Supplemental material

This Online Appendix is divided in two parts, the first one presenting the Health Technology Assessment (HTA) methods, while the second links diagnostic tests with the emerging field of personalized medicine.

1. Health Technology Assessment Methods

Health technology assessment (HTA) aims to offer a coherent framework for guiding treatment decisions by optimizing health outcomes within the constraints of limited resources. HTA can be conducted for all types of interventions: diagnostic, surgical, medical, behavioral or complex, which can include pharmaceutical and medical devices [16, 34]. In the analysis, an intervention is compared to one or more alternative interventions called comparator(s).

NICE established the Diagnostics Assessment Programme (DAP) in 2009 to assess innovative medical diagnostic technologies.There are four main types of economics evaluation methods, according to how health outcomes are measured and valued: cost-benefit analysis (CBA), cost-effectiveness analysis (CEA); cost-utility analysis (CUA); and cost-minimization analysis (CMA).[1] The table below summarizes how the effects are measured in each method. The literature also sometimes references at the fifth method, called cost consequences analysis (CCA).

Table 1. The different types of economic evaluation (Adapted from [1])

Cost–benefit analysis (CBA)	Effects are measured in monetary units
Cost-effectiveness analysis (CEA)[i]	Effects are measured in any other unit of effect, *e.g.*, life-years gained, deaths averted, jobs saved, treatment responders, units of a patient-reported outcome measure, …
Cost-utility analysis (CUA)	Effects are measured in QALYs (or less commonly DALYs), which are utilities summed over time
Cost-minimisation analysis (CMA)	Effects are not considered, just costs alone

One QALY (or Quality Adjusted Life Year) is equal to one year in perfect health. It ranges from 1 (perfect health) to 0 (dead). For example, a person with chronic

[1] Often the terminology can be confusing, as these terms are used in various ways by different authors and do not always accurately describe the nature of the research (*e.g.*, [14]).

disease in which they experience utility of 0.5, will have ½ a QALY in one year, and 1 QALY over two years. QALYs are thus a measure of the total amount of (quality-adjusted) health experienced by an individual over a period of time; so even though utility is measured on a scale from 0 to 1, the QALYs can range from 0 to the length of follow-up (in years).

We are now going to cover each method sequentially.

Cost-benefit analysis (CBA).

Cost-benefit analysis (CBA) is a type of comparative evaluation of interventions, characterized by assigning monetary values to their outcomes. This allows both benefits (outcomes) and costs to be measured in the same units. The core principle of cost-benefit analysis (CBA) asserts that a health intervention is considered desirable if its benefits outweigh its costs. Alternatively, this can be expressed using the benefit-cost ratio: the intervention should be approved if the ratio exceeds one. When choosing between multiple interventions, in addition to meeting the basic requirement, the preferred option should have the highest benefit-cost ratio. This indicates that the selected intervention delivers greater benefits per monetary unit spent compared to alternative uses of the funds.

In the field of healthcare evaluation, there has been a general challenge in expressing health outcomes (such as survival) in monetary terms. As a result, CBA is not as widely used in health-economic evaluations, despite its popularity in other fields. Guidelines of several EU countries (Finland, Portugal, Russia, Spain and Sweden) include CBA as a possible type of analysis, while some others (Belgium, Hungary, Italy, Norway) state that CBA is not a recommended type of analysis or that it should be used as a complementary analysis. It has been argued that QALYs are not appropriate when the health condition is acute, and that in this case an alternative approach such as willingness to pay should be employed. For example, the guidelines for Sweden suggest that a CBA may be used in situations where it is challenging to apply QALYs (*e.g.*, when an intervention is associated with severe pain over a short period of time) [16].

Cost-effectiveness analysis (CEA).

In contrast to CBA, CEA compares the relative costs and effects of different interventions without assigning monetary values to the effects. Each intervention is represented as a point on a two-dimensional graph, with the net input cost (in monetary units) on the vertical axis and the effect (measured in life-years gained, deaths averted, jobs saved, *etc.*) on the horizontal axis. To compare different interventions, it is crucial to use the same measure of health effects consistently throughout the analysis. Additionally, this approach is limited to a single health effect measure per study.

