HEcoPerMed

Health Economic Models for Personalised Medicine

Deliverable 2.1: Workshop 1 Report

GUIDANCE AND 3 CASE STUDIES ON HEALTH ECONOMIC MODELLING OF PERSONALISED MEDICINE

Report Information

Title:	GUIDANCE AND 3 CASE STUDIES ON HEALTH ECONOMIC MODELLING OF PERSONALISED MEDICINE
Authors:	Balázs Nagy, László Szilberhorn, Tamas Zelei
Version:	V1.0
Work package:	2
Date of publication:	December 16, 2020
Dissemination level:	Public
Nature:	Report

Project Information

Project acronym:	HEcoPerMed
Project full title:	Health Economic Models for Personalised Medicine
Starting date:	01.01.2019
Duration:	36 months
Coordinator:	Dr. Susanne Giesecke

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 824997.

GUIDANCE AND 3 CASE STUDIES ON HEALTH ECONOMIC MODELLING OF PERSONALISED MEDICINE

WORKSHOP REPORT

Contents

Date: 2020.09.24
Venue: Microsoft Teams
Organizer: Syreon Research Institute on behalf of HEcoPerMed consortium

Participants:	**Consortium:** Andreas Weinhäusel, Apostolos Tsiachristas, Balázs Nagy, Gábor Kovács, Heleen Vellekoop, László Szilberhorn, Manuela Kienegger, Maren Walgenbach, Matthijs Versteegh, Maureen Rutten-van Molken, Rositsa Koleva-Kolarova, Sarah Wordsworth, Simone Huygens, Susanne Giesecke, Tamas Zelei, Wolfgang Ballensiefen **External:** Adam Brooke, Adriána Papp-Zsólyom, Alexander Thompson, Alona Masheiko, Beate Jahn, Bethany Shinkins, Bjørn Oddvar Strøm, Chloe Thomas, David Glynn, Gurdeep Sagoo, Helen Dakin, Jacoline Bouvy, James O'Mahony, Jose Leal, Julia Bonastre, Katherine Payne, Ken Redekop, Oresta Piniazhko, Maarten Postma, Mark Pennington, Michael Gillett, Peter Murphy, Robert McQueen, Tomas Mlčoch, Valeriya Serediuk, Valesca Retel, Zsolt Gaál

1. BACKGROUND

On behalf of the Health Economics for Personalised Medicine (HEcoPerMed) consortium, Syreon Research Institute organized a one-day virtual workshop with international experts including representatives of healthcare payers, testing technology developers/manufacturers and health economists. The modelling guidance for personalised medicine (PM) that was developed in Work Package 1 was presented, as well as preliminary findings from its application in three case studies (Work Package 2). The original plan was to organize a face-to-face workshop. However, due to the COVID-19 outbreak the face-to-face meeting was converted to a virtual one.

2. AIM

The aim of the workshop was to solicit expert opinion that will be used to refine and validate the guidance and discuss the appropriateness of the modelling methods of the selected case studies. Results of the workshop are planned to be channelled back to finetune ongoing guidance development and modelling activities.

3. METHODS

Participants in this virtual workshop were identified from relevant literature and through the professional network of the HEcoPerMed members. The main selection criteria were familiarity and experience with personalised medicine modelling and knowledge of (at least) one of the case study topics, with efforts toward a balanced geographical and gender distribution of participants.
Participants were sent pre-reading materials one week before the meeting that contained the workshop agenda, information about the consortium, a description of the case studies and the draft paper on the guidance for the harmonization and improvement of economic evaluations of personalised medicine. The workshop took place on the 24th of September 2020. The morning session started with a keynote speech by Prof. Sarah Wordsworth, who presented the “Opportunities and challenges for health economics of personalised medicine”. Subsequently, Wolfgang Ballensiefen introduced “HEcoPerMed Consortium and ICPerMed”. Afterwards, Simone Huygens gave a detailed introduction to the “Guidance - Recommendations for economic evaluations of personalised medicine” (developed in Work Package 1). After the presentation the participants were invited to ask questions for clarification. The morning session ended with a brief overview by Balázs Nagy on the “Case study selection rationale”.
During the afternoon session, the participants were separated into three groups according to the case studies:

