Università degli Studi Mediterranea
di Reggio Calabria
Dipartimento di Giurisprudenza, Economia e Scienze Umane
Dottorato di ricerca in Diritto ed Economia
XXXIII ciclo

Three essays on experimental health economics

Tesi di Dottorato:
Domenica Romeo

Tutor:
Ch.mo Prof. Massimo Finocchiaro Castro

Coordinatore:
Ch.ma Prof.ssa Carmela Salazar

ANNO ACCADEMICO 2019-2020

Table of Contents

Introduction

This thesis is divided in three chapters, each referred to a different experiment, connected by a common theme which is medical doctors' decisions. The framed field experiments were conducted at the metropolitan hospital of Reggio Calabria, thanks to a formal agreement signed by the same hospital and the Mediterranean University. According to such agreement, the parties involved are committed to participating in the research project 'Experiments in health economics' and to sharing results which could be useful to the hospital for management policy. I have been, then, allowed by the interested chief administrators to enter in several departments for 3 months during 2018-2020 (e.g. genecology, orthopaedics, surgery, otolaryngology and so on) in close contact with the hospital physicians.

In the first experiment, I investigate for the first time the level of coordination among the Italian health professionals, by using a simple, portable and incentive compatible tool such as the coordination game, widely adopted in previous literature, though in other settings. By identifying areas where relational coordination plays an important role, I have chosen to focus on orthopaedics, paediatrics and oncology. For each of the afore mentioned departments, I designed a specific vignette (i.e., a description of a hypothetical situation which physicians could face on their job) and a set of four actions which a physician can take in response to that. In this framed field experiment, physicians are asked to evaluate the appropriateness of each of the possible actions on a scale of one to four in order to match the valuations of most respondents (i.e. modal answer). Data show that the frequency mean of coordination across the experiment is 52% and that coordination increases when physicians often exchange opinions and share positive feedbacks with colleagues. In addition, the presence of a leader in the medical department, facilitates coordination. Evidence provided by this study shows that coordination in the Italian health sector could be missing, which makes the need for action all the more acute.

In the second chapter, built on the same experiment as in Chapter 1, I test whether physicians act according to national guidelines. Adding a second step to the coordination game, by submitting the same scenarios proposed in the first experiment, physicians can receive information concerning what it should be done, in the specific situation, according to national guidelines (i.e. providing them with the written guidelines) and eventually change decisions. Results show that only 8% of the physicians did not want to know national guidelines content.

However, only 21% of the subjects (with a choice frequency of 23%) decided to change their appropriateness judgment after realizing they were in contrast with guidelines in either one or two vignettes. Overconfidence and perceived guidelines ambiguity could explain such a low value. Finally, the more physicians consult scientific sources (i.e. signal of open-mindedness), the more willing to accept suggestions, coming from guidelines reading, and to change their decisions when they are wrong. Since guidelines dissemination is shown to increase the level of coordination between physicians, hospitals should consider effective programmes to spread their knowledge.

Finally, in the third experiment, I test whether and to which extent the adoption of fee-for-service or salary system can induce physicians to practice patient dumping. In an artefactual field experiment, physicians facing the possibility of being sued for medical malpractice decide whether or not to provide medical services for patients with different state of health. Also, I check whether the introduction of the risk of being sued for a physician for having practiced dumping can have effect on his behaviour. Results show that dumping is more often observed under Salary than under Ffs. However, physicians seem to be insensitive to the introduction of dumping liability under the same incentive mechanism, though it seems to trigger a higher amount of services provided. Policy strategies concerning the incentive scheme may vary on the hospital purposes.

CHAPTER 1

Assessing relational coordination in the health sector: an experimental approach

Abstract

Coordination is the key to the success of any organization. In the healthcare sector, departments with higher level of coordination result in greater promptness and quality of care and lower mortality rates. Sharing of codes of conduct and procedures has been shown to reduce postoperative pain and improve postoperative functioning. For these reasons, enhancing coordination between health professionals is becoming central in many countries healthcare policy, including Italy. However, despite the various attempts to encourage care coordination, Italian health sector continues to be characterized by a high level of fragmentation which led to a waste of 2,58 millions of euros in 2017. In this study, I investigate for the first time the level of coordination among the Italian health professionals in the context of the hospitals of Reggio Calabria, by using a simple, portable and incentive compatible tool such as the coordination game (Krupka-Weber norm elicitation task), widely adopted in previous literature, though in other settings. Data show that the average frequency of coordination across the experiment is 52% and that coordination increases when physicians do often exchange opinions and share positive feedbacks with colleagues. In addition, the presence of a leader in the medical department, facilitates coordination. Evidence provided by this study shows that coordination in the Italian health sector could be missing, which makes the need for action all the more acute.

Introduction

Coordination is the key to the success of the organization (Webb, 1991): it ensures unity of action and integrates different activities, resources and structures in an organization context (Beuselinck et al., 2007).

Modern day organizations are becoming increasingly dynamic and complex: people working in an organization can have different backgrounds, work attitudes, knowledge and skills. For this reason, managers have to give a common direction to the efforts of all the individuals working in different departments to accomplish the objectives of the business (Lewis, 2006). According to Mangham (1986), when individuals are willing to coordinate to achieve a common objective, and benefit from the process of joint work and coordination, organizations reach their maximum efficiency. In the healthcare sector, coordination becomes even more important than in any other organization. Specifically, relational coordination, defined as the "sharing of codes of conduct and procedures[1]" among health professionals (Gittell, 2000) has been shown to improve some dimensions of performance, including emergency and intensity care (Fargason and Haddock, 1992). In other words, quality of care is significantly improved by relational coordination. In their study, Gittel et al. (2000) find that such a coordination (see footnote 1) significantly reduces postoperative pain and improves postoperative functioning[2]. Moreover, the average length of hospital stay is significantly shortened by the frequency of communication among care providers. Other studies (see e.g. Argote 1982, Baggs et al. 1992, Shortell et al. 1994) show that coordination in hospital emergency units improves promptness and quality of care and reduces mortality rates. On thecontrary, departments with low levels of coordination result in the worst performance[3].

1 Coordination in a broader sense encompasses frequency and accuracy in communication and even problem-solving capacity (Gittel et al., 2000).

[2] According to the National Veterans Administration Surgical Risk Study, coordination is said to reduce surgical morbidity rates, a state of illness or lack of mental or physical health (e.g. permanent disability) found during or after an operation (Jacobs et al., 2007).

[3] For example, according to the Country Health Profile (2017) cardiovascular diseases which are the leading cause of death among women in Luxembourg, and second to cancer for men are attributable to the lack of an integrated and coordinated system of care for chronic patients.

The purpose of this study is to measure the level of relational coordination among healthcare providers within the Italian health organization and, additionally, to monitor whether common practices are adopted in such a setting, developing a simple, portable and incentive-compatible tool, based on the Krupka-Weber norm elicitation task (see e.g. Gaechter et al. 2013; Barr et al. 2018; Burks and Krupka 2012). In order to detect whether a group of people share a common understanding of the practices and rules that apply to specific decision situations, Krupka and Weber (2008), for the first time, adopted a special type of coordination game in which people are incentivized to "coordinate" with others in evaluating what constitutes "appropriate behaviour" in a given situation. A similar approach can be used to investigate whether in the hospital physicians coordinate, facing the same scenario.

In fact, in recent years, enhancing coordination among health professionals is becoming central in many countries healthcare policy, especially (Romoren et al., 2011). For example, in France healthcare sector encouraged communication between health professionals and in particular between general practitioners and hospital specialists with the aim of improving care coordination (Simonet, 2008). Health networks were introduced in Germany and Switzerland to coordinate the activities of different health professionals responsible for a group of patients suffering from the same illness (Ratajcak, 1998). Scotland and Denmark incentivised collaborations between home care, primary care and acute (hospital) care to support integrated care-provided[4], and defined agreed goals and outcomes, as well as shared procedures and standards (European Union Report, 2017).

Italy, context of reference for the purpose of this study, tried to reinforce care delivery integration through collaborations between general practitioners and local sanitary units (Del Torso et al., 1997). More generally, in Italy new models of community care services (e.g. community care networks) were experimented to achieve greater coordination thanks to the 'Legge Balduzzi' (Law 189/2012), which facilitates the establishment of networks of healthcare providers (OECD Report, 2015).

[4] In this paper I adopt the same definition provided by the European Union Report (2017) for integrated care: 'Integrated care includes initiatives seeking to improve outcomes of care by overcoming issues of fragmentation through linkage or coordination of services of providers along the continuum of care.'

However, despite the various attempts to encourage care coordination, Italian health sector continues to be characterized by a high level of fragmentation (OECD Report, 2015). Due to a regionally based National Health Service[5] and in order to deal with the endemic factor, different regions adopt different rules and regulations, giving rise to different medical outcomes (Tedeschi, 2010). However, the management of the endemic factor exacerbated such differences, considered to be primarily responsible for a poor care coordination at a national level. In this regard, Gimbe Foundation[6] estimates that inadequate coordination in the health sector led to a waste of 2,58 millions of euros in 2017 (Cartebellotta et al., 2019).

According to a survey carried out by the subgroup of integrated care, Italy has developed a specific set of indicators to explicitly measure aspects of integrated care[7], which could allow to measure what is the current extent of coordination in the health sector (European Union Report, 2017). However, information infrastructure is not being utilised sufficiently, due to a missing link between different data and a poor use of the medical record (OECD Report, 2015). Thus, it becomes difficult to measure coordination inside the hospital. Therefore, the aim of this paper is to use a new approach, informed by recent, cutting-edge research in behavioural economics, to tackle the lack of information about the extent and nature of miscoordination in the Italian healthcare system.

