

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/208980>

Please be advised that this information was generated on 2020-09-09 and may be subject to change.

Anouck Kluytmans, Jaap Deinum, Kevin Jenniskens, Antonius Eduard van Herwaarden, Jolein Gloerich, Alain J. van Gool, Gert Jan van der Wilt and Janneke P.C. Grutters*

Clinical biomarker innovation: when is it worthwhile?

<https://doi.org/10.1515/cclm-2019-0098>

Received January 28, 2019; accepted June 2, 2019; previously published online July 9, 2019

Abstract

Background: Choosing which biomarker tests to select for further research and development is not only a matter of diagnostic accuracy, but also of the clinical and monetary benefits downstream. Early health economic modeling provides tools to assess the potential effects of biomarker innovation and support decision-making.

Methods: We applied early health economic modeling to the case of diagnosing primary aldosteronism in patients with resistant hypertension. We simulated a cohort of patients using a Markov cohort state-transition model. Using the headroom method, we compared the currently used aldosterone-to-renin ratio to a hypothetical new test with perfect diagnostic properties to determine the headroom based on quality-adjusted life-years (QALYs) and costs, followed by threshold analyses to determine the minimal diagnostic accuracy for a cost-effective product.

Results: Our model indicated that a perfect diagnostic test would yield 0.027 QALYs and increase costs by €43

per patient. At a cost-effectiveness threshold of €20,000 per QALY, the maximum price for this perfect test to be cost-effective is €498 (95% confidence interval [CI]: €275–€808). The value of the perfect test was most strongly influenced by the sensitivity of the current biomarker test. Threshold analysis showed the novel test needs a sensitivity of at least 0.9 and a specificity of at least 0.7 to be cost-effective.

Conclusions: Our model-based approach evaluated the added value of a clinical biomarker innovation, prior to extensive investment in development, clinical studies and implementation. We conclude that early health economic modeling can be a valuable tool when prioritizing biomarker innovations in the laboratory.

Keywords: biomarker innovation; diagnostic innovation; early health technology assessment; liquid chromatography mass spectrometry; primary aldosteronism.

Introduction

In laboratory medicine, the past decades have been an era of health technology innovation [1]. Inventions such as genome sequencing, liquid chromatography mass spectrometry (LC-MS) and random-access analyzers have considerably improved the personalized diagnosis and treatment of numerous medical conditions. However undeniable the advantages of health technology innovation are [2], the advances have coincided with a steep rise in health care expenditure [3, 4]. Studies predict that health care costs will consume around one-third of family incomes by the year 2040 [5, 6], which would challenge the accessibility of health care and the social solidarity to collectively cover costs [7]. In addition, it is not always clear whether innovations provide ‘true value for money’ [8]. The need for decision-making in innovation is therefore evident.

This is particularly true in the field of molecular biomarkers. Considerable progress in the fields of next generation sequencing and MS have enabled a detailed analysis of DNA, proteins and metabolites, yielding a wealth of biomarker candidates. However, many of these

*Corresponding author: Janneke P.C. Grutters, Radboud Institute for Health Sciences, Department for Health Evidence (Route 133), Radboud University Medical Center, PO Box 9101, 6500 HB Nijmegen, The Netherlands, E-mail: Janneke.Grutters@radboudumc.nl

Anouck Kluytmans and Gert Jan van der Wilt: Radboud Institute for Health Sciences, Department for Health Evidence (Route 133), Radboud University Medical Center, Nijmegen, The Netherlands

Jaap Deinum: Radboud Institute for Health Sciences, Department of Internal Medicine (Route 463), Radboud University Medical Center, Nijmegen, The Netherlands; and Department of Internal Medicine III, University Hospital Carl Gustav Carus at the Technische Universität Dresden, Dresden, Germany

Kevin Jenniskens: Julius Center for Health Sciences and Primary Care (Room van Geuns 5.18, Internal Mail HP GEUNS 5.02), UMC Utrecht, Utrecht, The Netherlands

Antonius Eduard van Herwaarden: Radboud Institute for Molecular Life Sciences, Department of Laboratory Medicine, Radboud University Medical Center, Nijmegen, The Netherlands

Jolein Gloerich and Alain J. van Gool: Radboud Institute for Molecular Life Sciences, Translational Metabolic Laboratory, Department of Laboratory Medicine, Radboud University Medical Center, Nijmegen, The Netherlands

candidate biomarkers do not progress to clinical application and the innovation gap between research and application is widening [9]. Efficient ways to calculate the added clinical value of biomarkers prior to their development could prove useful in the process of biomarker prioritization.

Health economic modeling provides tools to make such calculations, thereby supporting decision-making on which biomarker innovations to pursue and which to lay aside. The goal of health economic modeling is to provide insight in the costs and clinical benefits of a (novel) intervention, and compare them to standard of care. Its methods are increasingly being applied at the early stages of technology development to inform decisions regarding the further development of innovative ideas [10]. Even when there is no data on the diagnostic accuracy or health impact of the innovation, health economic modeling methods such as headroom analysis offer estimates of its potential value by assessing the room for improvement in current care practices [11].