- Testing for NTRK fusions and starting tumour-agnostic treatment with TRK inhibitors
- Upfront DPYD (dihydropyrimidine dehydrogenase deficiency encoding gene) testing (ToxNav) to personalise chemotherapy treatment for metastatic breast cancer
- Adjusting insulin treatment after genetic testing in Maturity-Onset Diabetes of the Young (MODY)

Each session started with an introduction to the selected case study and the planned modelling methodology. After the overview the participants were invited to ask questions for clarification.
In the next phase, a structured discussion was held regarding the application of the guidance for economic evaluations of personalised medicine to each case study, which was guided by pre-selected guidance items. The relevant items for each case study had been selected by the HEcoPerMed researchers prior to the workshop and were allocated based on each team's priorities, and aimed to ensure wide coverage of the guidance items. In the structured discussion, two questions were raised for each guidance item:

1. Given the context of this case study, does the recommendation seem complete and well-phrased?
2. Do you have any comments on the modelling approach we propose?

For question 1, the participants were asked to choose from three options: "Yes", "No, not complete" and "No, not well-phrased". Answers were managed with Mentimeter polling application, where each participant answered without knowing others' answers. Nonetheless, the answers of each participant were seen by the moderator of the sessions, which enabled the moderator to inquire for the rationale behind the chosen answer when necessary. The first question – which aimed to gather participants' reflections on the comprehensiveness and wording of the recommendations included in the guidance – was only opened for discussion in case of disagreement, where disagreeing participants were asked one-by-one to share their opinion. This method was used to avoid any peer pressure and to save time for meaningful discussion only in case of disagreements. The second question – which aimed to have participants' reflections on the (planned) modelling approach in the case studies – was opened for discussion in case of comments. In this case, participants were simply asked to provide their comments. These reflections were only used to facilitate and adjust the model development process and did not provide direct feedback on the applicability of the recommendations, therefore these are only briefly summarized in this report. The poll results and the comments of the participants are presented anonymously in the results section of this report.

4. RESULTS

4.1 General recommendations related to the guidance

This section presents the general recommendations by experts which emerged in the discussion that followed the presentation of the guidance for the harmonisation and improvement of economic evaluations of personalised medicine.

<u>Early HTA</u>

Expert: We often build early economic models when PM is first developed. Have you thought about which of these recommendations are needed in an early economic analysis (as a minimum)?

<u>Structure of guidance</u>

Expert: In the final guidance/document, it might be helpful to consider grouping and ordering the individual recommendations in the same order/categories as the CHEERS checklist.

Model type

Expert: Some PM issues are difficult to incorporate in the often-used Markov models. Would you propose different modelling techniques, such as patient-level modelling?

Consortium member: There is no single model most appropriate for PM modelling, as the usefulness of a model type depends on the specific decision problem. We therefore don't prescribe any single modelling technique.

Relationship between surrogate and final outcome

Expert: It is potentially a waste of information if only premature survival data from the trial is used (e.g. in cancer), while data on surrogate outcomes (e.g. tumour response) is disregarded.
It might be worth adding a recommendation stating that if surrogate modelling is used, the data used to estimate the relationship between surrogate and final outcome should be made explicit. Existing surrogate and final endpoints validation frameworks may support this (e.g. German Institute of Quality and Efficiency in Health Care (IQWiG). Validity of surrogate endpoints in Oncology; 2011; Lassere et al. 2012[1])

Consortium member: Agreed, the surrogate-final outcome relationship is important to consider. Another aspect to consider in addition to this is the genotype-phenotype relationship (e.g. how the level of PD-L1 expression relates to tumour response).

Expert: Agreed that it is worth exploring how we can make better use of trial data on surrogate outcomes.

Treatment effectiveness data

Expert: Are there modelling solutions to address limited data availability (such as in the small single-arm trials for TRK-inhibitors larotrectinib and entrectinib)?

Consortium member: While modelling can be used to address some of the challenges, the main solution to insufficient/inadequate data may be increased communication between regulators and national HTA agencies about what kind of data is needed.

