eunethta

EUROPEAN NETWORK FOR HEALTH TECHNOLOGY ASSESSMENT

4

SECOND DRAFT Reflection Paper

Personalised medicine and co-dependent technologies, with a special focus on issues of study design

Version of December 2015

The primary objective of this EUnetHTA JA2 WP 7 methodological reflection paper is to focus on methodological challenges encountered by HTA assessors while performing relative effectiveness assessments (REA) of (co-dependent) health technologies in the emerging field of personalised medicine (PM).

The reflection paper aims to contribute to the ongoing discussion, both within EUnetHTA and with external stakeholders, about appropriate methods for the evaluation of PM technologies and lays the initial groundwork for the development of a comprehensive guideline with detailed recommendations for HTA and REA in this field.

As such, this reflection paper presents non-binding views and statements of EUnetHTA network members on the topic of PM.

In no way does it represent the official opinion of the participating institutions or individuals.

This methodological reflection paper on “Personalised medicine and co-dependent technologies, with a special focus on issues of study design” has been developed by IQWiG - GERMANY

With assistance from draft group members from
HAS – FRANCE and OSTEBA – SPAIN

The reflection paper was reviewed by a group of dedicated reviewers from
AIFA – ITALY, NICE – UNITED KINGDOM, AETSA – SPAIN

Table of contents

Acronyms – Abbreviations

AETSA	Andalusian HTA Agency
AIFA	Italian Medicines Agency
C	Control
CF TDN	Cystic Fibrosis Therapeutics Development Network
DNA	Deoxyribonucleic acid
ECFS-CTN	European Cystic Fibrosis Society-Clinical Trials Network
EGAPP	Evaluation of genomic applications in practice and prevention
EGFR	Epidermal growth factor receptor
EU	European Union
HAS	French National Authority for Health
HER2	Human epidermal growth factor receptor type 2
HTA	Health technology assessment
HTAAP	Health Technology Assessment Access Point
IMI	Innovative Medicines Initiative
IQWiG	Institute for Quality and Efficiency in Health Care
LEA	Linked evidence approach
MSAC	Australian Medical Services Advisory Committee
NICE	National Institute for Health and Care Excellence
NSCLC	Non-small-cell lung cancer
OSTEBA	Basque Agency for HTA, Department of Health
PFS	Progression-free survival
PM	Personalised medicine
RCT	Randomised controlled trial
REA	Relative effectiveness assessment
T	Treatment
UK	United Kingdom
USA	United States of America

Summary

The relevance of personalised medicine (PM) for patient care and the health care system in general is a subject of intense discussion, often accompanied by great expectations and promises concerning the patient-relevant benefit of PM technologies. A sharp increase in the dissemination of PM into the health care system is therefore to be expected. The reflection paper aims to contribute to the ongoing discussion, both within EUnetHTA and with external stakeholders, about appropriate methods for the evaluation of PM technologies. It also aims to support the development of a future methodological EUnetHTA guideline on PM containing detailed recommendations on various aspects from the health technology assessment (HTA) perspective. The working definition of PM for this reflection paper is taken from a recent systematic review conducted to understand how the term “PM” is actually used in scientific practice: “PM seeks to improve stratification and timing of health care by utilizing biological information and biomarkers on the level of molecular disease pathways, genetics, proteomics as well as metabolomics” [1]. This definition clarifies two aspects: (i) PM refers to groups of patients (i.e. does not actually lead to truly individualised interventions), and (ii) these groups are defined by biological information and biomarkers. The authors of this reflection paper postulate that in general, the existing HTA methods are appropriate tools for the assessment of PM technologies and that no specific new methods need to be developed. However, existing methods may need to be adapted. The reflection paper discusses the impact of four main study designs on the benefit assessment of PM technologies: (i) randomise-all, (ii) enrichment, (iii) single-arm, and (iv) biomarker-strategy design. These designs were identified in a recent systematic review. Several modifications are available and can also be combined with each other (hybrid designs). Whether the use of a specific design is appropriate depends on what question is asked and whether the assumptions made to interpret the results are acceptable. In the assessment of PM, as in most medical technologies, randomised controlled trials (RCTs) represent the highest level of evidence. However, such studies are rarely conducted within the context of PM, except for the enrichment design, which has some limitations. Due to this lack of direct high-quality evidence, some HTA bodies adopt a so-called “linked evidence approach” in the evaluation of co-dependent PM technologies. Other HTA bodies regard this approach as an “interim solution” associated with various methodological weaknesses. To infer from the current paucity of RCT-based evidence that lower-quality study designs should be used to assess PM technologies increases the risk

of biased conclusions in HTA. On a system level this may reduce the incentive to conduct more studies with higher-level evidence.

1. Introduction

1.1. Problem statement

The relevance of personalised medicine (PM) for patient care and the health care system in general is a subject of intense discussion, often accompanied by great expectations and promises concerning the patient-relevant benefit of PM technologies [2-5]. A sharp increase in the dissemination of PM into the health care system is to be expected: An analysis from 2012 of the clinical pipelines of 21 leading pharmaceutical and biotechnology companies showed that between 12% and 50% were PM technologies [6]. In addition, substantial resources are being invested in PM. For instance, in the USA, President Obama has allocated $215 million of the 2016 budget to a major research programme, the “Precision Medicine Initiative”, focussing on the development of more and better treatments for cancer [7,8]. Furthermore, between 2007 and 2012 an estimated €1 billion was invested in the promotion of PM technologies within the context of the Seventh EU Framework Programme for Research and Technological Development (FP7) [9].

This raises questions about the challenges involved for HTA bodies regarding the assessment and reimbursement of PM technologies. For instance, diverging opinions exist on whether or not new HTA methods are needed in this regard (e.g. [10-17]). These issues need to be discussed within the HTA community.

Various definitions of individualised or personalised medicine are available [18]. A broad one is provided by the European Alliance for Personalised Medicine: “A targeted approach to the prevention, diagnosis and treatment of disease based on an individual’s specific profile” [19]. A German working group conducted a systematic review to understand how the term “PM” was actually used in scientific practice. On the basis of the definitions identified, they derived the following “precising definition”: “PM seeks to improve stratification and timing of health care by utilizing biological information and biomarkers on the level of molecular disease pathways, genetics, proteomics as well as metabolomics” [1]. This definition clarifies two aspects: PM refers to groups of patients (i.e. does not actually lead to truly individualised interventions), and these groups are defined by “biological information and biomarkers”. The diagnostic test (e.g. imaging or laboratory biomarker) used to identify a specific patient group and the subsequent therapeutic intervention are co-dependent technologies, that is, “their use needs to be combined (either sequentially or simultaneously) to achieve or enhance the intended clinical effect of

either technology" [20]. The diagnostic test is also referred to as a "companion diagnostic" [21,22].

The above definition of PM by the German working group [1] leads to the term "stratified" (or "precision") medicine [23]. The two terms " PM" and "precision medicine" are often used interchangeably (also in scientific literature) [24]. "However, there was concern that the word `personalised´ could be misinterpreted to imply that treatments and preventive measures are being developed uniquely for each individual; in precision medicine, the focus is on identifying which approaches will be effective for which patients based on genetic, environmental, and lifestyle factors" [25]. PM and precision medicine essentially have the same goal, namely, "finding the best possible medicine and therapy for each and every individual patient" [26]. Precision medicine can therefore be considered an expansion of the PM concept and some researchers argue that the most recent term, "precision medicine", should be preferred to the term "PM" [27].

Another term often used in connection with PM is targeted therapy, a concept mostly used in cancer treatment, which "relies on the existence of a defined molecular target…and/or on the existence of a biomarker able to identify the target population…" [28].

PM "does not literally mean the creation of drugs or medical devices that are unique to a patient" [29], such as prosthetics designed for individual patients or the individualised processing of human cells and tissues (e.g. autologous stem cell transplants or patient-specific cancer vaccines) [29]. This does not imply that such interventions do not require evidence-based assessment; however, they do not form part of the common understanding of PM as outlined in the definition above [1]. Furthermore, this definition does not cover the individualised selection of treatments by means of n-of-1 trials.

1.2. Objective(s) and scope of the reflection paper

The reflection paper aims to contribute to the ongoing discussion, both within EUnetHTA and with external stakeholders, about appropriate methods for the evaluation of PM technologies. It also aims to support the development of a future methodological EUnetHTA guideline on PM containing detailed recommendations on various aspects from the HTA perspective. The paper discusses the impact of four main study designs on the

benefit assessment of PM technologies. Its focus it thus rather narrow. The cost-effectiveness assessment of PM, as well as legal, ethical, and societal implications, are beyond the scope of the reflection paper. However, these aspects are discussed elsewhere, for example, by Gutiérrez-Ibarluzea on the basis of the EUnetHTA core model [30].