The CEA is often conducted by comparing a new intervention to an existing one or the standard of care. In this context, the analysis measures the *incremental*

cost of the assessed procedure (compared to the standard) and the *incremental* health outcome. The costs considered include the price and associated medical costs of the new treatment and the standard of care. An important factor to consider is savings from preventing adverse effects associated with the standard of care, which is particularly relevant in personalized medicine (see next sections). Each of these measures (incremental cost and effect) can be either positive or negative, leading to the four quadrants illustrated in Figure 1.

The CEA is often expressed in terms of *incremental cost-effectiveness ratio* (ICER). This represents the net input costs (in monetary units) required to achieve each additional unit of health outcome:

$$\text{ICER} = \frac{\text{the change in costs}}{\text{the change in effect}}.$$

The ICER corresponds to the slope of the line connecting the studied procedure to the (0,0) point in Figure 1. Interventions can be ranked by ICER from lowest to highest, with the most cost-effective intervention having the lowest ICER. It is important to note that the ICER can be negative, for instance, when a better outcome is reached at a lower cost, as shown in Quadrant 4 of Figure 1.

In CEA, an intervention is considered cost-effective if the ICER is below a given value called the *cost-effectiveness threshold*. This threshold can be determined by the maximum willingness to pay for an additional unit of health outcome, for instance. The use of a cost-effectiveness threshold is highly controversial, with ongoing debate about what the threshold should represent and the appropriate methodology for determining its value. [60, 37]. We will come back to this point when we discuss the Cost Utility Approach below.

An alternative measure to the ICER frequently used in the literature is the incremental net monetary benefit (NMB). This measure provides a more straightforward interpretation, as it eliminates the ambiguity of what a positive or a negative ICER means [1].[2] The NMB is calculated by multiplying incremental effects by willingness-to-pay (or cost-effectiveness threshold) and then subtracting the incremental costs. A positive MNB indicates that the treatment is cost-effective at the given willingness-to-pay threshold and therefore, is considered worthwhile compared to the comparator.

We can now discuss the various examples shown in Figure 1. The cost-effectiveness plane in Figure 1 has 2 axes that represent the incremental difference between the intervention group(s) and the comparator group. Differences in effects are represented on the horizontal axis, while difference in costs are shown on the vertical axis. The plane is divided into four quadrants. Procedures in Quadrant 2 (such as Example F) are clearly dominated by the comparator, as they incur higher costs for worse outcomes. Likewise, all procedures in Quadrant 4 dominate the comparator, as they incur lower costs for

[2] Obviously, better effects at lower costs are always desirable, and worse effects at higher costs never are, while both situations exhibit a negative ICER.

better outcomes. To determine whether examples in Quadrants 1 are 3 should be chosen, we must assess how much society is willing to pay for each additional unit of health outcome. This is represented by the "Willingness-to-pay" (WTP), or cost-effectiveness threshold on the figure. All examples below this line are considered worth implementing as they have a positive MNB, while those above the line are not worth implementing and exhibit a negative MNB. Note that positive NMB examples include procedures where the added benefit justifies the cost (such as Example B), but also procedures that result in worse health outcomes if the cost reduction is substantial enough (such as Example D). Similarly, procedures deemed not cost-effective can either increase health outcomes (but at too high a cost as in Example A) or decrease health outcomes, if the cost reductions are too small (as in Example E).