Specific applications of novel technologies are often proposed by biomarker scientists or clinicians who tend to be unfamiliar with health economic modeling concepts. Using the example of diagnosing primary aldosteronism (PA), we will illustrate how such early modeling methods can help clinicians and diagnostic test developers in the laboratory make evidence-informed decisions on whether or not to pursue a new diagnostic innovation.

In the field of PA, new concepts are being considered out of concern for the current diagnostic method, which centers around the aldosterone-to-renin ratio (ARR) test. First, renin's biological variability is a cause for concern. It is known that renin measurements vary depending on time of day, posture during sampling and patient characteristics [12]. The different recommended cut-off values for the ARR in the Endocrine Society Clinical Practice Guideline introduce further variability between laboratories [13]. Second, for reliable use of the ARR as a screening test – the Japan Endocrine Society even recommends screening all hypertensive patients for PA – the test's sensitivity should be very high to avoid false negatives [14]. Doubts regarding the ARR's suitability are evident from the many studies that investigated its diagnostic properties, indicating a repeated desire to evaluate the use of ARR and its cut-off value [15]. Third, optimization of the PA diagnosis is an active field of research, implying concerns regarding current diagnostic methods. This is illustrated by ENSAT-HT and PRIMAL, two ongoing clinical trials [16, 17], and the Berge et al. and Rehan et al. studies investigating the use of LC-MS for the analysis of aldosterone [18, 19]. The analyses conducted by Berge et al.

and Rehan et al. indicated that targeted MS could (semi-) quantitate renin and related isoforms, providing the basis for an innovative test for PA diagnosis with higher specificity and selectivity.

We believe novel tests for PA are highly relevant for an early health economic assessment to estimate their potential value prior to (further) development and clinical validation. Through decision analytic modeling we analyzed the room for improvement in current PA testing and determined what properties a new test should have (in terms of sensitivity, specificity and costs) for it to become cost-effective.

Materials and methods

Model overview and validation

To simulate the expected costs and health benefits of various diagnostic strategies, we constructed a health economic model (see the Glossary in Figure 1) that consists of a decision tree for the diagnostic part (Figure 2) and a cohort state-transition model (see the Glossary in Figure 1) to simulate the long-term health and economic consequences (Supplementary Material, Appendix Figure 1). The model was developed based on literature and input of various experts in the areas of hypertension, laboratory medicine and technology assessment. The model was constructed and analyzed using Microsoft Office Excel 2016.

Target population

The target population consists of patients diagnosed with resistant hypertension who are suspected of PA. We assume that prior attempts to treat their hypertension with antihypertensive medication have failed and that patients are being monitored by a hypertension specialist. For this population, the aim of diagnosis is not only to treat hypertension but to treat hyperaldosteronism and its adverse cardiovascular effects. The starting age of the patient cohort in our model is 40 years old, and a cohort-state transition model was created that reflects a lifelong perspective on costs and effects.

Model structure

In the decision tree, the patients are divided into those with and without PA. Patients with PA can have a (false) negative or (true) positive test result and those without PA a (true) negative or (false) positive test result. Every positive test result of the current (ARR) and new test is followed by a confirmatory saline-infusion (SI) test (which is considered one of four recommended reference tests for PA) [13]. In the case of a positive SI test, a CT scan plus adrenal

Term	Meaning
Decision model	A simplified representation of a certain health care setting whose purpose is to inform health-related decisions by simulating the health effects and costs incurred following a certain health care strategy (1).
Headroom	The maximum room for improvement within a certain health care strategy. Headroom is calculated by comparing a care strategy in current care to a hypothetical, perfect care strategy. The difference between the two indicates the room for improvement. Headroom consists of the potential gain in health effects, monetized at a cost-effectiveness threshold, plus or minus any difference in costs between the current and perfect care strategies. Headroom provides an indication of the maximum price at which a perfect innovation could be considered cost-effective.
Cohort state-transition model	A specific type of decision model, suited for simulating periodic outcomes in a hypothetical patient cohort. A cohort state-transition model consists of a hypothetical cohort of patients that is distributed across health states (e.g. healthy, ill) that are associated with a certain health effect and certain costs. At every periodic cycle, patients in the cohort may remain in their original health state or transition to another health state. The total health effects and costs of the cohort are recorded at each cycle, until the time horizon of the model has been met. The results of different care strategies may then be compared to determine which strategy yields the most favorable average cost-benefit balance (2).
Probabilistic sensitivity analysis (PSA)	One source of uncertainty in decision modeling stems from uncertainty regarding the input parameters of a model. The impact of such parameter uncertainty on model results can be investigated using probabilistic sensitivity analysis (PSA). In a PSA, the decision model is re-run with randomly sampled values from distributions of the parameters whose exact value is uncertain. Taken together, all runs of a PSA (e.g. 10,000) reveal the uncertainty in model results as a consequence of uncertainty in input parameter values. PSA can be used to provide simulated 95% confidence intervals of the results (3).
Quality-adjusted life-years	Quality-adjusted life-years (QALYs) are the product of the number of life-years spent in a certain health state and the utility of one's health (or quality of life) during that time. As a result, 10 QALYs may be interpreted as both 10 life-years spent in perfect health, and 20 life-years spent with a utility of 0.5 (4).
Utility	Utilities express the societal preference for being in a certain health state. Utilities are typically expressed on a scale from 0 (death) to 1 (perfect health).