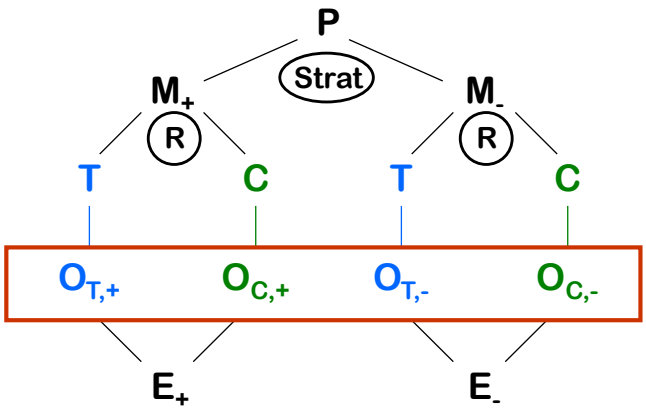
2. Study designs for the evaluation of biomarkers

This reflection paper discusses the four main categories of study designs aiming to assess the patient-relevant benefit of PM technologies: (i) randomise-all, (ii) enrichment, (iii) single-arm, and (iv) biomarker-strategy design. They were selected on the basis of a systematic review that aimed to identify study designs used in the evaluation of prognostic and predictive biomarkers and to develop a classification scheme [31]. To the best of our knowledge, this review is so far the only one of its kind. Several modifications of these study designs are available and can also be combined with each other (hybrid designs).

Enrichment, randomise-all, and biomarker-strategy designs are studies where different interventions or strategies are compared in parallel in groups of patients. Basically, the allocation of patients to groups does not necessarily have to be conducted in a randomised manner. However, randomisation is the only known form of allocation that generally results in equality of structure, and thus in the comparability, of study groups. Conversely, this means that for non-randomised studies a far greater effort is required to ensure sufficient similarity of structure. In addition, as greater variance can be expected, larger sample sizes are usually required for non-randomised studies than for randomised ones.

2.1. Randomise-all design

On the basis of the idea that the use of a biomarker can better select the patients for a certain treatment and hence improve patient-relevant outcomes, the interaction between diagnostic information provided by biomarkers and treatment effects needs to be evaluated. This requires a comparison of treatments within the different biomarker strata. Ideally this should be conducted by means of an RCT in which the biomarker test result remains blinded. As all patients are randomised to the different treatment options independently of their biomarker status, this type of design is referred to as a “randomise-all” design (see figure 1). Alternative terms are described in Tajik et al. [31].



Legend

C: Control; E: Effect; M: Marker status; O: Outcome; P: Patients; R: Randomisation; Strat: Stratification; T: Treatment; +: positive; -: negative

Figure 1: Randomise-all design

The following text describes the “simple” situation of a binary biomarker status (positive or negative) and two treatment options ([experimental] treatment [T] or control [C]). However, the basic principles can also be applied to more complex situations.

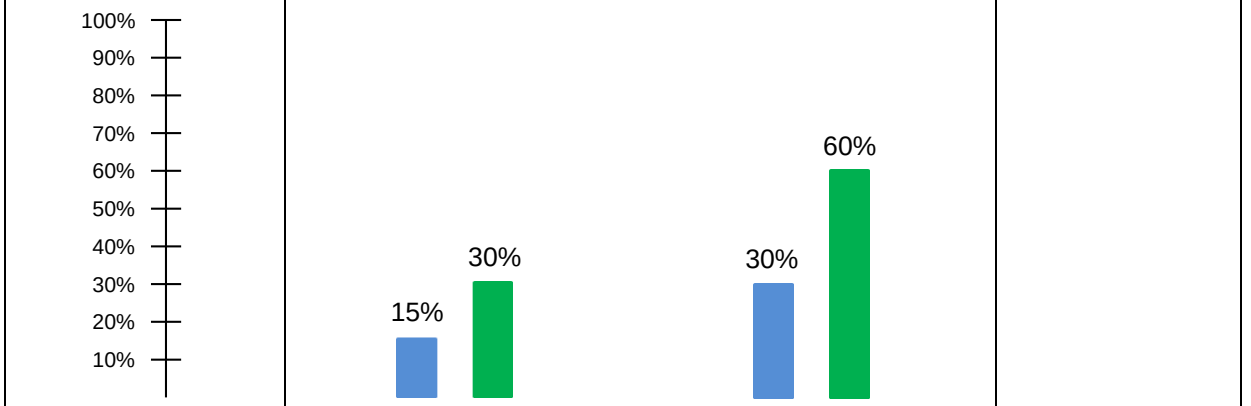
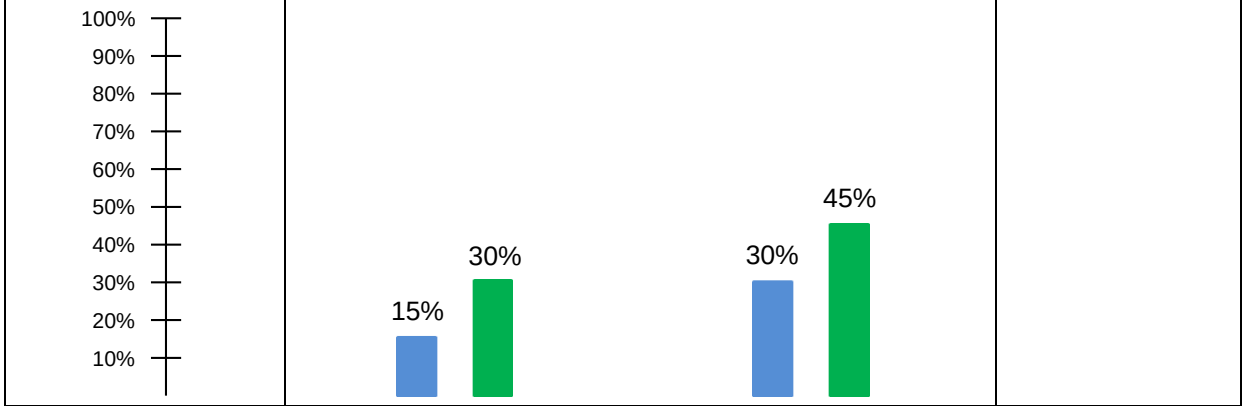
In the case of a qualitative interaction there is a reversal of effects, for instance, an advantage of T over C in biomarker-positive patients, but an advantage of C over T in biomarker-negative ones. A reversal of effects has rarely been shown in RCTs, presumably because, if such effects were assumed, a study comparing treatment alternatives in one of the two strata would not be regarded as appropriate, resource-efficient, or ethically justifiable in the first place. An example of such a reversal of effects is a study comparing first-line gefitinib therapy with standard chemotherapy (carboplatin plus paclitaxel) in patients with epidermal growth factor receptor (EGFR)-mutation-positive or negative advanced non-small-cell lung cancer (NSCLC) [32]. It is notable that the study was not primarily designed to detect an interaction between marker status and treatment effect, but to show the non-inferiority of gefitinib in the overall study population. Non-inferiority was shown; however, the survival curves of the two treatment groups crossed, which was ascribed to the interaction between marker status and treatment effect. In Europe, gefitinib was subsequently approved for patients with EGFR-positive NSCLC [33].

An interaction can also be described as qualitative if an advantage of T is shown over C in marker-positive patients, but no difference is shown in the marker-negative group. A well-known example is the benefit of antihormonal therapy (e.g. with tamoxifen) in patients with breast cancer depending on hormone-receptor status [34].

If treatment effects in the biomarker strata are in the same direction and only show quantitative differences, this represents a quantitative interaction. This situation is common and at the same time particularly problematical for the question of the benefit of a

biomarker-based choice of treatment. If the “personalised” treatment under investigation is only offered to patients in the subgroup where the greater treatment effect was shown, then a potentially beneficial treatment may be withheld from the remaining patients. This can only be justified if other reasons (e.g. adverse effects, costs) outweigh the (positive but smaller) treatment effects in the remaining patients. A further point should be noted: If a treatment effect is shown and the biomarker under investigation is of prognostic relevance for the outcome of interest, then at least a quantitative interaction always exists. This interaction is either on an absolute scale with a relatively constant treatment effect or on a relative scale with an absolutely constant treatment effect (see table 1). If a biomarker is to be used as a basis for the choice of treatment, a more than irrelevant quantitative interaction must exist. The demonstration of such an interaction requires a general consensus about the statistical methods to be applied.

