

Finocchiaro Castro, Massimo; Guccio, Calogero; Romeo, Domenica

Working Paper

A systematic literature review of 10 years of behavioral research on health services

Suggested Citation: Finocchiaro Castro, Massimo; Guccio, Calogero; Romeo, Domenica (2022) : A systematic literature review of 10 years of behavioral research on health services, ZBW - Leibniz Information Centre for Economics, Kiel, Hamburg

This Version is available at:

<https://hdl.handle.net/10419/266248>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

A systematic literature review of 10 years of behavioral research on health services

Massimo Finocchiaro Castro^{a,b}, Calogero Guccio^{c,b,*} & Domenica Romeo^{b,c}

^a *Department of Law and Economics, Mediterranean University of Reggio Calabria, Italy* ^b *Health* ^b *Econometrics and Data Group, University of York, UK*

^c *Department of Economics and Business, University of Catania, Italy*

Abstract

Experimental economics is, nowadays, a well-established approach to investigate agents' behavior under economic incentives. In the last decade, a fast-growing number of studies have focused on the application of experimental methodology to health policy issues. The results of that stream of literature have been intriguing and strongly policy-oriented. However, those findings are scattered between different health-related topics, making difficult to grasp the overall state-of-the-art. Hence, to make the main contributions understandable at a glance, we conduct a systematic literature review of laboratory experiments on the supply of health services. Of the 1,084 articles retrieved from 2011, 36 articles published in peer review journals have met our inclusion criteria. For them, we describe the different experimental settings, and we classify them according to the main area of interest. Finally, we provide some insights for future research in the field.

Keywords: laboratory experiments; supply of health services; systematic literature review; physicians' behavior; payment systems; health policy.

JEL: I11; C91; C92.

* **Corresponding author:** Calogero Guccio, Department of Economics and Business, University of Catania, Italy, guccio@unict.it.

1. Introduction

In the last decades, laboratory and field experiments have been designed to test theoretical models in many economic areas (Croson, 2005; Croson and Gächter, 2010), such as bargaining (Roth, 2020), auctions (Kagel and Levine, 2010), public good provision (Zelmer, 2003) and finance (Duxbury, 2015). More recently, economics experiments are becoming very popular to test individual and organizational decision-making related to health and healthcare (Galizzi and Wiesen, 2018). In fact, controlled environments, such as experimental laboratory, allow to test *ex-ante* the effects of health policy changes, like the introduction of new financial incentive scheme to physicians and minimize confounding effects when looking of causality nexus between variables. Such desirable feature becomes extremely relevant in health economics where agents' behavior could affect individual wellness, having legal consequences and being ethically sensitive.

The merit of experimental methods led to a fast-growing literature in health economics, addressing several topics: risk and time preferences (e.g. Galizzi et al., 2016), health insurance choices (e.g. Kairies-Schwarz et al., 2017); providers' incentives (e.g. Di Guida et al., 2019), altruism (e.g. Lee et al., 2019), competition (e.g. Brosig-Koch et al., 2017), professional norms (e.g. Kesternich et al., 2015), malpractice (e.g. Finocchiaro Castro et al., 2019), medicines price policies (e.g. Wettstein and Boes, S., 2022). Also, experimental designs in health economics may vary in several dimensions (Galizzi and Wiesen, 2018) such the wording of instructions (neutral vs. health-related), the type of participants (students, medical-students, or physicians) joining the experimental sessions (Kairies-Schwarz and Souček, 2020; Wang et al., 2020; Ge et al., 2021). Therefore, a comprehensive, systemic, and reader-oriented review of experimental health economics is needed. We believe our work can fill this gap in the literature.

At the same time, we are aware that other scholars have previously reviewed the existing literature, though with different purposes. Galizzi and Wielsen (2018) critically discuss the literature state-of-the-art, explaining the methodologies, debating potential areas of application of experiments to health, and thus suggesting scopes for further research. Also, Vlaev et al. (2019) summarize the available literature on the use of financial incentives to change health behaviors. However, whereas the former

could be classified as a methodological paper, rather than a review, the latter focuses on a very specific subject, lacking the comprehensiveness which we aim to achieve with our work. Finally, in a special issue of the *Journal of Economic Behaviour and Organization*, Cox, Green and Hennig-Schmidt (2016) provide an overview of laboratory experiments in four different topics of healthcare research (i.e., clinical decision support, physicians' incentives, healthcare systems and insurance, healthcare delivery, and public health), emphasizing in the conclusions all the strengths of experimental methods.

Hence, there is a lack of a comprehensive collection of the main contributions and their most relevant features from the supply-side perspective. For, we conduct a systematic literature review of the articles published in peer-reviewed journals from 2011, examining laboratory experiments in health economics looking at the supply of health services. Specifically, we focus on those studies using experimental methods to test predictions in the health sectors related to different perspectives in the provision of health services. The initial bibliographic metadata is drawn from the SCOPUS database. Of the 1,084 articles retrieved from 2011, 36 articles have met our quite selective inclusion criteria, which restrict the attention to laboratory experiments and to those experiments whose data have been gathered by merging laboratory and artefactual field experiments. Thus, we exclude field and online experiments¹, which do not allow the experimental control which is critical for their internal validity together with discrete-choice, control trials and questionnaire experiments based on hypothetical decisions only². Looking at the different areas of interest, we have detected one main topic macro-category (payment schemes), which covers a large portion of our dataset, although other research topics such as health insurance, competition and risk preferences will be discussed too.

The remainder of the paper is organized as follows. Section 2 presents our background and the method applied in the systematic literature review, showing some preliminary results on bibliographic "metadata"; Section 3 discusses the basic summary of the

¹ Thus, in our sample, we have excluded papers as Galizzi et al. (2016), who measure patients and physicians' risk-taking behaviour in a Greek hospital, because the experiments have been run outside the laboratory in a less controlled environment. At the same time, we have excluded from the sample papers as Kesternich et al. (2015), Li J. (2018), Lee et al. (2019), Kairies-Schwarz and Souček (2020), Wettstein and Boes (2022) because the experiments have been run over the internet.

² We excluded for example the online experiment by Ahlert et al. (2013).

selected papers; Section 4 describes the selected papers distinguished by topic. Finally, Section 5 provides some concluding remarks.

2. Background

2.1 Literature review method

In this section, we outline the method and selection criteria used to review the literature. We conduct a systematic literature review with the aim to provide an objective representation of the overall literature on our topic of interest. We first elaborate upon our selection criteria regarding the lab experimental approach in the field of health economics and looking at the supply side.

According to Greenhalgh (1997), a systematic review is nothing but an overview of primary studies which explicitly defines objectives, materials, and methods and has been conducted following an explicit and reproducible methodology. There are three main advantages for writing a systematic review: to summarize the existing evidence concerning a given topic; to detect any gap which leave space for future research; to provide a framework which help locate new research activities in appropriate positions (Kitchenham, 2004). Although it shares some peculiarities with a literature review in the traditional sense, a systematic review has to be looked at as a self-contained research project in itself which investigates a clearly defined issue (Denyer and Tranfield, 2009). Differently from systematic reviews, narrative reviews, which represent an alternative reviewing approach, do not indicate neither the databases and the methodologies followed to perform the review, nor the inclusion criteria used to extract the dataset, thus preventing other authors from replicating the study (Rother, 2007). For this result, we opt for a systematic review, following three main steps. First, we select the database to be investigated (for instance Scopus, Web of Science, PubMed, Google Scholar, etc.) and by looking at the papers, we detect the keywords which allow us to build our search string. Throughout this stage, we select the papers to be analyzed, by defining our inclusion/exclusion criteria. In the second stage, we provide a descriptive and content analysis of the papers included in our sample. Finally, we focus on each of the selected papers, summarizing its contents and comparing different experimental settings and findings from different studies. Figure 1 reports the main steps of the literature search and identification of studies.

<<Figure 1 about here>>

Several bibliographic databases, containing articles in peer-review journals and other types of publications, could potentially represent data sources to carry on a systematic review (for instance, Google Scholar, Web of Science (incl. MEDLINE), Scopus, EconLit, etc.). One relevant perspective to choose the most appropriate database is whether it is endowed with a classification system that leads to the balancing of two conflicting goals: 1) to gather a wide coverage of the most suitable outlets where to publish papers focusing on our topic; 2) to allow to differentiate among publication subjects (e.g., Cavalieri et al., 2019). Following the approach of Robinson and Botzen (2019), we opt for the SCOPUS database in conjunction with Google Scholar, applying a parallel check through snowballing. Indeed, SCOPUS database spans from the general field of health to more specialized fields of health economics and experimental economics, offering a quite accurate definition of the subject areas and a good coverage of citation data in scholar journals. Thus, we are confident that SCOPUS database covers an extremely large proportion of the different experimental approaches used in health economics. However, we will explain below how we have managed to be reasonably confident that our sample of papers is as inclusive as possible of the literature that meets our inclusion criteria.

2.2 Data collection

Our systematic literature review process moves to the second step, which is data collection. We ran a database search to inform this review in March 2022. As previously explained, we limit our research to SCOPUS using a search string which includes the words lab, experiment, physician, and economic³.

To guarantee the quality of the works selected, we consider articles written in English language only, published in peer-review journals, from 2011 onwards. Initially, we

³ The baseline implemented search string was: (TITLE (lab* AND experiment* AND physician*) OR TITLE (experiment* AND economic* AND physician*) OR KEY (lab* AND experiment* AND physician*) OR KEY (experiment* AND economic* AND physician*) OR ABS (lab* AND experiment* AND physician*) OR ABS (experiment* AND economic* AND physician*)) AND PUBYEAR > 2010 AND (LIMIT-TO (LANGUAGE, "English")) AND (LIMIT-TO (SRCTYPE, "j"))

find 1,084 papers that provide a broad selection of topics and subjects.

The selection criteria are based on types of studies, types of experimental approaches, and types of topics. For a study to be included, it must deal with health economics topics, to adopt the experimental methodology, look at the supply side, and to be a laboratory experiment. To be crystal clear in defining our inclusion rule, we have considered eligible only papers in which subjects in the lab, and eventually merged with artefactual field sessions, have been asked to provide health care services under different economic incentives. Consequently, we have considered all other settings ineligible. Thus, we have excluded all those papers not related with health economics issues and those not applying the experimental methodology, at least mainly, in a controlled laboratory setting. Furthermore, from the health economics experimental papers, we have excluded all the experiments run outside the laboratory (e.g., field and online experiments) as well as discrete-choice experiments, control random trials and questionnaire experiments, those based on hypothetical choices where the economic incentives to participants have been missing and experiments on health-related behavior⁴.