Figure 1: Glossary containing key terminology pertaining to early health economic modeling.

vein sampling (AVS) is used to subtype PA. Following the diagnosis, patients receive treatment: adrenalectomy and/or medication for those patients with a confirmed PA and continuation of antihypertensive medication for those not diagnosed with PA. Following treatment, patients have a probability to either become normotensive, with or without continued use of their medication, or remain hypertensive.

The long-term part of the model is shown in Supplementary Material, Appendix Figure 1. Note that reducing the risk of cardiovascular events is an important aim in the treatment of hyperaldosteronism. All patients are at risk of developing cardiovascular events (i.e. stroke, coronary artery disease, atrial fibrillation or heart failure).

The risk is lowest for the normotensive group, higher for the hypertensive group, and highest for the PA groups [20]. Once affected by a cardiovascular complication, patients move to a post-complication health state, which is associated with higher health care costs, a lower quality of life and higher risk of death compared to their starting health states. At every cycle, patients in the cohort may die from cardiovascular or other causes.

The decision tree and state-transition model are equally structured for the two strategies (ARR test and hypothetical test). Most parameter values remain the same; only the sensitivity and specificity is different; for the hypothetical test they are both 1, meaning there are no false positives or false negatives (see Model input). The

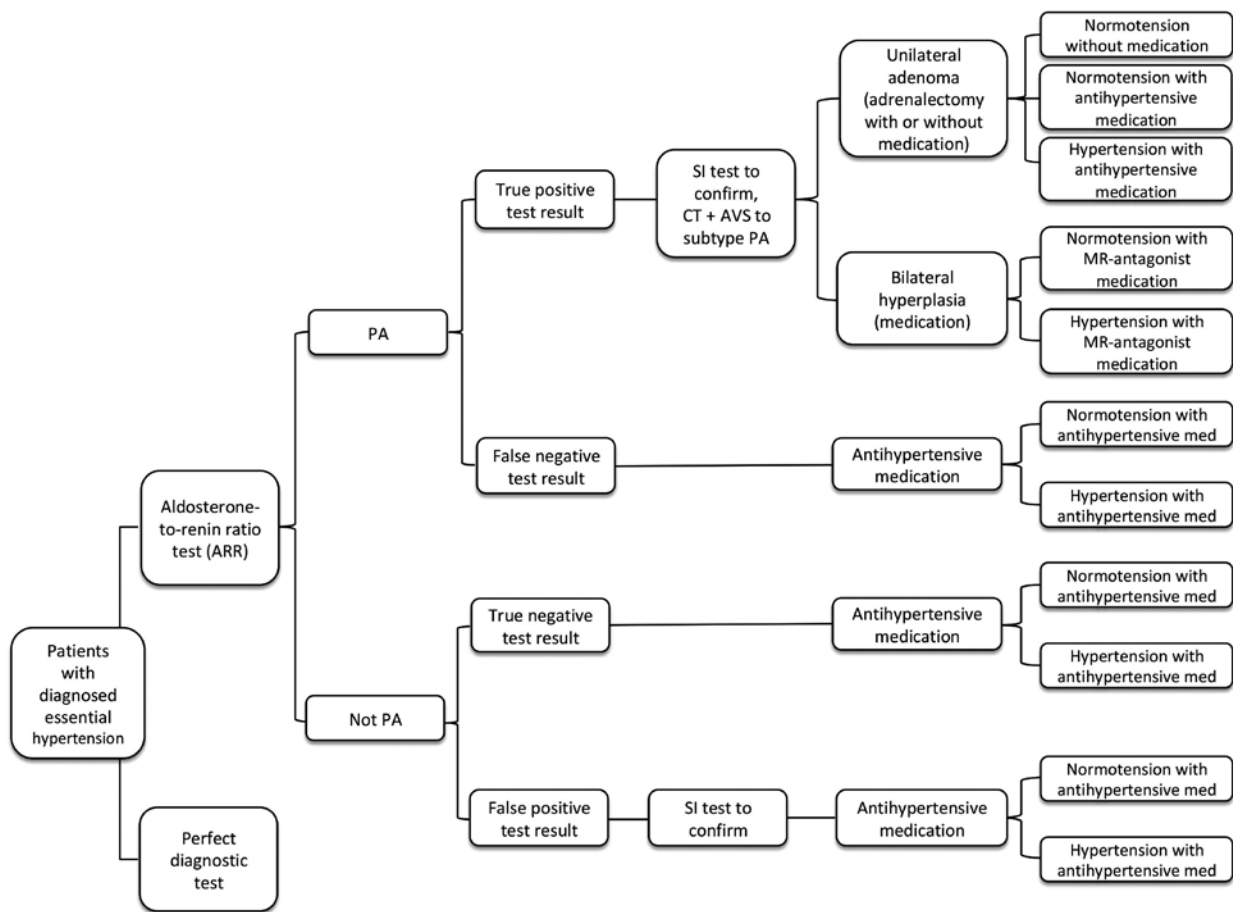


Figure 2: Decision tree for the diagnosis of primary aldosteronism (PA), comparing the current ARR test to a hypothetical diagnostic test, leading to one of five health states that form the starting point of the long-term part of the model that simulates yearly costs and health effects (see Supplementary Material, Appendix Figure 1).

model hereby simulates a scenario of perfect diagnostics; note that the probability of normotension after treatment remains the same.