Table 1: Quantitative interaction in dependence on the effect scale

Failure rate	Marker positive	Marker negative	Interpretation
			
Effect: relative risk reduction	50%	50%	No interaction
Effect: absolute risk reduction	15%	30%	Quantitative interaction
			
Effect: relative risk reduction	50%	33%	Quantitative interaction
Effect: absolute risk reduction	15%	15%	No interaction

= Treatment = Control

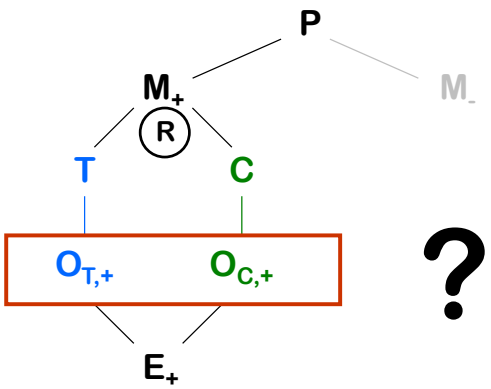
The advantage of a randomise-all design is that only this type of design is suitable to assess the individual components of a test-treatment strategy and their interaction. The following questions can be answered: “Is the treatment effective?” and “Is the biomarker sufficiently predictive?” The disadvantage is that the direct (physical or psychological) effects of the test (see Bossuyt and McCaffery [35]) on the patients cannot be assessed, as all patients are tested.

In principle, a randomise-all design can be applied retrospectively if the biomarker information is still available after completion of the treatment study [36]. However, two

main problems arise here: Firstly, a multiple testing problem is created that basically cannot be solved, but only restricted by the post-hoc formulation of a study objective in an amendment to the original study protocol (e.g. including a definition of thresholds). This type of study thus has a prospective-retrospective design [37]. Secondly, it must be ensured that the diagnostic tests used in the original study to determine the biomarkers and the subsequent treatments still represent current standards. It can be assumed that in view of the abundance of "omic" information, prospective-retrospective studies will play an increasing role in clinical research. To conduct these studies, data and samples from clinical studies not directly related to the original study objective will need to be made available and networks formed.

2.2. Enrichment design

In an enrichment design, only marker-positive (or sometimes only marker-negative) patients are included in a controlled treatment study, that is, the study inclusion criteria are extended by the marker status (see figure 2). Ultimately every clinical study with inclusion and exclusion criteria is thus a study with an enrichment design.



Legend

C: Control; E: Effect; M: Marker status;
O: Outcome; P: Patients; R: Randomisation; T: Treatment; +: positive; -: negative; ?: no data

Figure 2: Enrichment design

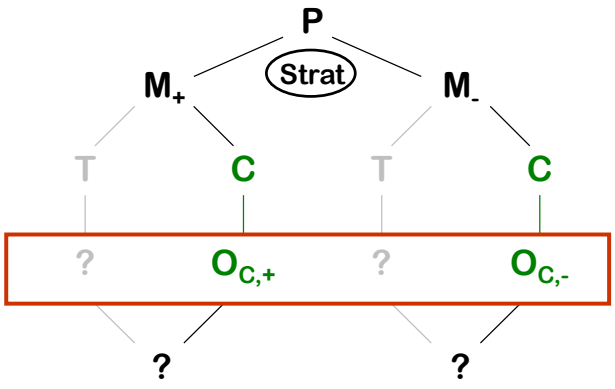
Compared with the randomise-all design, the information is missing on the effectiveness of treatment in patients with the excluded marker status. Instead, conclusions on the (unfavourable, insufficient or lacking) treatment effects in the excluded group are drawn on the basis of assumptions. However, as shown by the example of the impact of the human epidermal growth factor receptor type 2 (HER2) status in women with breast cancer, even highly plausible assumptions may be inaccurate or wrong. On the basis of

pathophysiological mechanisms, only patients with HER2-positive tumours were supposed to be treated in two pivotal studies of the monoclonal antibody trastuzumab, with beneficial effects on disease-free and overall survival [38]. About 10% of patients were subsequently reclassified as HER2-negative; however, the drug also appeared to show a benefit in these patients [39]. A benefit of trastuzumab was also shown in patients with HER2-negative primary tumours, but with circulating tumour cells expressing HER2 [40]. The US National Cancer Institute is currently conducting a study on trastuzumab therapy in patients with HER2-low breast cancer [41].

One advantage of a study with an enrichment design is that it requires a smaller sample size than other types of studies.

2.3. Single-arm design

A study is designated as single arm if only the prognostic value of a biomarker with regard to the occurrence of an unfavourable patient-relevant event is evaluated. In this type of study, different biomarker strata, not different treatment options, are compared (see figure 3). Such studies are typically conducted in patients receiving standard treatment. If no treatment is available, the study explores the “natural course” of the disease; however, this type of exploration is extremely rare nowadays, as some kind of treatment is usually offered, which may modify the course of disease.



Legend

C: Control; E: Effect; M: Marker status;
O: Outcome; P: Patients; Strat: Stratification;
T: Treatment; +: positive; -: negative;
?: no effect estimate

Figure 3: Single-arm design

In single-arm studies, similarly to studies with an enrichment design, half of the information required to answer the (decisive) question on interaction is missing, namely, on the prognosis in patients receiving experimental treatment and thus on the effectiveness of this treatment [42]. Therefore this type of study is basically not suitable to demonstrate the

benefit of a biomarker with regard to the optimal choice of treatment [43]. Such a benefit can only be directly derived if the study is able to identify patients with an extremely low (or high) risk of the unfavourable patient-relevant event, as these patients are highly unlikely to benefit from the new (or existing) treatment.

2.4. Biomarker-strategy design

In a biomarker-strategy design an RCT is conducted to determine whether a strategy using a biomarker to choose a treatment is superior to a conventional strategy without this biomarker (see figure 4). A clear advantage of this design is that the direct effect of applying the test in patients can be evaluated [35]. In addition, on a population level an unbiased estimate is provided of the effect of using a biomarker-based strategy to inform the choice of treatment.

The biomarker strategy design is typically used for the evaluation of screening tests [44,45] but can also be used for the evaluation of any other diagnostic test [46,47]. In particular, the question of the benefit of serial monitoring or other complex strategies can often be clarified only with this type of design [48,49].

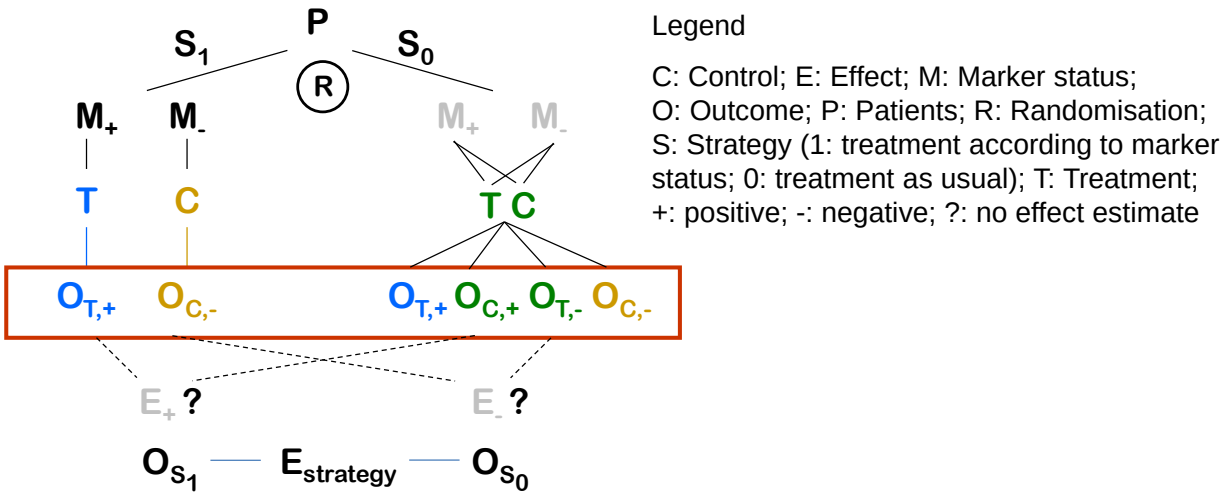


Figure 4: Biomarker-strategy design

A disadvantage of this design is that the test and treatment components of the biomarker strategy cannot be evaluated separately. A further disadvantage is that this design is less efficient than other designs (i.e. has less power and will require a larger number of patients), as both treatment groups include patients with similar characteristics who

received identical treatments. This diminishes the contrast between the groups [50]. In the extreme case – if the same number of marker-positive and marker-negative patients is included and all patients in the control group receive the same treatment – this may apply to half of the patient population.

