WORKING PAPERS

N° 1507

February 2024

"Economic Incentives to Develop and to Use Diagnostic Tests
A Literature Review"

David Bardey, Philippe De Donder and Vera Zaporozhets

Toulouse School of Economics

ECONOMIC INCENTIVES TO DEVELOP AND TO USE DIAGNOSTIC TESTS

A LITERATURE REVIEW[1]

David Bardey (U Los Andes and TSE, d.bardey@uniandes.edu.co))

Philippe De Donder (TSE and CNRS, philippe.dedonder@tse-fr.eu)

Vera Zaporozhets (TSE and INRAe, vera.zaporozhets@tse-fr.eu)

January 2024

Abstract

This survey deals with the economic academic literature on diagnostic tests, with a focus first on the determinants of the use of these tests by healthcare providers, and then on the incentives to develop new diagnostic tests. It is structured in four parts. The first part provides general results in this literature regarding how healthcare providers (mostly, physicians) react to the (explicit or implicit) incentives embedded in existing health institutions, and especially to payment schemes and reimbursement rules. The second part deals more specifically with the incentives to use diagnostic tests including, among them, biomarker tests. Both sections follow a positive approach, describing individual reactions to various incentives. The third section rather takes a normative approach and tries to ascertain which incentives should be given to providers to better use existing diagnostic tests. Finally, the fourth section studies the development of new diagnostic tests, both from the viewpoint of the health authorities (when should they be developed?) and of the industry (how to incentivize them to develop the right kind of test?).

Key words: Diagnostic tests, Healthcare systems, Incentives.

JEL Code: D86, H51 and I11.

[1] The authors acknowledge financial support from the French Agence Nationale de la Recherche under grant ANR-17-EURE-0010 (Investissements d'Avenir program).

1. Introduction

We survey the economic academic literature studying the incentives to develop and then to use diagnostic (and prognostic) tests. These tests encompass all the procedures used to reveal what ails patients, and/or the best suited treatment(s). We cover diagnostic tests in general, but with a specific focus on companion tests. Such tests come together with treatment(s), with the objective of determining the adequacy between the treatment(s) and the patient. Companion tests are especially important in the case of cancer, where they are often called biomarkers tests as they look for genes, proteins, and other substances (dubbed biomarkers or tumor markers) that can provide information about the specific cancer developed by an individual, and how to treat it. Biomarkers testing then falls within the realm of precision (or personalized) medicine.

Diagnostic tests are important tools for healthcare providers to establish or to obtain a more precise diagnostic. They are also important from a more macro-economic perspective, by allowing to reduce the amount of unnecessary health expenditures. In the US healthcare system, Shrank *et al.* (2019) estimate that these unnecessary expenditures represent approximately 25% of total health care spending. Part of this wasteful healthcare spending comes from overtreatments, that typically can be reduced thanks to an appropriate use of diagnostic tests.

Medical treatments are rarely safe and effective for everyone. Thus, in case of cancers for instance, biomarker testing brings several advantages. First, it improves the treatment efficacy and minimizes adverse effects as it allows to screen patients for responders and non-responders. Second, it allows for better-informed medical decisions as it helps health practitioners choose the medicine best fitted given the characteristics of the patient. Third, it allows to prevent/limit unnecessary interventions and overtreatment by excluding patients deemed unlikely to respond. According to D'Avó Luis and Seo (2021), the potential cost per life-year gained from biomarker-guided drugs (therapies that require biomarker testing before prescription) is below the threshold value used in the literature at which the intervention is considered cost-effective.

These companion tests are then playing an increasingly important role, both in enhancing the use of existing treatments and in the authorization of new ones. For instance, a review by the European Medicines Agency shows that approximately half of cancer drugs authorized over the 2015-2018 period required patients to be screened by a genetic test before determining their treatment (Antoñanzas *et al*., 2019).

This survey focuses on the individual decisions by healthcare providers to use diagnostic tests, and later on the incentives to develop them in the first place. As we shall see, the use of diagnostic tests by healthcare providers depends on several factors such as their degree of altruism, their time constraint, but also the payment

schemes they face. For instance, as long as doctors' time constraints are affected, positively or negatively, the use of diagnostic tests also impacts doctors' income and consequently, according to the payment scheme at work, may modify their trade-off between patients' welfare and the level of income achieved. While updating their knowledge to better use diagnostic tests results requires to invest time, generating an opportunity cost, once this cost is sunk, doctors can be more efficient in formulating their diagnostics, in terms of precision as well as in terms of rapidity. This investment decision not only depends on the type of payment scheme offered by payers, but also on doctors' intrinsic motivations (*i.e.* altruism toward their patients).

Section 2 studies how receptive healthcare providers (mainly doctors)'s behaviors are to incentives, and to the various forms of payment and reimbursement schemes (such as fee for service, capitation, salary or pay for performance). This section covers both theoretical and empirical contributions (including several based on laboratory experiments) and concludes that healthcare providers can indeed be incentivized to change their medical practice by these payment schemes.

Section 3 then moves to the specific case of diagnostic tests, mostly of the companion type. Fisher *et al.* (2003) and Brody (2010) point out that a significant proportion of medical testing decisions are deemed inappropriate, entailing either over- or underprovision. While the public attention often focuses on overprovision, cases of underprovision of diagnostic testing is prevalent in the medical literature (see for instance Newman-Toker *et al.* (2013), Singh *et al.* (2013), Zhi *et al.* (2013), O'Reilly (2014). Sollman (2015) estimates the economic impact of undertesting at as high as 38% of total healthcare expenditure.

The empirical studies reviewed here show that healthcare providers seem to exhibit a mixture of monetary and altruistic motivations (prescribing fewer tests when they know their patients' out-of-pockets costs, for instance). Diagnostic tests are more often used when they are well-known and easy to interpret. Other, theoretical, contributions have shown that healthcare providers could exhibit over-confidence, and rely too much on their expertise and too little on tests.

Both sections 2 and 3 take a positive, or descriptive, approach trying to assess physicians' reactions to incentives, especially according to the payments' schemes they face. These sections help us understand the trade-offs faced by doctors, and the role played by the payment schemes' incentives. Section 4 then adopts a normative approach, where the health authorities are looking for the optimal reimbursement scheme to attain their objectives. The theoretical contributions surveyed stress that any reimbursement rule involves trade-offs between objectives (for instance, rewarding good health outcomes may result in higher health expenditures). They also stress some counter-intuitive results, such as that even costless diagnostic tests should not be made mandatory, or that social welfare is not always increasing with the providers' degree of altruism. The U.S. practice of reimbursing separately biomarkers tests and the associated treatments is criticized

and rendered at least partially responsible for the dearth of biomarker tests used in practice.

Finally, section 5 follows a dynamic approach and surveys the (theoretical and empirical) literature on the incentives for the industry to develop an innovative test in the first place. The main dichotomy here is between tests developed at the same time as treatments (companion tests) or after the treatments. Not surprisingly, the prospects for the latter are quite bleak, especially when the decrease in market size when the test is introduced is not sufficiently compensated by an increase in the (often regulated) price. Instruments studied to promote the development of innovative tests (beyond higher prices) are pay-for-performance schemes (where the reimbursement depends on the success of the treatment) and procurement design rules.