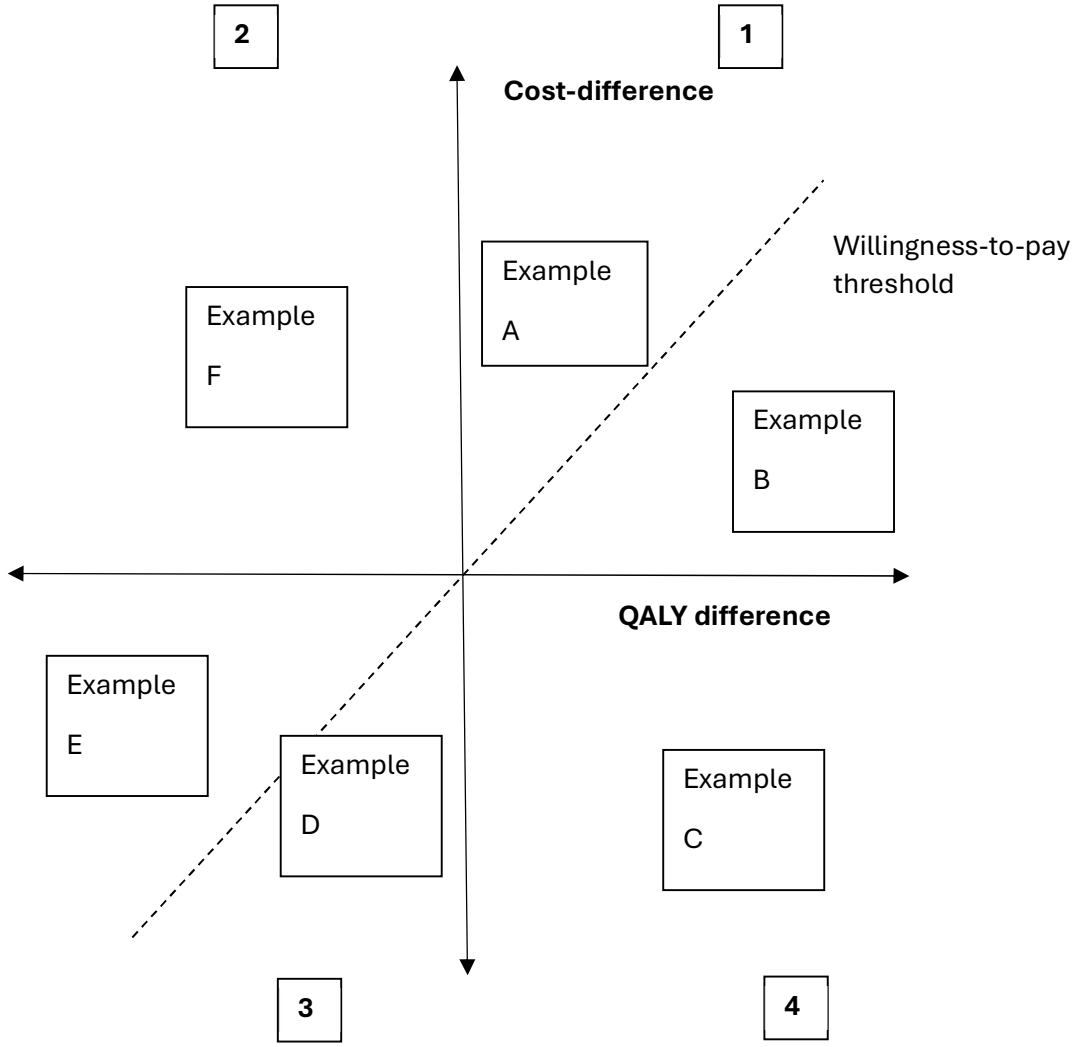

Figure 1: The cost-effectiveness plane.

Cost-utility analysis (CUA).

CUA is a form of CEA where the health outcomes are measured in terms of quality-adjusted life years (QALYs). Alternatively, CUAs sometimes use the

disability-adjusted life year (DALY) as another measure of disease burden. DALYs are expressed in terms of the number of years lost due to ill-health, disability or early death. They are calculated by summing the years lived with disability and the years of life lost. This measure combines life expectancy with the adjusted quality of life during a disease or disability for a population, making it as a societal measure of the disease burden. In contrast, QALYs tend to be more individual-focused, as they assess the quality and quantity of life for a specific person.

It is generally considered best practice to design a CUA, a single summary ratio that provides information on the incremental cost *per* QALY gained of a new technology compared to the current best practice. A CUA is a commonly used and recommended approach because it allows for comparisons across different health programs and policies using a common unit of measure [14]. However, a CUA is not always feasible or practical, especially when information about morbidity is not available, making it difficult to calculate QALYs.