2.5. Hybrid design

Several modifications of the above-mentioned study designs are available and can also be combined with each other; for example, an enrichment design in marker-positive patients with a single-arm design in marker-negative ones, or an enrichment design with a subsequent randomise-all design [31,51].

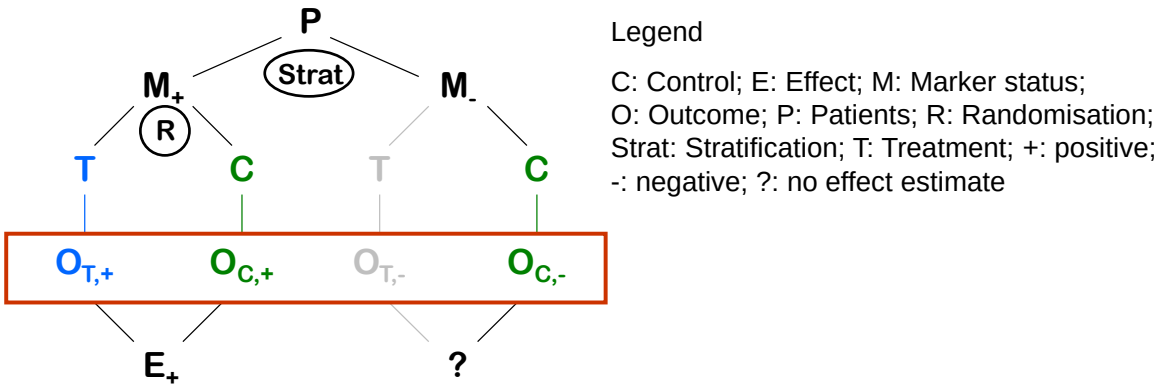
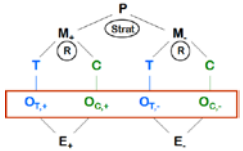
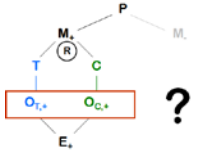
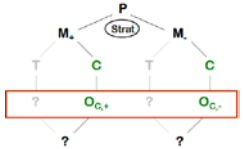
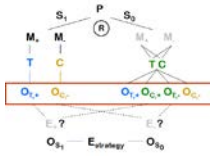
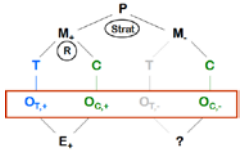


Figure 5: Example of a hybrid design (enrichment for M_+, single-arm for M_-)

In addition, it is possible to examine several biomarkers and several treatment options within the same study or compare various test-treatment strategies with each other. To present all of these additional options would go beyond the scope of this reflection paper. Table 2 summarises the possible conclusions that can be drawn from the study designs outlined above and presents examples from the literature (not restricted to molecular-biological markers).

385 Table 2: Study designs – Implications for possible conclusions

Design \ Allows conclusions with regard to …	Prognosis	Treatment effect	Treatment x marker interaction	Effect of testing (direct + indirect)	Examples
Randomise-all 	Yes	Yes	Yes	No	[32,34,52-55]
Enrichment 	In M_+ only	In M_+ only	No	No	[38-41,56-58]
Single-arm 	In C only	No	No	No	- [a]
Biomarker-strategy 	No	No	No	Yes	[46,59-61]
Example of a hybrid design (enrichment + single-arm) 	In M_-,C only	In M_+ only	If prognosis in M_-,C is excellent or very poor (qualitative interpretation)	No	[62-65]

386 a: No examples presented, as a vast number of prognostic studies are available in the literature [66-68].

387

2.6. Handling the situation of insufficient (direct) evidence for the benefit assessment of co-dependent technologies

Whether the use of a specific design is appropriate depends on what (relevant) question is asked and whether the assumptions made to interpret the results are acceptable.

Unfortunately, of the designs presented above, in practice often only (prognostic) single-arm studies or at best studies with an enrichment design are available for the evaluation of test-treatment strategies. However, conclusions on the benefit of a strategy can only be drawn from these study designs under stringent assumptions. In principle, two possibilities exist: if the assumptions required to determine a benefit are inappropriate (i.e. are not based on robust values), then a conclusion of no hint of a benefit is drawn. This is currently the standard approach applied, for example, by the German Institute for Quality and Efficiency in Health Care (IQWiG) [69]. Alternatively, for the (common) situation where there is a lack of direct evidence of the benefit of a diagnostic test, a “linked evidence approach” (LEA) may be applied. This approach was developed by the Australian Medical Services Advisory Committee (MSAC) [70-72] and “involves the narrative linking of evidence assessing components of a test-treatment pathway in order to predict the likely impact of testing on patient health outcomes” [72]. This means that results from diagnostic accuracy studies investigating a new test versus a reference test are linked to results from treatment studies in which the reference test served as an inclusion criterion [70]. However, the LEA requires the availability of a reference standard that generally aims to identify the same patients as the new test. But many new genetic or molecular tests do not fulfil this requirement, as their aim is to identify patients with new characteristics (e.g. previously unknown biomarker status).

In addition, the LEA may be affected by bias or variation; for instance, the characteristics (e.g. demographic features or disease severity) of the patients included in the diagnostic and treatment studies may differ [70]. As with an indirect comparison of interventions, these characteristics must therefore be comparable. However, a further problem arises, namely, how to define thresholds to determine whether comparability is sufficient. The situation is easier if a clinical trial has already shown a benefit for a particular biomarker. For an alternative biomarker, it would be sufficient to determine that it identifies the same population as the first biomarker, that is, to determine that the results of the two tests agree [73]. Methods for this type of evaluation are available [74]. However, measures of overall agreement should generally be avoided, and as with equivalence hypothesis

testing, thresholds must be defined to determine whether positive percent agreement is sufficient.

3. Discussion

The concept of PM, that is, the selection of a specific treatment option on the basis of patient characteristics (predictive biomarkers) that interact with treatment effects, is not really new [75] – medical interventions have always been chosen on the basis of some form of diagnosis, mostly “classical” pathological and clinical signs and symptoms. The novelty of the current approach merely consists in the increased use of information at a molecular level as stratification characteristics (hence the name “stratified” or “precision” medicine). In the opinion of the authors of this reflection paper, no specific new methods are required to assess the (diagnostic/prognostic and therapeutic) PM technologies. This is because a methodological framework has been available for decades to estimate treatment effects in the most unbiased manner possible and to estimate different treatment effects in subgroups. However, what is required is the conjoint assessment of these technologies and possibly a new regulatory framework.

PM can help to change subgroup analyses from being viewed as a questionable methodological approach (“Subgroups kill people“ [76]) to being viewed as a recognised one (“Not doing subgroup analyses has very probably killed more people“ [77]). PM can thereby help to ease the way for stratified therapeutic strategies with more targeted and effective treatments.

While PM encompasses a wide range of diagnostic, prognostic, monitoring, and other technologies, the general principle of benefit assessment (i.e. the comparison of the use versus non-use of the new PM technology) applies to all of them. The ultimate goal is to demonstrate clinical utility (see, for example, [48]).

Due to rapid medical progress, for example, in the field of genomic research in cancer, as already stated, some researchers call for novel assessment methods in HTA [11]. For instance, if one considers the use of PM in lung cancer treatment, which involves highly complex PM strategies with various treatment choices based on simultaneous testing of multiple biomarkers [78], it is obvious that the patient-relevant benefit of these highly complex strategies compared with the current standard strategy has to be demonstrated. However, this can be evaluated within adequately controlled trials [79-81], so that traditional HTA methods can be used for the assessment. This means that the current EUnetHTA methodological guidelines for the REA of pharmaceuticals could also be used

to assess intervention trials of PM technologies (e.g. the guidelines on assessment of the internal validity of clinical trials, the applicability of evidence, and the use of clinical endpoints [82]). However, existing methods may need to be adapted. For instance, the further development of statistical methods for subgroup analyses is desirable: there is a need for a framework to exclude an irrelevant quantitative interaction between the biomarker and the treatment, or – vice versa – to establish a relevant quantitative interaction. In addition, the combination of confirmatory approaches for testing treatment effects in an entire study population and in subgroups is currently being discussed, taking the problem of multiple testing into account [51,83]. However, these already existing methods are practically unused so far in studies of PM technologies.