Expert: Part of a model's purpose is to reveal uncertainty both in inputs and assumptions that need to be made in order to produce a result. Value of information analysis is useful tool that should be considered in these cases.

Allocating testing costs to subsequent treatments

Expert: Currently we have a black-and-white situation where treatments either get allocated 100% of associated testing costs (if a test is to be newly introduced into treatment pathways) or 0% of testing costs (if the test is already part of standard care). Practice might be more nuanced, as multiple treatments may be given after a test that is to be newly introduced. The argument is being made by

[1] Lassere MN, Johnson KR, Schiff M, Rees D. Is blood pressure reduction a valid surrogate endpoint for stroke prevention? An analysis incorporating a systematic review of randomised controlled trials, a by-trial weighted errors-in-variables regression, the surrogate threshold effect (STE) and the Biomarker-Surrogacy (BioSurrogate) Evaluation Schema (BSES). BMC Med Res Methodol 2012;12:27.

industry that assigning 100% of testing costs to the first treatment to come onto the market, punishes the “first mover”. Perhaps we need a better way of assigning cost.

Consortium member: An economic evaluation should reflect the decision problem at hand for decision-makers over a sufficiently long time horizon, which often requires allocating 100% of the costs of a newly introduced test to the treatment that requires its introduction. However, the effect of alternative ratios for the cost allocation can be investigated in sensitivity analysis.

Consortium member: Innovation can be rewarded using alternative governmental instruments (e.g. through research and innovation funds) instead of via the healthcare budget. The issue of cost allocation may be temporary and may disappear once broad gene testing becomes common practice.

What to include in calculation of testing costs?

Expert: Is there a recommendation concerning the costs of the test/intervention and what should be considered/included, e.g. cost of implementation, equipment, expected number of individuals using it, etc?

Perspective

Expert: You could consider recommending taking a societal perspective in sensitivity analysis, if the base case perspective is not societal (to be in line with Neumann et al.).

Consortium member: This recommendation would potentially not be specific enough to PM. If we believe that a sensitivity analysis with a societal perspective should always be included, this PM-specific guidance is maybe not the place to discuss that.

Expert: I think it's always important to remember how the decision threshold should be adjusted if the scope of the analysis is broadened. I think Neumann et al. did not recognise this. Conducting multiple scenario analyses can result in multiple estimates without any clear decision emerging.

4.2 Reflections and recommendations on specific guidance items

This section summarises the results of the Mentimeter polling on each guidance item presented and respective comments by experts.

–

#3 Item[2]: “Ensure that the position of the modelled testing and treatment options accurately reflect clinical practice.”

Mentimeter poll result:

	MODY	NTRK	TOXNAV	**SUM**
Yes	4	5	4	**13**
No, not complete	2	2	2	**6**
No, not well-phrased	0	1	0	**1**

Comments:

- Modellers should make clear

[2] The original numbering of the guidance items were kept for the report, i.e. #1 and #2 were not discussed during the afternoon sessions.

(1) how well included tests reflect all available options and
(2) where included tests fit in the testing pathway.

- The word “modelled” should maybe be replaced, as this doesn’t capture the key issue, i.e. to what extent what is “modelled” is relevant to the decision problem. Modellers should try to justify not only what they have done but also what they haven’t done.
- Clinical practice versus clinical guidelines or “optimal practice”
    - Modelled testing and treatment options don’t necessarily need to reflect clinical practice, as current clinical practice may not be “optimal practice”. It may even be divergent from clinical guidelines.
- When modelling screening interventions, modellers should take care of all relevant strategies for all patient subgroups, as otherwise (cost-)effective strategies might not be included in the model and consequently not identified by decision-makers.
- Consider using the term “diagnostic strategy” instead of “testing”.
- Consider adding “in the relevant study population and/or subpopulation” after “reflect clinical practice”.

#4 Item: “Ensure that the data on the predictive accuracy of a test is the latest available and obtained in a population that matches the modelled population.”