The value assigned to each unit of QALY (the cost-effectiveness ratio) can be based on different theoretical or methodological approaches, including the opportunity cost approach (supply-side approach) and the willingness to pay (WTP) approach, which is a demand-side approach [3, 36, 50]. The opportunity cost/supply-side threshold reflects the opportunity cost of allocating health system resources to a particular use, representing the forgone benefits that could have been achieved if those same resources were directed toward other activities. This implies that the threshold value represents the shadow price of the budget constraint.

The second approach, the WTP or demand-side approach, is based on the amount individuals (*e.g.*, the general public) are willing to pay for health improvements, or, less commonly, the willingness to pay of patients themselves. The WTP approach may be traced back to attempts to link CEA with CBA and welfare economics. If QALYs meet certain conditions that make them represent utility, and there is a single societal WTP for a QALY, then CEA can be reformulated in a way that is equivalent to the CBA [42].

Cost-minimization analysis (CMA).

CMA does not consider health outcomes and focuses solely on costs. Therefore, CMA can be conducted when it has been demonstrated that there is no difference in the effect between an intervention and its relevant comparator (*e.g.*, [16]).

Cost consequences analysis (CCA).

CCA is an economic evaluation in which disaggregated costs and a range of outcomes are presented, allowing decision-makers to form their own judgements on the relevance and relative importance of these factors to their decision-making context [14]. CCAs have been recommended for complex interventions with

multiple effects, as well as for public health interventions that provide a range of health and non-health benefits, which are difficult to measure using a common unit (NICE, 2013). The outcomes in CCA are not limited to health outcomes like QALYs and can include other relevant measures, such as non-health factors relevant for decision-makers. CCA can be especially valuable to funders who are more focused on patient-oriented outcomes and intervention costs. It may also be useful in feasibility or pilot studies, where it is unclear which costs and outcomes will be most relevant to future definitive trials. Despite its potential, CCA is considered as an underused method of economic evaluation [26].

CEA/CUA as the preferred assessment methods

Since the seminal work [63], CEA have been widely published in US medical literature covering a diverse range of drugs, devices and medical procedures. However, they have received a mixed reception within the US medical healthcare [36].[3] In response to concerns about methodological standards, the US Public Health Service convened a panel of 13 non-governmental scientists and scholars in 1993 to review the state of CEA and provide recommendations for its use and conduct in health and medicine. The primary goals were to enhance the quality of CEA and promote comparability across studies, as many published studies described as CEA employed "surprisingly different methods" [49]. In 2016, the Second Panel on Cost-Effectiveness in Health and Medicine updated the work of the original panel by reflecting on the evolution of CEA and its perspectives (see, for example, [36]). While it continues to recommend QALY as the preferred societal outcome measure, it acknowledges certain caveats in its application.[4]

Most EU countries recommend using the CUA as the primary type of economic evaluation. In some cases, such as France, Ireland and the Netherlands, it is also recommended to accompany CUA results with a cost-effectiveness analysis (CEA) that uses costs per life-year gained (LYG) as the outcome measure to enhance usability [16]. Additionally, other guidelines such as those in Belgium, Norway and Sweden stipulate that the CUA should be always accompanied by a CEA with costs per life-years gained as the outcome measure [16].

There are several controversial issues in the CEA (see [19] for a detailed discussion) such as:

(i) How do we include the indirect time-related costs of treatment or benefits?
(ii) Should CEA include future medical costs incurred during years of life "extended" by a current medical intervention?

[3] There is a certain reluctance towards using CEA within the US legislative system such as the specific prohibition in the Affordable Care Act about the use of QALY and cost effectiveness in allocation decisions, or the congressional prohibition on funding CEA. Those restrictions have led to limited public funding for these analyses [49].

[4] The panel members specifically acknowledge the problem with aggregating the outcomes across individuals. For instance, it is unlikely that many would consider the saving of 1 minute of life among 525 600 people as being equivalent to the saving a year of life in a single individual [49].

(iii) Does the applied in CEA measures of effectiveness (*e.g*., incremental life years) discriminate against older patients?
(iv) Is it possible to find an "optimal" threshold for cost-effectiveness ratios?

Regarding the latter point, the values assigned by this threshold vary significantly between countries. Decision-makers may employ either implicit or explicit threshold values. Explicit threshold values indicated that decision-makers have formally adopted and publicly disclosed the threshold, using it as a basis for decisions on resource allocation. In contrast, implicit thresholds are neither official or publicly disclosed but can be inferred retrospectively by analysing decision-making patterns within a specific healthcare system.