The academic literature that we cover adopts three types of methodologies: theoretical (development and solving of analytical models), empirical (exploitation of databases, mostly with the use of regressions) but also experimental (development of laboratory experiments). We try and make clear, for each paper, which type of methodology is used (some contributions developing, for instance, both an analytical model and either an empirical or experimental analysis building on this model). We summarize the key takeaways from our survey at the end of each section, and come back to them in the conclusion.

We first summarize the main results of the literature dealing with the consequences of financial incentives that healthcare providers may face. Next, we focus on the role of such financial incentives on the providers' decision to use diagnostic test and personalized medicine tools.

2. Healthcare providers and incentive schemes

In this section, we survey the main general results presented in the economic literature studying how healthcare providers' incentives, in particular doctors, are influenced by the payment schemes and reimbursement rules that they face. According to McGuire (2000), doctors can modify their medical practice through two types of behaviors, both related to their time constraint. On the one hand, doctors may adjust the total time allotted to their medical practice in general, or to each patient on average. On the other hand, they can modify the volume of services provided to each patient or during each event.

Along the first dimension, Showalter and Thurston (1997) take advantage of a reform of the tax system to study physicians' labor supply in the USA. They find that self-employed physicians are pretty sensitive to tax rates, suggesting that the physicians' labor supply depends on their income, and, consequently, on the payment scheme used. Batalgi *et al.* (2003) find similar results in the context of Norwegian hospitals.

All in all, the total time worked by physicians seems to depend on the remuneration obtained. In other words, and as it can be expected, physicians' labor supply shows a degree of elasticity.

As pointed out in Gosden *et al.* (1999), there seems to be a general agreement that fee-for-service (FFS henceforth) schemes favor supply-induced demand behaviors from healthcare providers. They show that these behaviors tend to generate inflation of heath care costs.[2] For instance, Brekke *et al.* (2017) study empirically how general practitioners (GP) respond to fee changes at the intensive margin. They use detailed administrative data covering all GPs in Norway during the period 2006–2011. Their results reveal that a higher consultation fee leads to more visits and lower treatment intensity. Fortin *et al.* (2021) obtain similar findings in Canada. Based on linked administrative and survey panel data, they study the labor supply behavior of physicians who could adopt either a standard fee-for-service contract or a mixed remuneration contract. Under the latter, physicians receive a *per diem* while the fee for services provided is reduced. These authors estimate a structural discrete choice model that incorporates service intensity (services provided per hour) and contract choice into a labor supply framework in order to control for the selection bias. Their results indicate that supply of services is reduced under a mixed payment contract. In particular, the number of hours spent seeing patients is less sensitive to incentives than the supply of services that include diagnostic tests.

On the contrary, capitation payment (CAP hereafter) and salary are two payment schemes which seem to be relatively effective to control health expenses, sometimes at the expense of the quality or quantity of services offered to patients Bardey *et al.* (2012). This negative effect on health care quality is likely to be stronger with a salary than with a capitation payment. Indeed, with a capitation payment, a lower quality may generate more events per patient, and this increasing number of events may affect the number of patients who need to visit their doctors, and then decrease doctors' income with this payment scheme. Moreover, if the demand is sufficiently sensitive to quality, *ceteris paribus*, physicians have more incentives to be cautious with the quality of services than when they receive a flat payment as a salary.

However, it is important to acknowledge that measuring the quality supplied is a complex issue since quality is not always observable, and consequently not always contractible. Moreover, even though healthcare providers may respond to financial incentives and their competitive environment, their decisions are also guided by intrinsic motivations and altruistic concerns. In such a context, as pointed out by Benabou and Tirole (2003), one must be cautious when introducing financial incentives since they may generate crowding out with intrinsic motivation. Typically,

[2] It is the case in a labor supply set-up including a time constraint, when the substitution effect dominates the income effect. In other words, when the FFS rate increases, doctors value more the time devoted to work and less the time devoted to leisure. See for instance Devlin and Sarma (2008).

these countervailing effects mean that remuneration schemes conveying explicit financial incentives may produce unintended quality outcomes.

Since it is usually complicated to consider in theoretical models as well as in empirical estimations healthcare providers' intrinsic motivation on the one hand, and the real quality provision on the other hand, several articles tackle these issues by running experiments aiming at replicating doctors' trade-offs. Green (2014) designs an experiment to compare the outcome generated by several prominent physicians' payment schemes including fee-for-service, capitation, salary, and payment for performance (P4P henceforth). This author shows that doctors' intrinsic motivations play a significant role in their decision-making. While retrospective payment schemes tend to 'crowd out' intrinsic motivations, under FFS or a blended scheme that combines FFS and P4P physicians provide a lower overall quality of services. Finally, the results of her experiments reveal that when physicians receive either a salary or a CAP they provide a higher overall quality of service.

This potential quality/quantity trade-off is analyzed by Lagarde and Blauuw (2021) in a real-effort experiment that replicates situations of multitasking environments where some of the outputs achieved are rewarded while others are not. More precisely, they design a health economics' laboratory experiment where they test the impact of physicians' financial incentives on quality and quantity outcomes according to the remuneration scheme used. In practice, the two activities are: a routine activity (medical data entry) and a cognitive activity that aims to capture the diagnosis elaboration. Subjects are randomly allocated to a control, gain or loss contract. Interestingly, the authors find that participants increase performance differently when patients are submitted to potential losses or when physicians are rewarded for the outcomes achieved. While patients' losses contribute to increase participants' performance through a greater attention that reduces the number of mistakes, bonuses tend to increase the time spent on the rewarded activity. Contrary to the prediction obtained from theoretical models of multitasking, the authors do no observe externalities, either negative or positive, on the non-incentivized activity (*i.e.*, the diagnosis task).

Green *et al.* (2022) find quite different results when studying other payment schemes often used in the healthcare sector. They design an experiment to focus on the difference between a flat rate and P4P on health outcomes. In line with Lagarde and Blauuw (2021), they find that compared to a flat rate, a payment for performance scheme increases the number of incentivized measures met, but this positive result comes at the expense of the quality of care through unintended effects on adherence to standards of care. In line with the general results of Benabou and Tirole (2003), this calls for more prudence when implementing payment for performance to remunerate physicians since this last negative result can be interpreted as a crowding out effect.

Finally, Byambadalai *et al.* (2023) provide an interesting theoretical model in which doctors choose health care quality according to their altruism level and their competitive environment. In contrast to the rest of the literature, these authors abstract from the payments schemes' properties to rather focus on the consequences of the competitive environment. In other words, these authors study how altruistic preferences are modified by markets' incentives. In addition to their theoretical setting, they conduct a laboratory experiment using a within-subject design. Subjects are asked to choose health care qualities for hypothetical patients in different market structures, from monopoly to quadropoly. Prices, costs, and patients' benefits are experimental incentive parameters.

When healthcare services are provided by a physician in a monopoly position, the subjects choose quality by trading off profits and altruistic patients benefits. In contrast, when the experiment replicates the institutional setting of a competitive environment, due to the uncertainty toward their competitors' altruism, each subject competes for patients by choosing qualities. The authors compute the Bayes-Nash equilibrium that describes subjects' quality decisions as functions of altruism. Using a nonparametric method, they estimate the population altruism distributions from the quality observed in this Bayes-Nash equilibrium in different markets and incentive configurations. They conclude that competition tends to reduce altruism, although duopoly and quadropoly equilibrium qualities are much higher than monopoly. Although markets crowd out altruism, the disciplinary powers of market competition are stronger. Counterfactuals corroborate the hypothesis that physicians' preferences can change according to markets' competition degree.