After reading titles, abstracts, and keywords, we have excluded 1,011 articles that have not met our inclusion criteria. The full texts of the remaining 73 articles have been read in parallel by two researchers and, in the case of disagreement, by a third one (the so-called benefit of the doubt rule). Only 32 papers have passed the final selection meeting our eligible criteria, mainly in terms of experimental settings.

Since our search strategy and our choice of the SCOPUS database may have missed some important references on the topic, we apply a parallel check through snowballing, using the reference list in each paper and imputing the citations by a generic search engine (i.e., Google Scholar). Doing so, other 4 papers have been added, after looking at the references of the 36 selected papers, leading to a final sample of 36 papers. Figure 2 depicts the PRISMA flow diagram following the steps of the systematic review which lead to the final sample.

<<Figure 2 about here>>

⁴ We are aware that the exclusion of important research such as online and field experiments as well as experiments on health-related behaviors is a critical decision for our systematic review and that other researchers may have opted for different solutions. However, we believe that our choice allows for a clearer picture of the results of a specific part of the literature that looks at the supply side.

Although we have been careful in constructing our sample, we recognize that a limited number of studies may have been excluded because the search did not retrieve them or because checking by references and snowballing did not identify them. Nevertheless, we are reasonably confident that the included studies provide a complete and updated overview of the literature regarding the laboratory experimental approach in the field of health economics.

3. Basic summary of the sampled publications

In this Section, we provide some descriptive statistics of the sampled publications. First, we show a synoptic table which collects all the reviewed papers, differentiating them by the sample size and the subject pool selected. Additionally, we discuss the trend of the papers by year, publishing journal and area of interest. Finally, we move to the authors' perspective, providing a summary of the most productive authors, their affiliations and the funding sources.

3.1. Synoptic table

Table 1 lists the 36 papers of our review and distinguishes them by the outlet, the topic, the sample size, and the employment of either medical students or physicians in the experiment. Numbers in parentheses in the last two columns indicate the specific number of that subject pool joining the experiment.

<<*Table 1 about here*>>

The average sample size of the selected papers is 184.08 (std. 125.56) ranging from a minimum of 23 to a maximum of 608. Nine studies do not employ neither medical students nor physicians. Physicians join six of the above-mentioned experiments and in four cases there have been subsamples of medical students too. Restricting our attention to physicians, on average 40.67 of them take part in the experiments (std. 36.38) ranging from a minimum of 4 to a maximum of 99.

The small number of studies including physicians is not surprising. In fact, experimenters often have difficulty in persuading physicians to take part in experimental laboratory sessions. As shown by Rahman et al. (2011) in the context of clinical trial, but easily applicable to our setting, physicians' unwillingness to join sessions is due to many participation barriers such as lack of time, lack of incentives and recognition, communication troubles, absence of any research experience and in some circumstances '*a scientifically uninteresting research question*', which makes them not involved at all. The same authors suggest the use of financial rewards to encourage doctors' participation. However, it must be taken into account that physicians' opportunity cost is very high, especially when compared with students' opportunity cost, which instead allows to keep the experimental budget low.

3.2. Analysis of search results

Figure 3 shows the number of published papers by year.

<<*Figure 3 about here*>>

The first paper was published in 2011 with a constant trend of one paper per year until 2013. Then, we move from three papers in 2014, to zero in 2015⁵. In 2016-2017 we witness the first peak with eight papers published in 2016 and seven in 2017. In fact, as shown by Cox, Green and Hennig-Schmidt (2016) in recent years the pros of experimental methods applied to the healthcare have emerged, driving many authors to employ the behavioral approach to investigate many research topics on this field. Finally, from 2018 up to the present year, we observe a volatile pattern maybe due to the exclusion of field experiments from our review which took place in that period. However, we expect a downward trend in the two-year period 2022-2023 due to the outbreak of Covid-19 in 2019-2020 which prevents experimenters from running sessions because of national restrictions. Figure 4 reports the documents per year by source.

<<*Figure 4 about here*>>

⁵ This is due to our sample restrictions which excluded both field and online experiments but also experiments focusing on nudges on the demand side.

We restricted the attention to the top three journals in terms of number of published papers: Journal of Economic Behavior and Organization (10 papers); Health Economics (10 papers); Social Science & Medicine (3 papers). The cited journals publish 64% of the sampled papers (31 over 36). Figure 5 mirrors Figure 4 to some extent, with two peaks in the two-year period 2016-2017 in correspondence of Journal of Economic Behavior and Organization and Health Economics, certainly due to the launches of the special issues ‘*Experimental and Behavioral Economics of Healthcare*’⁶ and ‘*Behavioural experiments in Health supplement*’⁷, respectively. Finally, Figure 4 differentiates documents by subject area.

<<Figure 5 about here>>

As expected, almost 29% of the selected papers falls under the scope of Economics, Econometrics and Finance. 25% of the published studies belongs to Medicine, followed by Business, Management and Accounting (20%), Social Science (8%) and Arts and Humanities (7%). The breakdown by the remaining areas is almost equitable.

3.3. Authors’ perspective

Since many authors have published several works on the topic, it may be worth to provide some summary statistics also on this regard. Figure 6 ranks the ten most prolific authors.

<<Figure 6 about here>>

Hennig-Schmidt is in first place with seven documents, followed by Wiesen with six papers who have co-authored five papers with her. To the third party parimerito place to Kairies Schwarz and Brosig-Koch with four papers each. Finally, three papers each are written by Di Guida, Godager, Gyrð-Hansen, Kokot, Oxholm and Waibel. Figure 7 depicts the top seven authors’ affiliations.

<<Figure 7 about here>>

⁶ <https://www.sciencedirect.com/journal/journal-of-economic-behavior-and-organization/vol/131/part/PB>

⁷ <https://onlinelibrary.wiley.com/toc/10991050/2017/26/S3>

Four German universities enter in such a special ranking together with two Scandinavian universities, the University of Oslo and the University of Southern Denmark (Syddansk Universitet), respectively and finally the Swiss Federal Institute of Technology in Zürich (ETH Zurich) with one of the last available places. Finally, Figure 8 ranks the top three funding sponsors.

<<Figure 8 about here>>

The Deutsche *Forschungsgemeinschaft*, a German public organization which finances research projects, takes the first place with eight papers, followed by both the Research Council of Norway (*Norges Forskningsråd*) which also supports research and innovation projects and the University of Southern Denmark with three studies each.

To provide an initial framing of the topics covered in our sample of papers we use a bibliometric approach based on the VosViewer software (Van Eck and Waltman, 2010), the results of which are depicted in Figures 9 and Figures 10. In Figure 9 the nodes and interconnections between the keywords employed by the authors are represented. Figure 10 depicts the nodes and interconnections between the words in the title and abstract of each paper.

<<Figure 9 about here>>

<<Figure 10 about here>>

In general terms, it is noted that the group of papers is relatively compact, and in both cases three clusters of different sizes are automatically identified by the software standard settings. Looking at the keywords we find a clear prevalence of the group of papers employing words related to payment systems (FFS, CAP, P4S, etc.). The other two clusters have "Health services" and "Health care delivery" as their main nodes. If we look at Figure 10 where the nodes and interconnections between the words in the title and abstract of each paper are depicted, we observe three clusters of which two are relatively similar in size and one is smaller. The two larger clusters have "payment schemes" and "medical treatment" as their main node, respectively, while the smaller cluster has "medical care" as its main node.

This first analysis seems to confirm what was said earlier about the clear prevalence of studies on payment schemes. Of course, the analyses provided in this section are

only descriptive in nature, and in the next section we will examine in detail the content and the approaches employed by our sample of papers

4. Review

In the following paragraphs, we will first focus on the papers investigating the role of payment schemes on physicians' decisions, accounting for 58.3% of our sample, and afterwards, on those works that cannot be grouped under one specific research subject.

4.1. The role of payment schemes

Measuring how physicians respond to payment schemes is the most common topic among experiments in health economics that look at the supply side. In our systematic review, 21 articles out of 36 deal with physicians' payment schemes. Table 2 summarizes the topic investigated and the main results of each paper.

<<*Table 2 about here*>>

Most of those experiments have been run with students, whereas few of them involved medical students (see for instance Bardey et al., 2021; Di Guida et al., 2019) and physicians (Reif et al., 2020).

Although not all papers assess payment schemes as the main objective of their research question, a relevant portion of works focus on payment schemes to test physicians' behavior in medical service provision. Hennig-Schmidt et al. (2011) have been the first to test the theoretical predictions introduced by Ellis and McGuire (1986)'s seminal model showing that physicians' treatment decisions are affected by payment systems. Thus, in their design, participants, acting as physicians, choose the amount of services to provide to standard patients under alternative payment schemes, capitation (CAP) and fee for service (FFS) respectively. Results, confirmed by Brosig Koch et al. (2016), comparing physicians and students' behavior, show that physicians tend to overtreat patients under FFS and to undertreat them under CAP. The above-mentioned experimental design has been replicated by several other authors to test the impact of different monetary incentives, such as mixed systems (Brosig-Koch, Hennig-Schmidt, Kairies-Schwarz and Wiesen, 2017), report cards (Green, 2014), and salary (Lagarde

and Blaauw, 2017). Similarly, Brosig-Koch, Kairies-Schwarz and Kokot (2017) test whether subjects' *ex-ante* preferences for either CAP or FFS, elicited through the strategy method (Selten, 1967), can justify their *ex-post* treatment decisions in the lab. Using the same payment schemes, Hennig-Schmidt et al. (2014) measure patient-regarding motivations among medical and nonmedical students, showing that medical students are more altruistic and prone to sacrifice their own profit compared to nonmedical ones. Still focusing on CAP and FFS, Keser et al. (2020) make participants progressively face a reduction in the lump sum payment (i.e., CAP) to test whether physicians react by customizing care at the individual patient level and whether this result is also observable under FFS. Additionally, drawing from Hennig-Schmidt et al. (2011)'s model, Finocchiaro Castro et al. (2019) introduce a random probability for a physician to be sued for malpractice to test its effect on medical service provision, whereas Martinsson and Persson (2019) propose a patient health benefit function to show how physicians' altruism varies on patients' medical needs. Finally, under the same design, Godager et al. (2016) prove that disclosing information of providers' performances to their peers benefits the quality of care under FFS.

