become apparent during the project is patient and public involvement (PPI) in benefit-risk assessment: It is possible that the benefit-risk balance of a treatment may vary depending on whose perspective is adopted, and there is a strong case for representing all stakeholders in the assessment process. There is work to be done to establish the extent to which PPI in benefit-risk assessment is desirable/feasible and to test specific methods for its application.

References

1. CIOMS Working Group IV. Benefit-risk balance for marketed drugs: Evaluating safety signals. http://www.cioms.ch/publications/g4-benefit-risk.pdf: CIOMS.

2. European Medicines Agency. Benefit-risk methodology project. http://www.ema.europa.eu/ema/index.jsp?curl=pages/special_topics/document_listing/document_listing_000314.jsp&mid=WC0b01ac0580223ed6; 2009.

3. Benefit-risk methodology project; Update on work package 5: Effects Table pilot (Phase I). (EMA/74168/2014); 2014. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Report/2014/02/WC500162036.pdf.

4. Structured Approach to Benefit-Risk Assessment in Drug Regulatory Decision-Making. Draft PDUFA V Implementation Plan; 2013. Available from: http://www.fda.gov/downloads/forindustry/userfees/prescriptiondruguserfee/ucm329758.pdf.

5. Annex 1 – Grant Agreement N° 115004. PROTECT - Description of work; 2013 [cited 2014 8/27]. Available from: https://eroombayer.de/eRoom/PH-GDC-PI-SID/IMI-PROTECT/0_10a702.

6. Mt-Isa S, Hallgreen CE, Wang N, Callréus T, Genov G, Hirsch I, Hobbiger SF, Hockley KS, Luciani D, Phillips LD, et al. Balancing benefit and risk of medicines: A systematic review and classification of available methodologies. Pharmacoepidemiol Drug Saf 2014;23(7):667-78.

7. Hunink M, Glasziou P, Siegel J, Weeks J, Pliskin J, Elstein A, Weinstein MC. Decision making in health and medicine: Integrating evidence and values. Cambridge: Cambridge University Press; 2001.

8. Hammond JS, Keeney RL, Raiffa H. Smart choices: A practical guide to making better decisions. Boston, MA: Harvard Business School Press; 2002.

9. Coplan PM, Noel RA, Levitan BS, Ferguson J, Mussen F. Development of a framework for enhancing the transparency, reproducibility and communication of the benefit-risk balance of medicines. Clin Pharmacol Ther 2011 02;89(2):312-5.

10. Levitan BS, Andrews EB, Gilsenan A, Ferguson J, Noel RA, Coplan PM, Mussen F. Application of the BRAT framework to case studies: Observations and insights. Clin Pharmacol Ther 2011 02;89(2):217-24.

11. Mussen F, Salek S, Walker S. Benefit-risk appraisal of medicines. John Wiley & Sons, Ltd; 2009.

12. Dodgson J, Spackman M, Pearman A, Phillips LD. Multi-criteria analysis: A manual. 2000.

13. Keeney RL, Raiffa H. Decisions with multiple objectives: Preference and value tradeoffs. New York: John Wiley; 1976.

14. Tervonen T, Figueira JR. A survey on stochastic multicriteria acceptability analysis methods. Journal of Multi-Criteria Decision Analysis 2008;15(1-2):1-14.

15. 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. Statist.Med. 2011;30(12):1419-28.

16. Lahdelma R, Hokkanen J, Salminen P. SMAA - stochastic multiobjective acceptability analysis. Eur J Oper Res 1998 04/01;106(1):137-43.

17. Holden WL, Juhaeri J, Dai W. Benefit-risk analysis: A proposal using quantitative methods. Pharmacoepidemiol Drug Saf 2003 10;12(7):611-6.

18. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. N Engl J Med 1988 06/30;318(26):1728-33.

19. Attia J, Page J, Heller RF, Dobson AJ. Impact numbers in health policy decisions. J Epidemiol Community Health 2002 08;56(8):600-5.