Based on the insights gathered from some physicians who supported me in designing the experiment, by identifying areas where relational coordination plays an important role, we have chosen to focus on orthopaedics, paediatrics and oncology. For each of the aforementioned departments, we designed a specific vignette[8] (i.e., a description of a hypothetical situation which physicians could face on their job) and a set of four actions which a physician can take in response to that. In this framed field experiment, physicians were asked to evaluate the appropriateness of each of the possible actions on a scale of one to four in order to match the modal valuation (i.e. the valutation with the highest frequency) for each vignette. Although

[5] Further details are provided in 'The Italian healthcare system' section.

[6] Italian Group for the Evidence-based medicine.

[7] Integration is measured both horizontally (i.e. between services which are on the same level of the healthcare process) and vertically (i.e. between organizations belonging to different levels of the hierarchy such as primary care and secondary care) (European Union Report 2017).

[8] Each vignette has been designed by a physician specialized in that specific discipline. Further details will be provided in section 5.1.

physicians are asked to individually rank the options, the incentive structure induces them to match the modal rating, trying to guess what their peers think for each of the alternatives.

Results show that the average frequency of coordination across the experiment is 52% and that coordination increases when physicians exchange opinions and share positive feedbacks with colleagues. In addition, the presence of a leader in the medical department facilitates coordination. Moreover, the results suggest that physicians who have been working longer for the same hospital are more likely to implement the outcomes prescribed by the guidelines.

Overall, evidence provided by this study shows that coordination in the Italian health sector could be missing, which makes the need for action all the more acute.

Literature review

This study contributes to at least two strands of literature[9]. On the one hand, since the main purpose is to test whether physicians agree on what is the most appropriate action to be taken in each situation, this study endorses the trend of the coordination games used in experimental economics to this end. On the other hand, it combines the existing literature on social norms, since it tries to identify clinical practice guidelines comparable to social norms in this setting, as it is shown in detail in the second chapter.

As outlined by Hermalin (1999), each company has its own social rules, derived from culture and tradition, and sets its objectives according to them. Despite being underestimated, such rules are meant to drive entrepreneurial choices and therefore need to be measured in order to explain how they affect actions.

Although research to elicit social norms inside the organizations was fast growing (see e.g. Perkins and Wechsler 1996; Kanazawa and Stil 2001), it relied on survey-only approach which lacked the reward medium needed to incentivize respondents to reveal their true beliefs/preferences. Moreover, many economists who used laboratory experiments to identify norms were not able to distinguish between norms referred to different groups or individuals or to link norms to subjects' behaviour (e.g. Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000).

[9] Further literature is provided in Chapter 2.

Krupka and Weber (2008) filled this gap by using a coordination game in which people are incentivized to "coordinate" with others in evaluating the appropriateness of different actions in a given situation. In both a standard and a bully variant version[10] of the dictator game, participants are asked to rank the alternative distributions of the pay- offs between themselves and another person, according to their perceived degree of social appropriateness on a four point scale, where 1 corresponds to 'very socially inappropriate', while 4 corresponds to 'very socially appropriate'. In order to be rewarded at the end of the experiment, subjects have to assess each of the options so as to match with the modal response provided by the other participants facing the same task. Because people are incentivized to report what the majority of respondents think is appropriate, and not their own personal opinion, the elicitation procedure allows to measure the extent to which there is a common understanding among a group of people of the appropriateness of behaviours in a specific situation[11]. In fact, results show that people care about social norms and want to behave consistently with them. Similar results are obtained in the experiment by Krupka et al. (2016). Social norms are elicited using the Krupka and Weber's approach (2008) for two different games, a Double Dictator Game and a Bertrand game, before letting participants really playing them. In the former, matched subjects simultaneously decide on the distribution of their endowment. In the latter, partnered subjects simultaneously report a number between 0 and 100 knowing that the person selecting the smallest number would get a payoff corresponding to that number, giving the other player a payoff of 0. Subjects are asked to ethically evaluate each of the possible actions available to a hypothetical individual facing each of the games. Data show that subjects are significantly affected by social norms in their decisions. The same procedure is applied in a three-person gift exchange game by Gaechter et al. (2013). In their experiment, a principal pays a wage to each of the two agents involved in the 'game'. Then, the agents sequentially decide the level of effort to put into the tasks. Subjects are asked to assess the appropriateness of a set of actions available to the different players (i.e. the principal and the two agents). Results from the norms elicitation task are consistent with social norms compliance.

[10] While in the standard version the decision maker initially receives $10 while another person gets $0, in the bully variant version, both the decision maker and the other person initially receive $5.

[11] Notice that the incentive mechanism so far described is common to all the experiments.
I am discussing through this section, from now on.

Another experimental contribution in this regard comes from Barr et al. (2018) who measure the social appropriateness of discrimination using the now known norms elicitation procedure. Participants are divided into two identity groups according to their nationality, either British or Chinese in the first treatment, and according to the colour of the ball they pick from an envelope, in the second treatment. Then they are described the allocator game, where an allocator decides how to distribute an amount of money between two players belonging to two different groups. Each subject reports how appropriate he considers each of the possible actions an allocator can take, so as to match the judgment of appropriateness of another randomly selected participant belonging to the same identity group. Results evidence that discrimination is perceived as socially inappropriate.

Finally, under the strong assumption that economic organizations are driven by social norms and that employers are likely to conform with them, Burks and Krupka (2012) use a methodology slightly different to that adopted in the experiments so far described. The novelty of their approach lies in the attempt to identify social norms using the elicitation technique first introduced by Krupka and Weber (2008) to link them to behaviour in the business setting of a large financial service firm in the U.S.A, which provides investment services to customers. In their field experiment, financial advisers and corporate leaders are presented with three different vignettes which describe hypothetical situations they could face on the job. Participants are asked to ethically evaluate each of the possible actions, which a financial adviser can adopt in response to the situation in order to match the ethical valuation of an anonymous other person facing the same scenario. Since the identity of the matched person varies throughout the experiment, the authors can make a distinction between the norms held by distinct teams within the corporate hierarchy (i.e. financial advisers and their managers). Results show a general tendency for corporate leaders to predict the ethical appropriateness valuations of their peers. Similar results are gleaned for financial advisers with an exception: extreme actions lead to some subtle misalignment across this group. When the financial advisers' interests could conflict with the firm objectives, some form of misalignment occurs between the corporate leaders' expectations about the financial advisers' norms and their actual norms.

While we draw from the above-mentioned literature for this experimental design, none of these experiments has been run in the health context. Therefore, to the best of our knowledge, this the first study to apply experimental methodology to measure the level of coordination in the

healthcare sector and to investigate, in the second chapter, whether physicians adhere to common practices or guidelines, which should preferably be applied to the specific clinical cases.

The Italian healthcare system

Since their birth, all Italians are enrolled in the National Health System (NHS), enacted in 1978, and have right to choose a paediatrician or GP in accordance with their place of residence. Starting from the 1999 reform and in order to delegate more authority to the regions, decentralization was adopted, giving rise to a regionally based National Health Service (Simonet, 2008). Italy was divided into several local sanitary units, responsible for the healthcare management (i.e., resources allocation to the hospitals, assessment of the effectiveness of local care providers). Such units were divided, in their turn, into small-scale autonomous sanitary enterprises providing care at the distinct level (Tedeschi, 2010). As a result, the allocation of the decision- making power was redefined. According to the Italian Constitution, the State continues to ensure access to free or low-cost healthcare, such as treatment at public hospitals, to all citizens and residents since their establishment. However, responsibility for healthcare is now shared between the State and the 21 regions (France et al., 2005). In particular, the Central Government is responsible for defining the general objectives and the leading principles of the national healthcare, while regions are in charge of the provision of health services.

Although decentralization was aimed to control healthcare expenditure and facilitate the implementation of public health programs, it also comes with a couple of drawbacks. Regions differ for demography, social and economic development, organization and healthcare spending. For this reason, national initiatives aimed at improving heathcare quality are not uniformly implemented at regional level. In particular, regions tend to adopt different rules, different regulations and different health-care infrastructure. However, central agencies are unable to monitor the different regional activities and the resulting administrative complexities

of a regionally based healthcare system lead to a waste of 2,48 millions of euros in 2017 (Cartebellotta et al., 2019).

The above-mentioned regional differences resulted in a clear North-South divide (Landrain, 2004). Typically, regions in the North provide better infrastructures and deliver higher quality healthcare, based on the available data. For this reason, a large number of patients moves from the south to the north of Italy to seek better care (OECD Report, 2015). Furthermore, the adoption of a regionally based National Health Service exacerbated coordination problems which Italian healthcare system suffers from (Tedeschi, 2010). An increasing number of patients complained about care coordination, reporting miscommunication between different professionals (e.g., a primary doctor and a specialist). In addition, they experienced some problems such as receiving conflicting information from different doctors or not having a recommended medical test (OECD Report, 2015).

Especially in recent years, there are many initiatives carried out by the Central Government intended to improve care quality and uniform procedures and rules so as to fight both North-South divide and the lack of coordination between different health organizations and different medical professionals. Public authorities tried to achieve better integration between GPs through networks, associations and other forms of grouping. Interprofessional collaborations between GPs, paediatricians, nurses and specialists were introduced into the National Agreements of 2009 (Tedeschi, 2010) in order to improve coordination and continuity of care for patients. Moreover, starting from 2005, the National Agreement for primary care for health professionals defined objectives reflecting evidence-based guidelines and professional standards. However, regions could introduce further objectives and were free to manage their health service provision, in respect for the national health plan.