Departing from Hennig-Schmidt (2011)'s design, Di Guida et al. (2019) investigate how physicians under FFS allocate services to patients with different responsiveness to treatments throughout 36 working days, highlighting that resource constraints might be a deterrent to overprovision. The influence of resource limitations on physicians' patient prioritization is confirmed by Oxholm et al. (2019) in their laboratory experiment, where medical students are incentivized by CAP, differentiated CAP (i.e., the fixed amount vary with the patients' needs) and salary. Similarly, Oxholm et al. (2021), distinguishing patients by treatment responsiveness, demonstrate how redistribution of services is stricter under pay for performance when resource constraints are at play.

Differently from previous works, Reif et al. (2020) account for the presence of an insurer who has to budget for physicians' cost of providing services to patients whose status of health can be misreported. Dishonesty is also investigated by Hennig-Schmidt et al. (2019) who underline the need to introduce audit probability to avoid fraudulent behavior in reporting information (i.e., in that context obstetricians report the birth weight), which determine reimbursement rates. Referral rates are instead the focus of Waibel and Wiesen (2021) who show that when referral fees are increased the number

of referrals raise regardless of the patient type. Differently, payment systems variations are used to assess hospital readmission rates by Cox et al. (2016a), demonstrating that pay for performance incentives together with decision support system drive to more cost-effective discharge decisions. The effectiveness of pay-for-performance is also confirmed in the laboratory experiment conducted by Bardey et al. (2021), assessing the impact of monetary incentives in the use of personalized medicine.

Finally, moving to another recurring topic, financial incentives are used as a tool to investigate how competition between providers affect physicians' provision behavior in Brosig-Koch, Hehenkamp and Kokot (2017). Patient-regarding preferences are, instead, analyzed in Wang et al. (2020) comparing medical students and real physicians and to show that such preferences are not significantly different across the subject pools.

4.2. Other topics in the provision of health services

There are several papers in our pool which cannot be inserted into a specific group, facing a variety of health topics such as resource allocation, health insurance decisions, competition and so on. Table 3 summarizes the specific topic investigated and the main results of each paper.

<<Table 3 about here>>

Focusing on the first theme, Ahlert et al. (2012) ask economics and medicine students to allocate a given amount to seven potential recipients varying in the quantity needed to obtain a positive payoff, either in a neutral or in a medical framework (i.e., the allocator is a physician and the recipient is a patient). Results show that economists are significantly affected by the experimental setting mimicking more often payoff-maximizers' behaviors in the neutral framing than in the medical one.

Starting from budget motivations due to the inception of Covid-19, Brendel et al. (2021) check how resource scarcity impacts on medical service provision. Medical and nonmedical students in the role of physicians decide how many services to provide to patients with varying characteristics, under different budget constraints. Results reveal that patients' health benefits decrease in response to more severe budget limitations,

receiving fewer services. To address the role of altruistic preferences in medical decisions, Kolstad and Lindkvist (2013) combine the results of a dictator game and medical students and nurses' responses to a questionnaire to investigate whether their social preferences affect their willingness to work in the public or private sector in Tanzania. Results show that medical students preferring to work in the public sector show more pronounced pro-social preferences than those opting for the private sector. In the same setting, Brock et al. (2014), merging a laboratory and field experiment, measure clinicians' generosity through a dictator game and their adherence to the protocols during patients' consultations. Data show a positive correlation between clinicians' generosity and the quality of provided care.

Moving to health insurance, Huck et al. (2016) investigates the effects of both insurance and competition on the interaction between patients and physicians. Patients, who pay the whole cost of the treatment or share the cost with all the other patients in the insurance condition, can choose whether or not to consult a physician and eventually which physician to refer to in the competition condition. The physicians, instead, choose the treatment to provide. Under the insurance condition, patients more frequently consult the physician, while physicians are more likely to overtreat the patients. The last result is mitigated when competition is introduced. The effects of market competition on medical treatments are assessed in Ge and Godager (2021). Participants acting as physicians select the medical services to provide under three different market conditions: monopoly, duopoly and quadropoly. Results show that participants are more patient-oriented in their decisions when competition is higher. Additionally, the authors' model seems to well predict the participants' behavioral changes. The outcomes of a merger among competing hospitals are discussed in Han et al. (2017). Participants in the role of a hospital head decide on the quality of services to provide to patients before and after eventually experiencing a merger. Participants' selection reveal that quality does not benefit from merging. Close to competition issue, Mimra et al. (2016) address the role of second consultation in a lab experiment where participants are randomly assigned the role of physicians or patients. The former decides whether or not to overtreat a patient, the latter can eventually ask for a second consultation at a high or low cost depending on the treatment. Overtreatment is mitigated under the second consultation condition. When search costs are reduced, patients overuse second opinions. Similarly, Martin-Lapoirie (2021) check how

teamwork among healthcare providers affects the individual precaution behavior under different liability scenarios. Subjects playing as healthcare professionals select the effort level for each consultation, while patients embodied by computers decide whether to refer twice to the same physician or to consult two different physicians. Results show that strict liability and the negligence rule (i.e. the physician has to compensate the patient only if he has demonstrated negligence in at least one consultation) lead to similar precaution behaviors. In their laboratory experiment, Angerer et al. (2020) investigate how introducing the possibility for physician of being monitored either randomly or upon the patient's request can avoid misbehaviors such as undertreatment, overtreatment, and overcharging. Data show that both endogenous monitoring and exogenous monitoring succeed quite well in reducing the level of undertreatment and overcharging observed and improve market efficiency. In order to improve decisions, Cox et al. (2016b) investigate how introducing clinical decision support system (CDSS) affect physicians and fourth-year medical students' hospital discharge decisions. Recommendations provided by CDSS contain patients' probability of readmission in case of an incorrect early discharge decision which is costly to the provider (i.e. a reduction of the maximum obtainable payoff). Results provide evidence for CDSS as an effective tool to improve discharge decisions.

Prescription behavior is the focus of Greiner et al. (2017), who test the possibility of separating prescription and treatment activities through a lab experiment. In the baseline condition the physician decides the prices for possible treatments, while the patient decides whether or not to consult the doctor and whether to undertake the suggested treatment. Under a different experimental condition, the patient interacts with two different doctors: the first one is only in charge of the prescription phase (for free), the second one only implements the treatment previously prescribed. Although this second condition results in a reduction of overtreatment, it reduces efficiency due to miscoordination between doctors involved.

Organ donation is addressed in Kessler and Roth (2014) and Herr and Normann (2016). In the former, college students play a game where they have to opportunity to register as organ donors, although instructions were neutrally framed, under different allocation rules. Results demonstrate that the presence of a loophole, where subjects can register to get a priority but simultaneously refuse donating organs, has a detrimental effect on the donation resulted by the priority rule. In the latter, medical

and nonmedical students first join several rounds of a donation game and, after having already tested it, they are asked to vote for the implementation of a priority rule in the last rounds of the game. Two-thirds of the participants show stronger preferences for the priority rule.

Finally, the last paper included focuses on participants' risk and time preferences measurement. As reported in the literature, risk preferences are domain-dependent (see e.g., Wölbert, and Riedl, 2013; Weber et al., 2002), and then, several authors prefer measuring risk across different contexts before drawing conclusions. For instance, in their laboratory experiment Arrieta et al. (2017) measure medical and nonmedical students risk preferences in deciding for others both in the monetary and health domain, using the Holt and Laury (2002) (HL)'s multiple price list method. Participants playing the role of a physician, who takes decisions in three different health contexts, have to choose the treatments to provide to patients. Depending on the context, health gains can be expressed in terms of years of life for a patient with varying health conditions or hours of pain alleviated. Results confirm that risk tendencies are health-context specific. Additionally, students with a medical background are found to be more risk-averse than their peers and surprisingly such attitude is exacerbated in the health domain.

5. Concluding remarks

Our study provides a systematic review of the literature, applying behavioral and experimental methods to health issues related to different perspectives in the provision of health services. This is not an easy task. Many studies have been incorrectly classified as '*experiments in behavioral health*', although their designs are not incentive-compatible and do not provide real consequences for participants (Gibson, 2021). Thus, of the 1,084 articles retrieved, published between 2011 and march 2022, only 36 articles meet our inclusion criteria. Specifically, we focus only on laboratory experiments, excluding online and field experiments (which lack the controlled environment required by experimental methodology to guarantee their internal validity), as well as discrete-choice experiments, control random trials and questionnaire experiments based on hypothetical choices due to the absence of any monetary incentive.

The selected papers have been first classified according to the object of analysis. A large portion of the 36 papers investigated the issue of payment schemes, whereas the remaining studies treat several different themes such as health insurance, organ donation and market competition, making it impossible to group them into specific categories. For each paper, we have checked the number of participants and whether they were student, medical students, or physicians, describing their experimental designs and main results.

The main aspect emerging from our systematic review of the provision of health services in the lab is the need to involve more physicians in health-related experiments to increase the external validity of the results. Although we are fully aware of the difficulty to gather physicians in a lab, due to their high opportunity cost, their awareness of medical procedure and their experience can make experimental results much more sound and able to provide health policy implications. We acknowledge that some researchers argue that that issue typically affects field experiments more than laboratory ones. However, some experimental papers show that choosing medical students, or even nonmedical students, to act as physicians in health-related decisions concerning patient treatment may affect the external validity of the results. In this regard, both Brosig-Koch et al. (2016) and Finocchiaro Castro et al. (2022) show that subjects' answers to incentivized choices vary on their background and that physicians more easily grasp the main incentives in the experimental designs. The authors conclude that experimenters need to carefully select their pools before testing any health economics prediction.

Another aspect raised by our systematic review is the poor connection between two fields of research: behavioral and experimental economics on the one side and health economics on the other side (Gibson, 2021). Such disconnection is confirmed by the lack of incentive compatibility typical of many discrete choice and online experiments. As suggested by Gibson (2021) some of the experiments carried out in specific areas (i.e., decisions about health-related behaviors such as smoking, diet, and alcohol drinking) can be improved with the introduction of behavioral consequences for the participants' stated preferences, providing the appropriate incentive compatible scheme for each area to be investigated.

