Research Article

For reprint orders, please contact: reprints@futuremedicine.com

Personalized Medicine

Lessons learned from the application of the HEcoPerMed guidance to three modeling case studies

Balázs Nagy*[,1,5], Tamás Zelei[1], Heleen Vellekoop[2], Simone Huygens[2], Matthijs Versteegh[2], Maureen Rutten-van Mölken[2,3], Rositsa Koleva-Kolarova[4], Apostolos Tsiachristas[4], Sarah Wordsworth[4] & László Szilberhorn**[,1]

[1]Syreon Research Institute, Budapest, Hungary

[2]Institute for Medical Technology Assessment, Erasmus University Rotterdam, 3000 DR, Rotterdam, The Netherlands

[3]Erasmus School of Health Policy & Management, Erasmus University Rotterdam, Rotterdam, The Netherlands

[4]Health Economics Research Centre, University of Oxford, Oxford, UK

[5]Center for Health Technology Assessment, Semmelweis University, Budapest, Hungary

*Author for correspondence: balazs.nagy@syreon.eu

**Author for correspondence: tamas.zelei@syreon.eu

Background: The HEcoPerMed consortium developed a methodological guidance for the harmonization and improvement of economic evaluations in personalized medicine. **Materials & methods:** In three therapeutic areas, health economic models were developed to scrutinize the recommendations of the guidance. **Results:** Altogether, 20 of the 23 recommendations of the guidance were addressed by the models. Seven recommendations were applied in all studies, six in two of the studies and seven in one of the studies. Recommendations with an essential role on the final conclusions of the analyses were identified in each study. **Conclusion:** The guidance was found to be best used as a tool to identify and prioritize issues, verify solutions and justify decisions during the economic analysis of personalized interventions.

First draft submitted: 23 March 2023; Accepted for publication: 24 May 2023; Published online: 11 September 2023

Keywords: cost–effectiveness • economic evaluation • genetic testing • guidance • modeling • personalized medicine • recommendations

Economic evaluations in health are used to support decisions about the efficient allocation of resources. Optimizing pathways of care, treatment selection and adjustment for stratified groups or individuals with the means of personalization (e.g., genetic testing, omics profiling) has great potential to improve patient outcomes and ensure efficient spending of scarce health care resources. However, features of personalized care have cast doubt on the appropriateness and applicability of current health economic modeling approaches that have been primarily used to analyse non-personalized interventions [1]. Conducting economic evaluations in the field of personalized medicine (PM) often creates challenges (e.g., availability of costs and health outcome data, accurate evaluation of test–treatment combinations, complexity of modeling potential real-life patterns) that are less frequently seen in traditional economic analyses [1].

The European Commission-funded consortium 'Health Economic Models for Personalized Medicine' (HEcoPerMed; https://hecopermed.eu/) developed the 'Guidance for the Harmonisation and Improvement of Economic Evaluations of Personalised Medicine' [2]. The guidance identifies topics that are deserving additional attention by health economic modelers and experts who develop or evaluate economic models to improve the consistency and quality across different models for personalized care. Some of the recommendations remind modelers and evaluators of good practices that are often neglected; others steer modelers and evaluators who are uncertain about how to proceed in the face of ongoing methodological and policy discussions [2].

To scrutinize applicability, verify usability and formulate considerations about the practical usefulness of the HEcoPerMed guidance, the consortium developed health economic models for three real-world case studies [3–5].

Future Medicine

This article discusses our experience regarding the feasibility of the guidance recommendations to the case studies and formulate practical recommendations on the real-world use of the guidance based on this experience.

Materials & methods

Case studies

The three case studies were conducted by the Erasmus University of Rotterdam, the University of Oxford and Syreon Research Institute. They were concerned with genetic profiling: two in the field of advanced cancer and one in diabetes. Each of the three research groups developed one core country-specific cost–effectiveness model, which was then adapted to the other two countries and jurisdictions [6–8]. Due to the different disease areas and different healthcare settings, these three economic models allowed for variation in the application of the guidance and highlight specific challenges in the space of health economic modeling of PM.

The first study examined the *NTRK* gene fusion testing among patients with locally advanced or metastatic solid tumors who have already received one or more lines of treatment. Patients who tested positive with different test combinations were treated with the *NTRK* inhibitor entrectinib. The comparator in this study was no genetic testing followed by standard of care for all patients. The time horizon of the model was 20 years, the cycle length was 1 month and the base case analysis was performed from the societal perspective of The Netherlands. An estimate of the prognostic value for survival and time to treatment discontinuation of the *NTRK* gene fusion was used to adjust analyses. The base case results showed that the incremental cost–effectiveness ratio (ICER) was well above the maximal acceptable cost–effectiveness. The scenario analysis showed that when the costs of testing for *NTRK* gene fusions were not considered the cost–effectiveness ratio was below the maximal acceptable ICER. More details are provided by Huygens *et al.* [3].