KEY TAKEAWAYS

Healthcare providers can react to changes in monetary incentives by modifying the total amount of time they spend practicing medicine, and/or by changing the volume of services per patient. As for the former, empirical studies have shown that providers' total labor supply varies with monetary incentives, for instance in the US and in Norway. As for the latter, there is general agreement that fee-for-service (FFS) reimbursement rules generate an inflation in health care costs, while capitation payments (CAP) and salaries contain those costs, sometimes at the expanse of quality.

Theoretical contributions show how intrinsic motivations (such as altruism) matter and may interfere with monetary incentives, for instance in the form of crowding out, with unintended effects on quality. The presence of such effects is confirmed by laboratory experiments.

We now survey the literature dealing more specifically with the use of diagnostic tests.

3. Physicians' incentives to use diagnostic tests

A relevant distinction in this section is between ambulatory care and healthcare delivered in hospitals. Regarding ambulatory care, in most healthcare systems, diagnostic tests are reimbursed by health insurance. Patients may face some out-of-pockets costs according to the generosity of their health insurance coverage, but physicians' decisions regarding diagnostic tests are not of a financial nature, as tests prices do not affect their income, at least directly, *i.e*. through the payment scheme. The physicians' trade-offs related to tests' use depend on i) their medical practice and the expected patients' outcomes, ii) non-medical incentives such as profit or revenue incentives, medical liability fears, and patient demand (which are among the reasons cited to explain deviation from the care guidance).[3] The first component of this trade-off is obviously related to physicians' altruism and how they value their patients' welfare, and crucially on the diagnostic precision's gain compared to the welfare cost generated by the test on patients. Physicians' non-medical incentives are more convoluted since they may include their own welfare, including their income and their workload, but also their patients' non-medical welfare.

For instance, Tierney *et al.* (1990) study how physicians' decisions may vary according to the information they have about their patients' out-of-pockets diagnostic tests costs. 121 physicians have been split in two groups. Those who belonged to the intervention group knew the out-of-of-pocket amount paid by patients, while those of the control group did not have access to this information. While during the 14 weeks preceding the experiment there were no differences between the physicians of the two groups, during the 26-weeks intervention period, the authors find that physicians of the treatment group ordered 14 percent fewer diagnostic tests. The authors conclude that displaying the costs of the test that patients must bear impacts the physicians' incentive to prescribe it. From this result, one may infer that adjusting the level of copayments, at least those related to diagnostic tests and displaying this information to physicians, can be helpful to implement the optimal number of tests.

Obviously, as we have seen with Byambadalai *et al.* (2023)'s result, the relative importance of the various types of non-medical incentives depends on the doctors' institutional environment. For ambulatory care, the density of doctors affects the competition intensity which, in turn, alters doctors' incentives to induce their patients' demand (Léonard *et al*., 2009), diagnostic tests being one way to implement this demand induction strategy. However, using tests beyond guidelines'

[3] See for instance Lee and Levy (2012) and Smith-Bindman *et al.* (2008).

recommendations not only increases the burden of health expenditures, but can also be detrimental to patients through the diagnostic therapeutic cascade (Deyo, 2002).[4]

At first sight, the trade-off faced by physicians when they determine if they prescribe or not some diagnostic tests is similar to the trade-off they face when they use personalized medicine tools. In both cases, the financial cost or incentives to prescribe them are the same and mainly depend on the payment and reimbursement schemes at work. Beyond that, a relevant difference is that while most popular diagnostic tests are easy to read and interpret for physicians since they have been included in the general medical practice and incorporated in standard medical guidelines, few doctors feel at ease with the diagnostic information conveyed by genetic or molecular tests. In other words, some differences may appear according to the type of tests used by physicians, but also to their initial background impacting the time they need to devote to interpret the tests' results. Consequently, one main difference between types of tests is that physicians have to devote time to professional formations in order to manage the information conveyed by personalized medicine tools, unlike for most of current diagnostic tests. This investment can be considered as a sunk cost which gives doctors the skills to reduce their variable cost when treating patients with personalized medicine tools (see Bardey *et al*., 2021).

The fee-for-service payment scheme is usually associated with (excessive) supply-induced demand (see for instance Bardey and Lesur [2006]). This may also apply to traditional diagnostic tests. For instance, Yip and Hsiao (2009) in a preliminary evaluation of the health system in China explain that inappropriate incentives as part of China's fee-for-service payment scheme have resulted in rapid cost increases, despite the provision of low-quality medical services. One symptom of such inefficiencies has been the inflation of diagnostic tests in China, which these authors consider as generated by this payment scheme. They review international experiences to suggest desirable changes for the Chinese health system. Even though they recognize that the change of payment scheme would not eliminate by itself all the inefficiencies, the authors claim that the current FFS scheme is responsible for most of unnecessary diagnostic tests.

Anaya *et al*. (2016) contend that the situation may be different for certain innovative diagnostic tests. More precisely, these authors claim that most of the time, diagnostic tests may allow physicians to save time establishing their diagnostic, implying in some circumstances fewer patients' visits, and consequently, under a FFS scheme, a lower income. They mention that other providers' payment schemes should be used to promote a better use of personalized medicine tools but without specifying which one.

[4] This expression is mostly used for cardiovascular troubles. It corresponds to the tight linkage between the diagnostic tests and the therapeutic intervention. See for instance Lucas *et al.* (2008).

Following this premise, Bardey *et al*. (2021) combine a theoretical model and an experiment to analyze how the different payment schemes affect providers' incentives to use personalized medicine tests. They show that payment for performance provides better incentives than fee-for-service or capitation payment schemes to promote the use of personalized medicine. Interestingly, after controlling for a potential selection bias, their experimental results show that physicians who have decided to bear the cost of personalized medicine tools acquisition are characterized by an important commitment device behavior. In other words, when physicians decide to invest their time to update their skills with the personalized medicine tools, they seem to be more devoted to their patients. In the light of these results, these authors claim that P4P or capitation payments would favor more the use of personalized medicine tools than FFS. Moreover, they recommend subsidizing partially the physicians' personalized medicine investments costs.

Obviously, the physicians' willingness to use diagnostic tests does not only depend on their ability to exploit the information they convey but also on their ability to identify the right diagnostic without this information. To shed light on this issue, Dai and Singh (2020) study the situation where a physician who is partially altruistic must establish his patient's diagnosis. Physicians can use their own medical ability or be supplemented by a perfect diagnostic test (*i.e*. that reveals the true condition of their patient). They are aware of their (high or low) diagnostic ability which constitutes their private information (*i.e.* there is adverse selection). In addition to their degree of altruism towards their patient, physicians also care about their own reputational payoff that depends on the peer perception of their ability to identify the right diagnostic. Physicians decide whether to use a test whose cost is borne by their patients. The authors show that a unique separating equilibrium exists in which the high-ability physicians only use their own ability to establish their diagnostic, while the low-ability physicians perform the test. A first inefficiency pointed out by the authors is that due to this sorting condition, it is possible that high-ability physicians skip the diagnostic test when it would be optimal to do it. Surprisingly, everything else equals, the high-ability physicians' tendency to underuse the diagnostic test increases in their altruism. The authors also point out that financial incentives (from malpractice lawsuit concerns for instance) may help fuel undertesting in the equilibrium.