20. Heller RF, Dobson AJ, Attia J, Page J. Impact numbers: Measures of risk factor impact on the whole population from case-control and cohort studies. J Epidemiol Community Health 2002 08;56(8):606-10.

21. Heller RF, Buchan I, Edwards R, Lyratzopoulos G, McElduff P, Leger SS. Communicating risks at the population level: Application of population impact numbers. BMJ 2003 11/15;327(7424):1162-5.

22. Verma A, Torun P, Harris E, Edwards R, Gemmell I, Harrison RA, Buchan IE, Davies L, Patterson L, Heller RF. Population impact analysis: A framework for assessing the population impact of a risk or intervention. J.Public Health (Oxf) 2012 03;34(1):83-9.

23. Chuang-Stein C, Entsuah R, Pritchett Y. Measures for conducting comparative benefit: Risk assessment. Drug Inf J 2008;42(3):223-33.

24. Korting H, Schafer-Korting M. The benefit-risk ratio. A handbook for the rational use of potentially hazardous drugs. Boca Raton: CRC Press LLC; 1999.

25. Payne JT, Loken MK. A survey of the benefits and risks in the practice of radiology. CRC Crit Rev Clin Radiol Nucl Med 1975;6(3):425-39.

26. Sassi F. Calculating QALYs, comparing QALY and DALY calculations. Health Policy Plan 2006 09;21(5):402-8.

27. Ried W. QALYs versus HYEs--what's right and what's wrong. A review of the controversy. J Health Econ 1998 10;17(5):607-25.

28. Gelber RD, Cole BF, Gelber S, Aron G. Comparing treatments using quality-adjusted survival: The Q-twist method. The American Statistician 1995 05/01;49(2):161-9.

29. Goldhirsch A, Gelber RD, Simes RJ, Glasziou P, Coates AS. Costs and benefits of adjuvant therapy in breast cancer: A quality-adjusted survival analysis. J Clin Oncol 1989 01;7(1):36-44.

30. Garrison LP, Towse A, Bresnahan BW. Assessing a structured, quantitative health outcomes approach to drug risk-benefit analysis. Health Aff (Millwood) 2007 05;26(3):684-95.

31. Lynd LD, Najafzadeh M, Colley L, Byrne MF, Willan AR, Sculpher MJ, Johnson FR, Hauber AB. Using the incremental net benefit framework for quantitative benefit-risk analysis in regulatory decision-making--a case study of alosetron in irritable bowel syndrome. Value.Health 2010 06;13(4):411-7.

32. Minelli C, Abrams KR, Sutton AJ, Cooper NJ. Benefits and harms associated with hormone replacement therapy: Clinical decision analysis. BMJ 2004 02/14;328(7436):371.

33. Lynd LD, O'Brien BJ. Advances in risk-benefit evaluation using probabilistic simulation methods: An application to the prophylaxis of deep vein thrombosis. J Clin Epidemiol 2004 08;57(8):795-803.

34. van Staa TP, Smeeth L, Persson I, Parkinson J, Leufkens HG. What is the harm-benefit ratio of cox-2 inhibitors? Int J Epidemiol 2008 04;37(2):405-13.

35. Lumley T. Network meta-analysis for indirect treatment comparisons. Stat Med 2002 08/30;21(16):2313-24.

36. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med 2004 10/30;23(20):3105-24.

37. Nixon RM, Bansback N, Brennan A. Using mixed treatment comparisons and meta-regression to perform indirect comparisons to estimate the efficacy of biologic treatments in rheumatoid arthritis. Stat Med 2007 03/15;26(6):1237-54.

38. Ryan M, Gerard K, Amaya-Amaya M. Using discrete choice experiments to value health and health care. Dordrecht, The Netherlands: Springer; 2008.

39. Ryan M, Hughes J. Using conjoint analysis to assess women's preferences for miscarriage management. Health Econ 1997;6(3):261-73.

40. Ryan M, Bate A, Eastmond CJ, Ludbrook A. Use of discrete choice experiments to elicit preferences. Quality in Health Care 2001 09;10:I55-60.