Adding to the national initiatives, in his report (2015) the Organization for Economic Co-operation and Development (OECD) made some recommendations to improve the quality of healthcare in Italy and encouraged the coordination of the different regional activities by central government agencies. The primary objective is supporting the adoption of guidelines at regional

level and monitoring their impact on the medical outcomes. According to the OECD (2015), the State should incentivize full adherence to guidelines[12], together with communication and teamwork between health professionals so as to facilitate care coordination in response to the challenge launched by the demographic and epidemiological change. However, despite the existence of national guidelines, each hospital implements such guidelines but maintains its own long-established common practices and protocols (France et al., 2005). For example, Mapelli and Lucioni (2003) reported that 92 ASL (local sanitary units) produced 377 guidelines on chronic diseases. Therefore, since internal guidelines could somehow differ from hospital to hospital, it is hard to uniform medical treatment at the national level and to achieve care coordination which this study tries to test.

Theoretical framework

In this section, I develop a simple model to formalise the relationship between evidence-based guidelines and personal norms into the physician's utility function. Although I largely adapt the utility framework first introduced by Krupka and Weber (2008) and then readjusted by Burks and Krupka (2012), since this model specifically targets physicians, I am taking a step forward for our experimental purposes.

To derive this framework, I assume that each physician in face of a specific clinical circumstance forms his own diagnostic hypothesis based on his own experience. According to Samuels and Ropper (2005), thanks to their clinical experience, physicians gain subtle skills which might not be captured by evidence-based measures. In fact, physicians seldom resort directly to guidelines in their clinical decision making, preferring to rely on their 'mindlines', guidelines-in-the-head, acquired over a lifetime and informed by their conversation with colleagues, their interactions with patients, their early training and their readings (Gabbay and le May, 2004). Supporting this view, in his study Elstad et al. (2010) show that clinical experience brings physicians social, behavioural and intuitive skills and knowledge which

[12] According to the Institute of Medicine (1990), guidelines assist practitioners about diagnostic and treatment modalities for specific clinical circumstances.

allow them to meet patients' needs and to compare present day patients with similar past patients. However, physicians are said to consult outcomes research too, depending on their experience level (Tannenbaum, 1994). Indeed, more experienced physicians are less willing to adhere to practice guidelines for several reasons. First of all, copious, ever-changing and often contradictory guidelines are hardly able to take into account all the factors which come into play when physicians have to make clinical decisions (Gabbay and le May, 2004). Moreover, the longer physicians have been in practice, the lower their ability to incorporate new guidelines. In particular, their experience leads them to understand types of patients as well as the progress of their disease and this contributes to increase physicians' cognitive rigidity (Choudhry et al., 2005). On the other hand, less experienced physicians who cannot depend on such acquired skills are more likely to welcome new information provided by practice guidelines (Elstad et al., 2010). However, since adherence to guidelines reduces the risk of being sued for malpractice, physicians strive to conform to them regardless of their experience level (Elstad et al., 2010). In fact, since practice guidelines define the standard benchmarks for medical treatments, they support the law governing medical malpractice in assessing physicians' conduct and hence eventually identifying cases of negligence (see Gabbay and le May, 2004 and Havighurst, 1991). In the end, even though physicians can get insurance to reduce their financial risks, they carry many non-insurable costs incurred for malpractice litigation such as psychic, time costs of legal proceeding and the risk of undermining their reputation (Currie and Macleod, 2006) which they try to minimize by following guidelines.

Although our focus is not on physician's mindlines measurement, the likely mismatch between them and the national guidelines, as suggested by the literature, could explain physician's lack of coordination.

Based on the above, I can derive a framework to model the relationship between physician's mindlines, which replace personal norms in the original Krupka and Weber's model (2008), practice guidelines, which epitomize group norms (and then social norms introduced in the previous sections), and physician's behaviour.

I assume that a physician's utility function is given by:

$$U(a_k) = V(\pi(a_k)) + N_p(a_k) + \gamma N_g(a_k) + \boldsymbol{\theta} N_n(a_k) \qquad (1)$$

where:

- $V()$ represents the value given to the monetary payoff which increases in $\pi(a_k)$;
- $N_p(a_k)$ indicates to which extent physician's action is appropriate according to his mindlines given by his clinical experience (i.e., his social, behavioural and intuitive skills and knowledge), epitomizing an internal norm;
- $N_g(a_k)$ reflects the degree of appropriateness of an action according to the hospital internal guidelines[13] which constitute the own workgroup's relevant norms (e.g. norms generally shared by colleagues);
- $\gamma \geq 0$ dictates physician's sensitivity to the adherence to internal guidelines;
- $N_n(a_k)$ refers to the degree of appropriateness of an action following national guidelines, which constitute the standard procedures used by the tort system to investigate on physician's diligence;
- $\boldsymbol{\theta}$ stands for physician's sensitivity to the obedience to national guidelines.

Even though, internal guidelines and national guidelines should coincide, in recent years, heterogeneity in the use of procedures for specific medical conditions has been documented at a national level (Switzer et al., 2003). In particular, the production of clinical guidelines had an exponential growth so much that physicians, patients and other stakeholders have to juggle several guidelines which vary in quality and are sometimes discordant (Camera civile di Firenze, 2017). As confirmed by Mapelli and Lucioni (2003), in Italy many ASL (Local sanitary units), based on their common practices, have given rise to their own local guidelines to be followed by their employees (e.g., Emilia Romagna), which in some cases deal with clinical circumstances not covered by national guidelines. For this reason, physicians have to bear in mind both national and internal guidelines, when both exist, while treating the patient, without

[13] Notice that in the final demographic survey I asked participants whether or not the hospital which they work for adopt his internal guidelines.

neglecting their own clinical experience. Whether all these norms do not overlap, physicians decide what is the best strategy to adopt depending on the peculiarity of the clinical case as well as their experience level (Elstad et al., 2010).

In a nutshell, according to the afore-mentioned model, a physician has higher utility from an action whenever he perceives his mindlines to be the closest possible to both internal guidelines and national guidelines when such '*norms*' do not totally match (see the second chapter for further discussion).

Hypotheses

On the basis of equation (1), we can derive the hypotheses to be tested in this experiment.

As stated previously, physicians take into account their mindlines in making important clinical decisions. Such mindlines might vary from physician to physician, being the results of different kinds of knowledge, experience and values (Gabbay and le May, 2004). However, when physicians realize that they are in disagreement with their respected colleagues, their mindlines morph. By using the ethnographic methods of anthropologists, Gabbay and le May (2004) investigate practitioners' behaviour and beliefs and show that physicians tend to change their mindlines while interacting with trusted colleagues. According to the same authors that is where coffee-room chat comes into play: physicians share their stories and their experiences so as to help each other solve clinical problems.

In summary, physicians do not like having a different view from that of their trusted colleagues on what is the right thing to do. As a consequence, since the experiment ask them to coordinate on what they believe that others consider most appropriate, I expect that physicians coordinate in judging the appropriateness of each of the actions proposed in this experiment as alternative solutions to solve the specific clinical case. Since physicians come from the same hospital, they face the same hospital internal guidelines. In particular, the task elicits physicians' beliefs about their peers' evaluations, identifying the actual 'norm' regulating them, either local or national guidelines, which is the same for any subject in such a setting (only local guidelines could vary, though it is not the case). Unlike what happens into Gabbay and le May's study (2004), here physicians are not allowed to talk each other and sharing their ideas. However, even though

participants are not able to exactly predict how each of their colleagues is going to answer to each of the proposed vignettes, they are able to form an idea of the most likely modal answer and they might not want to depart from it. Thus, for each of the actions I do not expect that physicians' valuations significantly differ, apart from their different medical specialty.

Behavioural hypothesis 1: *For each of the actions, physicians coordinate in judging their degree of appropriateness*.

As far as the second hypothesis is concerned, I already mentioned the role of practical guidelines into physicians' decision-making process. Even if they are not taken as directives, being unable to consider the multiple factors which play a part in clinical decisions, they generally identify recommended courses of action under certain circumstances (Institute of Medicine, 1990). Having been trained into the same national context, I can safely assume that all the 52 physicians are able to recognize the presence of the national guidelines in this experiment, since one of the four proposed actions to be taken for each of the vignettes has been set according to them. In particular, I expect that physicians coordinate in giving the same appropriateness judgment to the actions conformed with guidelines. Moreover, I expect that the frequency of reporting such actions to be 'very appropriate' is higher than the same frequency for any other action not prescribed by guidelines. In other words, the statement 'very appropriate' is assigned much more frequently to the actions prescribed by the national guidelines, than to the other actions.

Behavioural hypothesis 2: *For each of the actions prescribed by the national guidelines, physicians coordinate in judging their degree of appropriateness*.

Behavioural hypothesis 3: *When an action is prescribed by national guidelines the frequency of reporting such action to be 'very appropriate' is higher than the same frequency for any other action not prescribed by guidelines*.

Design

The experimental design consists of two modules. The first part assesses the level of physicians' coordination over the individual evaluation of the appropriateness of each action (module elicits national/internal guidelines through the coordination game). Then, the second and last module provides physicians' important information such as their medical specialty and their years of professional experience.

The order of the tasks does not vary across the subjects. In order to avoid both sequence effect and carryover effects (Charness et al., 2012) subjects are informed about their individual earnings of the first part only at the end of the experiment, once they have completed the questionnaire. In the following sections I am going to describe the first part in detail.

Coordination Game

To measure physicians' level of coordination and to detect national and internal guidelines, as it will be done in the second chapter, I used the technique introduced by recent experimental literature to elicit norms (see e.g. Gaechter et al. 2013; Barr et al. 2018; Burks and Krupka 2012). Such a technique combines the description of hypothetical situations through ad hoc vignettes with the coordination game structure.

After discussing with three physicians specialized respectively in orthopaedics, paediatrics and oncology[14], I designed three vignettes, one for each of the above-mentioned fields of healthcare. Moreover, I asked them to check the final version of such vignettes and to ex ante rank each of the possible actions, according to both national guidelines[15] and their personal experience. Finally, before running the experiment, I asked for the opinion of a general practitioner working

[14] For the sake of clarity, it should be specified that these physicians were external subjects who did not take part in the experiment but just cooperated to design it.