The second study examined routine upfront *DPYD* gene testing with a new multivariant genetic test called ToxNav® to personalize fluoropyrimidine-based chemotherapy (capecitabine/5-fluorouracil [5FU]) for metastatic breast cancer patients. The model simulated a cohort of women with metastatic breast cancer, starting with ToxNav testing and consecutive dose adjustment of chemotherapy regimens and another cohort receiving full dose of capecitabine/5FU treatment without prior ToxNav testing. The base case analysis was performed from the perspective of the NHS in England on a lifetime horizon. The study demonstrated that upfront *DPYD* testing with a multigene panel results in net cost savings and improves quality of life of metastatic breast cancer patients. Results were strongly influenced by the rate of adverse events associated with chemotherapy. Univariate and probabilistic sensitivity analyses demonstrated that the ToxNav would retain its dominant status over no genetic testing strategy. More details are provided by Koleva-Kolarova *et al.* [5].

The third case study examined screening for Maturity Onset Diabetes of the Young (MODY) in the diabetes population. The health economic model simulated the diagnostic and therapeutical pathways of care of diabetes and estimated long-term effects in diagnosed and undiagnosed MODY patients. Diagnosis of MODY was done on the basis of a risk stratification questionnaire, laboratory test and next-generation sequencing. The analysis was performed from a healthcare payer perspective considering direct medical costs and outcomes over 20 years in Hungary. The results demonstrated that in young (<35 years) diabetes patients, detection of MODY and switch to appropriate treatment reduces therapeutical costs and improves patients' quality of life by avoiding inconvenient and expensive insulin treatment and its complications. Moreover, by carefully selecting the target population for genetic testing, screening MODY turns into a cost-saving alternative compared with no screening. More details are provided by Kovacs *et al.* [4].

All three core models were adapted to transfer them to the other two countries. Details for all country adaptations of the three case studies have been published [6–8]. Although the adapted models were not directly checked against the recommendations, they helped draw general conclusions for the analysis.

Guidance recommendations

The 'Guidance for the Harmonisation and Improvement of Economic Evaluations of Personalised Medicine' was developed in 2019–2021 by the members of the HEcoPerMed consortium. In this process, a targeted literature review of methodological papers was performed to overview modeling challenges in PM. Expert interviews were then held to discuss best modeling practices in PM. A systematic literature review of economic evaluations of PM was also conducted to gain insight into current modeling practice. The findings were synthesized and used to develop a set of draft recommendations. The draft recommendations were discussed at a stakeholder workshop and

subsequently finalized. The original paper by Vellekoop [9] discusses the methods and the 23 recommendations of the guidance in detail; a summary of these is provided in Table 1.

During the development of the economic models for the three case studies the modeler groups recorded the challenges they faced and their relation with the recommendations of the guidance. In this process, a 1 day virtual workshop with 27 stakeholders (from different jurisdictions) was held to receive input on the guidance's practical applicability on the case studies with regards to content and clarity and also the face validity of the constructed models. The workshop entailed the presentation of the three health economic models and the structured discussion of the modeling challenges in relation to specific recommendations of the guidance. The participants had opportunity to comment both on the models and on the respective recommendations.

Once the models were finalized the three modeler groups presented if and how they managed to address the recommendations. Finally, the guidance items for each case study were labelled as follows:

- Applied: if the study managed to appropriately consider the recommendation.
- Not included: if the study did not manage to appropriately consider the recommendation.
- Not applicable: if a particular recommendation was not relevant to the case study.

The purpose of the exercise was not only to see if the recommendations are practically applicable but also to understand if the guidance sufficiently covers PM-specific challenges coming across in the three case studies. The modeler groups therefore checked and discussed whether they faced challenges beyond the coverage of the guidance and whether the guidance needed further amendment or adjustment (see Table 1).

Results

The three modeling case studies addressed 19 of the 23 recommendations. Seven recommendations (Nos. 1–5, 15 and 18) were applied in all case studies. Six recommendations (Nos. 8, 9, 11, 17, 21 and 22) were applied in two of the three case studies. Seven recommendations (Nos. 6, 12, 14, 16, 19, 20 and 23) were applied only in one of the three case studies. Three recommendations of the guidance (Nos. 7, 10 and 13) were not applicable in any of the case studies. The summary of case study evaluation results is presented in Table 1.

Addressing the recommendations

The study on testing for *NTRK* gene fusions managed to apply 12 of 16 applicable recommendations. The study on *DPYD* testing addressed 16 of 17 applicable recommendations. The MODY screening case study addressed 12 of 18 applicable recommendations. The detailed application of the guidance to each case study is presented in Appendix Tables 1–3. The summary on how the three case studies addressed the recommendations is as follows.

Recommendation 1: 'For economic evaluations of PM, use the standard perspective as recommended by national HTA [Health Technology Assessment] guidelines in the base case'.

All three analyses followed the recommendations of the national guidelines and used the perspective of the healthcare payer. The analysis for *NTRK*, in alignment with the national guideline, applied both the perspective of the society in the base case and the healthcare payer and of the society as a scenario analysis. Altogether, the modeler groups found appropriate, applying the perspective recommended by the national guidelines, and adding alternative perspectives to the analysis was not deemed necessary.

Recommendation 2: 'For economic evaluations of PM, use the standard discount rates as recommended by national HTA guidelines in the base case'.