41. Recommendations for the methodology and visualisation techniques to be used in the assessment of benefit and risk of medicines; 2013. Available from: http://www.imi-protect.eu/documents/HughesetalRecommendationsforthemethodologyandvisualisationtechniquestobeusedintheassessmento.pdf.

42. Sutton AJ, Cooper NJ, Abrams KR, Lambert PC, Jones DR. A bayesian approach to evaluating net clinical benefit allowed for parameter uncertainty. J Clin Epidemiol 2005;58(1):26-40.

43. Sarac SB, FAU RC, FAU RM, FAU HC, Soeborg TF, Colding-Jorgensen M FAU - Christensen, Per,K., FAU CP, Thirstrup SF, Mosekilde E. A comprehensive approach to benefit-risk assessment in drug development. Basic & Clinical Pharmacology & Toxicology JID - 101208422 1018.

44. Phillips LD. Benefit-risk modeling of medicinal products: Methods and applications. In: A. Sashegyi, J. Felli, R. Noel, editors. Benefit-risk assessment in pharmaceutical research and development. Boca Raton, FL: CRC Press; 2013.

45. Review of visualisation methods for the representation of benefit-risk assessment of medication: Stage 2 of 2 [Internet]; c2013 [cited 2014 8/27]. Available from: http://www.imi-protect.eu/documents/ShahruletalReviewofvisualisationmethodsfortherepresentationofBRassessmentofmedicationStage2A.pdf.

46. Hallgreen, C., et al. Literature review of visual representation of the results of benefit-risk assessments of medicinal products. Pharmacoepidemiology and Drug Safety 2015 doi: 10.1002/pds.3880.

47. ICH. Final concept paper, 15 December 2010: Periodic safety update reports for marketed drugs E2C(R2) and gap and potential improvement analysis of ICH E2C, E2E and E2F. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E2C/Concept_papers/E2C_R2__Final_Concept_Paper_December_2010.pdf: ICH; 2010.

48. Franson T, Bonforte P. Benefit-risk in the US regulatory context. In: A. Sashegyi, J. Felli, R. Noel, editors. Benefit-risk assessment in pharmaceutical research and development. Boca Raton, FL: CRC Press; 2013. .

49. IMI ADVANCE; 2014. Available from: http://www.advance-vaccines.eu/.

	Description	Features	Rationale for use
Frameworks			
PrOACT-URL[7,8]	Problem, Objectives, Alternatives, Consequences, Trade-offs, Uncertainty, Risk, and Linked decisions	A structured, qualitative approach, based on decision theory, of issues to consider in assessing the benefit-risk balance of a drug and its comparator(s).	All models used either PrOACT-URL or BRAT, or both, to guide the modelling process. They are similar in contributing transparency and structure to the benefit-risk balance. For the Warfarin team, BRAT helped to identify relevant clinical effects and to visualise magnitudes. The Rimonabant team found PrOACT-URL to be more comprehensive, and for the Rosiglitazone team, it encouraged a focus on value trade-offs.
BRAT[9,10]	Benefit Risk Action Team	A structured, qualitative approach to assessing benefit-risk, based on MCDA, similar to PrOACT-URL, and supported by a set of guidelines and a tool.	
MCDA[11-13]	Multi Criteria Decision Analysis	A quantitative methodology for integrating multiple benefit and risk criteria for a drug and its comparator(s), to provide for each option an index of the benefit-risk balance.	All teams chose MCDA in both waves (SMAA for warfarin) because of its comprehensiveness, accommodation of any effect metrics and value judgements, and support for trade-off weighting, all requirements for a fully quantitative model.
SMAA[14-16]	Stochastic Multi-criteria Acceptability Analysis	An extension of MCDA that formally incorporates uncertainty by replacing point estimates of a drug's effects with probability distributions.	The Telithromycin, Rimonabant (wave 2) and Warfarin teams chose SMAA so they could explicitly model uncertainty and explore the effects of different weighting systems.
Metrics			
NNT/NNH[17,18]	Number Needed to Treat/Number Needed To Harm	The reciprocal of the difference in proportions of patients experiencing a given effect between the treatment and control group.	Only rimonabant (wave 1) explored NNT/NNH and Impact Numbers because they are simple ways to communicate a single effect. They are not recommended as general tools for benefit-risk assessment because they do not consider the clinical relevance of the effects nor are trade-offs between effects considered.
IN[19-22]	Impact Numbers	An extension of NNT that takes account of the population being represented and other aspects of context.	