Healthcare providers' decisions may be different for hospitals' care since the payment schemes' properties may exhibit different properties. As for ambulatory medicine, the use of retrospective reimbursement on the one hand, and fee-for-service scheme on the other, favor the therapeutic cascade that may contribute to increase health expenditure. However, as mentioned by Allen (2015), other payment schemes that aim to align incentives between payers and providers can also be implemented to contain healthcare costs. We will develop these payment schemes in the next section.

KEY TAKEAWAYS

The general results detailed in section 2 also apply to the case of diagnostic tests. For instance, the FFS reimbursement rule in China has been shown empirically to generate an inflation of diagnostic tests. Other empirical studies have shown that healthcare providers seem to exhibit a mixture of monetary and altruistic motivations (prescribing fewer tests when they know their patients' out-of-pockets costs, for instance). Diagnostic tests are more often used when they are well-known and easy to interpret. Lab experiments have revealed that the (sunk) investment in time and effort necessary to use certain tests (for instance PM tests) plays the role of a commitment device, increasing the subsequent use of those tests. Finally, theoretical contributions have shown that healthcare providers could exhibit over-confidence and rely too much on their expertise and too little on tests.

4. Reimbursement rule and payment schemes: diagnostic test and normative approach

This section builds on the previous one by taking a normative approach, surveying the papers looking at what would be the optimal way to reimburse healthcare providers in order to decentralize the optimal testing behavior.

Ghamat *et al.* (2018) study an issue similar to Dai and Singh (2020) (see end of previous section) in a slightly different set-up and with a more normative approach. In particular, the authors examine performance-based payment contracts to promote the optimal use of an optional diagnostic test for newly diagnosed cancer patients. As in Dai and Singh (2020), they model the interaction between two parties—a healthcare payer and a physician who also benefits from a private information. While the adverse selection is on the physician's ability in Dai and Singh (2020), this private information concerns the patients' characteristics in Ghamat *et al.* (2018). In addition to this adverse selection issue, Dai and Singh (2020) consider that the physician's effort to establish his diagnostic is not contractible, opening the door to moral hazard behaviors. Because of this information asymmetry, the authors show that it is not optimal to incentivize the physician to use a diagnostic test for all patients, even if the test is costless. The intuition behind this result is that when doctors test everybody, the payer has to increase the payment to them to ensure their participation constraint. This result implies that because of the contract cost, it would be inefficient to implement a policy that would make diagnostic testing compulsory. Interestingly, the authors show that the physician is not always able to take advantage of his private information. As in Dai and Singh (2020), social welfare is not always increasing in the physician's degree of altruism.

Carroni *et al.* (2023) revisit the physician-patient agency problem in the game theoretic class of persuasion models. They consider the situation where a patient

visits a physician after having experimented symptoms but without knowing whether she is really ill. These authors do not study the optimal payment scheme since they consider the situation in which the physician is paid by a fee-for-service. In such a context, the demand induced by FFS corresponds to the situation in which the physician convinces his healthy patient that she is ill. The patient suffers a health loss in two situations, when being treated despite being healthy, and when untreated despite being ill.

The authors assume that sick patients are heterogeneous with respect to a parameter that captures their willingness to receive a treatment. Patients characterized by low values of this parameter are reluctant to receive a treatment, while patients with high values, *ceteris paribus*, are more prone to receiving a treatment. The physician observes the value of this parameter and chooses accordingly and strategically the type of diagnostic test he will order to his patient, *i.e.* its precision defining the proportions of errors of type I and II. A key feature of the model is that the physician has discretion in deciding whether to recommend a test and, if so, in choosing its accuracy. Since patients decide whether to follow the treatment recommended by the physician, the diagnostic test here is a mean to convey information to the patient to convince him to follow her physician's recommendation. In other words, the patients with intermediate values of the willingness to receive a treatment typically need the information of a diagnostic test to accept the treatment.

Equipped with this set-up, the authors study the consequences of different policies. First, they consider a regulation that imposes minimum standards on the sensitivity and specificity of the test (*i.e.*, respectively the rate of false positives and false negatives). Their results show that a regulation that would limit the rate of false negatives does not have any impact since without regulation physicians already have incentives to recommend diagnostic tests with low false negatives. On the contrary, a regulation that would limit the rate of false positives is welfare enhancing because it reduces the induced demand by physicians, and, consequently, unnecessary treatments.

The authors also investigate if a regulation that would impose the use of diagnostic tests increases welfare. Their results are mixed and depend on the test cost. Finally, they explore a regulation that modifies the physicians' financial incentives. They show that countervailing effects are at play. On one hand, increasing the reimbursement per patient allows to reduce the number of untreated patients who would need a treatment. On the other hand, it contributes to increasing the number of unnecessary treatments which are costly and harmful to patients. Interestingly, due to the fact that diagnostic tests convey information to patients, the authors find that that the first effect dominates and that varying the physicians' financial incentives is a way to increase the social welfare.

Brandt and Cassou (2023) develop a set-up where a social planner contracts with a profit-maximizing hospital to decentralize the provision of diagnostic tests and treatments. Patients are characterized by primary symptoms of various severity. Healthcare providers receive this costless signal and then decide whether their patients have to do an imperfect diagnostic test. When a diagnostic test is done, the healthcare provider updates his belief in a Bayesian way. Once a diagnostic test has been done, it is assumed that the test results and the treatment chosen are verifiable and contractible. Since the social planner cannot verify the patients' primary symptoms when no diagnostic test has been done, the healthcare provider benefits from a private information that can be used to increase rents.

Brand and Cassou (2023) focus on prospective payment schemes which are typically used in hospitals. They derive optimal contracts within the class of prospective payments that can be interpreted as care Pathway-Related-Group (PRG hereafter) payments, namely a set of transfers for each possible situation (*i.e.* a combination of the patients' primary symptoms and diagnostic test result, if any). They show that prospective payments set to the average cost are not incentive compatible and may induce too many tests or overtreatment. Their results also reveal that to be incentive compatible, a PRG system must involve cross-subsidies. Another way to ensure that the payment scheme is incentive compatible is to reward good health outcomes but at the cost of higher health expenditures. Finally, the authors argue that due to the complexity of this system of incentive compatible prospective payments with cross-subsidies, a system of cost reimbursement (*i.e.*, a retrospective payment) may be preferred in practice.

Mougeot and Naegelen (2022) also tackle this issue focusing on the allocative and efficiency trade-off in the presence of shadow cost of public funds. They consider a costly test and two actors: a laboratory producing one or two drugs, one of them with a companion test, and the Health Authority in charge of the regulation. There is a population of patients diagnosed by a doctor who behaves as their perfect agent. While a standard treatment provides effective benefits to a part of the population, the other fraction only partially responds to it. Without the companion test, the doctor cannot know which patient is responsive to the drug.