In all three analyses, standard discount rates of costs and benefits were used according to the national guidelines: 4.0, 3.5 and 3.7% for costs and 1.5, 3.5 and 3.7% for benefits for the *NTRK* (The Netherlands), *DPYD* (UK) and MODY (Hungary) studies, respectively. The modeler groups deemed to discount costs and benefits similarly to the recommendations of the respective national guidelines; lower (than recommended by the guidelines) or hyperbolic discount rates (see Vellekoop *et al.*) were not regarded more feasible than the rates recommended in national guidelines.

Recommendation 3: 'Identify all relevant test–treatment pathways and justify why the pathways included in the model were selected'.

In the *NTRK* model all relevant test–treatment pathways, including four testing strategies: RNA-NGS test for all tumor types, immunohistochemistry (IHC) test for all tumor types, IHC test followed by RNA-NGS in patients with a positive IHC test result for all tumor types and stratified test strategies depending on the

Table 1. Summary table of HEcoPerMed recommendations and their application to the three case studies.

Guidance items	Recommendations	Application to the case studies		
		NTRK	ToxNav	MODY
Perspective and Discounting	1. For economic evaluations of PM, use the standard perspective as recommended by national HTA guidelines in the base case	Applied	Applied	Applied
	2. For economic evaluations of PM, use the standard discount rates as recommended by national HTA guidelines in the base case	Applied	Applied	Applied
Test–Treatment Pathways	3. Identify all relevant test–treatment pathways and justify why the pathways included in the model were selected	Applied	Applied	Applied
	4. When treatment requires the use of a test to stratify patients, include in the model the (downstream) costs and health outcomes of testing for both individuals who test (false) positive and individuals who test (false) negative	Applied	Applied	Applied
	5. Ensure that the data used to estimate the diagnostic accuracy of a testing technology are appropriate to the patient population in the model	Applied	Applied	Applied
	6. 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 on cost–effectiveness results using a sensitivity analysis	Not applicable	Not applicable	Applied
	7. When multiple tests are modeled in sequence, consider the interdependence between test results	Not applicable	Not applicable	Not included
	8. If there is a notable risk of increased morbidity or mortality as a result of waiting periods, incorporate in the model the costs and health outcomes due to the waiting periods	Applied	Applied	Not applicable
	9. Confirm that the assumed testing costs are accurate in the setting of interest and consider possible variations in costs across laboratories	Applied	Applied	Not included
	10. If relatives of index patients become eligible for genetic testing when the index patients test positive for a specific genetic marker, include the costs and health outcomes of testing relatives in the economic evaluation of the index patients	Not applicable	Not applicable	Not included
Effectiveness Data	11. Where possible, use effectiveness data from trials with two (or more) alternative treatment strategies	Not applicable	Applied	Applied
	12. When surrogate outcomes are used to estimate final outcomes, specify which data sources were used to estimate the relationship between surrogate and final outcomes and justify any assumptions made about the relationship	Not applicable	Not applicable	Applied
	13. When the effectiveness of the comparator is estimated using external data, account for a possible time trend in the effectiveness	Not included	Not applicable	Not applicable
	14. When the effectiveness of the comparator for patients with a specific genetic marker is estimated using external data, account for the prognostic value of the genetic marker and differences in its prevalence across the different data sources	Applied	Not applicable	Not applicable
	15. Specify which data sources were used to estimate the association between the genetic marker(s) of interest and clinical outcomes and justify any assumptions made about the association	Applied	Applied	Applied
Extrapolating Survival	16. When extrapolating survival data beyond the study period, use expert opinion alongside statistical fit to choose the survival model	Not included	Applied	Not applicable
	17. When extrapolating survival data beyond the study period, account for any excess mortality and morbidity among long-term survivors	Applied	Applied	Not applicable
Additional Elements of Value	18. Only include elements of value recommended by national HTA guidelines in the base case. If additional elements of value are included in a sensitivity analysis, ensure possible elements of negative value are equally considered and included for both the intervention and the comparator	Applied	Applied	Applied
Incorporating Compliance	19. Include parameters reflecting patient and clinician compliance in economic evaluations for decision-makers who require cost–effectiveness results under realistic circumstances	Not included	Applied	Not included
	20. When including patient and clinician compliance in economic evaluations, confirm that the assumed compliance is accurate in the setting of interest and consider possible variation in compliance across societal groups	Not included	Applied	Not included
Uncertainty Analysis	21. When expert judgement is used to estimate values for the input parameters in the model, synthesize the elicited values into a probability distribution to be included in a sensitivity analysis	Not included	Applied	Applied
	22. Identify uncertainties in structural assumptions and decisions and investigate their impact on cost–effectiveness results through a sensitivity analysis. Parameterize structural aspects where possible	Applied	Not included	Applied
Managed Entry Agreements	23. If a managed entry agreement is being considered for an intervention, include its conditions in the model evaluating the intervention	Not included	Applied	Not included

HTA: Health Technology Assessment; MODY: Maturity Onset Diabetes of the Young; PM: Personalized medicine.

NTRK fusion prevalence and TRK wild-type protein expression of the tumor types for different patient categories were incorporated. In the *DPYD* model the place of the genetic testing was determined based on literature and consultations with clinical experts. In the MODY model, two patient stratification scenarios were distinguished after looking at current clinical practice.