[15] Even though these physicians had no reason to lie, I compared their rankings with the written national guidelines for preventive purposes.

as a paramedic at an emergency department. He confirmed that the scenarios were realistic and easy understandable to any physician regardless of his medical specialty.

After developing the vignettes, describing hypothetical dilemmas, which physicians could face in their job, we handed them out to the participants. Physicians were presented with three scenarios and a set of four actions which a practitioner can take in response to them. Subjects were asked to indicate whether each of the four possible actions was 'very inappropriate', 'somehow inappropriate', 'somehow appropriate' or 'very appropriate' on a scale of one to four. In the instructions, it was specified that an action is appropriate when an average physician believes that it the correct thing to do given the specific situation. Although physicians were required to individually rank the options, the incentive structure induces them to match the modal rating, trying to guess what their peers think for each of the alternatives. In fact, to determine whether coordination is achieved, for each of the alternatives, the individual rating is compared with the modal rating for that option and if the twos overlap coordination is successful.

The vignettes were written from the perspective of a physician facing a patient suffering from a particular disorder or disease.

For simplicity, in this section I am going to focus on the first of the proposed vignettes. In this vignette, participants read about a patient who sustained a tear to the anterior cruciate ligament together with a meniscus injury. Since his knee suffers from a superabundant joint effusion, the attending physician must decide how to deal with such effusion.

After having familiarized themselves with the scenario, participants were asked to evaluate the appropriateness of each of the 4 alternate actions to take in order to mitigate the effusion. The options are given below, in order from least to most appropriate according to our ex ante collaborators' opinions. The most appropriate action corresponds to that suggested by national guidelines as confirmed by the indications proposed by the main Italian Orthopaedic Institutes (e.g. Rizzoli, Bologna; Regional Hospital, Lugano)[16].

1. Surgery treatment.
2. Recommending ice and rest.

[16] http://www.ior.it/curarsi-al-rizzoli/ginocchio-lesione-del-legamento-crociato-anteriore

3. Prescribing magneto-therapy applications and cryotherapy.
4. Performing joint aspiration (arthrocentesis).

The above-mentioned actions were presented to physicians in a random order. Practitioners were asked to proceed with the ranking of the actions for each of the three proposed vignettes. Table 1 provides an example of the form used by subjects to complete the tasks, which shows the way participants had to report their answers.

Table 1: Example of the response form

	Prescribing magneto -therapy and cryotherapy	Recommending ice and rest	Surgery treatment	Performing joint aspiration
1 Very inappropriate			X	
2 Somehow inappropriate	X			
3 Somehow appropriate		X		
4 Very appropriate				X

Physicians had to judge the appropriateness of the actions so as to match the response of others (similar to Krupka and Weber, 2008) in order to obtain the reward. Moreover, after completing all the vignettes, physicians are asked to indicate for each of the case how certain they are of being correct (similar to Kovacs et al., 2020) in order to assess their confidence level[17]. Finally,

[17] Overconfidence physicians are found to ignore decision support tools (i.e. guidelines) (Baumann et al., 1991)

once all the participants had completed the tasks, one of the 12 possible actions[18] was randomly selected and after having compared all the physicians' responses with the modal valuation, subjects were privately paid for correct matches. Each participant whose response coincided with the response of the vast majority of participants received a 10-euro meal ticket[19].

A very crucial issue concerns the external validity of the experiment. Even though, vignettes seem to be very realistic it is quite a leap assume that physicians facing the same situations for real would behave exactly taking the action judged as the most appropriate in the specific situation. In particular, in a real scenario other factors may play a crucial role making physician opt for a different action. For this reason, results cannot be generalized and the order in which physicians rank the actions according to their degree of appropriateness must not be taken as the order in which he would opt for each of the solutions in the real life. However, if coordination is not reached in such a controlled and simplified environment, it seems to be very implausible that coordination could be reached in hospital where the environmental variables become more pressing. As a result, poor coordination which has been found through this experiment could proxy for clinical environments, within certain limitations.

Procedure

The experiment was conducted at the main hospital of Reggio Calabria, thanks to an agreement signed by the same hospital and the Mediterranean University. According to such agreement, the parties involved are committed to participating in the research project 'Experiments in health economics' and to sharing results which could be useful to the hospital for management policy. I was, then, allowed by the interested chief administrators to enter in several departments (e.g. genecology, orthopaedics, surgery, otolaryngology and so on) to run the experiment. To join the experiment, during their coffee-breaks, physicians came, individually or in group, to the meeting room, to avoid any interference with the working schedule[20].

[18] Notice that for each vignette, there were 4 possible actions.

[19] https://research.utoronto.ca/compensation-reimbursement-research-participants

[20] To avoid any interaction between physicians who already joined the experiment and physicians who were about to participate, physicians enter in the experimental room through one door and then exit trough another

Although the financial incentive could be retained relatively low given the average income of the subject pool, their intrinsic motivation should be already enough to incentivize their performance in the experiment[21]. In fact, according to Gneezy and Rustichini (2000) when an activity has a motivation its own, in this case contributing to medical research, introducing a monetary reward can have a detrimental effect on performance. As a result, under certain conditions, introducing a compensation contingent on performance crowds out any endogenous incentive which instead the experimenter wishes to elicit (McKeganey, 2001).

Before starting the experiment, I have measured physicians' attitude towards risk. In fact, coordination games require that each player tries to guess the other players' behaviour (Heinemann et al., 2009). For this reason, subjects' risk attitude can affect the outcome of the coordination game, like what happens in a lottery (Neumann and Vogt, 2009).

To estimate physicians' degree of risk aversion, I asked them to complete a questionnaire taken from Holt and Laury (2002) with hypothetical rewards (see e.g. Galizzi et al., 2016). The questionnaire consisted of 10 hypothetical choices between a safer lottery called A and a riskier lottery called B. Payoff and probabilities were distributed such that the number of times a subject chose lottery A could be used to estimate his attitude towards risk. According to Holt and Laury (2002), as the probability associated to the high payoff outcome increases, subjects should shift from option A to option B.

The results of the questionnaire are slightly different from that obtained by Holt and Laury (2002): 44% of the subjects opted for the safer option for the first four choices, when the probability of the higher payoff was low, and then switched to the riskier option without ever moving back to lottery A. Approximately 14% of the subjects showed extreme risk aversion, selecting lottery A for all the options but the last one. On the other hand, 29% of the subjects chose lottery B for all the options. This result is justified by the presence of most surgeons in the pool (roughly 70%), who usually show greater risk seeking, as reported by Pikkel et al. (2016). Positive correlation was found between age and risk taking ($\rho<0.01$), as well as female

door (i.e. to avoid behavioural spillovers).

[21] They know they are contributing to an important study, whose results could benefit the whole society.

and risk taking ($\rho<0.01$). Finally, for the sake of completeness, 7 subjects showed inconsistent preferences.

After completing the above-described questionnaire, participants played the Coordination Game. The whole experiment was run using pen and paper and lasted approximately 15 minutes per subject. As far as the coordination game is concerned, each vignette can be viewed as an independent session and each action from the same vignette is comparable to a different period.

At the end of the experiment, only one action from the Coordination Game was randomly selected to determine subjects' payoffs. While payoffs were calculated, subjects were asked to complete a demographic questionnaire, asking about physicians' years of experience, job satisfaction and their relationship with colleagues. In the end, physicians matching the modal answer received their 10-euro meal ticket from the cafeteria[22]. The average reward for participants was €5.19.

Results

In this section I am going to focus on the data resulting from this experiment. I employed both non-parametric and regression analyses to test how coordination changes across the vignettes and to detect which variable can impact physicians' likelihood to coordinate.

Descriptive analysis and non-parametric tests

Before discussing the coordination experimental results, it is appropriate to specify when it may be said that coordination is reached. According to the structure of this experiment, I assume that coordination is achieved when a physician's appropriateness judgment matches the modal valuation for the specific action considered, which corresponds to the judgment given by most participants. In other words, I checked whether subjects judged in the same way each of the actions. Such a choice is aimed to preserve the external validity of the experiment. In particular,

[22] The above-mentioned incentive is reasonably salient for two reasons. First, cafeteria is the only hospital internal alternative available to physicians. Although there are some external cafés, walking distance from the hospital, their opportunity cost may be high (physicians would have to push out and walk for 15 minutes). Additionally, according to the regulation, the internal cafeteria must charge discounted rates (20% less) to the hospital's employees.

subjects giving the same valuations can be seen as colleagues who generally work together (i.e. because they are in charge of the same department) and have to make a decision and to agree on it (i.e. while they are prescribing a therapy to the patient who is about to be discharged). Therefore, coordination is measured action by action for each subject.

Figure 1 shows how the average frequency of coordination changes across the vignettes, leading to a final average level of 0.52, represented by the green dashed line. As it is possible to observe, coordination average frequency reaches approximately the same value in each vignette, going from 0.5 in the meniscus injury scenario to 0.53 in the oral cavity cancer scenario. However, the second vignette results in the highest observed level of coordination (0.54). This result is reasonable considering the sample distribution: 36% of physicians were either gynecologists or neonatologists and then specialized in the woman breast fissures scenario.