The authors study the pricing policy implementing the second-best allocation that is determined by the trade-off between the allocative efficiency and the distributional effects in the presence of a shadow cost of public funds,[5] first when there exists only one drug, and then in the presence of two drugs and of a companion test. They point out that the optimal prices are higher when the personalized medicine treatment is prescribed (which occurs when the effectiveness of the new treatment is higher than the average effectiveness of the standard treatment). However, the use of

[5] The concept of shadow cost of public funds is used in public economics/regulation theories to capture, usually in partial equilibrium environments, the idea that taxes introduce distortions and inefficiencies (Laffont and Tirole, 1993). Roughly speaking, the collection of 1$ generates a social cost of *X,* dubbed the shadow cost of public funds.

personalized medicine is welfare enhancing only if the cost of the companion test is low enough.

In a short qualitative article, Allen (2015) explains that the majority of American payers reimburse separately biomarker tests and the associated treatments. According to this author, this institutional arrangement does not seem to be optimal in the sense that few biomarker tests are prescribed. Allen (2015) comments that the emergence of more bundled payments is likely to be a change in the right direction since it involves some risk transfer between payers and providers where the latter become residual claimants of the expenditure reduction made possible by diagnostic tests.

Finally, Allen (2015) stresses the two main reasons preventing the generalization of biomarker tests currently: first, the fact that many tests identify biomarkers involved in the patients' disease for which there does not exist a specific drug yet, and second, even when personalized treatments exist, biomarkers may struggle determining which one is the most adapted to any specific case.

KEY TAKEAWAYS

The theoretical contributions surveyed here mostly model a situation of adverse selection, where healthcare providers have some private information (on their own ability, or on the patients' health situation) not accessible to the other actors. Some also add a moral hazard element, with a hidden action taken by the provider (such as her diagnostic effort). They stress that any reimbursement rule involves trade-offs between objectives (for instance, rewarding good health outcomes may result in higher health expenditures). They also stress some counter-intuitive results, such as that even costless diagnostic tests should not be made mandatory, or that social welfare is not always increasing with the providers' degree of altruism.

Beyond the optimal reimbursement rules, contributions to this literature also study the regulation of tests' characteristics (such as minimum level of specificity and/or sensitivity) when they are chosen by the healthcare provider, or the optimal contracts between hospitals and the health authority. Finally, the U.S. practice of reimbursing separately biomarker tests and the associated treatments is criticized and rendered at least partially responsible for the dearth of biomarker tests used in practice.

Up to now, we have assumed that a diagnostic test exists, and we have looked at both the positive and normative properties of various payment schemes. In the next section, we take a more dynamic approach and analyze the desirability and the incentives to develop new tests.

5. Dynamic incentives and diagnostic tests

The decision to associate a diagnostic test to a drug can occur at two moments of the drug's clinical development: either before/during clinical development (pre-approval case), or after marketing authorization (post-approval case). In the post-approval case, when the drug is already in the market, the incentives for additional R&D are limited. However, in the pre-approval case biomarker tests may increase the likelihood of drug approval. The mostly commonly cited example is the case of metastatic melanoma, for which the treatment vemurafenib (Zelboraf) was developed together with the COBAS BRAF V600E test and received simultaneous FDA marketing approval. It turned out to be the fastest FDA approval in history.

From the perspective of a pharmaceutical firm, the introduction of a companion diagnostic test post-approval causes the decline in the number of potential patients and hence lowers revenue *per* drug. As a result, if drug prices do not increase as a consequence of testing, the gross drug sales will drop. Given the above considerations, there has been increasing interest in the combination of drugs and companion diagnostic biomarker-based tests, to increase the probability of approval of the bundle drug-companion test on the one hand, and to increase the price of this bundle on the other.

Many cancer drugs were initially launched without a biomarker test. Regulatory authorities such as the US Food and Drug Administration (FDA) and the European Medicines Agency (EMA) have been actively encouraging the use of biomarker tests in the development and use of prescription drugs.[6] In such a context, Antoñanzas *et al.* (2019) report that in the three preceding years, approximately half of the cancer drugs authorized by the EMA required patients to be screened by a genetic test before determining their treatment. Gromova *et al.* (2020) estimate that around 65% of drugs approved by EMA and FDA between 2015 and 2019 have been associated with at least one biomarker in their development program.[7] However, pharmacogenomic tests are not yet widely available (*e.g.*, Alcenat *et al.*, 2021). In addition to their costs, one potential reason for their scarcity is the complexity and imprecision of biomarker predictions, feeding in to the lack of doctors' update to incorporate new medical knowledges.

[6] See Gromova *et al.* (2020). Note that the different regulatory requirements can also create challenges when it comes to the development of therapeutics and companion diagnostic tests. For instance, in the US, marketing approval for drugs and diagnostics is performed by a single agency, the FDA. However, this is not the case in the EU. The European Medicine Agency regulates the marketing approval for drugs, while each EU member state Notified Body monitors the performance standards of diagnostic tests.

[7] Gromova *et al.* (2020) point out that immunosuppressants, immunostimulants, drugs used in diabetes, antithrombotic drugs, antineoplastic agents and antivirals are the medical specialities which have developed most drugs that include one or several biomarkers.

There are also situations where, in the absence of diagnostic tests to identify responders to the treatment, the drug has insufficient value to payers. For example, Nebacumab, a treatment for sepsis, has been shown not to be cost-effective without the test and, therefore has been withdrawn from the market (Danzon and Towse, 2002). Similarly, one may anticipate that some drugs that are already off-patents and that have been substituted by a new generation of drugs may benefit from a second life thanks to biomarker tests as long as such tests identify patients for whom the older drugs may be more efficient than more recent ones. In other words, biomarkers tests not only facilitate the development of new and more efficient treatments, but they may also increase the matching quality with all available treatments, including older generations of treatments that may be found efficient for some specific biomarkers.[8] In such a case, it is worth pointing out that the use of biomarker tests is not always associated with costly medical treatments.

Health Authorities are usually concerned by the following aspects. In the absence of diagnostic tests, there is little or no stratification and therefore, there are more potential side effects (for non-responders). In the opposite case, the responders benefit from incremental health outcomes but there are possibly increasing costs due to higher prices (and fewer consumers). A noteworthy feature of biomarker tests is a tendency to treat small niches of patients, defining new rare conditions (prevalence lower than 40/100,000) for which new orphan drugs are marketed at nearly unaffordable prices (Graf von der Schulenburg and Frank, 2015). It is likely that this tendency will continue in the future, in parallel with the use of personalized medicine and the development of diagnostic tests (Antoñanzas *et al.*, 2019).

In a theoretical model, Brekke *et al.* (2023) study the incentives for pharmaceutical companies to develop biomarker tests that perfectly reveal the adequacy of the corresponding drug for a patient, and the impact of such tests on the equilibrium market allocations in the presence of two pharmaceutical firms and a single health plan. In their model, the two drug producers first choose whether to develop a test and then, the price of the drug they produce. Observing these decisions, the insurance plan decides which (if any) of the two drugs to include in its health plan and let physicians decide which drug (if any) to prescribe to each patient. The two drugs differ both vertically and horizontally, with each patient being defined by her location in a two-dimensional Hotelling space. This information is not known to anyone, unless a biomarker test is developed.