Recommendation 4: 'When treatment requires the use of a test to stratify patients, include in the model the (downstream) costs and health outcomes of testing for both individuals who test (false) positive and individuals who test (false) negative'.

In all three case studies, the entire spectrum of downstream costs and outcomes related to testing and treatment for both patients who tested positive and negative, including both false-positive and false-negative subjects, were considered.

Recommendation 5: 'Ensure that the data used to estimate the diagnostic accuracy of a testing technology are appropriate to the patient population in the model'.

In the *NTRK* model, the sensitivity and specificity of the IHC test was based on tumor-specific estimates from the literature while genetic testing was assumed to be 100% accurate. In the *DPYD* model, the diagnostic accuracy of testing was taken from Lee *et al.* [10], which applies for the case of predominantly Caucasian population. In the MODY model, genetic testing was assumed to be 100% accurate, whereas the MODY risk calculator's accuracy was based on the best matching scientific evidence [11]. The sensitivity and specificity of the autoantibody testing was based on scientific evidence [12].

Recommendation 6: '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 on cost–effectiveness results using a sensitivity analysis'.

Cut-off values are by nature not associated with genetic testing: consequently cut-off values for Next Generation Sequencing-RNA panel test of *NTRK* fusions and for DNA panel test to detect *DPYD* mutations were not applied. Nevertheless, for the case of identifying potential MODY patients with a risk assessment questionnaire, a 40% cut-off value was used (i.e., with 40% probability of having MODY the sensitivity [87%] and specificity [88%] of the questionnaire were maximized), whereas other cut-off values were also tested in scenario analyses.

Recommendation 7: 'When multiple tests are modelled in sequence, consider the interdependence between test results'.

In the *NTRK* model, the sensitivity and specificity of the IHC test did not influence the test characteristics of the subsequent genetic test because 100% sensitivity and specificity was assumed. In the *DPYD* study, modeling test sequence was not applicable. In the case of MODY, it was not possible to allow for the interdependence between the test results because of the lack of sufficient data.

Recommendation 8: 'If there is a notable risk of increased morbidity or mortality as a result of waiting periods, incorporate in the model the costs and health outcomes due to the waiting periods'.

The risk of increased mortality of patients when waiting for test results was considered in the *NTRK* analysis. Waiting to undergo the *DPYD* test and for its results (and the associated delay in receiving chemotherapy) was built into the *DPYD* economic model because it had a potential to impact on survival and quality of life in cancer treatment. Risk of increased mortality during testing was not relevant for the MODY analysis.

Recommendation 9: 'Confirm that the assumed testing costs are accurate in the setting of interest and consider possible variations in costs across laboratories'.

In the *NTRK* analysis difference in costs reported by different hospitals (specialized in treating different tumors and having their own price setting policies) were considered. In the *DPYD* model, the cost of genetic testing was determined for the UK setting and was in particular fit to the local settings of the Oxford University Hospitals NHS Trust. Due to lack of local data, cost of genetic testing did not reflect local prices in the MODY model, and the cost of genetic testing was assumed to be similar to the cost applied in the UK (€716).

Recommendation 10: 'If relatives of index patients become eligible for genetic testing when the index patients test positive for a specific genetic marker, include the costs and health outcomes of testing relatives in the economic evaluation of the index patients'.

Genetic testing in relatives of index patients were not deemed relevant for the *NTRK* and the *DPYD* analyses. Relatives of index patients were considered but not modeled in the MODY analysis because screening children was deemed not feasible during the time horizon of the analysis, siblings (if treated with insulin and younger than age 35) were assumed to be in the target group anyway and parents of screened subjects were older than the target group (see more about this challenge in the Discussion).

Recommendation 11: 'Where possible, use effectiveness data from trials with two (or more) alternative treatment strategies'.

The design of the clinical data supporting the analysis was an important challenge for both the Dutch and the UK models. In the case of *NTRK*, data on entrectinib were only available from the single-arm trial. Hence, a synthetic control group had to be constructed. In the *DPYD* model, for reduced and standard dose of capecitabine/5FU in (non-)mutation carriers modelers used estimates from a published cost–effectiveness analysis that fitted a Weibull distribution to progression-free and overall survival data from a randomized controlled trial. In the MODY analysis, to model patient pathways, several sources of randomized clinical trial data were applied (see more in references [3–5]).

Recommendation 12: When surrogate outcomes are used to estimate final outcomes, specify which data sources were used to estimate the relationship between surrogate and final outcomes and justify any assumptions made about the relationship'.

For the *NTRK* and *DPYD* analyses, surrogate outcomes were not necessary to model patient pathways. In the MODY analysis, data sources on surrogate outcomes were specified in several items of the diabetes patient pathway, based on scientific literature; there was special attention on HbA1c differences and complication specific outcomes regarding HbA1c (see more in reference [4]).

Recommendation 13: 'When the effectiveness of the comparator is estimated using external data, account for a possible time trend in the effectiveness'.

In the *NTRK* analysis, external data for the comparator was relatively recent (2012–2020), and thus a time trend in effectiveness was not accounted for, although it could have been investigated. In the case of *DPYD* testing, external data to build the comparator arm was not necessary. Issues around the proper use of effectiveness data for the comparator were not relevant in the MODY analysis.