Hence, in this systematic review, we have attempted to offer an overview, as comprehensive as possible, of a strand of literature dealing with issues related to the provision of healthcare services. This is an area that has grown significantly in the last ten years and, to the best of our knowledge, had not yet received a thorough and update overview. Although our review shows that the role of incentives related to payment systems is the most investigated strand, there is still room for analysis. For example, it is still poorly understood how in P4P systems physician behavior is influenced by base payment (FFS or CAP), how patient characteristics influence treatment decisions, and which payment system design features could potentially influence treatment decisions and to improve the quality of care for different types of patients.

Additionally, although many areas of research have been explored using laboratory experiments, other areas remain still untreated. For example, to the best of our knowledge, no study has investigated waiting lists from the perspective of healthcare providers yet, though the subject has been widely treated in the health economics literature. Another promising and as yet little explored area of research concerns the behavior of providers when there are peaks in demand or under extreme conditions such as pandemic situations.

Finally, some limits of our systematic review are worth mentioning. First, the literature selection process might be limited by the exclusion of some relevant articles which are not contained into SCOPUS database. Additionally, we might have missed other studies due to our keyword selection or to the restricted time span. Second, despite having used all the precautions specific to the systematic review approach to allow for replicability, a certain degree of discretion cannot be neglected. Excluding online and field experiments as well as experiments on health-related behavior is a critical decision for our systematic review. Consequently, despite transparently explaining the reasons behind our choices, we are conscious that other researchers may have opted for different solutions.

References

- Ahlert, M., Felder, S., & Vogt, B. (2012). Which patients do I treat? An experimental study with economists and physicians. *Health Economics Review*, 2(1), 1-11. doi:10.1186/2191-1991-2-1.
- Ahlert, M., Funke, K., & Schwettmann, L. (2013). Thresholds, productivity, and context: An experimental study on determinants of distributive behaviour. *Social Choice and Welfare*, 40(4), 957-984.
- Angerer, S., Glätzle-Rützler, D., & Waibel, C. (2021). Monitoring institutions in healthcare markets: Experimental evidence. *Health Economics*, 30(5), 951-971.
- Arrieta, A., García-Prado, A., González, P., & Pinto-Prades, J. L. (2017). Risk attitudes in medical decisions for others: An experimental approach. *Health Economics*, 26, 97-113.
- Bardey, D., Kembou, S., & Ventelou, B. (2021). Physicians' incentives to adopt personalised medicine: Experimental evidence. *Journal of Economic Behavior and Organization*, 191, 686-713.
- Brendel, F., Einhaus, L., & Then, F. (2021). Resource scarcity and prioritization decisions in medical care: A lab experiment with heterogeneous patient types. *Health Economics*, 30(2), 470-477.
- Brock, J. M., Lange, A., & Leonard, K. L. (2016). Generosity and prosocial behavior in healthcare provision evidence from the laboratory and field. *Journal of Human Resources*, 51(1), 133-162.
- Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N., & Wiesen, D. (2016). Using artefactual field and lab experiments to investigate how fee-for-service and capitation affect medical service provision. *Journal of Economic Behavior and Organization*, 131, 17-23.
- Brosig-Koch, J., Hehenkamp, B., & Kokot, J. (2017). The effects of competition on medical service provision. *Health Economics*, 26, 6-20.
- Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N., & Wiesen, D. (2017). The effects of introducing mixed payment systems for physicians: Experimental evidence. *Health Economics*, 26(2), 243-262.
- Brosig-Koch, J., Kairies-Schwarz, N., & Kokot, J. (2017). Sorting into payment schemes and medical treatment: A laboratory experiment. *Health Economics*, 26, 52-65. doi:10.1002/hec.3616.
- Cavalieri, M., Cristaudo, R., & Guccio, C. (2019). Tales on the dark side of the transport infrastructure provision: a systematic literature review of the determinants of cost overruns. *Transport Reviews*, 39(6), 774-794.

- Cox, J. C., Green, E., & Hennig-Schmidt, H. (2016). Experimental and behavioral economics of healthcare. *Journal of Economic Behavior and Organization*, 131, A1-A4.
- Cox, J. C., Sadiraj, V., Schnier, K. E., & Sweeney, J. F. (2016a). Incentivizing cost-effective reductions in hospital readmission rates. *Journal of Economic Behavior and Organization*, 131, 24-35. doi:10.1016/j.jebo.2015.03.014.
- Cox, J. C., Sadiraj, V., Schnier, K. E., & Sweeney, J. F. (2016b). Higher quality and lower cost from improving hospital discharge decision making. *Journal of economic behavior & organization*, 131, 1-16.
- Croson, R. (2005). The method of experimental economics. *International Negotiation*, 10(1), 131-148.
- Croson, R., & Gächter, S. (2010). The science of experimental economics. *Journal of Economic Behavior & Organization*, 73(1), 122-131.
- Denyer, D., & Tranfield, D. (2009). Producing a systematic review. In D. A. Buchanan & A. Bryman (Eds.). *The Sage handbook of organizational research methods*, 671–689. Sage Publications Ltd.
- Di Guida, S., Gyrd-Hansen, D., & Oxholm, A. S. (2019). Testing the myth of fee-for-service and overprovision in health care. *Health Economics*, 28(5), 717-722.
- Duxbury, D. (2015). Behavioral finance: insights from experiments I: theory and financial markets. *Review of Behavioral Finance*, 7(1), 78-96.
- Finocchiaro Castro, M. F., Ferrara, P. L., Guccio, C., & Lisi, D. (2019). Medical malpractice liability and physicians' behavior: Experimental evidence. *Journal of Economic Behavior and Organization*, 166, 646-666. doi:10.1016/j.jebo.2019.08.006.
- Finocchiaro Castro, M., Guccio, C., & Romeo, D. (2022). *An assessment of physicians' risk attitudes using laboratory and field data.*, 22(26). HEDG, c/o Department of Economics, University of York.
- Galizzi, M. M., Miraldo, M., Stavropoulou, C., & Van Der Pol, M. (2016). Doctor–patient differences in risk and time preferences: A field experiment. *Journal of Health Economics*, 50, 171-182.
- Galizzi, M. M., & Wiesen, D. (2018). Behavioral experiments in health economics. In: Hamilton, J. H., Dixit, A., Edwards, S., and Judd, K. Oxford research encyclopedia of economics and finance. Oxford research encyclopedias. Oxford University Press, Oxford, UK.
- Ge, G., & Godager, G. (2021). Predicting strategic medical choices: An application of a quantal response equilibrium choice model. *Journal of Choice Modelling*, 39, 106926. doi:10.1016/j.jocm.2021.100282.

- Gibson, J. (2021). Experimental evidence on behavioural nudges in health. In *A Research Agenda for Experimental Economics*, 119-134. Edward Elgar Publishing.
- Godager, G., Hennig-Schmidt, H., & Iversen, T. (2016). Does performance disclosure influence physicians' medical decisions? an experimental study. *Journal of Economic Behavior and Organization*, *131*, 36-46. doi:10.1016/j.jebo.2015.10.005
- Green, E. P. (2014). Payment systems in the healthcare industry: An experimental study of physician incentives. *Journal of Economic Behavior and Organization*, *106*, 367-378. doi:10.1016/j.jebo.2014.05.009.
- Greenhalgh, T. (1997). How to read a paper: Papers that summarise other papers (systematic reviews and meta-analyses). *Bmj*, *315*(7109), 672-675.
- Greiner, B., Zhang, L., & Tang, C. (2017). Separation of prescription and treatment in health care markets: A laboratory experiment. *Health Economics*, *26*, 21-35. doi:10.1002/hec.3575.
- Han, J., Kairies-Schwarz, N., & Vomhof, M. (2017). Quality competition and hospital mergers-An experiment. *Health Economics*, *26*, 36-51.
- Hennig-Schmidt, H., Selten, R., & Wiesen, D. (2011). How payment systems affect physicians' provision behaviour. An experimental investigation. *Journal of Health Economics*, *30*(4), 637-646. doi:10.1016/j.jhealeco.2011.05.001.
- Hennig-Schmidt, H., & Wiesen, D. (2014). Other-regarding behavior and motivation in health care provision: An experiment with medical and non-medical students. *Social Science and Medicine*, *108*, 156-165.
- Hennig-Schmidt, H., Jürges, H., & Wiesen, D. (2019). Dishonesty in health care practice: A behavioral experiment on upcoding in neonatology. *Health Economics*, *28*(3), 319-338.
- Herr, A., & Normann, H. T. (2016). Organ donation in the lab: Preferences and votes on the priority rule. *Journal of Economic Behavior and Organization*, *131*, 139-149.
- Huck, S., Lünser, G., Spitzer, F., & Tyran, J. R. (2016). Medical insurance and free choice of physician shape patient overtreatment: A laboratory experiment. *Journal of Economic Behavior and Organization*, *131*, 78-105.
- Kagel, J. H., & Levin, D. (2010). Auctions (experiments). In *Behavioural and experimental economics*, 14-22. Palgrave Macmillan, London.
- Kairies-Schwarz, N., Kokot, J., Vomhof, M., & Weßling, J. (2017). Health insurance choice and risk preferences under cumulative prospect theory—an experiment. *Journal of Economic Behavior & Organization*, *137*, 374-397.
- Kairies-Schwarz, N., & Souček, C. (2020). Performance pay in hospitals: An experiment on bonus–malus incentives. *International Journal of Environmental Research and Public Health*, *17*(22), 1-30. doi:10.3390/ijerph17228320.

Keser, C., Montmarquette, C., Schmidt, M., & Schnitzler, C. (2020). Custom-made health-care: An experimental investigation. *Health Economics Review*, 10(1) doi:10.1186/s13561-020-00299-4.

Kessler, J. B., & Roth, A. E. (2014). Loopholes undermine donation: An experiment motivated by an organ donation priority loophole in israel. *Journal of Public Economics*, 114, 19-28. doi:10.1016/j.jpubeco.2013.12.006.

Kesternich, I., Schumacher, H., & Winter, J. (2015). Professional norms and physician behavior: Homo oeconomicus or homo hippocraticus? *Journal of Public Economics*, 131, 1-11.

Kitchenham, B. (2004). Procedures for performing systematic reviews. *Keele, UK, Keele University*, 33(2004), 1-26.

Kolstad, J. R., & Lindkvist, I. (2013). Pro-social preferences and self-selection into the public health sector: Evidence from an economic experiment. *Health Policy and Planning*, 28(3), 320-327.