Solving this model, the authors obtain a rich set of insights. In case of a monopoly, the drug maker has an incentive to develop a test only if the market is not served in the absence of the test--*i.e.*, if the quality of the treatment is low. In that case, too few people are treated because of the pricing decision of the firm. It is worth stressing that the monopoly has no incentive to develop a test when its drug is anyway

[8] See Bardey *et al.* (2016) for a drugs' entry model where the drugs' price regulation has to take into account that more drugs available allow to reduce their adverse effect thanks to a better match.

prescribed in the absence of the test, because the decrease in the market size due to the test is not sufficiently compensated by the higher price (even though a test allows to increase the value of the drug to those it is prescribed to).

Competition drastically changes incentives and outcomes, as competitors have more incentives to develop a test than under monopoly. A low-quality drug producer has an incentive to develop a test, since it would otherwise be excluded from the market. This in turn induces the other firm to develop its own test. While the development of a biomarker test for the low-quality drug is welfare improving, this is not always the case for the high-quality drug due to the dampening of competition effect of the tests in the price-setting stage. An extension studies the case where the market is not fully covered when two tests are developed, and shows how results are affected as a function of the drug qualities.

In the same vein, Antoñanzas *et al.* (2018) study how to incentivize pharmaceutical companies to develop a biomarker test for a drug already marketed without such a test. This corresponds to the monopoly case studied by Brekke *et al.* (2023). They assume that the price of the drug is exogenous and cannot be changed after the biomarker test has been developed.[9] In that case, the health authorities need another instrument to incentivize the development of the test, since it will result in a decrease in the market size of the prescribed drug. They show how health authorities may use pay-for-performance instruments in that case. In the absence of a test, the authorities should fully penalize the drug maker when the treatment fails (by having the firm reimburse the totality of the price paid for the drug by the authority). Decreasing the size of this penalty when a test is used will then incentivize the development of such a test, even if this test is imperfect (decreasing but not driving to zero the fraction of patients who do not respond perfectly to the drug). They study how this optimal penalty rate is affected by exogenous elements such as the health value of the drug, the extent of its side-effects or the cost of monitoring the fraction of well-treated patients.

Dealing with this same issue, in a companion article, Antoñanzas *et al.* (2019) analyze the decision-making process of health authorities and pharmaceutical firms when dealing with treatment personalization. While in their companion article they assume that the drug price is fixed, they consider here that the health authority determines the drugs prices to maximize health net benefits.[10] Laboratories take as given drugs prices when deciding to invest in a test identifying the patients responsive to the drugs. The authors develop comparative static exercises to characterize the equilibrium features according to the price level and the drug's response rate. They show that the decision to develop a test depends more on the

[9] They claim that this the most common real-world situation, and provide the example of the drug panitumumab for patients with metastatic colorectal cancer.

[10] The authors consider a kind of payment for performance in the sense that the price is paid only when patients are well-treated.

drugs' response rate than on the price level. For relatively high values of the proportion of responders, health authorities do not incentivize the firms to search for biomarkers because on average the drug works. For low response rate levels, obviously, the drug is not adopted. Personalized medicine may occur for intermediate values of the proportion of responders.

These authors also compare the two prices set by the Health Authority according to the R&D decision. They point out that when the firm invests in R&D to stratify the patients' population, the price set by the Health Authority is not necessarily higher than the price when such stratification does not take place. More precisely, this comparison depends on the test price, the scale of adverse effects, the cost of producing the drug and the test precision. They show that when the cost of the test is rather low, the price the health authority sets when the firm carries R&D investment is higher than the price when the treatment is administered to all patients. Conversely, when the test is more expensive, the price set by the Health authority is higher without R&D expenses. Finally, for intermediate values of the test cost, the relationship between the drug price and the test price mainly depends on the treatment effectiveness.

Many researchers highlight the need for flexible and value-based pricing in order to reflect higher benefits of targeted treatment and to encourage pharmaceutical firms to develop drugs with biomarker testing (*e.g.*, Danzon and Towse [2002] and Garrison and Austin [2007]). Another possibility is to sponsor the research through an R&D subsidy (Hsu and Schwartz, 2008). Using a theoretical model that they calibrate for several diseases, Danzon and Towse (2002) conclude that testing will often be socially optimal, particularly if the proportion of non-responders is high, if serious adverse reactions can arise, or if the test is inexpensive.

Alcenat *et al*. (2020) study laboratories' incentives to increase their drugs' effectiveness in a moral hazard framework (*i.e.,* where the effort undertaken by the laboratory to increase its effectiveness is not observable by a health authority). They analyze the drug reimbursement contract of a laboratory that produces a new treatment that is associated to a genetic/biomarker test. In their model, the health authority can recommend either a standard treatment, or the use of a genetic/biomarker test to prescribe the most suitable treatment to each patient according to the test result. Their model reveals that the moral hazard informational structure impacts the optimal contract designed by the health authority when one of the two treatments dominates without genetic tests if the price of the new treatment is below a threshold. In these cases, moral hazard means that personalized medicine treatments will be less often implemented when effort is not observable than when it is observable. This is due to the fact that, *ceteris paribus*, the laboratory cannot fully internalize the benefits of its effort. In contrast, when the new treatment is preferred without genetic information and when the price of the new treatment exceeds the threshold, moral hazard does not impact the implementation of personalized medicine. However, these authors make two restrictive assumptions.

On the one hand, they assume that the companion test cost is equal to zero, and, on the other hand, their effort variable only impacts the probability that the personalized treatment is preferred over the standard treatment instead of considering that the personalized medicine treatment benefit increases in this effort variable.

Even though these two last articles provide useful insights regarding the policy that allows to implement the optimal use of personalized medicine in the presence of a companion test, in both cases and in contrast to Brekke *et al.* (2023) there is no competition between laboratories. In particular, the same laboratory provides both types of treatments (standard and personalized medicine one) and decides to develop or not a companion test. In practice, Allen (2015) points out that a lot of biomarker tests are developed by independent structures. In an interesting note, Scott-Morton and Seabright (2013) analyze the lack of incentives of the pharmaceutical industry to develop companion tests simultaneously to innovative drugs. As it is often the case in R&D issues, their main point to explain this lack of incentives is related to the gap between the social and private values generated by biomarkers. More precisely, these authors argue that biomarkers generate social value thanks to the demand reduction effect and the fact that they may avoid expensive health expenditure caused by ineffective medical treatments. Scott-Morton and Seabright show that this demand reduction effect dominates and causes a gap between the private and social value of biomarkers. They also explain that laboratories may benefit from some private information regarding the drugs' efficacy among subgroups. Thus, laboratories may fail to disclose such tests' information even when doing so would create large benefits for patients and healthcare providers. Finally, they discuss how procurement design and price regulation may succeed to align the private and social values generated by biomarkers.

KEY TAKEAWAYS

The incentives to develop a (companion) diagnostic test are weaker when the associated treatment has already been approved, because the subsequent decrease in the market share needs to be compensated by a sufficient increase in the treatment price. In contrast, the development of a diagnostic test may increase the likelihood that the companion drug is approved by the health authorities, and in some cases its price. Competition among laboratories increases the incentives to develop a companion test, compared to the monopoly setting. Incentives are also higher when the average efficiency of the treatment is low. At the same time, the development of companion diagnostic tests may dampen the price competition between laboratories. If the treatment price remains fixed after the development of a companion diagnostic test, pay-for-performance schemes (where the reimbursement depends on the success of the treatment) have to be used to incentivize the development of such tests. Alternatively, one can also use

procurement design or price regulation to align the private and social values generated by biomarkers, taking into account the private information laboratories have over the specific groups which can benefit from their new test.