For the National Health Service (NHS) in the UK, an explicit threshold has been set at £20,000/QALY, which can rise to £50,000 for life-threatening conditions [21, 9]. In the US, the cost-effective ratio typically ranges from $50,000/QALY to $150,000/QALY or even higher, depending on the individual or the disease. Hirth *et al*. [25] noted $50,000/QALY value was originally based on the estimated annual cost per QALY for the Medicare program for patients with chronic renal failure. In Sweden and the Netherlands, government authorities have recommended the thresholds of 500,000 SEK (approx. € 57,000) [50] and € 80,000 [6], respectively.

In 2011-2012 the French Haute Autorité de Santé (HAS) published a guideline on CEA, and a subsequent law mandated the use of CEA to determine pricing and reimbursement for new drugs and medical therapies [24]. However, until recently, no official cost-effectiveness threshold had been proposed to qualify ICERs. The study [57] proposes a method for estimating a value for a statistical QALY that can serve as reference for ICER thresholds in health assessment in France. The estimated reference values range from €147 093 to €201 398 per QALY, which are suggested as appropriate thresholds.

One of the major limitations to the widespread adoption of CEA is the lack of a globally accepted rule for setting relevant cost-effectiveness thresholds. The best-known recommendation is the WHO rule, which considers an intervention to be cost-effective if it results in a healthy year gained at a cost of less than three times the GDP *per* capita. In the past decade, several studies have challenged this rule by showing that cost-effectiveness thresholds are often substantially lower (less than 1 GDP per capita) in different countries, particularly in low- and middle-income countries. These studies argue that higher thresholds may lead to an unsustainable increase in health expenditure *per* capita.

A recent study [44] presents a conceptual framework for estimating cost-effectiveness thresholds and empirically derives them for 174 countries using World Bank data on country specific health expenditures and health outcomes from 2010 to 2019. The findings suggest that cost-effectiveness thresholds *per* QALY should range between US$87 for the Democratic Republic of Congo and $95 958 for the US. The study shows that that thresholds are less than 0.5 GDP *per* capita in 96% of low-income countries, 76% of lower-middle countries, 31% of upper-middle countries, and 26% of high-income countries. Overall, cost-

effectiveness thresholds *per* QALY are less than 1 GDP *per* capita in 168 (97%) of the 174 countries. For cost-effectiveness thresholds *per* life-year gained, the range is from $78 to 80 529, or between 0.12 and 1.24 GDP *per* capita with thresholds being less than 1 GDP *per* capita in 171 (98%) of the countries.

Recently, there has been a growing emphasis on societal perspective in healthcare, reflecting the viewpoint of decision-makers who aim to allocate health resources optimally across the entire population. In this context, when evaluating interventions, both health and non-health consequences should be considered, including effects on economic productivity, education, social services, criminal justice, housing or environment [36]. CBA is argued to be a more relevant economic evaluation method (*e.g.*, [7]) because it ensures that outputs are valued in monetary terms, making them directly comparable to costs, and helping to determine whether the expenditure is socially worthwhile. Additionally, CBA provides a broader social perspective by incorporating effects on everyone affected by an intervention both directly and indirectly. In contrast, CEA can identify interventions that are cost- effective but not necessarily socially worthwhile and it does not account for externalities. Furthermore, CUA uses a single threshold price that does not take into account the preference of patients who benefit from the interventions.

Roberts [49] highlights that the most significant issue with CEA is not its accuracy or consistency, but rather the misunderstanding and subsequent prohibitions of its use to improve resource allocation in US healthcare. According to Roberts, "trade-offs are unavoidable in the allocation of resources, and the methods of CEA render those trade-offs explicit and debatable," which is often a point of contention in healthcare decision-making.