Assumptions

We assume that SI tests are always conducted to confirm the PA diagnosis and that they have a sensitivity and specificity of 1. The assumed effectiveness of antihypertensive medication is based on a 50% treatment adherence rate [21]. Due to scarce data on long-term risk estimates for developing hypertension in this population, we assume that once patients are assigned to the normotension or hypertension health state, they will remain in that health state until they develop cardiovascular events or die. Regarding cardiovascular events, patients could only develop one complication during their remaining life-span due to scarce data on the probability and consequences of developing multiple cardiovascular events within this specific population.

Model input

All model inputs and their sources are listed in Supplementary Material, Appendix Table 1.

Transition probabilities

In the diagnostic part of our model, we assume a 15% prevalence of PA as reported by Jansen et al., which fits with the literature on PA in patients with essential hypertension [22, 23]. Because prevalence estimates differ across healthcare settings, we varied this parameter in a sensitivity analysis to investigate the relationship between PA prevalence and the headroom for a diagnostic test (see Analyses) [24]. For the ARR test, we assume a sensitivity of 0.89 and a specificity of 0.96 based on a meta-analysis conducted by Li et al. [15]. For confirmed PA patients we assume a probability of 0.47 of having an aldosterone-producing adenoma that qualifies for surgery from Shah and Deshpande [25]. The outcomes of unilateral adrenalectomy are based on a recent study by Williams et al., who reported that 37% of their patients became normotensive without further need for medication and another 47% following antihypertensive medication, while 16% remained hypertensive even with antihypertensive medication [26]. For all patients that did not have an aldosterone-producing adenoma suitable for surgery, we assume that the probability to become either normotensive or remain hypertensive is equal to the patients' adherence to their medication regimen, reflecting the notion that the medication itself is effective. Azizi et al. estimated medication adherence in this subgroup to be 0.5 [21].

In the long-term part of our model, we based the probability for our patients to develop one of four cardiovascular events on publicly available data on normotension and hypertension issued by the Dutch Ministry of Health. For the PA patients, odds ratios published by Monticone et al. were used to calculate their increased risk of cardiovascular events compared to hypertensive patients [20]. The probability of dying from cardiovascular causes was derived from various studies, where we distinguished between the year of the event and all subsequent years [27–30]. The probability of dying from other causes was based on publicly available data from Statistics Netherlands [31].

Utilities

To model health effects in terms of QALYs (see the Glossary in Figure 1), we retrieved utilities (see the Glossary in Figure 1) for every health state from the relevant literature. Recently it has been argued that, despite their limitations, QALYs do provide a useful metric for quantifying and comparing health effects associated with different health technologies (especially when evaluating multiple cardiovascular outcomes, as is the case in our study) [32]. QALYs were discounted at a rate of 1.5% annually to reflect the Dutch time preference for health effects [33].

Costs

In the model, costs were determined from a societal perspective and based on Dutch sources, comprising one-time costs of the various tests, the yearly costs of medication, and those associated with the four cardiovascular events. All costs were converted to 2016 prices and discounted at a rate of 4% [33].

Analyses

Based on the transition probabilities pertaining to the PA diagnosis, the cohort data were divided across health states (see Figure 2 and Supplementary Material, Appendix Figure 1). Using yearly cycles, we calculated the costs and QALYs proportional to either remaining in the starting health states or a progression to cardiovascular events or death. Next, we calculated the total costs and QALYs per patient at 10 years, 20 years and lifetime, and compared the costs and QALYs of the current test with those of the hypothetical test. We used these per-patient totals in our subsequent headroom [11], sensitivity and threshold analyses [34].

Our headroom analysis (see the Glossary in Figure 1) estimates the (financial) room for improvement in the current diagnostic pathway of patients with PA by comparing the ARR to a hypothetical, perfectly accurate test with a sensitivity and specificity of 1. The difference in QALYs between the current and the perfect diagnostic procedure is known as the effectiveness gap, or the health effects foregone due to diagnostic tests with imperfect accuracy in the current pathway. Headroom, then, is calculated by monetizing the effectiveness gap using a cost-effectiveness threshold – in our case €20,000 per QALY, following the Dutch recommendations for diagnostic and preventive interventions – and adding

or subtracting any differences in costs between the current and perfect diagnostic strategy. For example, an effectiveness gap of 0.1 QALY with an additional cost of €500 gives a headroom of $0.1 \times €20,000 - €500 = €1500$. The resulting headroom can be interpreted as the maximum price at which a perfectly accurate PA test could be considered cost-effective. Given the lifelong consequences of cardiovascular events a lifetime horizon on costs and health benefits is most appropriate. We will also report the main results at 10 and 20 years after the diagnosis.