Recommendation 14: 'When the effectiveness of the comparator for patients with a specific genetic marker is estimated using external data, account for the prognostic value of the genetic marker and differences in its prevalence across the different data sources'.

To adjust for the prognostic value of the *NTRK* gene fusion, the hazard ratios of *NTRK*+ patients for mortality and time to treatment discontinuation were estimated. Prognostic value of genetic markers for the comparator was not applicable for the case of *DPYD* and MODY testing.

Recommendation 15: 'Specify which data sources were used to estimate the association between the genetic marker(s) of interest and clinical outcomes and justify any assumptions made about the association'.

In the *NTRK* model, the prognostic value of *NTRK* gene fusions was based on the Hartwig Medical Foundation (HMF) database including metastatic cancer patients from 44 hospitals that had genomic profiling between 2012 and 2020. The hazard ratio of *NTRK*+ patients for mortality and time to treatment discontinuation was assumed constant across tumor types. In the *DPYD* model, data sources for association between genetic marker and clinical outcomes in the intervention arm was obtained from local data provided by the Oxford University Hospitals [13]. In the case of MODY, data sources about the association between genetic markers and clinical outcomes were specified for types of genetic mutations and therapies (and their effectiveness) based on the literature.

Recommendation 16: 'When extrapolating survival data beyond the study period, use expert opinion alongside statistical fit to choose the survival model'.

In the *NTRK* analysis, expert opinion alongside statistical fit to choose between the survival models was not applied. Until recently, the patients' *NTRK* status would have been unknown because no *NTRK* testing would have been done. It was therefore expected that experts (e.g., clinicians) would not be able to provide additional information on the most accurate survival models for *NTRK*+ and *NTRK*- patients separately. In the case of *DPYD* testing, the appropriate statistical model fit for the analysis of transition probabilities was chosen based on expert opinion and goodness of fit analysis. Because the therapy switch had a negligible effect on patient survival, changes in extrapolating survival related items were not applicable for the case of MODY.

Recommendation 17: 'When extrapolating survival data beyond the study period, account for any excess mortality and morbidity among long-term survivors'.

Excess mortality and morbidity were applied in the *NTRK* case because the data used for the analysis reflected metastatic patients' survival. Similarly, metastatic breast cancer is usually considered a terminal disease that is amenable to palliative but not curative intent, and thus excess mortality and morbidity among long-term survivors was covered by the data. Therapy switch has negligible effect on patient survival, so changes in extrapolating survival related items were not relevant in the MODY analysis.

Recommendation 18: 'Only include elements of value recommended by national HTA guidelines in the base case. If additional elements of value are included in a sensitivity analysis, ensure possible elements of negative value are equally considered and included for both the intervention and the comparator'.

No value elements beyond those recommended in the Dutch HTA guideline were included in the *NTRK* analysis. However, this guideline recommends the adoption of a societal perspective, as a result of which the costs of informal care and all healthcare costs during life years gained (related and unrelated to cancer) were incorporated. Productivity costs were not taken into account because the patients are in an advanced stage of cancer and have probably left the workforce. Similarly, no value elements beyond those recommended in the national HTA guideline were included in the *DPYD* and MODY analyses. However, in the *DPYD* analysis, the psychological benefits from knowing the results of the ToxNav test were considered, but because no differences were found between the tested and not tested cohorts, this element was not incorporated.

Recommendation 19: 'Include parameters reflecting patient and clinician compliance in economic evaluations for decision-makers who require cost–effectiveness results under realistic circumstances'.

For the case of *NTRK*, parameters specifically reflecting patient and clinician compliance were not included because such compliance was expected to be outstanding due to the severity of the disease and to the limited number of treatment alternatives. For the *DPYD* analysis, both clinicians' (96%) and patients' (100%) compliance to testing were estimated on the basis of the *DPYD* testing adherence reported by Tsiachristas *et al.* [13]. Full compliance to screening was considered in the MODY model because patients were expected to have strong incentive to improve their quality of life by switching from insulin therapy.

Recommendation 20: 'When including patient and clinician compliance in economic evaluations, confirm that the assumed compliance is accurate in the setting of interest and consider possible variation in compliance across societal groups'.

As mentioned at Recommendation 19, parameters reflecting patient and clinician compliance were not included in the case of *NTRK* testing. For the case of *DPYD* testing, compliance data were relevant to the local setting, and the applied values were within the ranges found in the literature. Due to the lack of evidence, possible variation in compliance across societal groups was not considered in the MODY analysis.

Recommendation 21: When expert judgement is used to estimate values for the input parameters in the model, synthesise the elicited values into a probability distribution to be included in a sensitivity analysis'.

Expert judgments were not synthesized into a probability distribution for the sensitivity analysis in the case of *NTRK*. Sensitivity analysis on key model parameters was taking into account the variability in literature findings and local data inputs in the *DPYD* analysis. Expert judgment was used to inform the cost of the ToxNav test, and the baseline value was varied in univariate and probabilistic sensitivity analyses to account for the uncertainty by using 20% range and a gamma distribution, respectively. Inputs in the probabilistic sensitivity analysis were crosschecked with experts in the MODY analysis.