Mentimeter poll result:

	MODY	NTRK	TOXNAV	**SUM**
Yes	3	3	3	**9**
No, not complete	3	2	0	**5**
No, not well-phrased	0	2	1	**1**

Comments:

- Consider changing “latest” to “best” available, as latest might not be equal to best available data (e.g. in terms of relevance to the country of interest, policy setting, quality and level of data synthesis).
- Because of sequencing of tests, you are probably not going to get populations that “match” the data strictly speaking, perhaps say “is appropriate to”.
- The uncertainty (resulting for example from variation in testing practice between labs) in the accuracy of tests should be considered.
- Consider adding examples to “predictive accuracy of a test” in brackets, like: (i.e. positive predictive value, negative predictive value, sensitivity, specificity)
- “Predictive accuracy” is affected not only by sensitivity and specificity of the test itself but also by human error and environmental factors.

#5 Item: “When different cut-off values are in use to determine test results, clearly define the cut-off value assumed in the base case. Investigate the effect of alternative cut-off values using sensitivity analysis.”

Mentimeter poll result:

	MODY
Yes	2
No, not complete	3
No, not well-phrased	1

Comments:

- According to expert suggestions the second sentence should state explicitly that we are interested in the effect of alternative cut-off values [on cost-effectiveness outcomes]

- When there is no consensus on the cut-off value, it may be appropriate to investigate the effect of alternative cut-off values in the base case analysis.

Cut-off values are also important in case of risk scores that are used to obtain a prior probability of a patient having a condition and exploring alternative cut-offs on the probability scale as decision rules for progressing to genetic testing._

#7 Item: "Confirm that the assumed cost and predictive accuracy of the test are accurate in the setting of interest and consider possible variation in cost and predictive accuracy of testing across laboratories."

Mentimeter poll result:

	TOXNAV
Yes	5
No, not complete	0
No, not well-phrased	0

No comments.

#8 Item: "If there is a risk of increased morbidity or mortality during waiting periods, incorporate relevant waiting periods in the model."

Mentimeter poll result:

	NTRK	TOXNAV	**SUM**
Yes	6	3	9
No, not complete	1	0	1
No, not well-phrased	1	2	3

Comment:

- Consider adding "a known and quantifiable risk" instead of just "risk".
- Consider adding "and any underlying disease progression"

#9 Item: "When a treatment requires the use of a test to stratify patients, include the (downstream) costs and health outcomes of testing for both individuals who test positive and individuals who test negative in the model."

Mentimeter poll result:

	MODY
Yes	4
No, not complete	2
No, not well-phrased	1

Comments:

- Although including downstream costs and health outcomes of testing seemed too obvious to be included to some experts, other experts were in favour of keeping this recommendation in the guidance.
- According to experts the recommendation should explicitly point out the true and false negative and positive cases, e.g. include into brackets.

#10 Item: "Include the costs and health outcomes of testing relatives of index patients with inheritable genetic mutations in the model."

Mentimeter poll result:

	MODY
Yes	1
No, not complete	2
No, not well-phrased	3

Comment:

- Consider adding “where possible” or “where there is adequate evidence on clinical trajectories of relatives with and in the absence of testing“ at the beginning to make it less strict

#13 Item: “When effectiveness of a treatment for a patient population with a specific gene mutation is estimated using historical data, account for the prognostic value of the gene mutation and its distribution in the historical cohort.”

Mentimeter poll result:

	NTRK
Yes	6
No, not complete	0
No, not well-phrased	1

Comment:

- The experts suggested changing “historical” to “external” as the issue of unknown prognostic value is not only relevant to historical data, this is a relevant issue when using any kind of “external data” to estimate effectiveness in the control group.
- We could say “broader group of patients than the subgroup with a specific gene mutation” instead of “historical cohort”.
- This sounds a bit unclear and maybe mixing up 1) the hazard ratio (prognostic value) to apply to the baseline risk & 2) Any evidence for a differential treatment effect in the subgroup with a gene mutation

#16 Item: “Do not include additional elements of value in the base-case analysis. When additional elements of value are included in scenario analyses, ensure possible elements of negative value have been equally considered and are included in the analysis of both intervention and comparator if relevant.”

Mentimeter poll result:

	TOXNAV
Yes	5
No, not complete	0
No, not well-phrased	0

No comments.