Many of the parameters in our model suffer from a degree of uncertainty. A PSA (see the Glossary in Figure 1) was conducted using 10,000 samples from beta distributions [34]. Standard errors for probabilistically varied parameters are listed in Supplementary Material, Appendix Table 1. The results of the PSA are reported using the percentile method, yielding ranges similar to 95% confidence intervals (CIs). Furthermore, a univariate sensitivity analysis was conducted to investigate the individual impact of the following parameters: the sensitivity and specificity of the ARR (varied from 0.22 to 1 and 0.56 to 1, respectively, based on the ranges reported in a meta-analysis by Li et al. [15]), the prevalence of PA (varied from 5% to 25%), and the cost-effectiveness threshold (varied from €20,000 to €80,000). The ARR sensitivity and specificity values were selected because of the heterogeneous figures reported in the literature and the PA prevalence rate and cost-effectiveness threshold based on potential differences across health care settings.

A novel test for PA is unlikely to be perfect and costless, hence we also performed a multivariate threshold analysis to investigate trade-offs between a novel test's sensitivity, specificity and cost. In this analysis both sensitivity and specificity were varied from 0 to 1 in steps of 0.1.

Ethical approval

Ethical approval: the conducted research is not related to either human or animals use.

Results

Headroom analysis

The average QALYs per patient of the ARR strategy with a lifetime horizon were 21.249 vs. 21.276 for the perfect diagnostic strategy (Table 1). The average lifetime costs per patient were €17,779 vs. €17,822. Compared to the current test, a perfect test would yield 0.027 QALYs – or nearly 10 days in perfect health – at a cost increase of €43 per patient. At our cost-effectiveness threshold of €20,000 per QALY gained, the resulting headroom is $0.027 \times €20,000 - €43$, or €498 (95% CI: €275–€808). When shorter time horizons of 10 and 20 years are considered, the potential headroom decreases, predominantly as a result of less QALYs that are to be gained.

Table 1: Results of the headroom analysis as averages per patient.

Horizon	Current test		Perfect test		Increments		Headroom (95% CI)
	QALYs	Costs	QALYs	Costs	QALYs	Costs	
10 years	7.721	€6867	7.724	€6915	0.003	€48	€19 (€9–€62)
20 years	13.692	€12,245	13.703	€12,280	0.011	€35	€181 (€90–€320)
Lifetime	21.249	€17,779	21.276	€17,822	0.027	€43	€498 (€275–€808)

Headroom was calculated with a cost-effectiveness threshold of €20,000 per QALY.

**Figure 3:** Tornado plot for the univariate sensitivity analyses, lifetime perspective on costs and effects.

Parameters of interest are listed on the y-axis, ranges specified in brackets. The x-axis contains the headroom values, with bars showing the result across the range of each parameter.

Univariate sensitivity analyses

The lifetime headroom results were further investigated whereby key parameters of interests were varied one by one (Figure 3). As an example, when the sensitivity of the current ARR test was changed from 0.89 to a perfect sensitivity of 1, the headroom was near 0, meaning the potential value of even a perfect diagnostic test would be limited. When an ARR test sensitivity of 0.22 was used (lowest point estimate from studies in the meta-analysis), headroom increased to €3500.

The sensitivity of the current test had the largest impact on the headroom estimate, followed by the cost-effectiveness threshold. PA prevalence had some impact on the headroom, indicating that an improved test would be most valuable in care settings with relatively speaking many PA patients. The impact of the specificity of the current test on the headroom results was negligible.

Multivariate threshold analysis

The results of the multivariate threshold analysis, in which the sensitivity and specificity of the novel test were varied simultaneously, are presented in Figure 4. To exemplify, a novel PA test for hypertensive patients with a sensitivity

of 0.90 and a specificity of 1 may cost up to €51 more than the current test and still be considered cost-effective. Note that for the current test we assumed a sensitivity of 0.89 and a specificity of 0.96 [15]. Figure 4 shows that a novel test for diagnosing PA may be worth the extra expenditure when its sensitivity is at least 0.90 and its specificity is at least 0.70.

Discussion

We applied headroom and threshold analyses to illustrate how early health economic modeling can support decisions regarding biomarker innovations. Focusing on a hypothetical novel test as an alternative to the currently used ARR test in the diagnostic pathway of PA, we found that there was some room for improvement. Additional research is needed to determine whether a headroom of roughly €500 per patient provides realistic opportunities for the R&D of innovations such as LC-MS techniques. This will further depend on the costs of research, development, implementation and maintenance/calibration, sharing of equipment between centers, and many other factors.