Recommendation 22: 'Identify uncertainties in structural assumptions and decisions and investigate their impact on cost–effectiveness results through a sensitivity analysis. Parameterise structural aspects where possible'.

Structural uncertainties of the *NTRK* model were incorporated into the sensitivity analysis with three scenarios using different assumptions about the inclusion and exclusion of testing costs and consequences. Deterministic sensitivity analysis and probabilistic sensitivity analysis were performed but no structural sensitivity analysis was conducted in the *DPYD* model. Structural sensitivity analysis was conducted by adapting two screening pathways in the MODY case.

Recommendation 23: 'If a managed entry agreement is being considered for an intervention, include its conditions in the model evaluating the intervention'.

Managed entry agreements or other forms of risk sharing were not considered for the economic analysis of *NTRK* testing, but the headroom price of the RNA-NGS test to detect an *NTRK*-fusion (i.e., the price at which the cost–effectiveness ratio remains below the maximal acceptable threshold) was calculated. The application of a managed entry agreement to the cost of ToxNav was considered for the budget impact analysis of the DPYD testing. Managed entry agreements were not considered for the economic analysis of MODY testing.

Discussion

The 'Guidance for the Harmonisation and Improvement of Economic Evaluations of Personalised Medicine' developed by the HEcoPerMed consortium was applied on three real-world case studies. The 23 recommendations provided comprehensive coverage of the challenges of the economic analysis of personalized interventions. The

modeler groups did not identify new challenges that were not in the guidance and were unique to the field of PM. It also was not felt there were inadequate or inoperable items in the guidance.

The majority (63–83%) of the recommendations were relevant for each case study, and the majority (63–94%) of the applicable recommendations were successfully addressed by the modeler groups. Crucial ones in each study with essential impact on the final conclusions of the analysis were also identified. These items were as follows.

In the case of testing for *NTRK* gene fusions, the inclusion of all downstream costs and health outcomes (No. 4) had a vital impact on the conclusions. When only the costs and outcomes of those identified by the genetic test were considered in the model, the ICER remained below the willingness-to-pay threshold (threshold of €80,000/quality-adjusted life year [QALY] > €38,658/QALY). On the contrary, when accounting for the costs of testing and outcomes of all tested patients, the ICER increased to €169,957/QALY, which changed the conclusion on entrectinib from cost-effective to not cost-effective. Another crucial point of the *NTRK* analysis was that only single-arm trial data were available to model the effect of entrectinib (No. 11). The Dutch HMF database was used to estimate a synthetic control arm that simulates the prognostic value of *NTRK* fusions. Statistical methods to match the study populations from the HMF database and the entrectinib trials (e.g., propensity score matching) could not be applied, which was deemed as inescapable limitation of the model.

In the case of testing for *DPYD* mutations the appropriate combination of available clinical (No. 11) and real-world (No. 15) data from the site of the research made it possible to generate high quality evidence and justify cost savings due to genetic testing. Clinical effectiveness of reduced and standard dose of capecitabine/5FU was supported by published literature. Data on the association between the genetic markers and clinical outcomes were obtained from local real-world data sources provided by the Oxford University Hospital [13].

In the case of genetic testing for MODY mutations appropriate conclusions were strongly driven by identifying all relevant test-treatment pathways (No. 3) and by depicting structural uncertainties in the modeled patient pathways (No. 22). Placing an autoantibody laboratory test in the diagnostic pathway between the MODY risk calculator and the genetic testing dramatically changed the cost–effectiveness results. The use of the laboratory test prevented the excess use of genetic testing and sufficiently reduced the average cost per patient, which turned screening from cost-effective (i.e., QALY gains for socially acceptable incremental costs) to a cost saving alternative.

There were numerous recommendations in each study which were not applicable or not included in the analysis. Three items (Nos. 7, 10, 13) were neither included nor applicable in any of the three case studies. Of the three items two referred to data challenges.

Recommendation 13 asks the analyst to account for a possible time trend when using external (e.g., historical) data for the comparator arm of the analysis. For the case of *NTRK* testing, the external data used in the comparator arm were very recent (2012–2020), hence time trends were not likely to cause any bias in the analysis. For the cases of DPYP and MODY testing, comparator data issues were not relevant.

Recommendation 7 asks to consider the interdependence between test results, if necessary. In the *NTRK* model the sensitivity and specificity of the first (IHC) test did not influence the results of the subsequent genetic test. In the *DPYD* study, modeling test sequence was not applicable. In the case of MODY, no sufficient data were available to consider interdependency between the different test results.

The third recommendation that was not used in any of the analyses (No. 10) discusses accounting for relatives of index patients who are potentially eligible for genetic testing. This challenge was not relevant for the genetic testing of *NTRK* fusion and *DPYD* mutations. For the case of testing MODY mutations, this challenge was considered but not included in the analysis for the following reasons. Parents of the index patients, due to being on insulin treatment for longer period, were not eligible to switch therapy (according to clinical experts). Siblings were good candidates for MODY testing; however, most of them fall into the current screening strata – being under 35 and on insulin. Children of the index patients were in scope, too; however, the implications (gaining benefits and saving costs) of their diagnosis would largely fall out of the 20-year time horizon of the analysis. The guidance says that in such cases, other models that do have the potential to include the impact of genetic testing on future generations are necessary. On the basis of these considerations, Recommendation 10 was not applied to the case of MODY.