	Description	Features	Rationale for use
BRR[23-25]	Benefit Risk Ratio	The ratio of the magnitude of the most beneficial favourable effect to the magnitude of the most unfavourable effect.	Efalizumab, Telithromycin and Rimonabant (wave 1) chose BRR for its simplicity and ability to trade-off the two effects.
QALY[26,27]	Quality Adjusted Life Years	A health-outcome index describing a patient's level of health on several generic criteria and how an intervention might improve the states of health over time.	None of the teams applied QALYs or the Q-TWiST approach because their focus on specific health outcomes does not take into account the many favourable and unfavourable effects that make up the benefit-risk balance for the drugs modelled in PROTECT.
Q-TWiST[28,29]	Quality-adjusted Time Without Symptoms and Toxicity	A combination of QALYs from three states of a patient undergoing cancer therapy.	
INHB[30-32]	Incremental Net Health Benefit	An extension of QALYs that looks at the difference between benefits of two treatment options minus the difference between risks.	As INHB is defined at present only for cancer treatments, it was not applicable to any of the drugs considered in PROTECT.
Estimation techniques			
PSM[33, 34]	Probabilistic Simulation Method	Uncertainty about each of the effects is represented by a probability distribution; Monte Carlo analysis shows the benefit-risk balance distribution for each drug and for the drug-comparator difference.	All wave 2 case studies chose this technique to see its impact on the benefit-risk balance, to explore scenarios about the effects that could not be seen in the deterministic models and to accommodate uncertainties of lower-quality data.
ITC/MTC[35-37]	Indirect Treatment Comparison/Mixed Treatment Comparison	Compares two treatment effects where direct evidence is unavailable. Uses the link of each treatment to the placebo for calculating the variance of the difference in effects.	Natalizumab and Rimonabant chose ITC/MTC in wave 2 either to deal with the heterogeneity of the data sources or to enable comparison with other active treatments.
Utility survey technique			
DCE[38-40]	Discrete Choice Experiment	Effect weights are derived from a person's preferences between pairs of combinations of health levels across favourable and unfavourable effects. Requires thinking about trade-offs between the effects.	Only Rimonabant used DCE, and this was to elicit preferences from patients.

Table 1. The 13 methodologies, their features and an explanation of how they were used in the case studies.

Drug	Indication
Efalizumab	Moderate to severe plaque psoriasis
Telithromycin	Mild to moderate community-acquired pneumonia (CAP), acute exacerbation of chronic bronchitis (AECB), and acute sinusitis (ABS) in patients of 18 years and older, as well as tonsillitis/pharyngitis caused by Streptococcus pyogenes in adults and adolescents, as an alternative when beta-lactam antibiotics are not appropriate
Natalizumab	Relapsing remitting multiple sclerosis
Rimonabant	Weight loss in obese or overweight patients with co-morbidities
Rosiglitazone + metformin	Type II diabetes
Warfarin	Ischemic stroke in patients with atrial fibrillation

Table 2. Drugs forming the basis of the PROTECT Benefit-Risk case studies

(a)

(b)

Key:

Figure 1. The structure and critical path for applying the methods in the case studies. (a) generic template, (b) efalizumab case study.

Figure 2. A tree diagram for the efalizumab example. The same tree was used in the descriptive PrOACT-URL and BRAT frameworks, and with the quantitative MCDA methodology. See Table 3 for an explanation of the abbreviations used.