Headroom analyses can similarly be applied to other indications and target populations to obtain an estimate of the room for improvement and the potential value of

Specificity of the new test	Sensitivity of the new test										
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0	–€ 4.122	–€ 3.675	–€ 3.228	–€ 2.781	–€ 2.334	–€ 1.887	–€ 1.440	–€ 993	–€ 546	–€ 99	€ 348
0.1	–€ 4.107	–€ 3.660	–€ 3.213	–€ 2.766	–€ 2.319	–€ 1.872	–€ 1.425	–€ 978	–€ 531	–€ 84	€ 363
0.2	–€ 4.092	–€ 3.645	–€ 3.198	–€ 2.751	–€ 2.304	–€ 1.857	–€ 1.410	–€ 963	–€ 516	–€ 69	€ 378
0.3	–€ 4.077	–€ 3.630	–€ 3.183	–€ 2.736	–€ 2.289	–€ 1.842	–€ 1.395	–€ 948	–€ 501	–€ 54	€ 393
0.4	–€ 4.062	–€ 3.615	–€ 3.168	–€ 2.721	–€ 2.274	–€ 1.827	–€ 1.380	–€ 933	–€ 486	–€ 39	€ 408
0.5	–€ 4.047	–€ 3.600	–€ 3.153	–€ 2.706	–€ 2.259	–€ 1.812	–€ 1.365	–€ 918	–€ 471	–€ 24	€ 423
0.6	–€ 4.032	–€ 3.585	–€ 3.138	–€ 2.691	–€ 2.244	–€ 1.797	–€ 1.350	–€ 903	–€ 456	–€ 9	€ 438
0.7	–€ 4.017	–€ 3.570	–€ 3.123	–€ 2.676	–€ 2.229	–€ 1.782	–€ 1.335	–€ 888	–€ 441	€ 6	€ 453
0.8	–€ 4.002	–€ 3.555	–€ 3.108	–€ 2.661	–€ 2.214	–€ 1.767	–€ 1.320	–€ 873	–€ 426	€ 21	€ 468
0.9	–€ 3.987	–€ 3.540	–€ 3.093	–€ 2.646	–€ 2.199	–€ 1.752	–€ 1.305	–€ 858	–€ 411	€ 36	€ 483
1	–€ 3.972	–€ 3.525	–€ 3.078	–€ 2.631	–€ 2.184	–€ 1.737	–€ 1.290	–€ 843	–€ 396	€ 51	€ 498

Figure 4: Results of the multivariate threshold analysis.

The Figure contains model results indicating the monetary loss or gains associated with a hypothetical new test for different specificity (y-axis) and sensitivity (x-axis) combinations. The red area indicates a net loss for the new test and the green area its additional value compared to the current test.

specific applications of new health technologies – diagnostic or otherwise – before committing resources to their further development and clinical studies. Our analyses may be especially useful in the early stages of biomarker innovation [35].

Decisions to adopt (or discard) innovative technological applications are particularly complicated if there is debate about the diagnostic accuracy of the ARR test, currently applied in clinical practice. While most studies investigating the current ARR test for diagnosing PA in patients with essential hypertension found that its specificity and sensitivity are high, some studies showed that they may be as low as 0.56 and 0.22, respectively [22, 36]. Through our model-based approach we were able to assess these discrepancies between studies by including them in a sensitivity analysis, showing how uncertainty regarding the ARR test's sensitivity and specificity influence the headroom.

Several choices and assumptions we made in the modeling process are worth mentioning. Due to the lack of data, we assumed that patients would either remain normotensive or hypertensive without allowing for transitions between the two. Because in reality hypertensive patients can become normotensive with a reduced risk of developing cardiovascular events, our headroom may be overestimated. We also assumed that false negatives would never be identified as having PA, while in reality the persistence of hypertension will likely prompt further testing and, in some cases, the eventual diagnosis of PA. Although a delayed diagnosis is costly as well, our assumption may have resulted in an overestimation of the headroom in that it underestimates the current detection rate of PA. Likewise, the model may overestimate the extra risk of cardiovascular events in PA patients because they were computed relative to the risks in a hypertensive

population that is likely to also contain undiagnosed PA patients.

Another choice we made is that we reduced the complex and multifaceted nature of the diagnostic process to one specific test and assumed that all other tests in the diagnostic pathway are perfectly accurate. While debatable [37], doing so allowed us to single out the effectiveness gap of the ARR test itself and ignore the inaccuracies of tests that are beyond the scope of our headroom analysis. In addition, we based the effectiveness of the medication regimen in false negatives and non-PA patients on a 50% treatment adherence. If the effectiveness is lower due to lower adherence to medication in daily practice, our headroom is underestimated and vice versa. Finally, the effect of detecting PA will depend on the type of hypertensive patient that is being referred for PA diagnosis. Patients that have had therapy resistant hypertension for over a decade may profit less from treatment than patients who have been hypertensive for a shorter period of time – due to increased endocrine damage over time. In our model we synthesized evidence about treatment effects as if the target population was always identical, but this assumption is open to debate.

We acknowledge that these – and other – assumptions may affect the support for our model. However, our goal was to illustrate how an exploratory analysis could illustrate whether or not a (diagnostic) innovation might be worthwhile and why. Our model can easily be adapted to incorporate different assumptions or model inputs, for example, when new data becomes available, which makes it a versatile tool to assess the cost-effectiveness of proposed innovations. It should be noted that we adopted a societal perspective, whereas the decision to invest in novel biomarker tests is usually taken at the level of the individual laboratory. While it is possible to adopt the perspective of the latter when doing early cost-effectiveness

analyses, we argue that it is desirable to make these R&D decisions from a societal perspective on the innovation's costs and effects.