Altogether, the three outstanding items of the guidance (Nos. 7, 10 and 13) remained valid and eligible to be applied for other health economic analyses. In the case of Recommendation 7, Garrison *et al.* provide an example about 'expanded reflex' testing versus standard testing for *HER2* mutations for breast cancer patients [14]. In case of Recommendation 10, there are good examples on familial hypercholesterolemia and *BRCA*-positive breast cancer patients whose identified inheritable pathogenic mutation can initiate testing of relatives who are at potential risk of the same mutation. In the case of Recommendation 13, there are a number of examples in the area of rare diseases

where, due to ethical and epidemiological (low prevalence/incidence) reasons, only data from single-arm studies are available.

It is important to acknowledge that this qualitative review of the guidance, based on three case studies, is neither exclusive nor conclusive. However, it provides good examples on applying the guidance recommendations in practice. Our positive hands-on experience on the applicability of the items confirmed that the conceptual, systematic and multilayer approach we had been carrying out during the guidance development was sufficient. Despite our studies obviously differing in the medical/healthcare settings, data and methods, they all confirmed the appropriateness and practical usefulness of the guidance.

Through our earlier review [2] we had identified 22 existing methodological papers that question, among other things, the methodological challenges specific to PM. We subsequently used this information as basis for interviews with experts. As a result, the guidance provides in-depth discussion and takes a stance in those discussions on important model assumptions. Hence, the benefit of the guidance is that it is more than a checklist, it is a critical appraisal of data use and assumptions when developing cost–effectiveness models for PM.

A noteworthy limitation is that we primarily relied on the experience of three independent modeling teams and there were overlaps across the individuals working on the original guidance and on the respective case studies. However, this circumstance was rather helpful than aggravating: guidance developers were able to directly test the translation of theory (guidance) into practice (case studies). On the other hand, we acknowledge it might have been challenging for those taking part in both the guidance development and the modeling to remain impartial and provide fair critique on the guidance. This potential weakness was intended to be controlled with the involvement of colleagues taking part in either the modeling or the guidance development exercise, but not in both; also external stakeholders of the workshop were acting as peer reviewers. Through the three case studies, we could not cover the entire spectrum of possible challenges in the field of PM, and (similar to the scope of the guidance) narrowed the focus of the modeling exercise to studies on genetic testing. Also other medical fields could be subject to future PM-related modeling studies on the guidance.

Conclusion

This article demonstrates that the guidance can be applied as a tool for analysts to identify and prioritize issues, verify solutions, formulate arguments and carefully justify decisions throughout the economic analyses of personalized interventions. Our case studies aim to serve as examples of such practices. Instead of taking the guidance as a fully comprehensive and compulsory checklist, it is advised to be looked at as a supporting instrument steering the economic evaluation and decision-making process to assess appropriately the economic aspects of personalized interventions in healthcare.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: www.futuremedicine.com/doi/suppl/10.2217/pme-2023-0040

Author contributions

The Dutch model was evaluated by M Rutten-van Mölken, H Vellekoop, M Versteegh and S Huygens. The UK model was evaluated by R Koleva-Kolarova, A Tsiachristas and S Wordsworth. The Hungarian model was evaluated by L Szilberhorn, B Nagy and T Zelei. The draft manuscript was prepared by B Nagy, L Szilberhorn and T Zelei. L Szilberhorn, T Zelei and B Nagy interpreted the three models' results and developed the draft manuscript. S Huygens, M Vellekoop, M Rutten-van Mölken, R Koleva-Kolarova, S Wordsworth and A Tsiachristas reviewed and commented on the draft manuscript. Based on the comments B Nagy, T Zelei and L Szilberhorn finalized the manuscript. All authors read and approved the current version of the manuscript.

Financial & competing interests disclosure

The HEcoPerMed project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 824997. The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes

employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

Data sharing statement

The datasets supporting the conclusions of this article are included in the main text and in the supplementary materials. All other datasets used and analyzed in the current study are available from the corresponding author upon request.