Furthermore, in the early stages of the (concomitant) development of biomarker-based diagnostics and biomarker-targeted treatments, (new) adaptive designs may play a role in Phase II studies in the efficient selection of promising biomarker-treatment combinations for subsequent Phase III studies [51].

The previously separate development and assessment of diagnostic tests and therapeutic interventions is therefore becoming increasingly obsolete. However, a conjoint assessment is not the current standard approach [84]. In the EU and the USA, the market access of diagnostics and drugs is still regulated separately. The regions differ in their way of handling companion diagnostics: As summarised by Byron et al. from the UK National Institute for Health and Care Excellence (NICE), in the EU “the licensed indication of a pharmaceutical may require the use of a companion diagnostic but the specific test for determining the mutation status is not normally stipulated.” In contrast, in the USA, “the regulatory process examines the suitability of a specific test for selecting patients for treatment with the corresponding pharmaceutical, and this test is stipulated in the licensing” [73]. The current regulatory processes can lead to the situation that, “even in cases of co-developed [drug-diagnostic] combinations drug reimbursement does not necessarily imply diagnostic reimbursement” [85]. Some governmental and HTA bodies have reacted to the current challenges: For instance, NICE has developed a policy for considering companion diagnostics using its Technology Appraisal and Diagnostics Assessment Programme [73]. The Australian government supported Merlin et al’s development of a national 5-component framework (“context, clinical benefit, evidence translation, cost-effectiveness, and financial impact”) for the HTA of PM [86]. Its introduction was accompanied by the establishment of the Health Technology Assessment

Access Point (HTAAP) to coordinate the assessment of companion diagnostics within the HTA process.

In the development and assessment of biomarkers, the “traditional” criteria of technical-analytical validity and clinical validity are still currently described as important steps in the pertinent methods papers [12,87-90]. However, the following problems should be noted: Firstly, according to the experiences of the Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group, evidence on the analytical validity of genomic biomarkers is sparse [91], that is, obtaining reliable and unbiased estimates to determine analytical validity is particularly challenging [91]. For instance, a search in bibliographic databases is generally insufficient to identify all of the relevant information; additional searches for unpublished data from companies and laboratories also need to be conducted. Secondly, there are specific problems with the analytical “gold standard”, for example, in the case of tumour gene-expression profiling in breast cancer, where no real gold standard existed at the time of evaluation [91,92]. Thirdly, the interpretation of genomic results may be hampered by DNA damage following formalin fixation [93]. And fourthly, considering clinical validity, there may be differences between the central lab test used for the (Phase 3) clinical trial and the test used in a routine (“real-life”) lab setting. In clinical practice the results of the companion diagnostic may differ from the test used in central lab conditions [94]. Beyond these problems, there is a general consensus that clinical utility must ultimately be proven [89,95,96]. For a better understanding of this issue, it is crucial to distinguish between prognostic and predictive biomarkers [90].

In principle, RCTs represent the highest level of evidence in the assessment of biomarkers for the stratification of patients. However, such studies are rarely conducted. To infer from this current paucity of RCT-based evidence that lower-quality study designs should be used to assess PM technologies increases the risk of biased conclusions in HTA. On a system level this may reduce the incentive to conduct more studies with higher-level evidence.

In general, the more a patient population in a clinical study is stratified, the smaller the population for analysis will become. It may thus be difficult to recruit the number of patients required to reach robust conclusions. However, this problem applies not only to randomised studies, but to all study designs. The first approach to solve this problem should thus avoid resorting to study designs of a lower evidence level. Instead, the creation of multi-institutional and international networks is needed to create a broader

basis for patient recruitment [97]. Commendable examples include the Cystic Fibrosis Therapeutics Development Network (CF TDN) [98] and the European Cystic Fibrosis Society-Clinical Trials Network (ECFS-CTN) [99].

In addition, data from completed studies should be made publicly available so that the interaction between a specific biomarker and the results of treatment can be analysed, also retrospectively (prospective-retrospective design). However, this requires the establishment of structures for storing and processing samples of patients included in the studies (biomaterial banks). Furthermore, the identification of new treatment targets and the implementation of the corresponding treatment approaches do not necessarily have to result in smaller patient populations. For instance, in oncology the role of basket and umbrella studies is currently being discussed. The former include studies of multiple tumour types harbouring the same mutation; the latter include studies of a single tumour type harbouring different mutations [100,101]. In both types of studies the targeted therapy is compared with the current standard therapy.

And finally: if “personalised” treatment effects are comparatively large – as is claimed by some advocates of PM – the corresponding trials might require only a small sample size and a limited amount of resources [102].

Annotation: Several projects, some in collaboration with EUnetHTA or EUnetHTA member organizations, are currently being conducted within the Innovative Medicines Initiative (IMI) launched by the EU and the pharmaceutical industry [103]. The aim of the initiative is to speed up the development and market access of promising medical technologies. A discussion of the IMI projects (e.g. ADAPT-SMART [104], GetReal [105,106]) would go beyond the scope of this reflection paper. It remains to be seen to what extent their results will have consequences for the assessment of PM technologies or will lead to the development of innovative study designs that could markedly help improve the evidence base.

Annexe 1. Definitions of key terms and concepts

According to the Biomarker Definitions Working Group, a **biomarker** is defined as a "characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention" [107]. Biomarkers are used as diagnostic tools for the identification of patients with a disease, as tools for staging of disease, as indicators of disease prognosis, and for prediction and monitoring of clinical response to an intervention [107].

The aim of a **diagnostic test** is to distinguish between individuals with and those without a particular disease. The quality of a diagnostic test is described by means of diagnostic accuracy measures such as sensitivity and specificity or likelihood ratios [108]. Together with the known or assumed prevalence of a disease, these measures can be used to calculate predictive values describing the likelihood that a given test result correctly identifies an individual as having or not having a disease.

A **prognostic test** identifies "patients with differing risks of a specific outcome, such as progression or death" [109]. The relationship between the test result and the frequency of an outcome is typically described by means of epidemiological measures such as the relative risk, hazard ratio, or odds ratio. In principle, the above-mentioned diagnostic accuracy measures can also be used for this purpose.

A **predictive test** "predicts the differential efficacy (benefit) of a particular therapy based on the marker status" [109]. This capability of a test is described as the interaction between the marker status (i.e. the test result) and the benefit of treatment. If no interaction exists, all patients benefit to the same extent from treatment, independently of the test result. In the case of a quantitative interaction, all patients benefit from treatment; however, the extent of benefit differs depending on the test result. In the case of a qualitative interaction, only a particular subgroup benefits from treatment, whereas the other subgroup experiences no benefit or even harm. Only the results of predictive tests can guide the choice of treatment; purely prognostic tests cannot [109]. However, a test may be both prognostic and predictive or it may be predictive without being prognostic [110].

The **clinical validity** of a test, i.e. its diagnostic or prognostic accuracy, is the "ability to detect or predict the associated disorder" [111]; the **clinical utility** of a test is the "ability to

affect clinical decisions and to improve patient outcomes in clinical practice" [111]. Hence, clinical utility describes the consequences of applying the information obtained from a medical test to the care of patients. Such consequences are not necessarily beneficial to patients, even if the information applied is (formally) correct [35].