Lagarde, M., & Blaauw, D. (2017). Physicians' responses to financial and social incentives: A medically framed real effort experiment. *Social Science and Medicine*, 179, 147-159. doi:10.1016/j.socscimed.2017.03.002.

Lee, Y. S., Kim, S. Y., & Chae, Y. (2019). Distribution of monetary incentives in health insurance scheme influences acupuncture treatment choices: An experimental study. *PLOS ONE*, 14(6): e0218154.

Li, J. (2018). Plastic surgery or primary care? altruistic preferences and expected specialty choice of U.S. medical students. *Journal of Health Economics*, 62, 45-59.

Martin-Lapoirie, D. (2021). Teamwork in health care and medical malpractice liability: An experimental investigation. *European Journal of Law and Economics*, 54(2), 251-282.

Martinsson, P., & Persson, E. (2019). Physician behavior and conditional altruism: The effects of payment system and uncertain health benefit. *Theory and Decision*, 87(3), 365-387.

Mimra, W., Rasch, A., & Waibel, C. (2016). Second opinions in markets for expert services: Experimental evidence. *Journal of Economic Behavior and Organization*, 131, 106-125.

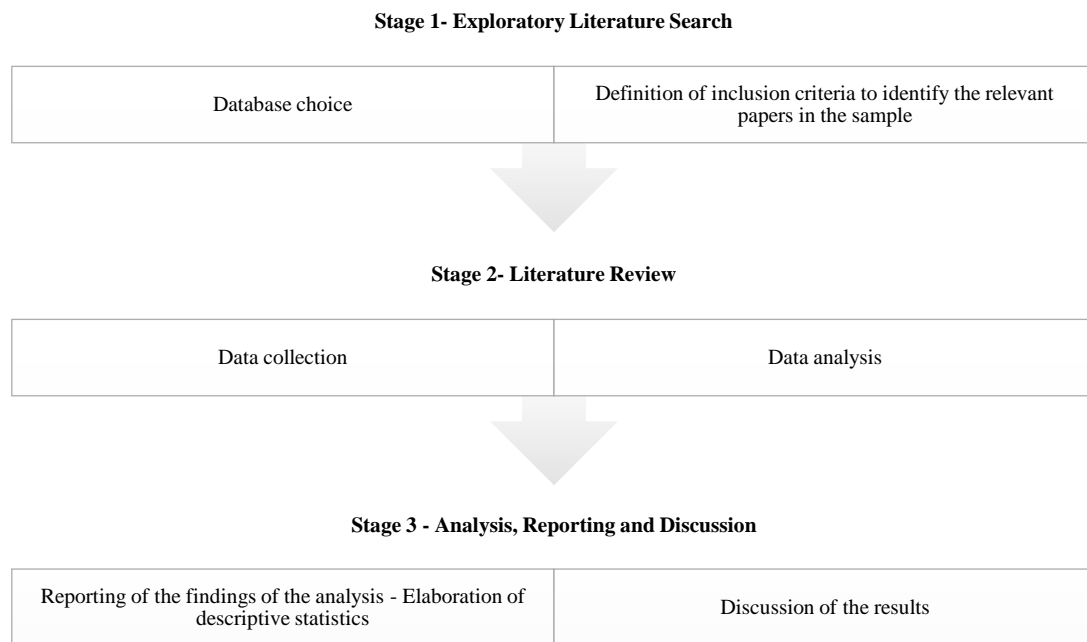
Oxholm, A. S., Di Guida, S., Gyrd-Hansen, D., & Olsen, K. R. (2019). Taking care of high-need patients in capitation-based payment schemes—an experimental investigation into the importance of market conditions. *Applied Economics*, 51(47), 5174-5184.

Oxholm, A. S., Di Guida, S., & Gyrd-Hansen, D. (2021). Allocation of health care under pay for performance: Winners and losers. *Social Science and Medicine*, 278

- Rahman, S., Majumder, M. A. A., Shaban, S. F., Rahman, N., Ahmed, M., Abdulrahman, K. B., & D'Souza, U. J. (2011). Physician participation in clinical research and trials: issues and approaches. *Advances in medical education and practice*, 2, 85.
- Reif, S., Hafner, L., & Seebauer, M. (2020). Physician behavior under prospective payment schemes—evidence from artefactual field and lab experiments. *International Journal of Environmental Research and Public Health*, 17(15), 1-37.
- Robinson, P. J., & Botzen, W. J. W. (2019). Economic experiments, hypothetical surveys and market data studies of insurance demand against low-probability/high-impact risks: A systematic review of designs, theoretical insights and determinants of demand. *Journal of Economic Surveys*, 33(5), 1493-1530.
- Roth, A. E. (2020). 4. Bargaining Experiments. In *The handbook of Experimental Economics*, 253-348. Princeton University Press.
- Rother, E. T. (2007). Systematic literature review vs narrative review. *Acta Paulista de Enfermagem*, 20(2), v-vi.
- Selten, R. (1967). Ein Oligopolexperiment mit Preisvariation und Investition. In H. Sauermann (Ed.), *Beiträge zur experimentellen Wirtschaftsforschung*. 136-138. Tübingen: Mohr.
- Van Eck, N.J., Waltman, L., (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* 84, 2, 523–538.
- Vlaev, I., King, D., Darzi, A., & Dolan, P. (2019). Changing health behaviors using financial incentives: a review from behavioral economics. *BMC public health*, 19(1), 1-9.
- Waibel, C., & Wiesen, D. (2021). An experiment on referrals in health care. *European Economic Review*, 131, 103612.
- Wang, J., Iversen, T., Hennig-Schmidt, H., & Godager, G. (2020). Are patient-regarding preferences stable? Evidence from a laboratory experiment with physicians and medical students from different countries. *European Economic Review*, 125, 103411.
- Wölbelt, E., & Riedl, A. (2013). *Measuring Time and Risk Preferences: Reliability, Stability, Domain Specificity*, 4339. CESifo Working Paper.
- Wettstein, D. J., & Boes, S. (2022). How value-based policy interventions influence price negotiations for new medicines: An experimental approach and initial evidence. *Health Policy*, 126(2), 112-121.
- Zelmer, J. (2003). Linear public goods experiments: A meta-analysis. *Experimental Economics*, 6(3), 299-310.