	Name	Description	Units	Raptiva	Placebo
Favourable Effects	PASI75	Percentage of patients achieving 75% reduction in baseline PASI[1] at week 12.	%	29.5	2.7
	PGA	Percentage of patients achieving Physician's Global Assessment[2] clear/almost clear at week12.	%	29.5	5.1
	DLQI	Dermatology Life Quality Index[3]. Mean change from base score.	Change score	5.8	2.1
Unfavourable Effects	Severe infections	Proportion of patients experiencing infections serious enough to require hospitalisation.	%/100 ptyrs	2.83	1.4
	PML	Number of cases of progressive multifocal leukoencephalopathy.	number	3	0

[1]PASI is a measure of the average redness, thickness and scaliness of the lesions (each graded on a 0-4 scale), weighted by the body region and the area affected. PASI range is from 0 to 72.

[2]PGA is a seven point scale with 7 being clear, 6 almost clear, 5 mild, 4 mild to moderate, 3 moderate, 2 moderately severe and 1 severe psoriasis.

[3]DLQI is a 10-item quality of life index scored by the patient on a four-point scale (0-3).

Table 3. Effects table for the efalizumab example.

Figure 3. Flowchart indicating the difference between quantitative and qualitative benefit-risk assessments, with recommended methods

Favourable Effects	PGA	14.6	
Favourable Effects	PASI75	13.4	
Favourable Effects	DLQI	8.9	
Unfavourable Effects	Serious Infections	-1.8	
Unfavourable Effects	PML	-17.6	
		17.5	

Figure 4. Results of a quantitative analysis (using MCDA) for the efalizumab example. The difference display shows the contribution of the weighted difference between drug and placebo for each effect. Right-extending (green) bars favour the drug and left-extending (red) bars favour the placebo, for a 17.5 total difference (out of 100) in favour of efalizumab.

Figure 5. Sensitivity analysis for the efalizumab case study, showing the effect of changing the weight on the PML criterion in MCDA. The vertical red line represents the current weight of 18.5 (out of a total of 100 for all five criteria). The intersections of that line with the slanting red and green lines define the 17.5 difference noted in Figure 4. If the PML weight is increased beyond 32, then the benefit-risk balance favours the placebo.

IMI PROTECT Benefit-Risk Assessment Roadmap and Recommendations

1- Planning: We recommend using a descriptive framework such as BRAT or PrOACT-URL to structure each benefit-risk assessment. A set of benefits and risks should be chosen that covers the full range of treatment effects, and represented visually using a tree diagram to indicate the hierarchy. A table template ('effects table' or 'source table') should be prepared, to represent the data that are required to be collected.

2- Evidence Gathering and Data Preparation: Assessors should review all available evidence and select data that are sufficient to and appropriate for the decision problem. The table template must be completed highlighting where data are available or missing for example by colour-coding missing data. The tree diagram and table produced initially may need to be revised in the light of available data.

3- Analysis: The analysis should be appropriate to the complexity of the task. Simple descriptive methods may suffice for everyday benefit-risk assessments, while quantitative decision models can provide additional clarity for more complex problems. When a quantitative benefit-risk assessment approach is used, stakeholders' value preferences and the benefit-risk magnitudes (by criteria and overall) should be represented by suitable bar graphs (particularly useful is the 'difference display'), dot plots or line graphs to promote accurate point reading, local and global comparisons, and judging trade-offs among alternatives. Care should be taken to avoid double counting events or effects in any analysis.

4- Exploration: All benefit-risk assessments should include a sensitivity analysis of some kind. Where benefits and risks are finely balanced, quantitative decision models facilitate the execution and communication of sensitivity analyses by clearly setting out the respective impacts of effects uncertainty and preference uncertainty on the results. The visual representations which should be used at this stage are distribution plots, line graphs, forest plots or tornado plots to provide comprehensive overview of the benefit-risk analysis allowing better-informed decisions.