Annexe 2. Bibliography

1. Schleidgen S, Klingler C, Bertram T, Rogowski WH, Marckmann G. What is personalized medicine: sharpening a vague term based on a systematic literature review. BMC Med Ethics 2013; 14: 55.
2. Ioannidis JP. Expectations, validity, and reality in omics. J Clin Epidemiol 2010; 63(9): 945-949.
3. American Society of Clinical Oncology. Accelerating progress against cancer: ASCO's blueprint for transforming clinical and translational cancer research [internet]. 2011 Nov [cited: 2015 Apr 09]. Available from: http://www.asco.org/sites/default/files/blueprint.pdf.
4. Personalized Medicine Coalition. The case for personalized medicine [internet]. 2014 [cited: 2015 Apr 09]. Available from: http://www.personalizedmedicinecoalition.org/Userfiles/PMC-Corporate/file/pmc_the_case_for_personalized_medicine.pdf
5. Miller FA, Hayeems RZ, Bytautas JP, Bedard PL, Ernst S, Hirte H et al. Testing personalized medicine: patient and physician expectations of next-generation genomic sequencing in late-stage cancer care. Eur J Hum Genet 2014; 22(3): 391-395.
6. Fugel HJ, Nuijten M, Postma M. Stratified medicine and reimbursement issues. Front Pharmacol 2012; 3: 181.
7. White House - Office of the Press Secretary. Fact sheet: President Obama's Precision Medicine Initiative [internet]. 2015 Jan 30 [cited: 2015 Apr 15]. Available from: https://www.whitehouse.gov/the-press-office/2015/01/30/fact-sheet-president-obama-s-precision-medicine-initiative.
8. Collins FS, Varmus H. A new initiative on precision medicine. N Engl J Med 2015; 372(9): 793-795.
9. Commission E. Commission staff working document. Use of '-omics' technologies in the development of personalised medicine. [internet]. [cited: 2015 Dec 19]. Available from: http://ec.europa.eu/health/files/latest_news/2013-10_personalised_medicine_en.pdf.
10. Windeler J, Lange S. Benefit assessment of personalised interventions: methodological challenges and approaches to a solution. In: Sandow V, Vollmann J, Wäscher S, Schildmann J (Ed). The ethics of personalised medicine: critical perspectives [in press]. Surrey: Ashgate; 2015. p. 175-185.
11. Becla L, Lunshof JE, Gurwitz D, Schulte in den Bäumen T, Westerhoff HV, Lange BM et al. Health technology assessment in the era of personalized health care. Int J Technol Assess Health Care 2011; 27(2): 118-126.
12. Teutsch SM, Bradley LA, Palomaki GE, Haddow JE, Piper M, Calonge N et al. The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Initiative: methods of the EGAPP Working Group. Genet Med 2009; 11(1): 3-14.
13. Ransohoff DF, Khoury MJ. Personal genomics: information can be harmful. Eur J Clin Invest 2010; 40(1): 64-68.
14. Ziegler A, Koch A, Krockenberger K, Grosshennig A. Personalized medicine using DNA biomarkers: a review. Hum Genet 2012; 131(10): 1627-1638.
15. De Leon J. Evidence-based medicine versus personalized medicine: are they enemies? J Clin Psychopharmacol 2012; 32(2): 153-164.
16. The Academy of Medical Sciences. Realising the potential of stratified medicine [internet]. [cited: 2015 Dec 29]. Available from: https://www.acmedsci.ac.uk/viewFile/51e915f9f09fb.pdf.
17. Food and Drug Administration. Paving the Way for Personalized Medicine. FDA's Role in a New Era of Medical Product Development [internet]. [cited: 2015 Dec 29]. Available from:

637 http://www.fda.gov/downloads/scienceresearch/specialtopics/personalizedmedicine/ucm37
638 2421.pdf.
639 18. European Science Foundation. Personalised medicine for the European Citizen:
640 towards more precise medicine for the diagnosis, treatment and prevention of disease
641 (iPM) [internet]. 2012 Nov [cited: 2015 Apr 09]. Available from:
642 http://www.esf.org/fileadmin/Public_documents/Publications/Personalised_Medicine.pdf.
643 19. European Alliance for Personalised Medicine. What does personalised medicine
644 promise [internet]. [cited: 2015 Apr 09]. Available from: http://euapm.eu/what-does-
645 personalised-medicine-promise.
646 20. Australian Government - Department of Health. Co-dependent and hybrid technologies
647 [internet]. 2011 Nov 29 [cited: 2015 Apr 14]. Available from:
648 http://www.health.gov.au/internet/hta/publishing.nsf/Content/co-1.
649 21. National Institute for Health and Clinical Excellence. Diagnostics Assessment
650 Programme manual [internet]. 2011 Dec [cited: 2015 May 05]. Available from:
651 http://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-diagnostics-
652 guidance/Diagnostics-assessment-programme-manual.pdf.
653 22. Food and Drug Administration. In vitro companion diagnostic devices: guidance for
654 industry and Food and Drug Administration staff [internet]. 2014 Aug 06 [cited: 2015 Apr
655 14]. Available from: http://www.fda.gov/ucm/groups/fdagov-public/@fdagov-meddev-
656 gen/documents/document/ucm262327.pdf.
657 23. Katsnelson A. Momentum grows to make 'personalized' medicine more 'precise'. Nat
658 Med 2013; 19(3): 249.
659 24. Jameson JL, Longo DL. Precision medicine--personalized, problematic, and promising.
660 N Engl J Med 2015; 372(23): 2229-2234.
661 25. Lister Hill National Center for Biomedical Communications, U.S. National Library of
662 Medicine, National Institutes of Health, Department of Health & Human Services. Genetics
663 home reference - Your guide to understanding genetic conditions – Handbook - Help Me
664 Understand Genetics [internet]. [cited: 2016 Jan 13]. Available from:
665 http://ghr.nlm.nih.gov/handbook.pdf.
666 26. Genome Alberta. “Personalized medicine” renamed “precision medicine” [internet].
667 [cited: 2016 Jan 13]. Available from: http://genomealberta.ca/blogs/personalized-
668 medicine-renamed-precision-medicine.aspx.
669 27. Committee on a Framework for Development a New Taxonomy of Disease; Board on
670 Life Sciences DoEaLS, National Research Council. Toward Precision Medicine. Building a
671 Knowledge Network for Biomedical Research and a New Taxonomy of Disease [internet].
672 [cited: 2015 Dec 19]. Available from: http://www.nap.edu/catalog/13284/toward-precision-
673 medicine-building-a-knowledge-network-for-biomedical-research.
674 28. Marquet P, Longeray PH, Barlesi F, Ameye V, Auge P, Cazeneuve B et al.
675 Translational research: precision medicine, personalized medicine, targeted therapies;
676 marketing or science? Therapie 2015; 70(1): 11-19.
677 29. President’s Council of Advisors on Science and Technology. Priorities for personalized
678 medicine [internet]. 2008 Sep [cited: 2015 Apr 09]. Available from:
679 https://www.whitehouse.gov/files/documents/ostp/PCAST/pcast_report_v2.pdf.
680 30. Gutiérrez-Ibarluzea I. Personalised Health Care, the need for reassessment: a HTA
681 perspective far beyond cost-effectiveness Ital J Public Health 2012; 9(4). e8653.1-9.
682 31. Tajik P, Zwinderman AH, Mol BW, Bossuyt PM. Trial designs for personalizing cancer
683 care: a systematic review and classification. Clin Cancer Res 2013; 19(17): 4578-4588.
684 32. Mok TS, Wu YL, Thongprasert S, Yang CH, Chu DT, Saijo N et al. Gefitinib or
685 carboplatin-paclitaxel in pulmonary adenocarcinoma. N Engl J Med 2009; 361(10): 947-
686 957.
687 33. European Medicines Agency. Iressa: European public assessment report [internet].
688 2009 Jul 22 [cited: 2015 Apr 26]. Available from:

http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Public_assessment_report/human/001016/WC500036361.pdf.
34. Early Breast Cancer Trialists' Collaborative Group. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet 2005; 365(9472): 1687-1717.
35. Bossuyt PM, McCaffery K. Additional patient outcomes and pathways in evaluations of testing. Med Decis Making 2009; 29(5): E30-38.
36. Duffy MJ, Crown J. Companion biomarkers: paving the pathway to personalized treatment for cancer. Clin Chem 2013; 59(10): 1447-1456.
37. Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. J Natl Cancer Inst 2009; 101(21): 1446-1452.
38. Romond EH, Perez EA, Bryant J, Suman VJ, Geyer CE Jr, Davidson NE et al. Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. N Engl J Med 2005; 353(16): 1673-1684.
39. Paik S, Kim C, Wolmark N. HER2 status and benefit from adjuvant trastuzumab in breast cancer. N Engl J Med 2008; 358(13): 1409-1411.
40. Georgoulias V, Bozionelou V, Agelaki S, Perraki M, Apostolaki S, Kallergi G et al. Trastuzumab decreases the incidence of clinical relapses in patients with early breast cancer presenting chemotherapy-resistant CK-19 mRNA-positive circulating tumor cells: results of a randomized phase II study. Ann Oncol 2012; 23(7): 1744-1750.
41. National Cancer Institute. Chemotherapy with or without trastuzumab after surgery in treating women with invasive breast cancer [internet]. In: ClinicalTrials.gov. 2015 May 04 [cited: 2015 May 06]. Available from: https://clinicaltrials.gov/ct2/show/NCT01275677.
42. Hingorani AD, Windt DA, Riley RD, Abrams K, Moons KG, Steyerberg EW et al. Prognosis research strategy (PROGRESS) 4: stratified medicine research. BMJ 2013; 346: e5793.
43. Mandrekar SJ, Sargent DJ. Design of clinical trials for biomarker research in oncology. Clin Investig (Lond) 2011; 1(12): 1629-1636.
44. Nelson HD CA, Humphrey L, Fu R, Pappas M, Daeges M, Griffin J. Screening for Breast Cancer: A Systematic Review to Update the 2009 U.S. Preventive Services Task Force Recommendation. Rockville, MD: Agency for Healthcare Research and Quality. [internet]. [cited: 2015 Dec 19]. Available from: www.uspreventiveservicestaskforce.org/Home/GetFileByID/1923.
45. Schröder FH, Hugosson J, Roobol MJ, Tammela TL, Zappa M, Nelen V et al. Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. Lancet 2014; 384(9959): 2027-2035.
46. Van Tinteren H, Hoekstra OS, Smit EF, Van den Bergh JH, Schreurs AJ, Stallaert RA et al. Effectiveness of positron emission tomography in the preoperative assessment of patients with suspected non-small-cell lung cancer: the PLUS multicentre randomised trial. Lancet 2002; 359(9315): 1388-1393.
47. Moulton CA, Gu CS, Law CH, Tandan VR, Hart R, Quan D et al. Effect of PET before liver resection on surgical management for colorectal adenocarcinoma metastases: a randomized clinical trial. Jama 2014; 311(18): 1863-1869.
48. Soletormos G, Duffy MJ, Hayes DF, Sturgeon CM, Barak V, Bossuyt PM et al. Design of tumor biomarker-monitoring trials: a proposal by the European Group on Tumor Markers. Clin Chem 2013; 59(1): 52-59.
49. Panidou ET, Trikalinos TA, Ioannidis JP. Limited benefit of antiretroviral resistance testing in treatment-experienced patients: a meta-analysis. Aids 2004; 18(16): 2153-2161.
50. Mandrekar SJ, Grothey A, Goetz MP, Sargent DJ. Clinical trial designs for prospective validation of biomarkers. Am J Pharmacogenomics 2005; 5(5): 317-325.

740 51. Mandrekar SJ, Sargent DJ. Drug designs fulfilling the requirements of clinical trials
741 aiming at personalizing medicine. Chin Clin Oncol 2014; 3(2): 14.
742 52. Cannon CP, Weintraub WS, Demopoulos LA, Vicari R, Frey MJ, Lakkis N et al.
743 Comparison of early invasive and conservative strategies in patients with unstable
744 coronary syndromes treated with the glycoprotein IIb/IIIa inhibitor tirofiban. N Engl J Med
745 2001; 344(25): 1879-1887.
746 53. Neumann F, Kastrati A, Miethke T, Pogatsa-Murray G, Mehilli J, Valina C et al.
747 Treatment of Chlamydia pneumoniae infection with roxithromycin and effect on neointima
748 proliferation after coronary stent placement (ISAR-3): a randomised, double-blind,
749 placebo-controlled trial. Lancet 2001; 357(9274): 2085-2089.
750 54. Olaussen KA, Dunant A, Fouret P, Brambilla E, Andre F, Haddad V et al. DNA repair
751 by ERCC1 in non-small-cell lung cancer and cisplatin-based adjuvant chemotherapy. N
752 Engl J Med 2006; 355(10): 983-991.
753 55. Karapetis CS, Khambata-Ford S, Jonker DJ, O'Callaghan CJ, Tu D, Tebbutt NC et al.
754 K-ras mutations and benefit from cetuximab in advanced colorectal cancer. N Engl J Med
755 2008; 359(17): 1757-1765.
756 56. Chapman PB, Hauschild A, Robert C, Haanen JB, Ascierto P, Larkin J et al. Improved
757 survival with vemurafenib in melanoma with BRAF V600E mutation. N Engl J Med 2011;
758 364(26): 2507-2516.
759 57. Lipworth BJ, Basu K, Donald HP, Tavendale R, Macgregor DF, Ogston SA et al.
760 Tailored second-line therapy in asthmatic children with the Arg^{16} genotype. Clin Sci (Lond)
761 2013; 124(8): 521-528.
762 58. Rosell R, Carcereny E, Gervais R, Vergnenegre A, Massuti B, Felip E et al. Erlotinib
763 versus standard chemotherapy as first-line treatment for European patients with advanced
764 EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label,
765 randomised phase 3 trial. Lancet Oncol 2012; 13(3): 239-246.
766 59. Mallal S, Phillips E, Carosi G, Molina JM, Workman C, Tomazic J et al. HLA-B*5701
767 screening for hypersensitivity to abacavir. N Engl J Med 2008; 358(6): 568-579.
768 60. Tonino PA, De Bruyne B, Pijls NH, Siebert U, Ikeno F, Van' t Veer M et al. Fractional
769 flow reserve versus angiography for guiding percutaneous coronary intervention. N Engl J
770 Med 2009; 360(3): 213-224.
771 61. Cree IA, Kurbacher CM, Lamont A, Hindley AC, Love S. A prospective randomized
772 controlled trial of tumour chemosensitivity assay directed chemotherapy versus physician's
773 choice in patients with recurrent platinum-resistant ovarian cancer. Anticancer Drugs 2007;
774 18(9): 1093-1101.
775 62. Palareti G, Cosmi B, Legnani C, Tosetto A, Brusi C, Iorio A et al. D-dimer testing to
776 determine the duration of anticoagulation therapy. N Engl J Med 2006; 355(17): 1780-
777 1789.
778 63. Bartlett J, Canney P, Campbell A, Cameron D, Donovan J, Dunn J et al. Selecting
779 breast cancer patients for chemotherapy: the opening of the UK OPTIMA trial. Clin Oncol
780 (R Coll Radiol) 2013; 25(2): 109-116.
781 64. Zujewski JA, Kamin L. Trial assessing individualized options for treatment for breast
782 cancer: the TAILORx trial. Future Oncol 2008; 4(5): 603-610.
783 65. Rutgers E, Piccart-Gebhart MJ, Bogaerts J, Delaloge S, Veer LV, Rubio IT et al. The
784 EORTC 10041/BIG 03-04 MINDACT trial is feasible: results of the pilot phase. Eur J
785 Cancer 2011; 47(18): 2742-2749.
786 66. Alonzo TA. Standards for reporting prognostic tumor marker studies. J Clin Oncol
787 2005; 23(36): 9053-9054.
788 67. Albert MA. Biomarkers and heart disease. J Clin Sleep Med 2011; 7(5 Suppl): S9-11.
789 68. Ioannidis JP, Panagiotou OA. Comparison of effect sizes associated with biomarkers
790 reported in highly cited individual articles and in subsequent meta-analyses. JAMA 2011;
791 305(21): 2200-2210.