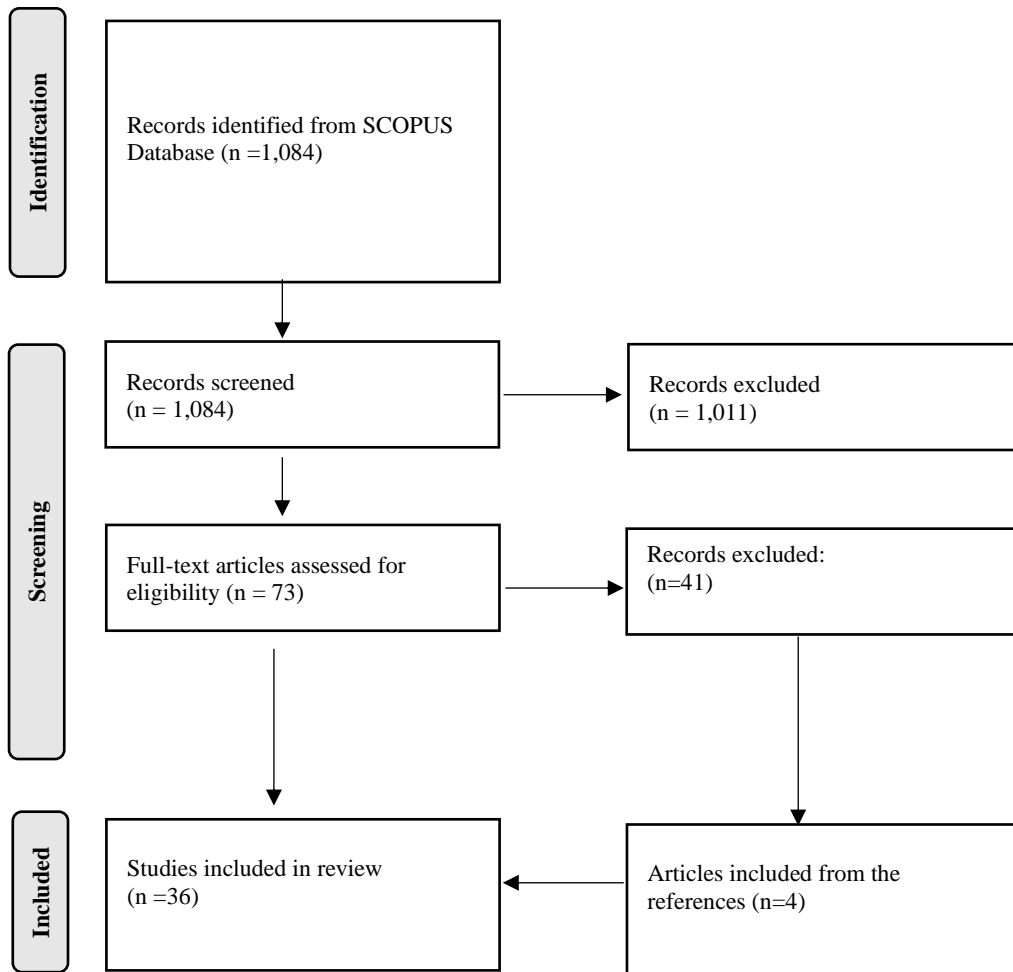
TABLES AND FIGURES

Figure 1 – Main steps of the literature search and identification of studies



Source: our elaboration

Figure 2 - PRISMA flow diagram of the systematic literature review



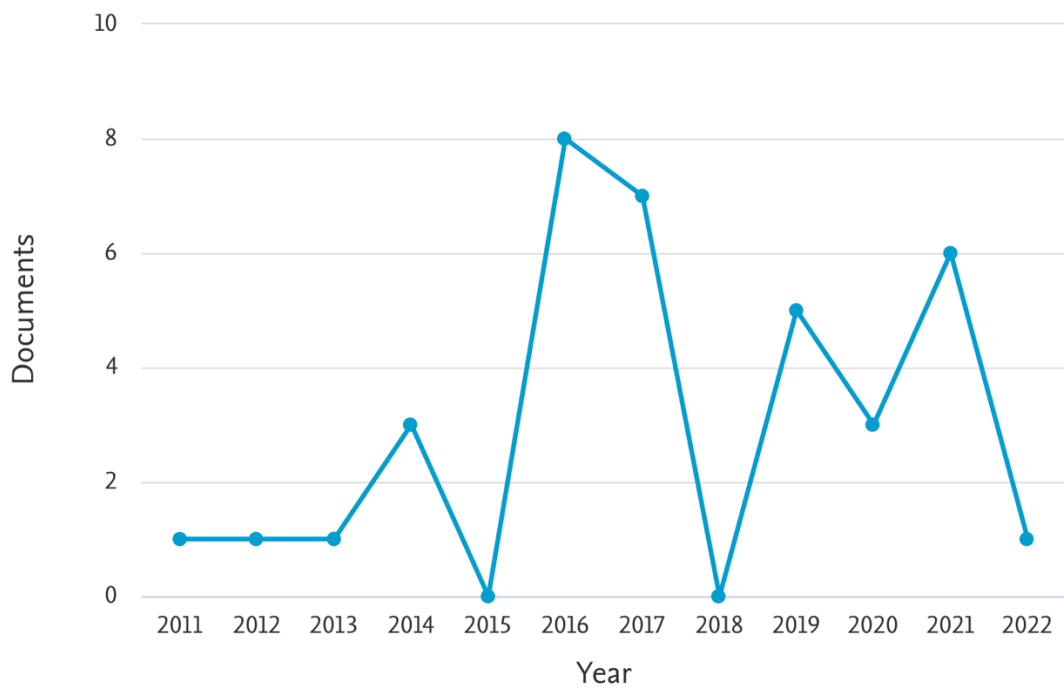
Source: our elaboration

Table 1 - Main features of the sample

Code	Authors	Journal	Topic	Sample size in the lab	Medical Students in the lab	Physicians in the lab
1	Ahlert et al. (2012)	Health Economics Review	Distributive behavior	136	yes (22)	no
2	Angerer et al. (2021)	Health Economics	Monitoring	424	no	no
3	Arrieta et al. (2017)	Health Economics	Risk preferences	257	yes (178)	no
4	Bardey et al. (2021)	Journal of Economic Behavior and Organization	Payment schemes	95	yes (95)	no
5	Brendel et al. (2021)	Health Economics	Medical service provision	174	yes (13)	no
6	Brock et al. (2016)	Journal of Human Resources	Prosocial behavior	149	no	yes (71)
7	Brosig-Koch et al. (2016)	Journal of Economic Behavior and Organization	Payment schemes	242	yes (76)	yes (29)
8	Brosig-Koch, Hehenkamp, Kokot (2017)	Health Economics	Payment schemes	178	no	no
9	Brosig-Koch, Hennig-Schmidt, Kairies-Schwarz, Wiesen (2017)	Health Economics	Payment schemes	213	yes (32)	no
10	Brosig-Koch, Kairies-Schwarz, Kokot (2017)	Health Economics	Payment schemes	185	yes (28)	no
11	Cox et al. (2016a)	Journal of Economic Behavior and Organization	Payment schemes	209	yes (209)	no
12	Cox et al. (2016b)	Journal of Economic Behavior and Organization	Discharge decisions	125	yes (105)	yes (20)
13	Di Guida et al. (2019)	Health Economics	Payment schemes	38	yes (38)	no
14	Finocchiaro Castro et al. (2019)	Journal of Economic Behavior and Organization	Payment schemes	106	yes (21)	yes (4)
15	Ge and Godager (2021)	Journal of Choice Modelling	Market competition	136	no	no
16	Godager et al. (2016)	Journal of Economic Behavior and Organization	Payment schemes	51	yes (51)	no
17	Green (2014)	Journal of Economic Behavior and Organization	Payment schemes	136	no	no
18	Greiner et al. (2017)	Health Economics	Prescription behavior	300	no	no
19	Han et al. (2017)	Health Economics	Hospital mergers	353	no	no
20	Hennig-Schmidt et al. (2011)	Journal of Health Economics	Payment schemes	42	yes (42)	no
21	Hennig-Schmidt and Wiesen (2014)	Social Science and Medicine	Payment schemes	86	yes (42)	no
22	Hennig-Schmidt et al. (2019)	Health Economics	Payment schemes	98	yes (51)	no
23	Herr and Normann (2016)	Journal of Economic Behavior and Organization	Organ donation	192	yes (21)	no
24	Huck S. et al. (2016)	Journal of Economic Behavior and Organization	Medical insurance	336	no	no
25	Keser C. et al. (2020)	Health Economics Review	Payment schemes	23	yes (23)	no
26	Kessler and Roth (2014)	Journal of Public Economics	Organ donation	608	no	no
27	Kolstad and Lindkvist (2013)	Health Policy and Planning	Prosocial behavior	80	yes (40)	no
28	Lagarde and Blaauw (2017)	Social Science and Medicine	Payment schemes	132	yes (132)	no
29	Martin-Lapoirie (2021)	European Journal of Law and Economics	Teamwork	120	yes (14)	no
30	Martinsson and Persson (2019)	Theory and Decision	Payment schemes	130	no	no
31	Mimra et al. (2016)	Journal of Economic Behavior and Organization	Second consultations	420	yes (8)	no
32	Oxholm et al. (2019)	Applied Economics	Payment schemes	55	yes (55)	no
33	Oxholm et al. (2021)	Social Science and Medicine	Payment schemes	143	yes (143)	no
34	Reif et al. (2020)	Journal of Environmental Research and Public Health	Payment schemes	126	no	yes (21)
35	Waibel and Wiesen (2021)	European Economic Review	Payment schemes	252	yes (50)	no
36	Wang et al. (2020)	European Economic Review	Payment schemes	277	yes (178)	yes (99)

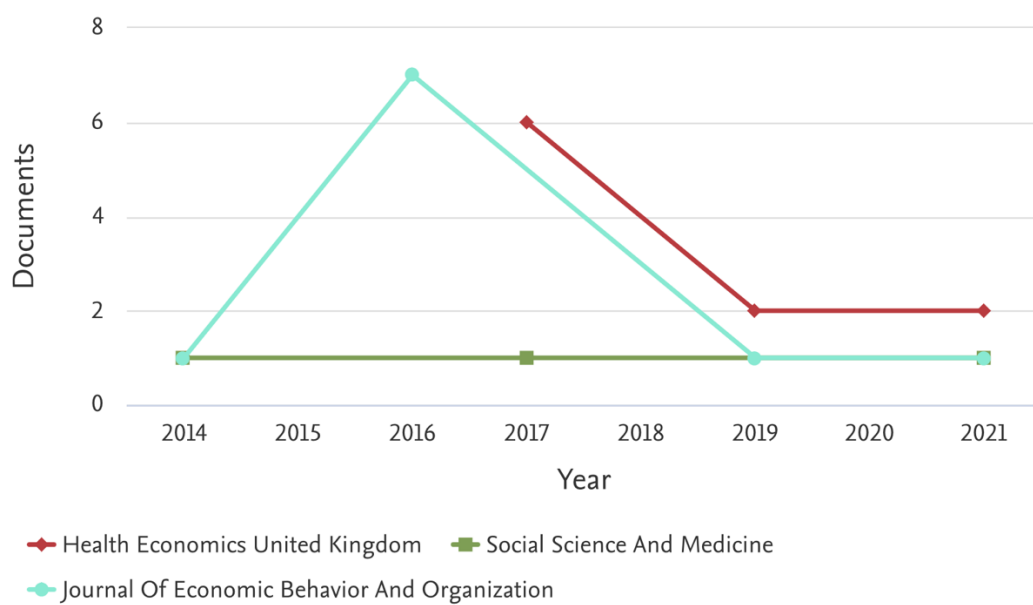
Source: our elaboration on Scopus database

Figure 3 - Documents by year



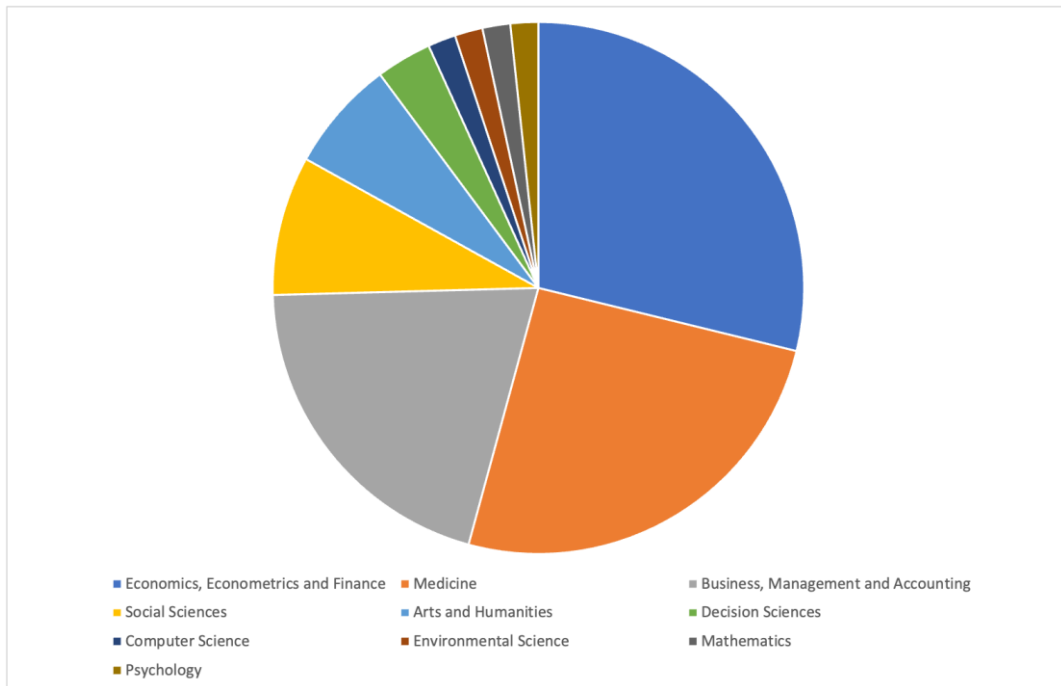
Source: Scopus' elaboration on our dataset