5- Conclusion and dissemination: Adopting a formal structure for a benefit-risk assessment is an effective way to improve the overall transparency and communicability of the process and facilitate robust decision making.
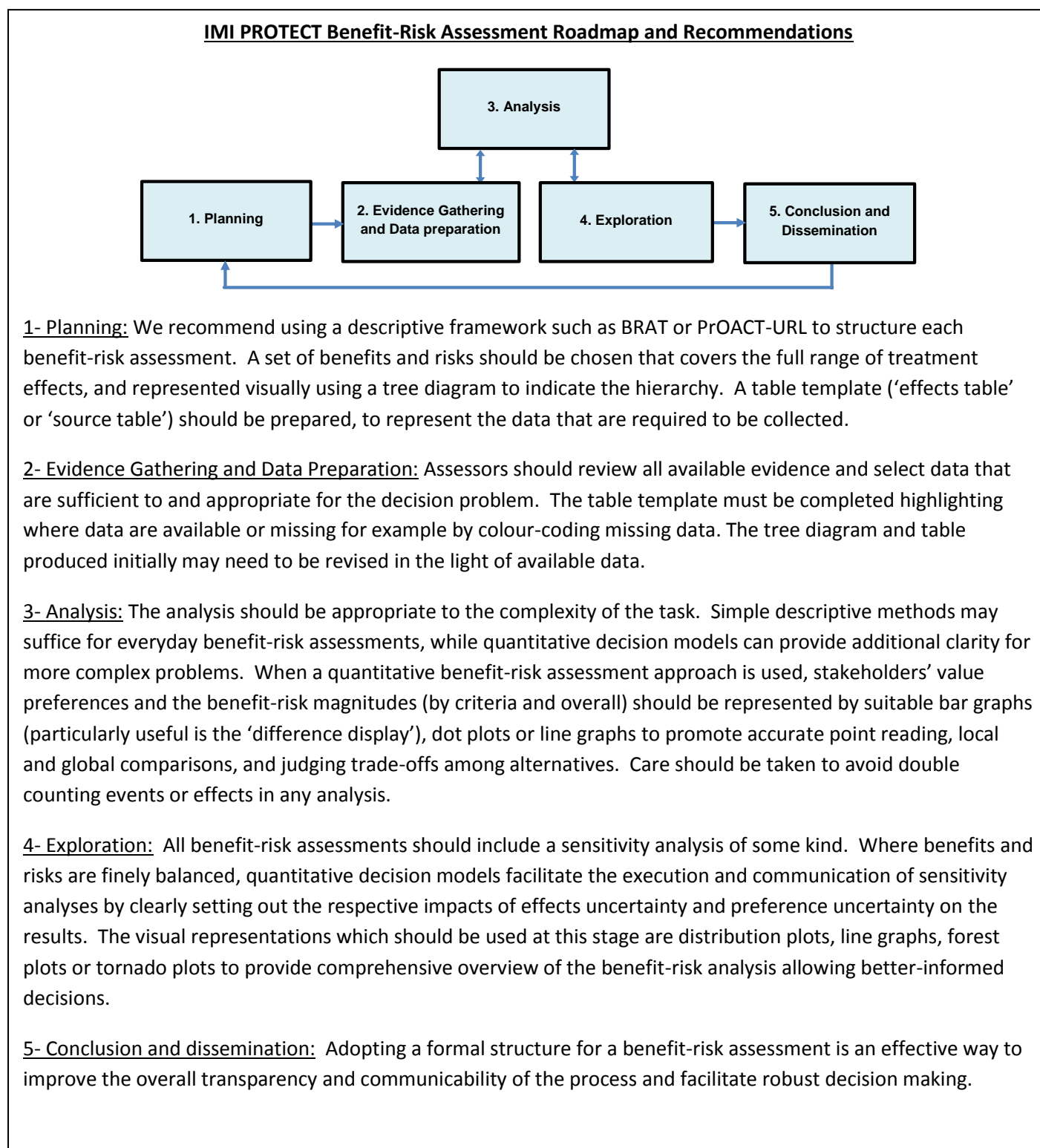
Figure 6. The five-stage Roadmap and Recommendations for benefit-risk assessments