Thus, when it comes to clinical biomarker innovation, when is it worthwhile? We have illustrated that the answer to this question is more complex than a simple yes/no. As indicated, many factors determine whether the room for improvement in current care and the potential value of a specific diagnostic innovation are 'large enough'. The analyses presented in this paper provide a starting point from which these factors may be investigated for specific novel biomarker technologies; if there is no room for improvement to begin with, or if it is known beforehand that the required diagnostic accuracy cannot be achieved, there would be no point in pursuing innovation. This may prevent a considerable waste of R&D resources that can now be spent on other, more (potentially) valuable innovations. If early health economic modeling shows that an innovation is potentially worthwhile, the estimate of its potential added value may be updated anytime new information becomes available – such as costs or diagnostic accuracy – by simply adding this information to the model. Similarly, research may show that current diagnostic tests are more or less accurate than previously believed, the impact of which can also be quickly processed by adding these new insights to the model.

In conclusion, we believe early health economic modeling methods such as headroom analysis can help biomarker scientists and clinicians decide on innovations before committing considerable resources. On a larger scale, models like ours could help increase a biomarker's clinical and monetary yield, and allow for efficient research budget allocation towards the most promising innovations. More applications of early modeling through collaborations between health economists and clinical experts will illustrate their benefits and help further their accessibility.

Acknowledgments: We are indebted to Mirre Scholte for her independent and thorough validation of our model and analyses. We thank colleagues from EATRIS for the constructive discussions on this topic. Finally, we thank our anonymous reviewers for their valuable input on this manuscript.

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Research funding: None declared.

Employment or leadership: JD participates in ENSAT-HT (European Union Horizon 2020 project, grant agreement No 633983).

Honorarium: None declared.

Competing interests: The funding organization(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

References

1. Blume SS. *Insight and industry: on the dynamics of technological change in medicine*. Cambridge, MA: MIT Press, 1992.
2. Cutler DM, McClellan M. Is technological change in medicine worth it? *Health Aff (Millwood)* 2001;20:11–29.
3. Bodenheimer T. High and rising health care costs. Part 2: technological innovation. *Ann Intern Med* 2005;142:932–7.
4. Sorenson C, Drummond M, Khan BB. Medical technology as a key driver of rising health expenditure: disentangling the relationship. *Clinicoecon Outcomes Res* 2013;5:223.
5. Keehan SP, Stone DA, Poisal JA, Cuckler GA, Sisko AM, Smith SD, et al. National health expenditure projections, 2016–25: price increases, aging push sector to 20 percent of economy. *Health Aff (Millwood)* 2017;36:553–63.
6. Health pays off – between choice and solidarity. The future of care. The Hague, The Netherlands: CPB Netherlands Bureau for Economic Policy Analysis, 2013. Available at: <https://www.cpb.nl/sites/default/files/publicaties/download/cpb-policy-brief-2013-03-future-health-care.pdf>.
7. Saltman RB. Health sector solidarity: a core European value but with broadly varying content. *Isr J Health Policy Res* 2015;4:5.
8. Lippi G, Mattiuzzi C, Cervellin G. No correlation between health care expenditure and mortality in the European Union. *Eur J Intern Med* 2016;32:e13–4.
9. van Gool AJ, Bietrix F, Caldenhoven E, Zatloukal K, Scherer A, Litton JE, et al. Bridging the translational innovation gap through good biomarker practice. *Nat Rev Drug Discov* 2017;16:587–8.
10. IJzerman MJ, Koffijberg H, Fenwick E, Krahn M. Emerging use of early health technology assessment in medical product development: a scoping review of the literature. *Pharmacoeconomics* 2017;35:727–40.
11. McAtter H, Cosh E, Freeman G, Pandit A, Wood P, Lilford R. Cost-effectiveness analysis at the development phase of a potential health technology: examples based on tissue engineering of bladder and urethra. *J Tissue Eng Regen Med* 2007;1:343–9.
12. Tomaschitz A, Pilz S. Aldosterone to renin ratio – a reliable screening tool for primary aldosteronism? *Horm Metab Res* 2010;42:382–91.
13. Funder JW, Carey RM, Mantero F, Murad MH, Reincke M, Shibata H, et al. The management of primary aldosteronism: case detection, diagnosis, and treatment: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab* 2016;101:1889–916.
14. Nishikawa T, Omura M, Satoh F, Shibata H, Takahashi K, Tamura N, et al. Guidelines for the diagnosis and treatment of primary aldosteronism – The Japan Endocrine Society 2009. *Endocr J* 2011;58:711–21.
15. Li X, Goswami R, Yang S, Li Q. Aldosterone/direct renin concentration ratio as a screening test for primary aldosteronism: a meta-analysis. *J Renin Angiotensin Aldosterone Syst* 2016;17:1–8.