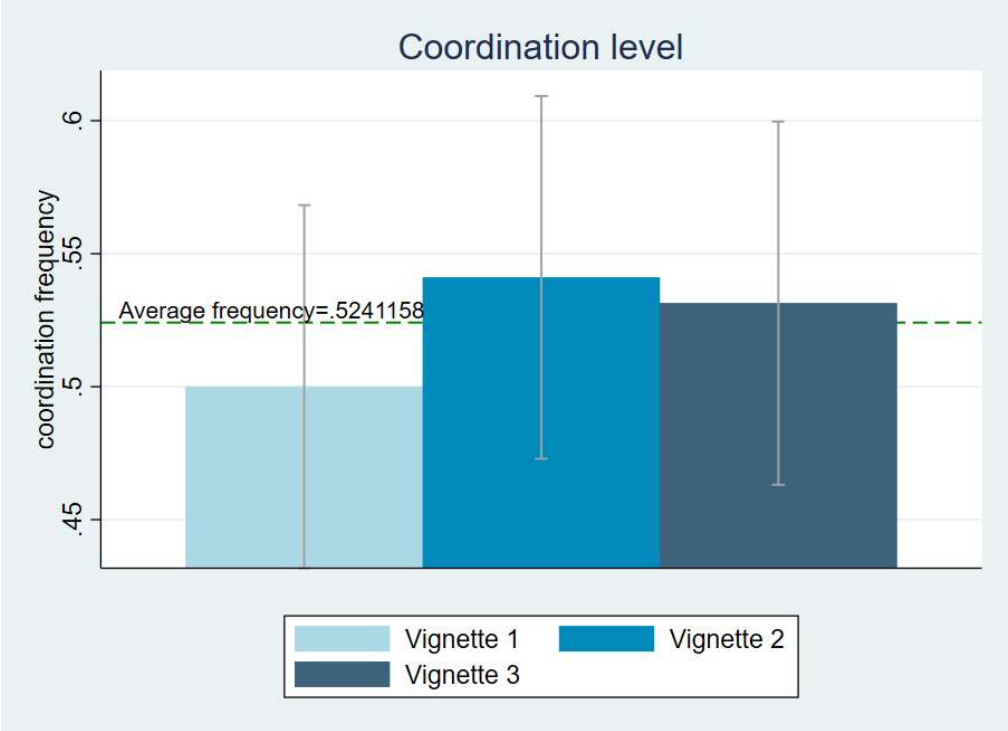


Figure 1: Average frequency of coordination across the experiment with confidence intervals

In particular, looking at each vignette I can check how each of the actions contributes to coordination. Figure 2 displays the average frequency of coordination for the meniscus injury scenario broken down by action. The orange bar represents the course of action suggested by national guidelines (see footnote 22). The level of coordination for such action is significantly above the average of the vignette, which indicates that subjects recognize the guideline, despite the heterogeneity of the pool of physicians[23], reported in Table 6 of Appendix B. This assumption is also confirmed by the highest number of perfect appropriateness judgments reported for such action (81%). Ice and rest placed second for both coordination and level of appropriateness, being generally considered a more conservative therapy. In fact, any physician, regardless of his specialty, is likely to prescribe ice and rest which are said to be immediately effective after an injury (Borra et al., 2015).

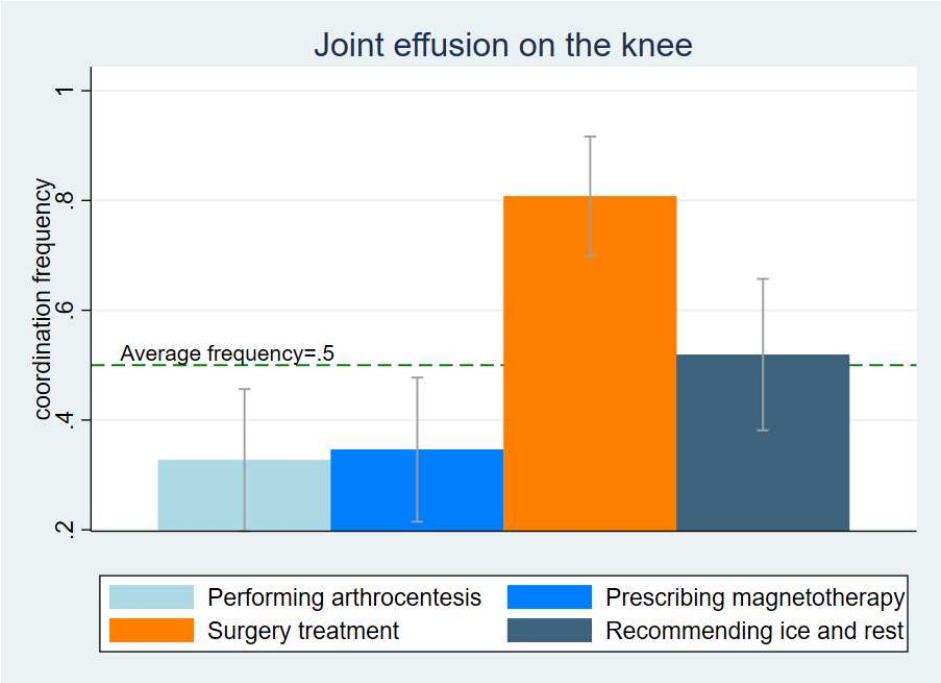


Figure 2: Average frequency of coordination in the first vignette with confidence intervals

[23] Only 8 physicians were specialized in orthopaedics, which this vignette specifically refers to, while the other practitioners were specialized in various branches of medicine (i.e. pediatrics, anesthesia).

As previously mentioned, the highest level of coordination is observed in the second vignette whose average frequency is illustrated in Figure 3. Here physicians face a situation where a 28-year-old woman suffers from breast fissures. As usual, the orange bar corresponds to the treatment suggested by national guidelines (Ministero della salute, 2019) which reconciles physicians, given the high level of coordination. On the other hand, the grey bar counterbalances the good level of coordination of the other actions.

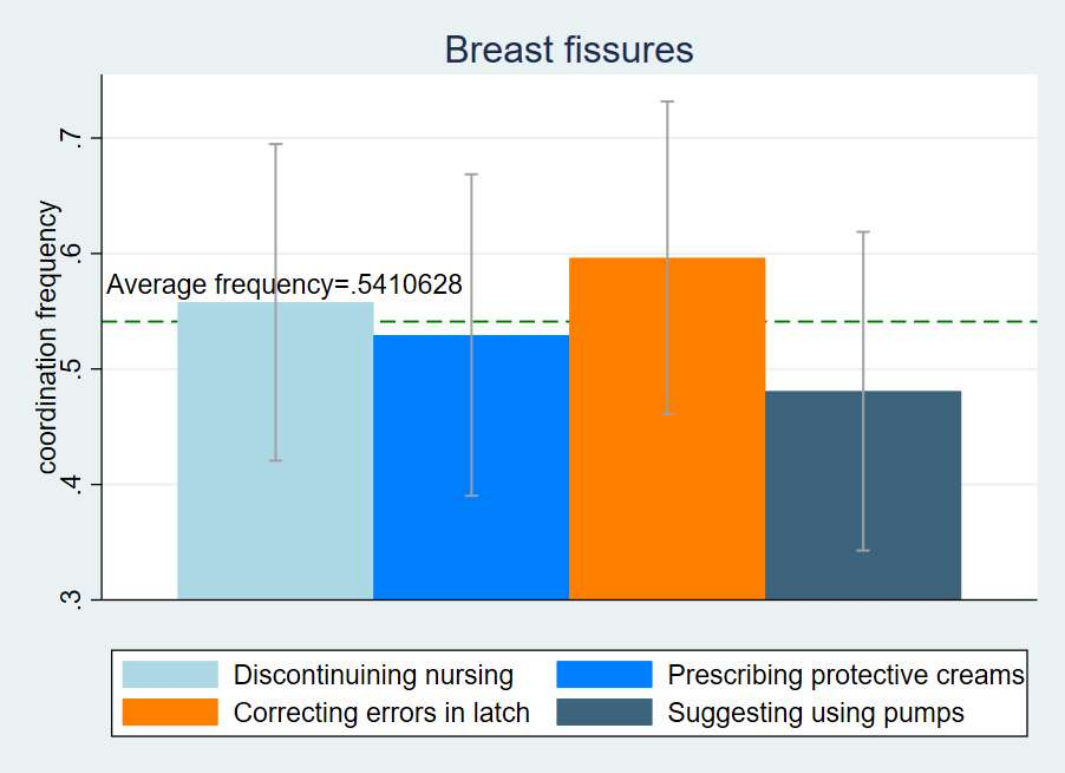


Figure 3: Average frequency of coordination in the second vignette with confidence intervals

Focusing in detail on this vignette, a slight gender effect can be detected which may be responsible for the observed level of coordination. In fact, valuations of appropriateness vary between men and women across the actions as figure 4 reports, and this difference is significant according to the Mann-Whitney test (p-value<0.05). In such a scenario, it is very unlikely that men, not specialized in paediatrics or childcare and who did not witness similar episodes, understand the 28-year-old woman's pain. Thus, as the lower right box displays, men could

assign the highest degree of appropriateness to a very unadvisable treatment such as ‘suggesting using pump’.

Figure 4: Gender difference in the second vignette

Finally, figure 5 shows the average frequency of coordination of the last vignette (i.e. portraying a patient suffering from an oral cavity cancer), where the highest level of coordination is reached in the second action which is ‘chemo- therapy’, reported to be ‘appropriate’ by 62% of the subjects.