Figure 4 - Documents per year by source



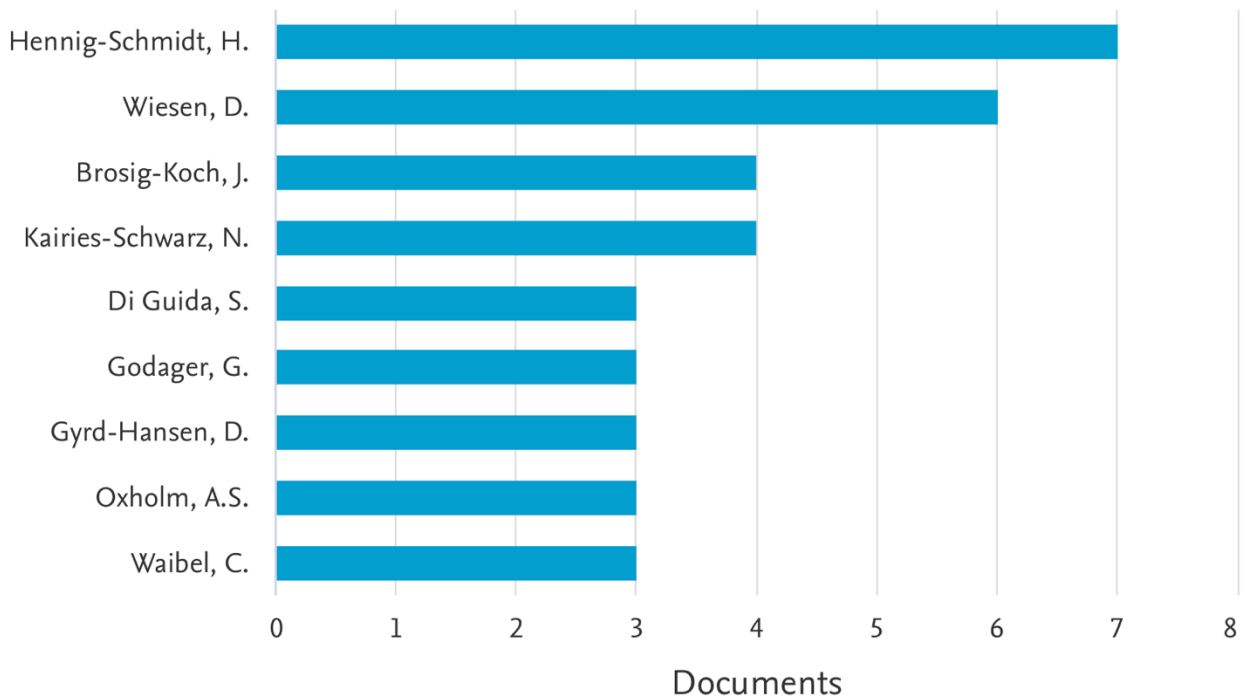
Source: Scopus' elaboration on our dataset

Figure 5 - Documents by subject area



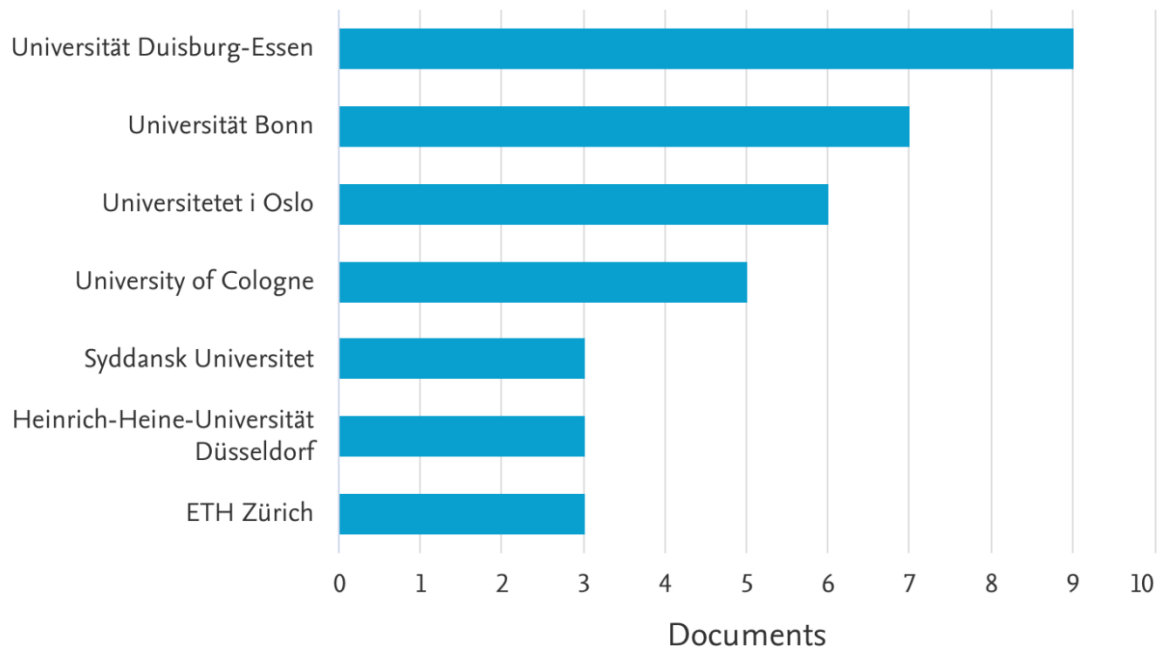
Source: Scopus' elaboration on our dataset

Figure 6 - Authors' productivity in terms of number of papers



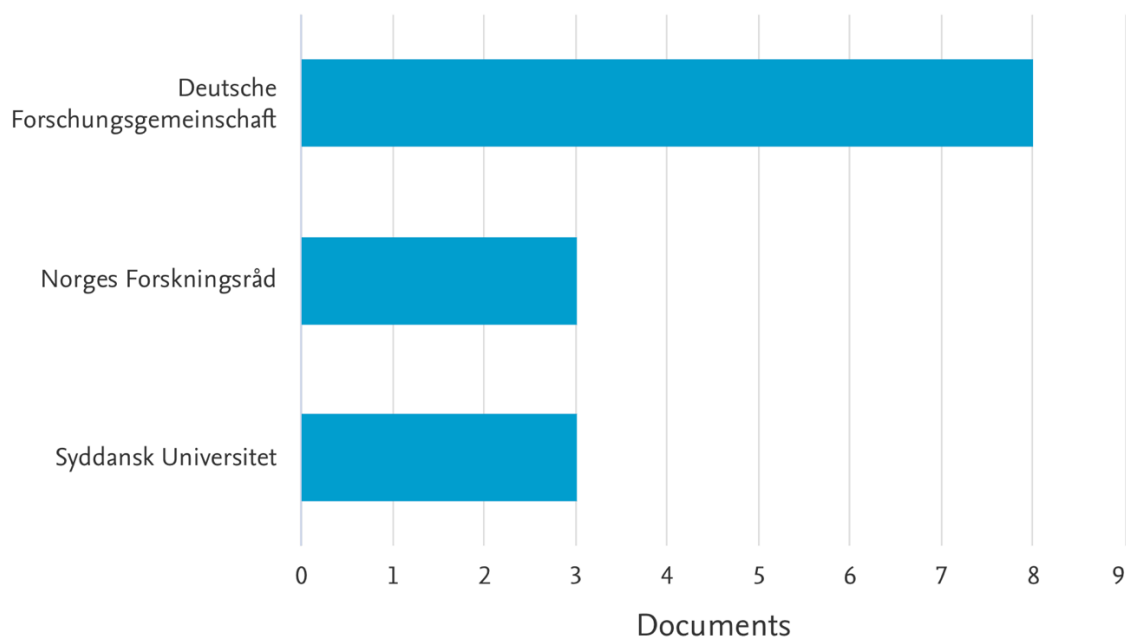
Source: Scopus' elaboration on our dataset