Summary points

- The European Commission-funded consortium Health Economic Models for Personalised Medicine (HEcoPerMed) developed the 'Guidance for the Harmonisation and Improvement of Economic Evaluations of Personalised Medicine'.
- To scrutinize applicability, verify usability and formulate considerations about the practical usefulness of the HEcoPerMed guidance, the consortium developed health economic models for three real-world case studies.
- In this article, we discuss our experience regarding the feasibility of the guidance recommendations to the case studies and formulate practical recommendations on the real-world use of the guidance based on this experience.
- The three case studies were conducted by the Erasmus University of Rotterdam, the University of Oxford and Syreon Research Institute. The studies were concerned with genetic profiling: two in the field of advanced cancer and one in diabetes.
- Each of the research groups developed one core country-specific cost–effectiveness model that was then adapted to the other two countries and jurisdictions. Due to the different disease areas and different healthcare settings, the three economic models allowed for variation in the application of the guidance and highlight specific challenges in the space of health economic modeling of personalized medicine (PM).
- During the development of the economic models, the modeler groups recorded the challenges they faced and discussed their relation with the recommendations of the guidance. The purpose of the exercise was to see if the recommendations are practically applicable and also to understand if the guidance sufficiently covers PM-specific challenges coming across in the three case studies.
- The modeler groups checked and discussed whether they faced challenges beyond the coverage of the guidance and whether the guidance needed further amendment or adjustment.
- The three modeling case studies addressed altogether 19 of the 23 recommendations. Seven recommendations (Nos. 1, 2, 3, 4, 5, 15 and 18) were applied in all case studies. Six recommendations (Nos. 8, 9, 11, 17, 21 and 22) were applied in two of the three case studies. Seven recommendations (Nos. 6, 12, 14, 16, 19, 20 and 23) were applied only in one of the three case studies. Three recommendations (Nos. 7, 10 and 13) were not applicable in any of the case studies.
- The majority (63–83%) of the recommendations were relevant for each case study, and the majority (63–94%) of the applicable recommendations were successfully addressed by the modeler groups. Crucial ones in each study with essential impact on the final conclusions of the analysis were also identified.
- The recommendations provided comprehensive coverage of the challenges of economic analysis of personalized interventions.
- New challenges that were not in the guidance and were unique to the field of PM were not identified. It also was not felt there were inadequate or inoperable items in the guidance.
- This work demonstrates that the guidance can be applied as a tool for analysts to identify and prioritize issues, verify solutions, formulate arguments and carefully justify decisions throughout the economic analyses of personalized interventions.
- Our case studies aim to serve as examples of such practices. Instead of taking the guidance as a fully comprehensive and compulsory checklist, it is advised to be looked at as a supporting instrument steering the economic evaluation and decision-making process to appropriately assess the economic aspects of personalized interventions in health care.

References

1. Terkola R, Antoñanzas F, Postma M. Economic evaluation of personalized medicine: a call for real-world data. *Eur. J. Health Econ.* 18(9), 1065–1067 (2017).
2. Vellekoop H, Huygens S, Versteegh M *et al.* Guidance for the Harmonisation and Improvement of Economic Evaluations of Personalised Medicine. *Pharmacoeconomics* 39(7), 771–788 (2021).
3. Huygens S, Vellekoop H, Versteegh M *et al.* Cost-effectiveness analysis of treating patients with NTRK-positive cancer with the histology-independent therapy entrectinib. *Value Health* 26(2), 193–203 (2022).
4. Kovács G, Nagy D, Szilberhorn L *et al.* Cost-effectiveness model of genetic testing strategies for Maturity Onset Diabetes of the Young (MODY). *Personalized Med.* (2023).

5. Koleva-Kolarova R, Vellekoop H, Huygens S *et al.* Cost-effectiveness of extended DPYD testing before fluoropyrimidine chemotherapy in metastatic breast cancer in the United Kingdom. *Personalized Med.* (2023).
6. Vellekoop H, Huygens S, Versteegh M *et al.* Cost-effectiveness analysis of different NTRK-testing strategies followed by the histology-independent therapy entrectinib for cancer patients in three European countries. *Personalized Med.* (2023).
7. Szilberhorn L, Zelei T, Vellekoop H *et al.* Cost-effectiveness and budget impact analysis of screening and testing strategies for Maturity Onset Diabetes of the Young (MODY) in three European countries. *Personalized Med.* (2023).
8. Koleva-Kolarova R, Vellekoop H, Huygens S *et al.* Budget impact and transferability of cost-effectiveness results of extended DPYD testing prior to fluoropyrimidine chemotherapy in metastatic breast cancer in different European healthcare systems.
9. Vellekoop H, Huygens S, Versteegh M *et al.* Guidance for the Harmonisation and Improvement of Economic Evaluations of Personalised Medicine. *Pharmacoeconomics* 39(7), 771–788 (2021).
10. Lee LYW, Starkey T, Sivakumar S *et al.* ToxNav germline genetic testing and PROMinet digital mobile application toxicity monitoring: results of a prospective single-center clinical utility study-PRECISE study. *Cancer Med.* 8(14), 6305–6314 (2019).
11. Shields B, Mcdonald TJ, Ellard S *et al.* The development and validation of a clinical prediction model to determine the probability of MODY in patients with young-onset diabetes. *Diabetologia* 55(5), 1265–1272 (2012).
12. Mcdonald TJ, Colclough K, Brown R *et al.* Islet autoantibodies can discriminate maturity-onset diabetes of the young (MODY) from type 1 diabetes. *Diabetic Med.* 28(9), 1028–1033 (2011).
13. Tsiachristas A, Vallance G, Koleva-Kolarova R *et al.* Can upfront DPYD extended variant testing reduce toxicity and associated hospital costs of fluoropyrimidine chemotherapy? A propensity score matched analysis of 2022 UK patients. *BMC Cancer* 22(1), 458 (2022).
14. Garrison LP Jr, Lalla D, Brammer M *et al.* Assessing the potential cost-effectiveness of retesting IHC0, IHC1+, or FISH-negative early stage breast cancer patients for HER2 status. *Cancer* 119(17), 3113–3122 (2013).