#17 Item: “Include parameters reflecting patient and clinician compliance in economic evaluations for decision-makers who require cost-effectiveness results under realistic circumstances.”

Mentimeter poll result:

	MODY	TOXNAV	SUM
Yes	5	5	10

No, not complete	1	0	1
No, not well-phrased	0	0	0

Comments:

- It was found that there might be uncertainty around the applicability of this recommendation for all PM cases. It is still important to include it but with possible remarks/reflection on the context of decision making
- The experts suggested to include **uptake** and compliance as in some evaluations this could make a difference. Expert elicitation may be used to obtain estimates of uptake and compliance.

#18 Item: "Confirm that the assumed compliance is accurate in the setting of interest and consider possible variation in compliance across societal groups."

Mentimeter poll result:

	MODY
Yes	5
No, not complete	1
No, not well-phrased	0

Comment:

- There could be variation of compliance between, for example some trial data and real setting, e.g. in many drug clinical trials it could be as high as 95% while in real clinical practice it could be 80% so this might also be included.

#19 Item: "When expert opinion is used to estimate quantitative model parameters, synthesise the opinions of the experts into a probability distribution to be included in sensitivity analysis."

Mentimeter poll result:

	NTRK
Yes	5
No, not complete	0
No, not well-phrased	2

Comment:

- Be careful with language around expert "opinions". The overarching term is expert judgement, with qualitative judgements called expert "opinion" and quantitative judgements called expert "values". Perhaps say something like "When experts are used to value input parameters, …"

#20 Item: "Identify uncertainties in structural assumptions and decisions and investigate their impact on cost-effectiveness results through sensitivity analysis. Parameterise structural aspects where possible."

Mentimeter poll result:

	NTRK
Yes	7
No, not complete	0
No, not well-phrased	0

No comments.

4.3 Case study specific recommendations in relations to the guidance or Reflections on the selected modelling methodologies

MODY

It was suggested to consider every diabetic patient below 35 years of age to be able to reach more GCK MODY patients. Regarding the MODY calculator cut-off point, next to using the clinically recommended one, the “cost-effective” cut-off point could as well be determined as a result of the modelling process. Testing relatives of patients highly depends on MODY type however modelling these additional patients were not recommended for the base case. It was suggested to investigate a number of scenarios on screening strategies and respective clinical pathways, to make sure that the optimal strategy is adapted to the jurisdiction of interest.

NTRK

It was mentioned that, when weighting studies that estimate diagnostic sensitivity and specificity, we should ensure to weight according to the number of samples tested positive, not the total number of samples tested. In estimating sensitivity and specificity for a test (e.g. IHC), we were encouraged to be specific about the technology used (e.g. Abcam or Cell Sign Tech). While the experts generally agreed with our proposed method for modelling the testing phase, it was mentioned that we might want to account for the fact that NTRK testing may happen a long time before TRK inhibitors are given to the patient. E.g. lung cancer patients might receive NGS early in their treatment process and test positive for NTRK but receive other, earlier-line treatments first. We were also encouraged to consider more explicitly how to account for heterogeneity in treatment effect across tumour locations.

TOXNAV

With regards to placement of the ToxNav test in the clinical pathway, the experts thought that the place of the test was reasonable and clinically justifiable. It was suggested that we could consider extrapolating the follow-up in the model to 5/10 years, or lifetime, which we have previously also considered. It was also advised to try and validate the predictive accuracy of the ToxNav test in another cohort/setting. Another suggestion was to compare the 19 variant ToxNav test to a test with fewer variants, e.g. 4, in terms of time to obtain the results from the test and cost which we have planned to do by comparing data of DPYD genetic testing from 2019 to 2020 that we obtained from local oncology directorates. Experts also advised to consider patients’ compliance as well as clinicians’, which we will incorporate in the model using hospital data.

CONTACT

Dr. Susanne Giesecke
Senior Researcher
+43 50550-4564
Susanne.Giesecke@ait.ac.at

Dr. Manuela Kienegger
Researcher
+43 50550- 4530
Manuela.Kienegger@ait.ac.at

www.hecopermed.eu