69. Institute for Quality and Efficiency in Health Care. General methods: version 4.2 [internet]. 2015 Apr 22 [cited: 2015 Dec 19]. Available from: https://www.iqwig.de/download/IQWiG_General_Methods_Version_%204-2.pdf.
70. Medical Services Advisory Committee. Guidelines for the assessment of diagnostic technologies [internet]. 2005 Aug [cited: 2015 Apr 14]. Available from: http://www.health.gov.au/internet/msac/publishing.nsf/Content/D81BE529B98B3DB6CA2575AD0082FD1B/$File/Guidelines%20for%20the%20assessment%20of%20diagnostic%20technologies%20Sept%202005.pdf.
71. Merlin T, Lehman S, Hiller JE, Ryan P. The "linked evidence approach" to assess medical tests: a critical analysis. Int J Technol Assess Health Care 2013; 29(3): 343-350.
72. Merlin T. The use of the ‘linked evidence approach’ to guide policy on the reimbursement of personalized medicines. Per Med 2014; 11(4): 435-448.
73. Byron SK, Crabb N, George E, Marlow M, Newland A. The Health Technology Assessment of companion diagnostics: experience of NICE. Clin Cancer Res 2014; 20(6): 1469-1476.
74. Food and Drug Administration. Statistical guidance on reporting results from studies evaluating diagnostic tests: guidance for industry and FDA staff [internet]. 2007 Mar 13 [cited: 2015 Apr 26]. Available from: http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071287.pdf.
75. Meyer UA. Welcome to the Journal of Personalized Medicine: A New Open-Access Platform for Research on Optimal Individual Healthcare. J Pers Med 2011; 1(1): 1-4.
76. Van Gijn J. Extrapolation of trials data into practice: where is the limit? Cerebrovasc Dis 1995; 5(3): 159-162.
77. Rothwell PM. Treating individuals 2: subgroup analysis in randomised controlled trials; importance, indications, and interpretation. Lancet 2005; 365(9454): 176-186.
78. Kris MG, Johnson BE, Berry LD, Kwiatkowski DJ, Iafrate AJ, Wistuba II et al. Using multiplexed assays of oncogenic drivers in lung cancers to select targeted drugs. JAMA 2014; 311(19): 1998-2006.
79. Abrams J, Conley B, Mooney M, Zwiebel J, Chen A, Welch JJ et al. National Cancer Institute's Precision Medicine Initiatives for the new National Clinical Trials Network. Am Soc Clin Oncol Educ Book 2014: 71-76.
80. Tourneau CL, Kamal M, Alt M, Verlingue L, Servois V, Sablin MP et al. The spectrum of clinical trials aiming at personalizing medicine. Chin Clin Oncol 2014; 3(2): 13.
81. Le Tourneau C, Delord JP, Goncalves A, Gavoille C, Dubot C, Isambert N et al. Molecularly targeted therapy based on tumour molecular profiling versus conventional therapy for advanced cancer (SHIVA): a multicentre, open-label, proof-of-concept, randomised, controlled phase 2 trial. Lancet Oncol 2015; 16(13): 1324-1334.
82. EUnetHTA. Methodological guidelines [internet]. [cited: 2015 May 11]. Available from: http://www.eunethta.eu/eunethta-guidelines
83. Baker SG, Kramer BS, Sargent DJ, Bonetti M. Biomarkers, subgroup evaluation, and clinical trial design. Discov Med 2012; 13(70): 187-192.
84. EPEMED. The European Personalised Medicine Association. Personalised Medicine in Europe – Enhancing patient access to pharmaceutical drug-diagnostic companion products. [internet]. [cited: 2015 Dec 19]. Available from: http://www.epemed.org/online/www/content2/104/107/910/pagecontent2/4339/791/ENG/EpemedWhitePaperNOV14.pdf.
85. Cohen JP, Felix AE. Personalized medicine's bottleneck: diagnostic test evidence and reimbursement. J Pers Med 2014; 4(2): 163-175.
86. Merlin T, Farah C, Schubert C, Mitchell A, Hiller JE, Ryan P. Assessing personalized medicines in Australia: a national framework for reviewing codependent technologies. Med Decis Making 2013; 33(3): 333-342.

87. Veenstra DL, Piper M, Haddow JE, Pauker SG, Klein R, Richards CS et al. Improving the efficiency and relevance of evidence-based recommendations in the era of whole-genome sequencing: an EGAPP methods update. Genet Med 2013; 15(1): 14-24.
88. Canadian Agency for Drugs and Technologies in Health. Evaluation frameworks for genetic tests [internet]. 2012 Jul [cited: 2015 Apr 26]. (Environmental Scans; volume 37). Available from: https://www.cadth.ca/sites/default/files/pdf/EvalFrameworksGeneticTesting-es-37_e.pdf.
89. Center for Medical Technology Policy. Evaluation of clinical validity and clinical utility of actionable molecular diagnostic tests in adult oncology [internet]. 2013 May 01 [cited: 2015 Apr 09]. Available from: http://www.cmtpnet.org/docs/resources/MDX_EGD.pdf.
90. Ludwig Boltzmann Institute. Procedural guidance for the systematic evaluation of biomarker tests: version 1.2 [internet]. 2014 Nov [cited: 2015 Apr 09]. (LBI Decision Support Documents; volume 77). Available from: http://eprints.hta.lbg.ac.at/1041/1/DSD_77.pdf.
91. Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group. The EGAPP initiative: lessons learned. Genet Med 2014; 16(3): 217-224.
92. Recommendations from the EGAPP Working Group: can tumor gene expression profiling improve outcomes in patients with breast cancer? Genet Med 2009; 11(1): 66-73.
93. Wong SQ, Li J, Tan AY, Vedururu R, Pang JM, Do H et al. Sequence artefacts in a prospective series of formalin-fixed tumours tested for mutations in hotspot regions by massively parallel sequencing. BMC Med Genomics 2014; 7: 23.
94. Van den Bulcke M SML, Salgado R, De Quecker E, De Schutter H, Waeytens A, Van Den Berghe P, Tejpar S, Van Houdt J, Van Laere S, Maes B, Hulstaert F. Next generation sequencing gene panels for targeted therapy in oncology and haemato-oncology. Health Technology Assessment (HTA) Brussels: Belgian Health Care Knowledge Centre (KCE). 2015. KCE Reports 240. [internet]. [cited: 2015 Dec 19]. Available from: https://kce.fgov.be/sites/default/files/page_documents/KCE_240_NGS_targeted_therapy_Report.pdf.
95. Redekop WK, Mladsi D. The faces of personalized medicine: a framework for understanding its meaning and scope. Value Health 2013; 16(6 Suppl): S4-9.
96. Febbo PG, Ladanyi M, Aldape KD, De Marzo AM, Hammond ME, Hayes DF et al. NCCN Task Force report: evaluating the clinical utility of tumor markers in oncology. J Natl Compr Canc Netw 2011; 9(Suppl 5): S1-32.
97. Griggs RC, Batshaw M, Dunkle M, Gopal-Srivastava R, Kaye E, Krischer J et al. Clinical research for rare disease: opportunities, challenges, and solutions. Mol Genet Metab 2009; 96(1): 20-26.
98. Goss CH, Mayer-Hamblett N, Kronmal RA, Ramsey BW. The cystic fibrosis therapeutics development network (CF TDN): a paradigm of a clinical trials network for genetic and orphan diseases. Adv Drug Deliv Rev 2002; 54(11): 1505-1528.
99. De Boeck K, Bulteel V, Tiddens H, Wagner T, Fajac I, Conway S et al. Guideline on the design and conduct of cystic fibrosis clinical trials: the European Cystic Fibrosis Society-Clinical Trials Network (ECFS-CTN). J Cyst Fibros 2011; 10 Suppl 2: S67-74.
100. Mandrekar SJ, Dahlberg SE, Simon R. Improving Clinical Trial Efficiency: Thinking outside the Box. Am Soc Clin Oncol Educ Book 2015: e141-147.
101. Biankin AV, Piantadosi S, Hollingsworth SJ. Patient-centric trials for therapeutic development in precision oncology. Nature 2015; 526(7573): 361-370.
102. Amir E, Seruga B, Martinez-Lopez J, Kwong R, Pandiella A, Tannock IF et al. Oncogenic targets, magnitude of benefit, and market pricing of antineoplastic drugs. J Clin Oncol 2011; 29(18): 2543-2549.
103. Innovative Medicines Initiative. The Innovative Medicines Initiative [internet]. [cited: 2016 Jan 13]. Available from: http://www.imi.europa.eu/.

104. ADAPT-SMART. Accelerated Development of Appropriate Patient Therapies: a Sustainable, Multi-stakeholder Approach from Research to Treatment-outcomes [internet]. [cited: 2015 Dec 30]. Available from: http://adaptsmart.eu/wp-content/uploads/2015/09/ProjectOverview-IMI2-ADAPTSMART.pdf.
105. Innovative Medicines Initiative. The GetReal Consortium [internet]. [cited: 2015 Dec 30]. Available from: http://www.imi-getreal.eu/.
106. Nordon C, Karcher H, Groenwold RHH, Ankarfeldt MZ, Pichler F, Chevrou-Severac H et al. The “Efficacy-Effectiveness Gap”: Historical Background and Current Conceptualization [in press]. Value in Health 2015 (in press).
107. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther 2001; 69(3): 89-95.
108. Memorandum for the evaluation of diagnostic measures. J Clin Chem Clin Biochem 1990; 28(12): 873-879.
109. Sargent DJ, Conley BA, Allegra C, Collette L. Clinical trial designs for predictive marker validation in cancer treatment trials. J Clin Oncol 2005; 23(9): 2020-2027.
110. Hayes DF, Trock B, Harris AL. Assessing the clinical impact of prognostic factors: when is "statistically significant" clinically useful? Breast Cancer Res Treat 1998; 52(1-3): 305-319.
111. Ginsburg GS, Kuderer NM. Comparative effectiveness research, genomics-enabled personalized medicine, and rapid learning health care: a common bond. J Clin Oncol 2012; 30(34): 4233-4242.