To conclude this section, we note that, although CEA/CUA are the assessment methods recommended by many health authorities across the world, their economic foundations are still being assessed by (theoretical) economists [20, 8, 33]. For instance, Garber [19] demonstrates that CEA can provide an appropriate tool for choosing among health interventions within a standard utility maximization framework. Meltzer *et al*. [33] consider theoretical grounds of CEA in constrained optimization, highlighting issues such as objectives to be maximized, constraints to be considered, resources consumed, and opportunities foregone. Even when the objective is to maximize health, there are multiple questions to consider as how to measure and combine effects on survival and health-related quality of life; how to measure costs; and how to treat effects that are uncertain or that occur over time. Other topics covered include theoretical issues that arise in the QALY model and their links to individual utility.

2. Diagnostic and companion tests

It is important to link diagnostic tests with the emerging field of personalized medicine even though there is no consensus on the definition of personalized medicine across countries and in academic literature. At one extreme, all

medicine can be dubbed as "personalized" since medicine has always been concerned with the individual needs, with an aim of diagnosing the specific ailment of a given individual, and the best treatment to be applied in that particular case. At the other extreme, the narrowest definition of personalized medicine requires the creation of drugs or medical devices that are unique to a patient.[5] The typology proposed in [58] is helpful in understanding this variance. They propose the patient therapeutic continuum illustrated on Figure 2 below. Most medicines have been prescribed empirically, meaning that either they work for all patients, or, if response rates are variable, there is no way to identify patients who are likely to respond well to a specific treatment. Advances in understanding the mechanisms underlying both the diseases and the drug responses have allowed to better match patients with treatments. At the extreme, the treatment is individualized, custom produced using the patient's own fluid, cells or tissue to seed production. This is what they call "individual medicine". In between these extremes, there is a growing number of cases where some marker exists which indicates whether a given patient is likely to react to a therapy. They propose to call these situations "stratified medicine" (rather than the more ambiguous term of personalized medicine). More precisely, "in stratified medicine, a patient can be found to be similar to a cohort that has historically exhibited a differential therapeutic response using a biomarker that has been correlated to that differential response" (p1).

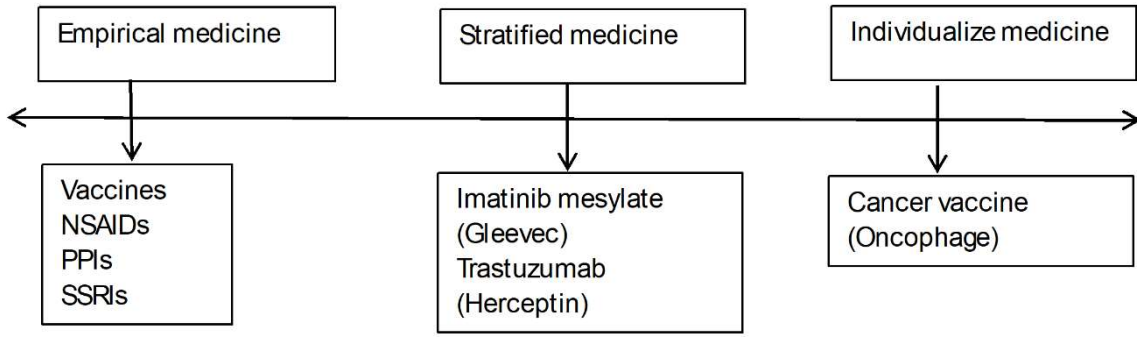

Figure 2. The patient therapeutic continuum. Individualized medecine, such as cancer vaccines that are based on a particular patient's tumour, represents one end of a continuum of patient therapy. Empirical medicine is at the otherend of this continuum. Some agents work for almost all relevant patients, such as non-steroidal anti-inflammatory drugs (NSAIDs), whereas other may only work for a subset of patients but no method is available to identify these patients, such as with antidepressants. In between is the field of stratified medicine, in which a patient can be found to be like a cohort that has historically showed a differential therapeutic response to a particular therapy using a clinical biomarker that has been correlated to that differential response. For example, the anticancer drug trastuzumab (Herceptin) shows

[5] The cancer vaccine Oncophage is an example of individualized medicine. To produce this vaccine, tumor cells are taken from a patient during surgery. A heat-stock protein and its associated peptides, which represent a unique 'signature' of the patient cancer, are then isolated from the tumor cells and formulated into a vaccine for administration after the recovery of the patient from surgery. This vaccine, which is only suitable for the patient from whom it is derived, stimulates an immune response that attacks tumor cells remaining after the surgery [58].

superior efficacy in breast cancer patients with HER2/neu-positive cancer. PPIs, proton-pump inhibitors; SSRIs selective serotonin-reuptake inhibitors (see Trusheim *et al.*, [58]).