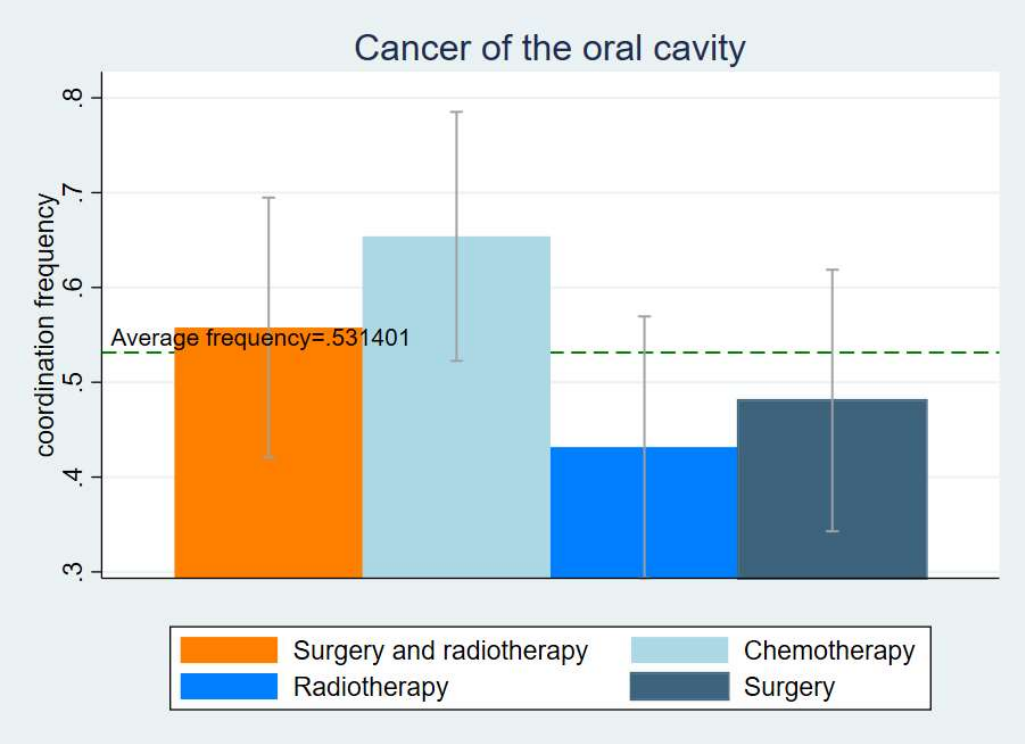


Figure 5: Average frequency of coordination in the third vignette

Such result has its foundation in recent literature on cancer treatments, where chemotherapy is expected to reduce the size of the tumor and to facilitate conservative treatment surgery[24]. Going further into details, figure 6[25] illustrates how each of the actions is assessed according to the appropriateness judgment. In particular, 'only surgery' stands out for the action with the highest likelihood of being judged as 'very inappropriate'. Similarly, the therapy corresponding to the guideline 'Surgery and radiotherapy' results in the highest probability of being judged as 'very appropriate'.

[24] http://www.arquivosdeorl.org.br/additional/acervo_port.asp?id=228

[25] Appendix B provides the same figures for the other two vignettes.

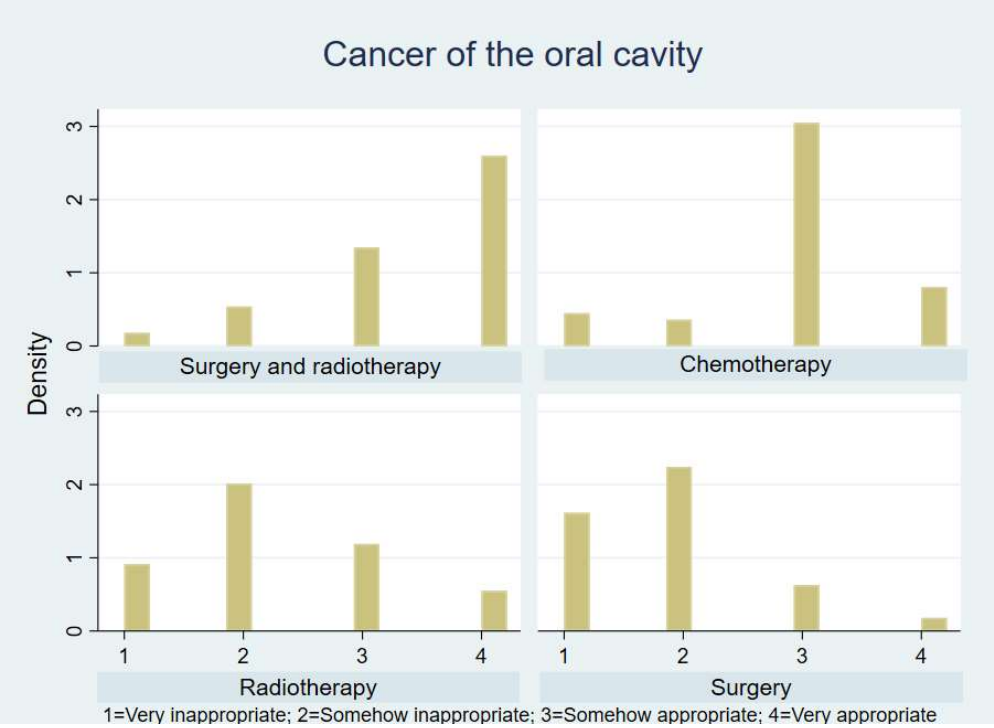


Figure 6: Density of the appropriateness judgment for action for the third vignette

Comparing the evaluations distribution across the actions, differences are significant according to the Kruskal Wallis test (p-value<0.001). The same results have been obtained for the two remaining vignettes (for the first vignette p-value< 0.001, for the second vignette p-value< 0.001[26]).

In the end, by briefly focusing on the courses of actions prescribed by national guidelines (which instead is the focus of the second chapter) for each of the proposed scenarios, it can be noticed from figure 7 that such actions result in very high average appropriateness ratings. In particular, by comparing the distributions of the statement 'very appropriate' for such actions

[26] Notice that in addition to the Kruskal Wallis test which compares all the actions at once, the ranksum test has been performed to compare the actions in twos for each vignette. Results are all significant at the 1% level.

and for their alternatives, not only their difference results to be significant according to the Median test (p-value<0.001) but also most of the valuations of the guidelines are greater than the average.

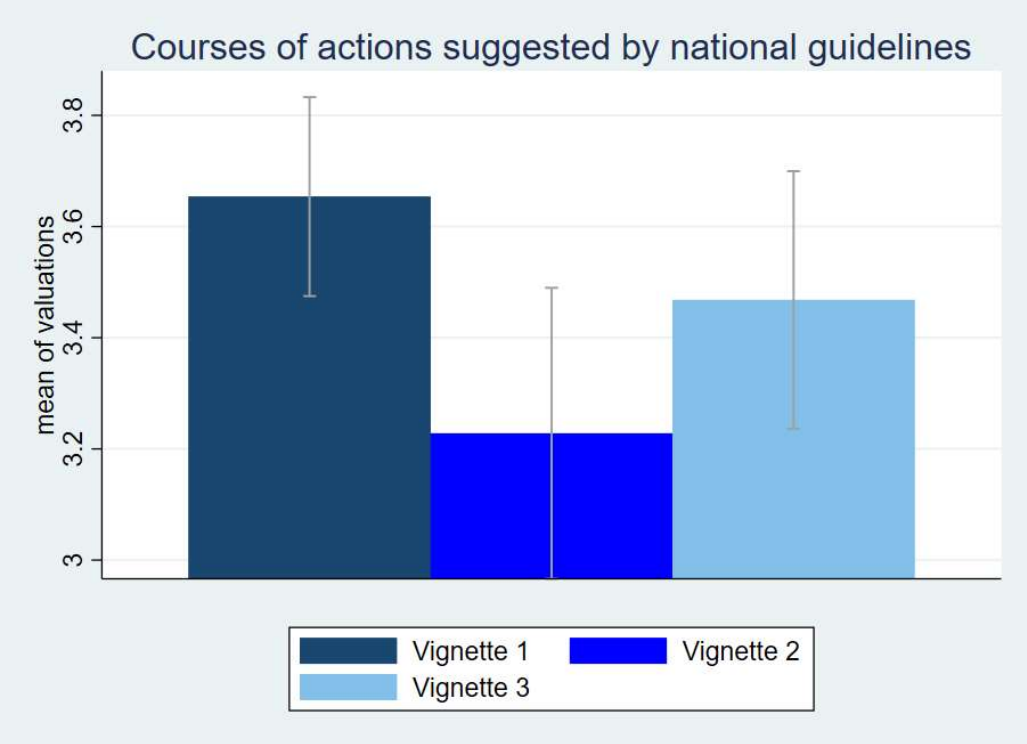


Figure 7: Means of valuations for the courses of action suggested by national guidelines with confidence intervals

However, the frequency of coordination observed in figure 8 is not very high for these alternatives which may signal an important problem such as lack of communication between parties involved in the decision making, or different mindlines given by different backgrounds and training.

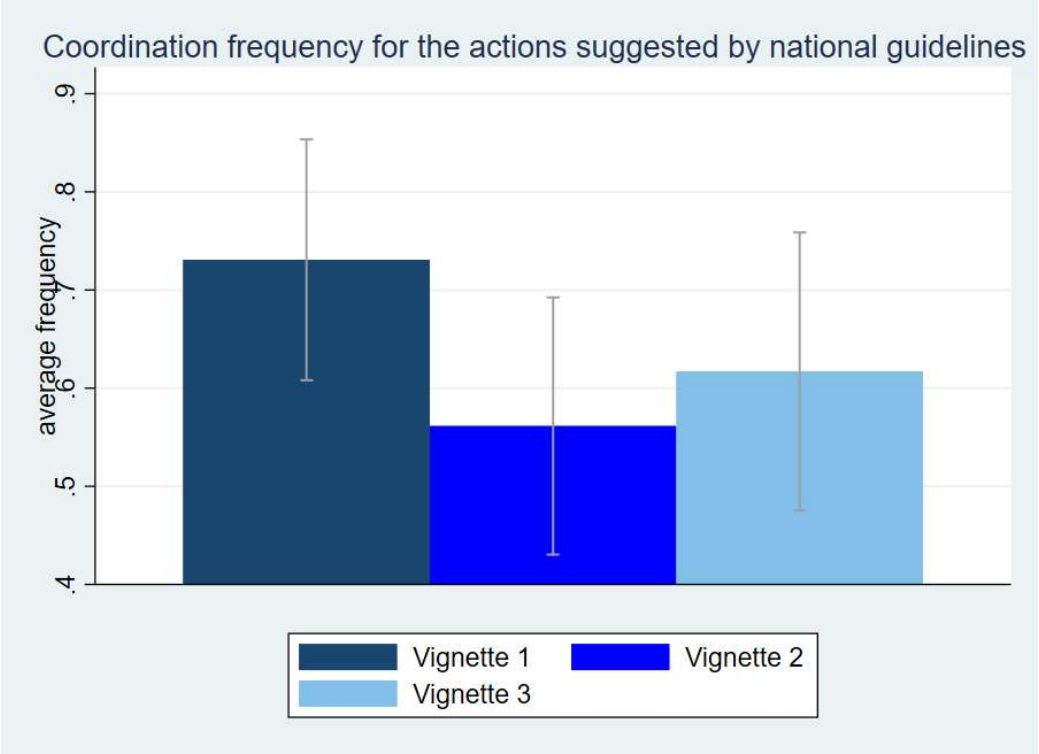


Figure 8: Average frequency of coordination for the courses of action suggested by national guidelines with confidence intervals

Regression analysis

I employed logit regressions on an experimental dataset of 624 observations, since each subject makes 12 choices.