6. Conclusion

This survey has focused on both the (financial) incentives to develop innovative tests, and the decision by healthcare providers to use available tests. As for the latter, we have started by reviewing the literature dealing with how receptive healthcare providers' behaviors, in general, are to incentives. The (theoretical, empirical, and experimental) literature concludes that healthcare providers are indeed sensitive to monetary incentives (in their labor supply decisions), but also that other considerations, such as altruism, do matter. Non-monetary motivations have to be taken into consideration, and especially the potential crowding-out of altruism by monetary incentives.

This being said, there is a general agreement that fee-for-service (FFS) payment scheme generate an inflation in health care costs, while capitation payments (CAP) and salaries contain those costs, sometimes at the expanse of quality. This need not always be the case for diagnostic tests. Some empirical studies do indeed find that FFS rules generate too many diagnostic tests, as in China for instance. But the empirical literature reviewed in the Introduction makes a strong case that under-utilization of diagnostic tests is a first-order problem. For instance, these tests may allow to save time, generating fewer doctors' visits and therapeutic acts, and are then discouraged by FFS schemes.

The mixed motivations of health providers play an important role here as well, with experimental studies showing that altruistic considerations induce them to prescribe fewer tests when they are made aware of their patients' out-of-pockets costs, for instance. This in turn means that the care setting (ambulatory *vs* hospitals) does matter, as different rules for out-of-pocket costs usually apply.

Diagnostic tests are more often used when they are well-known and easy to interpret. Laboratory experiments have revealed that the (sunk) investment in time and effort necessary to use certain tests (for instance personalized medicine tests) plays the role of a commitment device, increasing the subsequent use of those tests. Finally, theoretical contributions have shown that healthcare providers could exhibit over-confidence and rely too much on their expertise and too little on tests.

Finding the optimal incentive schemes then requires a very good understanding of the objectives and constraints faced by the healthcare providers. The theoretical models surveyed here usually assume some adverse selection (with an informational advantage for the healthcare providers), sometimes topped up with moral hazard (a hidden action taken by these agents). These models show that it is

general not advisable to make the use of diagnostic tests mandatory (even when they are costless). They also criticize the U.S. choice of reimbursing separately biomarker tests and the associated treatments, a practice contributing to the insufficient use of biomarker tests. We have also mentioned contributions directly addressing the regulation of the diagnostic tests' characteristics, such as their minimum level of specificity and/or sensitivity.

As for the development of tests, incentives are very different (and stronger) before the approval of the linked treatment, than after. In the former case, the companion test increases the likelihood that the drug is approved by restricting its use to those with the highest probability of responding. In the latter case, the development of the companion test decreases the market size and thus requires a sizeable increase in the drug's price to prevent overall profit from decreasing. The literature has established that the latter obstacle to post-approval test development is especially effective for monopolies, and less so when there is competition between innovators. Incentives to develop the test are also higher when the average efficiency of the treatment is low. At the same time, the adoption of tests may dampen the price competition between innovators, resulting in higher equilibrium prices. The development of companion tests may benefit older drugs which are off-patent, so that these tests are not always associated with costly treatments. Finally, one can use either pay-for-performances schemes, and/or procurement design and price regulation to incentivize the development of companion tests.

References

Alcenat, S., Marechal F and F. Naegelen, 2021, “Implementation of personalized medicine in a context of moral hazard and uncertainty about treatment efficacy”, *International Journal of Health Economics and Management*, vol 21, 81–97.

Allen, T., 2015, “Payment for Cancer Biomarker Testing”, *Arch Pathol Lab Med*, 39(3):300-4.

Anaya J-M, Bardey, D., Duarte-Rey, C., Sarmiento-Monroy, J-C, Castiblanco, J. and A. Rojas-Villarraga, 2016, "Personalized medicine. Closing the gap between knowledge and clinical practice", *Autoimmunity Reviews*, vol 15, Issue 8, 833-842.

Antoñanzas, F., Rodríguez-Ibeas, R., and C. Juárez-Castelló, 2018, “Personalized medicine and pay for performance: should pharmaceutical firms be fully penalized when treatment fails?”, *Pharmacoeconomics*, 36, 733-743.

Antoñanzas, F., Rodríguez-Ibeas, R., and C. Juárez-Castelló, 2019, “Pre-approval incentives to promote adoption of personalized medicine: a theoretical approach”, *Health Economics Review*, 9:28, 2-10.

Bardey, D., Cremer, H. and J-M Lozachmeur, 2012, "Doctors' Remuneration Schemes and Hospital Competition in a Two-Sided Market", *The B.E. Journal of Economics and Analysis & Policy*, Issue 1, vol 13.

Bardey D. and R. Lesur, 2006, "Supply Induced Demand, *Ex post* Moral Hazard and Optimal Health Insurance Contract", *Annals of Economics and Statistics*, 07/1, 279-293.

Bardey D., Lozachmeur, J-M and B. Jullien, 2016, "Health Insurance and Diversity of Treatment: A Policy Mix Perspective", *Journal of Health Economics*, vol 47, 40-53, 2016.

Bardey, D, Kembou Nzalé, S. and B. Ventelou, 2021, “Physicians’ incentives to adopt personalised medicine: Experimental evidence”, Journal of Economic Behavior and Organization 191, 686–713.Batalgi, B., Bratberg, B, and T Holmas, 2003, “A Panel Data Study of Physicians' Labor Supply: The Case of Norway”, Bergen University Working paper.

Batalgi, B., Bratberg, B, and T Holmas, 2003, “A Panel Data Study of Physicians' Labor Supply: The Case of Norway”, Bergen University Working paper.

Benabou, R. and J. Tirole, 2003, “Intrinsic and Extrinsic Motivation”, *The Review of Economic Studies*, Vol. 70, No. 3 pp. 489-520.

Bobinac A, Van Exel NJ, Rutten FF and WB. Brouwer, 2010. “Willingness to Pay for a Quality-Adjusted Life-Year: The Individual Perspective”, *Value Health* 13(8):1046-55.

Brekke, K. H. Helge, K. Monstad and O. Straume, 2017. "Do treatment decisions depend on physicians' financial incentives?," Journal of Public Economics, Elsevier, vol. 155(C), pages 74-92.

Brekke, K., Morten Dalen D. and O. Straume, 2023, Competing with precision: incentives for developing predictive biomarker tests, *Scandinavian Journal of Economics*, forthcoming.

Brody H, 2010, “Medicine’s ethical responsibility for healthcare reform—The top five list”, *New England J. Medicine*, 362(4):283–285.

Byambadalai, U., Ma, A. and D. Wiesen, 2023, “Changing Preferences: An Experiment and Estimation of Market-Incentive Effects on Altruism”, *Journal of Health Economics, forthcoming.*

Carroni E, Pignataro G. and L. Siciliani, 2023, “Persuasion In Physician Agency”, Discussion Papers in Economics, University of York.