16. U.S. National Library of Medicine. ENSAT-HT trial [Internet]. Available at: <https://clinicaltrials.gov/ct2/show/NCT02772315>. Accessed: 29.01.2018.
17. U.S. National Library of Medicine. PRIMAL trial [Internet]. Available at: <https://clinicaltrials.gov/ct2/show/NCT03105531>. Accessed: 29.01.2018.
18. Berge C, Courand P-Y, Harbaoui B, Paget V, Khettab F, Bricca G, et al. Decreased plasma prorenin levels in primary aldosteronism: potential diagnostic implications. *J Hypertens* 2015;33: 118–25.
19. Rehan M, Raizman JE, Cavalier E, Don-Wauchope AC, Holmes DT. Laboratory challenges in primary aldosteronism screening and diagnosis. *Clin Biochem* 2015;48:377–87.
20. Monticone S, D'Ascenzo F, Moretti C, Williams TA, Veglio F, Gaita F, et al. Cardiovascular events and target organ damage in primary aldosteronism compared with essential hypertension: a systematic review and meta-analysis. *Lancet Diabetes Endocrinol* 2018;6:41–50.
21. Azizi M, Pereira H, Hamdidouche I, Gosse P, Monge M, Bobrie G, et al. Adherence to Antihypertensive Treatment and the Blood Pressure-Lowering Effects of Renal Denervation in the Renal Denervation for Hypertension (DENERHTN) Trial. *Circulation* 2016;134:847–57.
22. Jansen PM, van den Born B-J, Frenkel WJ, de Bruijne EL, Deinum J, Kerstens MN, et al. Test characteristics of the aldosterone-to-renin ratio as a screening test for primary aldosteronism. *J Hypertens* 2014;32:115–26.
23. Raizman JE, Diamandis EP, Holmes D, Stowasser M, Auchus R, Cavalier E. A renin-ssance in primary aldosteronism testing: obstacles and opportunities for screening, diagnosis, and management. *Clin Chem* 2015;61:1022–7.
24. Käyser SC, Dekkers T, Groenewoud HJ, van der Wilt GJ, Carel Bakx J, van der Wel MC, et al. Study heterogeneity and estimation of prevalence of primary aldosteronism: a systematic review and meta-regression analysis. *J Clin Endocrinol Metab* 2016;101:2826–35.
25. Shah B, Deshpande S. Assessment of effect of diabetes on health-related quality of life in patients with coronary artery disease using the EQ-5D questionnaire. *Value Health* 2014;3:67–72.
26. Williams TA, Lenders JW, Mulatero P, Burrello J, Rottenkolber M, Adolf C, et al. Outcomes after adrenalectomy for unilateral primary aldosteronism: an international consensus on outcome measures and analysis of remission rates in an international cohort. *Lancet Diabetes Endocrinol* 2017;5:689–99.
27. Benjamin EJ, Wolf PA, D'Agostino RB, Silbershatz H, Kannel WB, Levy D. Impact of atrial fibrillation on the risk of death: the Framingham Heart Study. *Circulation* 1998;98:946–52.
28. Hankey GJ, Jamrozik K, Broadhurst RJ, Forbes S, Burvill PW, Anderson CS, et al. Five-year survival after first-ever stroke and related prognostic factors in the Perth Community Stroke Study. *Stroke* 2000;31:2080–6.
29. Goldberg RJ, Ciampa J, Lessard D, Meyer TE, Spencer FA. Long-term survival after heart failure: a contemporary population-based perspective. *Arch Intern Med* 2007;167:490–6.
30. Höfer S, Benzer W, Oldridge N. Change in health-related quality of life in patients with coronary artery disease predicts 4-year mortality. *Int J Cardiol* 2014;174:7–12.
31. Statistics Netherlands. General mortality rates [Internet]. Available at: <https://www.cbs.nl/en-gb/our-services/methods/statistical-methods/output/output/life-tables>. Accessed: 08.06.2017.
32. Neumann PJ, Cohen JT. Qalys in 2018 – advantages and concerns. *J Am Med Assoc* 2018;319:2473–4.
33. Guideline for conducting economic evaluations in health care. Diemen, The Netherlands: National Health Care Institute, 2015.
34. Briggs AH, Weinstein MC, Fenwick EA, Karnon J, Sculpher MJ, Paltiel AD. Model parameter estimation and uncertainty analysis: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group – 6. *Med Decis Making* 2012;32:722–32.
35. Postmus D, Graaf G, Hillege HL, Steyerberg EW, Buskens E. A method for the early health technology assessment of novel biomarker measurement in primary prevention programs. *Stat Med* 2012;31:2733–44.
36. Fischer E, Beuschlein F, Bidlingmaier M, Reincke M. Commentary on the Endocrine Society Practice Guidelines: consequences of adjustment of antihypertensive medication in screening of primary aldosteronism. *Rev Endocr Metab Disord* 2011;12:43–8.
37. Cornu E, Steichen O, Nogueira-Silva L, Küpers E, Pagny J-Y, Grataloup C, et al. Suppression of aldosterone secretion after recumbent saline infusion does not exclude lateralized primary aldosteronism – novelty and significance. *Hypertension* 2016;68:989–94.

Supplementary Material: The online version of this article offers supplementary material (<https://doi.org/10.1515/cclm-2019-0098>).