Table 2 below provides a summary of all the variables being considered. Variables result from physicians' answers to the final questionnaire (available in Appendix A). Since data are obtained from multiple observations per physician, errors are cluster correlated (Cameron et al., 2008). Therefore, cluster robust standard errors are used. Additionally, dummy variables for each vignette are introduced to control for vignette effects.

Table 2: Summary statics

VARIABLES	mean	sd	min	max	obs
Male	0.596	0.491	0	1	624
Age	51.50	8.763	31	66	624
Years of service	12.58	10.48	1	40	624
Satisfaction	2.804	0.793	1	4	612
Size	13.42	8.756	5	35	624
Collegues'advice	2.769	0.609	1	4	624
Positive influence	2.827	0.643	1	4	624
Negative influence	2.462	0.746	1	4	624
Team	3.077	0.703	1	4	624
Results sharing	3.192	0.622	1	4	624
Leader	0.288	0.453	0	1	624
Whatsapp	0.653	0.476	0	1	588
Training	3.038	0.588	2	4	624
Brainstorming	3.115	0.467	2	4	624
Updating	3.250	0.515	2	4	624
Research	0.462	0.499	0	1	624
National guideline	3.692	0.462	3	4	624
Internal guideline	3.520	0.500	3	4	600
Colleagues guidelines	3.255	0.555	2	4	612

Collegues'judgment	3.212	0.600	2	4	624
Department coordination	3.077	0.806	1	4	624
Insurance	1	0	1	1	624
Speciality	0.186	0.389	0	1	624
Confidence	3.821	0.863	1	5	604
Coordination	0.524	0.500	0	1	622
Guidelines	0.250	0.433	0	1	624
Request guidelines	.92308	.2667	0	1	624
Valuations	2.690	1.076	1	4	622
Risk Seeking	0.207	0.405	0	1	624

First of all, in table 3 I estimated a logit to test how each variable affects physicians' probability of coordinating. Numbers reported in the tables correspond to coefficients, whose interpretation is not straightforward as in the case of a standard multiple regression, because the magnitude of the effect on probability depends on both the values of the other explanatory variables and the starting value of such variables.

The dependent variable, 'coordination', is a dummy variable which assumes the value 1 whether the subject coordinates with his partner and 0 otherwise. *Speciality* is a dummy variable which refers to physicians' medical department, and specifically it assumes the value 1 when physician's training regards the field of the vignette (e.g. for a pediatrician, specialty takes the value one for the four actions proposed in the second vignette 'breast fissures' and 0 for the remaining 8 actions). *Leader* is also a dummy variable equal to 1 if subjects report that when they work in team there is a member whose opinion weights the most (e.g. head of the department or first operator). *Risk seeking* is also a dummy variable which takes the value 1 for risk taking subjects. *Request guidelines* refers to whether physicians, after filling in the vignette, ask for guidelines support. *Whatsapp* is a dummy variable referred to whether physicians have a whatsapp group with their colleagues to discuss patients' treatments. Finally, positive

(negative) influence indicates whether and to which extent a colleague's similar (divergent) view affects physician's decision[27].

Table 3: Logit for coordination

VARIABLES	(1) First model	(2) Second model	(3) Third model
confidence	-0.133	-0.0901	-0.0594
	(0.118)	(0.103)	(0.121)
speciality	0.852***	0.858***	0.910***
	(0.264)	(0.251)	(0.257)
leader	0.131	0.364***	0.375***
	(0.199)	(0.140)	(0.139)
male		-0.00685	0.000203
		(0.230)	(0.229)
age		-0.0293*	-0.0292*
		(0.0155)	(0.0152)
yearsofservice		0.00241	0.00255
		(0.0133)	(0.0131)
updating		-0.183	-0.180
		(0.152)	(0.154)
whatsapp		-0.154	-0.159
		(0.243)	(0.241)
riskseeking		-0.554***	-0.557***
		(0.158)	(0.159)
positiveinfluence		0.323**	0.324**
		(0.129)	(0.129)
negativeinfluence		-0.389***	-0.387***
		(0.136)	(0.135)
vignette1			-0.181
			(0.297)
vignette2			-0.229
			(0.285)
Constant	0.412	2.448**	2.435**
	(0.474)	(1.155)	(1.167)
Observations	604	584	584

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

[27] As we observe, all the variables are categorical. Table 7 in Appendix B provides a summary of all the variables used.

As we would expect, it is easier to achieve coordination whether physicians are specialized in the object of the vignette than otherwise. In fact, each type of doctor undertakes training and education which allows him to treat patients with specific diseases[28]. As a result, it is more likely that a physician agrees with another colleague specialized in the same field than with a colleague from a different department.

Reasonably, the presence of a leader in the team increases the probability that physicians coordinate. Evidence shows that the presence of a leader inside an organization facilitates coordination, overcoming possible misalignments between employers' objectives (Bolton et al., 2012). Additionally, leadership skills are found to benefit patients and improve healthcare organization (Rotenstein et al., 2018).

Risk seeking negatively affects coordination. The rationale stems from physicians' preference to following their own ideas, taking risks, instead of coordinating with others.

Positive influence increases physicians' probability to coordinate. Intuitively, since physicians prefer agreeing with their colleagues than not, receiving positive feedbacks promotes shared understanding and contributes to group cohesion and coordination (Slof et al., 2010). Specifically, according to Janssen and Bodemer (2013), consensus knowledge triggers successful collaboration. However, although sharing opinions is essential for coordination (Bromme et al., 2005), problems could arise when ideas do not match. This could explain why negative influence reduces the likelihood of coordination. In particular, if a physician takes a colleague's view into proper consideration when they initially do not agree, this may lead to divergence of interpretation of the clinical case which may even worsen coordination.

As a next step, I isolated coordination on national guidelines and in particular we focused our attention on the three courses of action suggested by national guidelines, one for each vignette.

[28] https://www.medicalnewstoday.com/articles/types-of-doctors

Table 4: Logit for coordination on guidelines

VARIABLES	First model	Second model	Third model
confidence	0.202	0.374*	0.341
	(0.184)	(0.221)	(0.247)
speciality	1.013**	1.460**	1.696**
	(0.496)	(0.603)	(0.766)
leader	-0.263	0.525	0.501
	(0.418)	(0.417)	(0.406)
male		-0.943	-0.836
		(0.612)	(0.640)
age		-0.138***	-0.134***
		(0.0402)	(0.0394)
yearsofservice		0.0933***	0.0894***
		(0.0296)	(0.0290)
updating		-0.770*	-0.796**
		(0.407)	(0.406)
whatsapp		0.133	0.0159
		(0.579)	(0.611)
riskseeking		-1.341**	-1.365**
		(0.544)	(0.560)
positiveinfluence		0.852**	0.864**
		(0.371)	(0.373)
negativeinfluence		-0.480	-0.508
		(0.352)	(0.369)
vignette1			0.407
			(0.587)
vignette2			-0.623
			(0.589)
Constant	-0.369	6.597**	6.817**
	(0.806)	(2.698)	(2.734)
Observations	151	146	146

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The estimation results illustrated in table 4 proves that physicians' years of service for that specific hospital increases their probability of coordinating on judging guidelines. By estimating the marginal effect of the variable of interest, we see that on average one additional year of service would predict a 2% increase in the probability of coordination. In fact, experience is one of the many contributing factors to coordination (Dezso et al., 2012). More specifically, working for many years with the same people creates team familiarity which

boosts communication and coordination skills, improving team performance (Faraj and Sproull, 2000). On the other hand, any one year increase in the age reduces physicians' likelihood to coordinate by 3% (i.e. computing the marginal effect), partially contradicting the above reasoning. However, despite a positive correlation between age and years of service, it is not so obvious that older physicians have been longer in practice in the same hospital or in the same department (indicative is the case of a 64-year old physician who has been working into the hospital for two years). Since these two variables address different aspects, their effect on coordination may not overlap, especially if an old physician has worked for different hospitals and/or departments. As we already know, physicians experience (regardless the hospital and the department they work for) results in their cognitive rigidity (Choudhry et al., 2005). Consequently, more experienced physicians are less willing to conform with guidelines, which could explain such negative sign.

The output reported above shows that men are less likely to coordinate on guidelines than women. This result can be easily interpreted looking at subjects' responses to the final questionnaire. In particular, men declared to be less willing to share results of diagnostic tests or outcomes of surgical procedures with their colleagues. Moreover, they said to seldom participate in training courses and brainstorming sessions which could jeopardize coordination.