Dai, T. and S. Singh, 2020, “Conspicuous by Its Absence: Diagnostic Expert Testing Under Uncertainty”, Marketing Science, Vol. 39, No. 3, 540–563.

Danzon P. and A. Towse, 2002, “The economics of gene therapy and of pharmacogenetics”, 5(1):5-13.

D′Avó Luís, A. and M. Seo, 2021, “Has the development of cancer biomarkers to guide treatment improved health outcomes?”, The European Journal of Health Economics, 22:789–810.

Devlin, R.A. and Sarma, S., 2008. Do physician remuneration schemes matter? The case of Canadian family physicians. *Journal of health economics*, *27*(5), pp.1168-1181.

Deyo, R., 2002, “Cascade Effects of Medical Technology”, Annual Review of Public Health**,** Vol. 23:23-44.

Eeckhoudt, L., 2002. *Risk and medical decision making* (Vol. 14). Springer Science & Business Media.

Fisher ES, Wennberg DE, Stukel TA, Gottlieb DJ, Lucas FL, Pinder EL, 2003, “The implications of regional variations in Medicare spending. Part 1: The content, quality, and accessibility of care”, *Ann. Internal Medicine*, 138(4):273–287.

Fortin, B., Jacquemet, N., and B. Shearer, 2021, Labour supply, service intensity, and contracts: Theory and evidence on physicians, *Journal of Applied Econometrics*, *36*(6), pp.686-702.

Ghamat, S., Zaric, G. and H. Pun, 2018, “Contracts to Promote Optimal Use of Optional Diagnostic Tests in Cancer Treatment”, Production and Operations Management, Vol. 27, No. 12, pp. 2184–2200.

Garrison L and M. Austin, 2007, “The economics of personalized medicine: a model of incentives for value creation and capture”, Drug Information Journal, 41: 501-509.

Gosden, T., Pedersen, L., and D. Torgerson, 1999, “How should we pay doctors? A systematic review of salary payments and their effect on doctor behaviour”, Quarterly Journal of Medecine, 92:47–55.

Graf von der Schulenburg, J. and M. Frank, 2015, “Rare is frequent and frequent is costly: rare diseases as a challenge for health care systems”, European Journal of Health Economics, 16(2):113-8.

Green, E., 2014, “Payment Systems in the Healthcare Industry: An Experimental Study of Physician Incentives”, Journal of Economic Behavior & Organization.

Green, E., Peterson, K. Markiewicz, K., O’Brienc, J and N. Arring, 2022, “The Impact of a Merit-Based Incentive Payment System on Quality of Healthcare: A Framed Field Experiment”, working paper.

Gromova, M., Vaggelas, A., Dallmann, G. and D. Seimetz, 2020, “Biomarkers: Opportunities and Challenges for Drug Development in the Current Regulatory Landscape”, *Biomark Insights*.

Hsu, J. and and E. Schwartz, 2008, “A model of R&D valuation and the design of research incentives”, Insurance: Mathematics and Economics, Volume 43, Issue 3, 350-367.

Hunter, R. and Shearer, J., 2014. Cost-consequences analysis-an underused method of economic evaluation. *National Institute for Health Research*, pp.4-5.

Kalia, M., 2015, "Biomarkers for personalized oncology: recent advances and future challenges." Metabolism 64.3: S16-S21.

Laffont J-J and J Tirole, 1993, A Theory of Incentives in Procurement and Regulation. MIT Press.

Lagarde, M. and D. Blauuw, 2021, “Effects of incentive framing on performance and effort: evidence from a medically framed experiment”, Journal of the Economic Science Association, 7, p. 33–48.

Lee, D. and F. Levy, 2012, “The sharp slowdown in growth of medical imaging: an early analysis suggests combination of policies was the cause” Health Affairs ;31(8), 1876–1884.

Léonard, C., Stordeur, S. and D. Roberfroid, 2009, “Association between physician density and health care consumption: A systematic review of the evidence”, Health Policy, vol 91, Issue 2, 121-134.

Lucas, F.L., Siewers, A.E., Malenka, D.J. and Wennberg, D.E., 2008. Diagnostic-therapeutic cascade revisited: coronary angiography, coronary artery bypass graft surgery, and percutaneous coronary intervention in the modern era. *Circulation*, *118*(25), pp.2797-2802.

McGuire T, 2000, Physician Agency, Chap 9, Handbook of Health Economics.

Mougeot M and F. Naegelen, 2022, “Incentives to implement personalized medicine under second-best pricing”, Health Economics, 31, 2411-2424.

Newman-Toker DE, McDonald KM, Meltzer DO, 2013, “How much diagnostic safety can we afford, and how should we decide? A health economics perspective”, *BMJ Quality Safety* 22(S2):ii11–ii20.

O’Reilly KB, 2014, “Sin of omissions: When tests fly under the radar”, *CAP Today*, 28(5):60–66.

Scott-Morton, F and P. Seabright, 2013, “Research into biomarkers: how does drug procurement affect the design of clinical trials?”, Health Management, Policy and Innovation 1(3): 1-15.

Sevim, D. and Felder, S., 2022. Decision thresholds for medical tests under ambiguity aversion. *Frontiers in Health Services*, *2*, p.825315.

Showalter M. H. and N. K. Thurston, 1997, "Taxes and labor supply of high-income physicians," *Journal of Public Economics*, 66(1), 73–97.

Shrank, W., Rogstad, T. and N. Parekh, 2019, Waste in the US Health Care System: Estimated Costs and Potential for Savings, *JAMA*, 322(15):1501-1509.

Singh H, Giardina TD, Meyer AND, Forjuoh SN, Reis MD, Thomas EJ, 2013, "Types and origins of diagnostic errors in primary care settings", *JAMA Internal Medicine* 173(6):418–425.

Smith-Bindman, R., Miglioretti, D. and E. Larson, 2008, "Rising Use of Diagnostic Medical Imaging in a Large Integrated Health System: The use of imaging has skyrocketed in the past decade, but no one patient population or medical condition is responsible", Health Affair, 27(6): 1491–1502.

Snowsill, T., 2023. "Modelling the Cost-Effectiveness of Diagnostic Tests," *PharmacoEconomics* 41:339–351

Sollman, P., 2015, "Achieving value: Community outreach and other strategies for labs", *Medical Laboratory Observer*, 47(12): 22–24.

Tierney, W., Miller, M. and C. McDonald, 1990, "The effect on test ordering of informing physicians of the charges for outpatient diagnostic tests", The New England Journal of Medicine, 1500-1504.

World Health Organization, 2011. Increasing access to diagnostics through technology transfer and local production.

Available at https://www.who.int/publications/i/item/9789241502375

World Health Organization, 2021. "The selection and use of essential in vitro diagnostics", Report of the third meeting of the WHO Strategic Advisory Group of Experts on In Vitro Diagnostics.

Available from: https://iris.who.int/bitstream/handle/10665/339064/9789240019102-eng.pdf?sequence=1

Yip, W. and W. Hsao, 2009, "China's health care reform: A tentative assessment", China Economic Review, 20, 613–619.

Zhi M, Ding EL, Theisen-Toupal J, Whelan J, Arnaout R, 2013, "The landscape of inappropriate laboratory testing: A 15-year metaanalysis", *PLoS One* 8(11): e78962.