Equipped with this terminology, Figure 3 allows us to make the link with diagnostic tests, as they appeared both within the realm of empirical medicine (to confirm a diagnostic) and of stratified medicine (to test for the treatment response). We will cover both tests in this survey.[6] Note that the treatment offered to the patient may or may not be individualized, and that we are not going to cover susceptibility tests, which allow to assess the probability that an individual may develop a disease in the future (see [4], for instance). Targeted therapy is the combination of test/treatment.

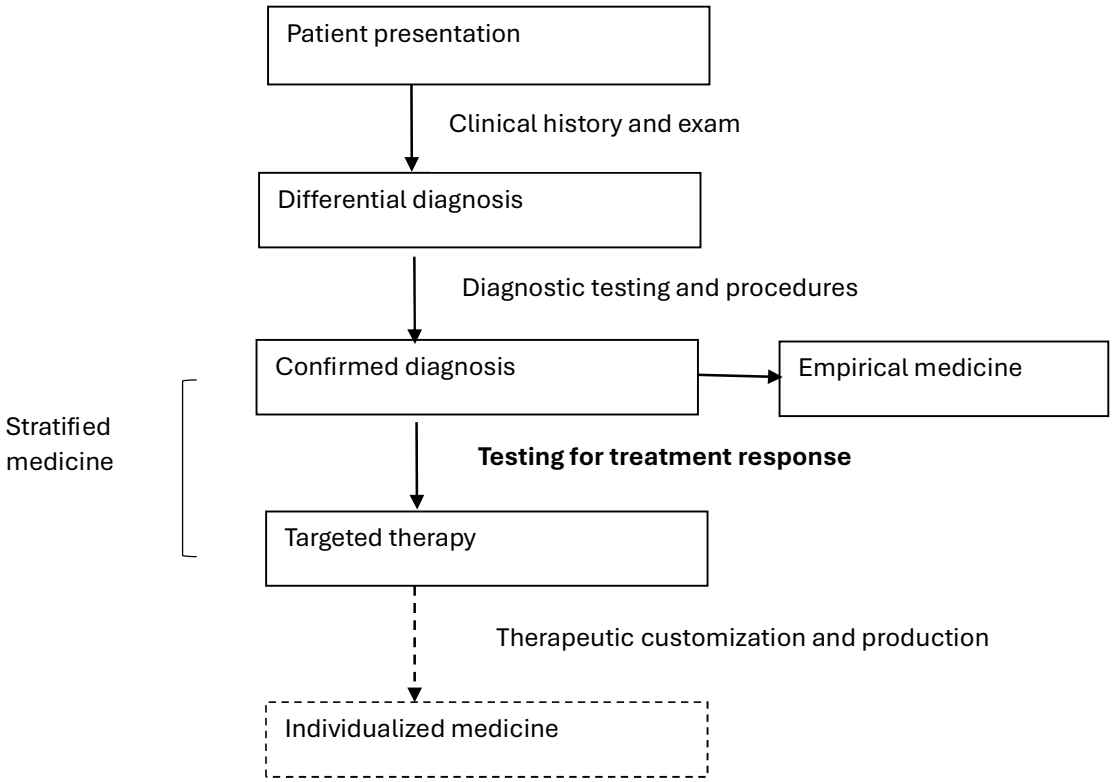

Figure 3. Stratified medicine in the clinical context. In empirical medicine, a differential diagnosis is made on the basis of patient history and physical assessment and following diagnosis confirmation from laboratory tests and clinical observation, a therapy is prescribed. Stratified medicine involves a further step in which a clinical biomarker is evaluated to associate a patient with a specific therapy. Ending this further, individualized medicine involves the customized production of the therapy (for example, using the patient's own cells), [58].

[6] See the Annex for more on how to link the various types of tests to several definitions of personalized medicine used in practice (especially in the U.S.A).