Figure 7 - Documents by affiliation



Source: Scopus' elaboration on our dataset

Figure 8 - Documents by funding sponsor



Source: Scopus' elaboration on our dataset

Table 2 – Studies in the sample that explore the role of payment schemes

Authors	Topic	Results
Bardey et al. (2021)	Personalized medicine and P4P, CAP and FFS	<ul style="list-style-type: none"> - Pay for performance (P4P) scheme incentivizes the adoption of personalized medicine compared to CAP and FFS. - Information on personalized medicine improves providers' performance regardless of the payment condition (though the effect is larger under FFS). - When information is costly, and once controlled for self-selection, personalized medicine increases the quality of care.
Brosig-Koch et al. (2016)	FFS and CAP	<ul style="list-style-type: none"> - Participants overtreat patients under FFS and undertreat them under CAP. - Physicians are less sensitive to payment incentives than students and medical students.
Brosig-Koch, Hehenkamp, Kokot (2017)	Competition between physicians and payment schemes (CAP and FFS)	<ul style="list-style-type: none"> - Competition reduces deviations from patient-optimal treatment, mitigating both overprovision in FFS and underprovision in CAP. - Competition effects depend on patient characteristics and payment conditions. - Tacit collusion arises from the repeated competition, particularly under FFS payment.
Brosig-Koch, Hennig-Schmidt, Kairies-Schwarz, Wiesen (2017)	CAP, FFS and mixed	<ul style="list-style-type: none"> - Participants overtreat patients under FFS and undertreat them under CAP, though to a less extent compared to predictions under profit-maximization assumptions. - Mixed payment schemes significantly reduce deviation from patient-optimal treatment level and improve patients' health benefits. - Altruistic behavior towards the patients varies on participants background (medical students are found more altruistic than nonmedical students).
Brosig-Koch, Kairies-Schwarz, Kokot (2017)	Physicians' preferences for FFS or CAP	<ul style="list-style-type: none"> - Most participants prefer FFS to CAP, regardless of their previous experience with of the two payment schemes. - Subjects preferring FFS are ex-ante and ex-post less patient-oriented than those choosing CAP.
Cox et al. (2016a)	Hospital readmissions rates and P4P	<ul style="list-style-type: none"> - While the use of deferred profit-sharing bonus (BO) payments increases readmission rates, the employment of instantaneous profit-sharing bundled (BU) payments does not. - The probability of overall unplanned readmissions increases when a patient is discharged earlier than recommended. - Pay-for-performance mechanism decreases hospital length of stay, though this effect is more marked when physicians are provided with evidence-based discharge criteria (clinical support system).
Di Guida et al. (2019)	FFS with different fee sizes, patient types and market conditions	<ul style="list-style-type: none"> - Patients are generally overtreated under FFS, though the effect varies on the patient's type. - Decreasing the fee size reduces overprovision, regardless of the patient's type. - Harmed patients, undergoing a decrease in the benefit of care due to overprovision, are at less risk of excess treatment. - Resource constraint mitigates overprovision compared to resource abundance.
Finocchiaro Castro et al. (2019)	Medical malpractice liability and payment schemes (CAP, FFS, and mixed)	<ul style="list-style-type: none"> - Participants overtreat patients under FFS and undertreat them under CAP. - The introduction of malpractice liability increases the amount of services provided regardless of the payment structure. - Subjects with a medical background react more aggressively to liability, especially under CAP. - The provision of medical services is also affected by the patient's severity of illness.
Godager et al. (2016)	Performance disclosure and FFS	<ul style="list-style-type: none"> - When performance information on physicians' performance is disclosed to peers, participants' likelihood of providing services in accordance with the medical norm or with the maximization of the joint benefit (patient and provider's) significantly increases.
Green (2014)	Financial incentives and P4P, CAP and FFS	<ul style="list-style-type: none"> - Under retrospective payments systems (i.e. FFS and FFS with P4P) physicians provide more medical services than under prospective payment systems (salary, CAP, CAP with P4P and CAP with report card). - Focusing on the quality of services, retrospective payment systems rule out participants' intrinsic motivations for patient's wellbeing.
Hennig-Schmidt et al. (2011)	FFS and CAP	<ul style="list-style-type: none"> - Participants overtreat patients under FFS and undertreat them under CAP. - Overprovision and underprovision behavior depend on patient's degree of illness. Under FFS, physicians overserve patients in a good and intermediate health status, while more severe patients and intermediate ones are underserved under CAP. - Patients' benefit loss is larger for patients in a good health status under FFS compared to CAP, while the opposite is true for intermediate and severely ill patients.
Hennig-Schmidt and Wiesen (2014)	Patient-regarding behavior and FFS and CAP	<ul style="list-style-type: none"> - Medical students are more patient-oriented in their provision behavior, eventually sacrificing their own profit in favor of patients' welfare. - Nonmedical students provide a higher (lower) volume of medical services under FFS (CAP) compared to medical students. - Medical students show to be influenced by the patients' health benefit in their decisions, while nonmedical students only look at their own payoffs.
Hennig-Schmidt et al. (2019)	Fraudulent behavior in healthcare and DRG reimbursement rates	<ul style="list-style-type: none"> - In the absence of audit mechanisms dishonest behavior (obstetrics' misreporting of child's birth weight which increases the obstetrics' payoff) is observed in three-quarters of decisions. - Introducing a random audit probability together with a fine reduces upcoding behavior when dishonest report is detectable.

Keser C. et al. (2020)	Custom-made healthcare and FFS and CAP	<ul style="list-style-type: none"> - Participants overtreat patients under FFS and undertreat them under CAP. Overprovision is higher for patients in good health-status, while underprovision is higher for high-severity patients. - Physicians customize their care according to the payment method. - Physicians are not affected by an ex-ante payment reduction in CAP. - Patients are worse off under CAP than under FFS.
Lagarde and Blaauw (2017)	Social incentives and FFS, CAP and salary	<ul style="list-style-type: none"> - Regardless of patient's benefits, FFS produces the highest quantity of output, and CAP the lowest. - Risk-adjusted CAP avoids patients cream-skimming. - The highest quality of output is achieved under salary, followed by CAP. - Social incentives (benefits to patients linked to the quality of work) improves providers' performance.
Martinsson and Persson (2019)	Altruism and FFS and CAP	<ul style="list-style-type: none"> - Physicians' attitude towards risk and ambiguity affects their provision of medical treatments under CAP. - Most physicians are altruistic towards the patients, though the degree of altruism varies on patients' need for care. - Both pure altruism and pure selfishness are more often observed under FFS than under CAP.
Oxholm et al. (2019)	Market conditions and CAP	<ul style="list-style-type: none"> - Underprovision of care in capitation-based scheme to severely ill patients exists regardless of resources availability. - High-severity patients benefit the most from a fixed salary to providers under resource abundance, while no difference between patient types is detected under resource constraint. - Physicians take better care of patients when differentiations to CAP are introduced, regardless of resource availability.
Oxholm et al. (2021)	Medical service provision and CAP and P4P	<ul style="list-style-type: none"> - Patients who can potentially reach the health target benefit from pay-for-performance system by receiving additional care, compared to patients without any potentiality who are provided with less care. - Physicians redistribute care between patients in P4P under resource constraint.
Reif et al. (2020)	Financial incentives and FFS and CAP	<ul style="list-style-type: none"> - Payment systems affect physicians' reporting and provision behavior. - Physicians care about the payoffs of a third party which funds medical service provision and reduce the amount of services provided to save costs for the third agent. - Participants are more patient-oriented in the medical framing, compared to the neutral framing.
Waibel and Wiesen (2021)	Referral rates	<ul style="list-style-type: none"> - Diagnostic effort does not vary on the referral fees. - High referral rates increase referrals regardless of physicians' degree of altruism. - Compared to a baseline condition without referral fees, the introduction of medium-size referral rates incentives referrals for barely altruistic primary care providers of severely ill patients.
Wang et al. (2020)	Patient-regarding preferences and FFS and CAP	<ul style="list-style-type: none"> - Patient-regarding motivations do not differ across the subject pools (Chinese and German physicians, Chinese medical students). - Experience makes physicians rational in the decision-making.

Source: our elaboration on Scopus database

Table 3 – Other topics in the provision of health services explored in the sample

Authors	Topic	Results
Ahlert et al. (2012)	Distributive behavior	<ul style="list-style-type: none"> - Economists behave more often as payoff maximizers in the neutral framing than in the medical one. - Physicians are less sensitive to the experimental setting but are more willing to maximize their payoffs in the medical frame. However, they are generally more likely to maximize the number of recipients or to behave according to the Rawlsian rule.
Angerer et al. (2021)	Monitoring	<ul style="list-style-type: none"> - In the absence of liability and verifiability, undertreatment and overcharging are detected. - Both endogenous monitoring and exogenous monitoring reduce the level of undertreatment and overcharging observed and improve market efficiency.
Arrieta et al. (2017)	Risk preferences	<ul style="list-style-type: none"> - Risk tendencies are health-context specific. - Students with a medical background are more risk-averse, especially in the health domain. - When subjects decide for others, introducing benefits to the third party reduces the level of risk aversion.
Brendel et al. (2021)	Resource scarcity and prioritization decisions	<ul style="list-style-type: none"> - Most participants allocate a constant portion of their budget to patients and then reduce the amount of services provided in response to significant budget reductions. - Most physicians provide equal benefits between patients. - The less resources are available the less are patient benefits.
Brock et al. (2016)	Generosity and prosocial behavior	<ul style="list-style-type: none"> - Physicians who show prosocial behavior in the laboratory are also generous in their normal practices. - Physicians provide better performances in response to peers' monitoring and encouragement. This effect is equally shared between generous and ungenerous clinicians.
Cox et al. (2016b)	Hospital discharge	<ul style="list-style-type: none"> - Clinical decision support systems (CDSS) in the form of recommendations on patients' discharges reduce readmissions rates and patients' length of stay. - CDSS promote time efficiency in making discharge decisions and improve participants' performance as measured in terms of experimental earnings. - Subjects perform better under time constraint. - Subjects appreciating CDSS are more likely to provide better discharge decisions.
Ge and Godager (2021)	Strategical decision-making and market competition	<ul style="list-style-type: none"> - Patients benefit from larger competition. - Higher competition drives to individuals' deterministic behavior.
Greiner et al. (2017)	Prescription behavior	<ul style="list-style-type: none"> - Separating prescription and treatment activities between physicians reduces overtreatment behavior and increases patient's willingness to accept severe treatments. - When physicians provide diagnosis free of charge, undertreatment is observed, together with patients' reluctance to accept mild treatments.
Han et al. (2017)	Quality competition and hospital mergers	<ul style="list-style-type: none"> - Average quality following a merger is lower than pre-merger quality. - Participants' choices on quality level are significantly higher than predicted for pure profit-maximizer hospitals (maybe due to altruistic behavior towards the patients). - When the merger leads cost synergies, average quality choices increase compared to the scenario without synergies. - Results do not change between individual and team decisions.
Herr and Normann (2016)	Organ donation	<ul style="list-style-type: none"> - Two-thirds of the participants show stronger preferences for the priority rule (a reciprocal rule which prioritizes registered donors). - Priority rule increases donors' registrations. - Medical students register more often as donors and opt for the priority rule more frequently compared to students with other backgrounds. - When asked about the desirability of the priority rule in the field, participants of the donation experiment vote for such rule more often than non-participants.
Huck S. et al. (2016)	Medical insurance	<ul style="list-style-type: none"> - Under the insurance condition (patient shares the cost of the treatment with all the other patients), patients more frequently consult the physician, while physicians are more likely to overtreat the patients. - The introduction of competition (patient can choose which physicians to be assigned to) mitigates excess consultations and overtreatment. - When insurance and competition interact, efficiency increases and patients are likely to receive the required treatment.
Kessler and Roth (2014)	Organ donation	<ul style="list-style-type: none"> - When information about the others' donations and use of the loophole rule (i.e. subjects can register to get a priority but simultaneously refuse donating organs) are made public, subjects are less willing to donate. - The priority allocation rule increases donations.
Kolstad and Lindkvist (2013)	Prosocial behavior and self-selection in the public sector	<ul style="list-style-type: none"> - Medical students preferring to work in the public sector show more pronounced prosocial preferences than those opting for the private sector.
Martin-Lapoirie (2021)	Teamwork and Medical Malpractice	<ul style="list-style-type: none"> - Strict liability and the negligence rule (i.e the physician has to compensate the patient only if he has demonstrated negligence in at least one consultation) lead to similar precaution behaviors. - Healthcare providers choose a positive precaution behavior even in the absence of liability.

		<ul style="list-style-type: none"> - Strict liability is less efficient than the negligence rule in reducing deviations from social optimal precaution level. - Teamwork reduces precaution behavior under liability rules.
Mimra et al. (2016)	Second consultations	<ul style="list-style-type: none"> - Introducing the possibility for a patient of asking for a second opinion at a cost reduces overtreatment. - Reduced search costs incentivize second consultations and improves market efficiency. The reduction in the cost of treatment outweighs the increase in search costs.

Source: our elaboration on Scopus database