Finally, I tested which variable can influence physicians' likelihood to judge the courses of action prescribed by guidelines as 'very appropriate' actions. As before, I restricted the dataset to the three courses of action suggested by national guidelines. Table 5 displays the output of the probit where the dependent variable 'Conformity with guidelines' takes the value 1 whether the subject reports the action to be very appropriate.

Table 5: Probit for judging guidelines as 'very appropriate'

VARIABLES	(1) First model	(2) Second model	(3) Third model
confidence	-0.0248	-0.0682	-0.177
	(0.214)	(0.253)	(0.276)
speciality	0.00840	0.311	0.471
	(0.375)	(0.350)	(0.409)
riskseeking	-0.110	-0.403	-0.489
	(0.448)	(0.454)	(0.468)
leader		0.162	0.157
		(0.394)	(0.414)
team		-0.487*	-0.541**
		(0.261)	(0.276)
positiveinfluence		0.775***	0.855***
		(0.269)	(0.294)
negativeinfluence		-0.311	-0.310
		(0.346)	(0.369)
yearsofservice		0.0637**	0.0660**
		(0.0251)	(0.0271)
age		-0.0908***	-0.0938***
		(0.0295)	(0.0322)
male		-1.009***	-1.086***
		(0.373)	(0.413)
vignette1			1.343**
			(0.554)
vignette2			0.287
			(0.473)
Constant	0.695	5.455***	5.493***
	(0.882)	(1.739)	(1.906)
Observations	150	150	150

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

I added one more variable with respect to the previous estimations, team which comes from subjects' answers to the question 'How often decisions about a patient are taken as a result of a team's valuation?' on a four-point scale, where 4 stands for the highest frequency, namely. As we notice, working in team is reported to decrease the probability of acknowledging guidelines appropriateness, which might be counterintuitive. In fact, we should expect that teamwork facilitates the diffusion of national guidelines and the adherence to them, which happens if the components of the medical equipe who already follow them try to convince their colleagues

about the guidelines appropriateness. On the contrary, if members of the team have different ideas and different mindlines and share them with their colleagues, actions not really conforming with guidelines could be accepted as the right solution to adopt given their success in similar previous clinical cases. As far as the other significant variables are concerned, the same reasons given for the previous probit can be still applicable.

Conclusions

This framed field experiment investigates the level of coordination of the Italian healthcare sector using the coordination game. By designing three vignettes describing different clinical cases and a set of four actions, which a physician can take in response to them, I asked 52 physicians to evaluate the appropriateness of each of the possible actions on a scale of one to four in order to match the modal valuations. Since one of the proposed actions corresponded to that suggested by national guidelines, I was also able to test whether common practices are adopted in the context of the main hospital of Reggio Calabria, where the experiment took place.

Results show that the overall average level of coordination across the experiment is 52%. The regression analyses evidence that coordination increases when physicians do often exchange opinions and share positive feedbacks with colleagues. In addition, the presence of a leader in the medical department, either a primary doctor or a professor outside of the team, facilitates coordination. Furthermore, when physicians have been working longer for the same hospital, the probability that physicians share guidelines and judge them as the most appropriate actions to adopt in the specific clinical case increases.

Even though, in recent years Italy distinguished itself for the initiatives aimed at promoting integrated care through the creation of networks and cooperatives between different health professionals for example in Emilia Romagna, Lombardy and Piemonte (European Union Report, 2017), there is still a lot to do. More generally, collaborations between general practitioners and specialists together with more formal integrated path among professionals belonging to different levels of care organizations should be encouraged to fight lack of coordination. Furthermore, since sharing ideas and feedback is said to increase the level of

coordination, teamwork should be incentivized through brainstorming and planning sessions. In this regard, leaders must be trained to be able to manage a team, facilitating collaboration and communication between different members.

All these measures aimed at increasing coordination could solve another important problem revealed by the experimental results which is physicians' dissatisfaction with the hospital they work for. Employees' satisfaction leads to the success of an organization and must be guaranteed. When an employee is satisfied with his firm and feels like an integral part of it, he is more likely to accept and follow the corporate norms and procedure (Kim and Lee, 2007) which in our case are epitomized by national guidelines. In fact, a positive correlation, though not significant, is found between work satisfaction and national guidelines compliance.

The above-mentioned actions which directly intervene on the hospital organization could improve the relationship between physicians and the hospital they work for and coordination could benefit from better organized structures and above all more satisfied employees.

Finally, since the experiment addresses an increasingly discussed topic it could be worth it to replicate it, adding other possible vignettes and including physicians from further departments. Access to hospital data could help identifying areas which require greater intervention in terms of coordination and for which further scenarios and so vignettes could be imagined.

Another possible extension could be adapting the same experimental design to a context where healthcare is provided at a national level. As we previously explained, Italian National Health Service is regionally based, and this could cause lack of coordination between different professionals. On the contrary, nationally based healthcare systems should be characterized by higher level of coordination and larger spread of national guidelines. Therefore, comparing results from different countries could be useful to detect solutions to overcome poor coordination of the healthcare sector. Moreover, additional vignettes could be designed in order to represent scenarios where different subjects (e.g. physicians and nurses) are involved in the situation and have to make a decision. This could allow to measure coordination between people who are in the same medical team or in different teams, and among people who are at the same or different level of the organization (e.g. nurses evaluating a nurse or a physician). In addition, by varying the matching answer (i.e. by distinguishing modal answers for different categories) similarly to Burks and Krupka (2012) the coordination game would allow to distinguish between norms held by different groups inside the same organization.

References

Argote, L., (1982), Input uncertainty and organizational coordination in hospital emergency units, *Administrative Science quarterly*, 27, 420-434

Baggs, J.G., Ryan, S.A., Phelps, C.E., Richeson, J.F., & Johnson, J.E., (1992), The association between interdisciplinary collaboration and patient outcomes in a medical intensive care unit, *Heart lung: the journal of critical care*, 21(1), 18-24

Barr, A.,& Lane, T., & Nosenzo D., 2018, On the social appropriateness of discrimination, *Journal of Public Economics*, Volume 164, 153-164

Baumann, A. O., Deber, R. B., & Thompson, G. G., (1991), Overconfidence among physicians and nurses: the 'micro-certainty, macro-uncertainty'phenomenon, *Social science & medicine*, *32*(2), 167-174

Beuselinck, E., Verhoest, K., & Bouckaert, G., (2007), Reforms of Central Government Coordination in OECD-Countries for Cross-National Unifying Processes? In: *Cultural Aspects of Public Management Reform*, 77–109, Eds. Kuno Schedler & Isabella Proeller, Amsterdam: JAI Press

Bolton, G. E., & Ockenfels, A., (2000), ERC: A theory of equity, reciprocity, and competition. *American economic review*, 90(1), 166-193

Bolton, P., Brunnermeier, M. K., & Veldkamp, L., (2012), Leadership, coordination, and corporate culture, *Review of Economic Studies*, 80(2), 512-537

Borra, V., De Buck, E., & Vandekerckhove, P., (2015), RICE or ice: what does the evidence say? The evidence base for first aid treatment of sprains and strains. In: *Filtering the information overload for better decision*, Ab- stracts of the 23rd Cochrane Colloquium, 2015 3-7 Oct, Vienna, Austria, John Wiley & Sons

Bromme, R., Hesse, F. W., & Spada, H., (2005), Barriers, biases and opportunities of communication and cooperation with computers: Introduction and overview, In Barriers and biases in computer-mediated knowledge communication, 1-14, Springer, Boston, MA

Burks, S., & Krupka E., (2012), A Multi-Method Approach to Identifying Norms and Normative Expectations within a Corporate Hierarchy: Evidence from the Financial Services Industry, *MANAGEMENT SCIENCE*, Volume 58, No. 1, 203–217

Camera civile di Firenze, (2017), Le Corti Fiorentine, *Rivista di diritto e procedura civile*, n.3/17

Cameron, A. C., Gelbach, J. B., & Miller, D. L., (2008), Bootstrap-based improvements for inference with clustered errors, The Review of Economics and Statistics, 90(3), 414-427

Cartabellotta, N., Cottafava, E., Luceri, R., & Mosti, M., (2019), 4 Rapporto GIMBE sulla sostenibilità del Servizio Sanitario Nazionale, Fondazione GIMBE, Available at : www.rapportogimbe.it

Charness, G., Gneezy, U., & Kuhn, M.A., (2012), Experimental methods: Between-subject and within-subject design, *Journal of Economic Behavior and Organization*, 81, 1-8

Choudhry, N.K., Fletcher, R.H., & Soumerai, S.B., (2005), Systematic re- view: the relationship between clinical experience and quality of health care, *Ann Intern Med*, 142(4), 260–273

Currie, J., & Macleod, W.B., (2006), First do no harm? Tort reform and birth outcomes, *National Bureau of Economic Research*, Nber Working paper series

Del Torso, S., Bussi, R., & DeWitt, T., (1997), Primary care pediatrics in Italy: Eighteen years of clinical care, research, and teaching under a national health service system, *Pediatrics99*, (1), January: E8

Dezso, C. L., Grohsjean, T., & Kretschmer, T., (2012), Coordination experience and team performance: Evidence from the electronic games industry, *Robert H. Smith School Research Paper No*, RHS, 06-136

European Union, (2017), Tools and methodologies to assess integrated care in Europe, *Report by the Expert Group on Health Systems Performance As- sessment*

Elstad, E. A., Lutfey, K. E., Marceau, L. D., Campbell, S. M., von dem Kne- sebeck, O,. & McKinlay, J. B, (2010), What do physicians gain (and lose) with experience? Qualitative results from a cross-national study of diabetes, *Social science & medicine*, 70(11), 1728-1736

Faraj, S., & Sproull, L., (2000), Coordinating expertise in software development teams, *Management science*, *46*(12), 1554-1568

Fargason, C.A. Jr., & Haddock, C.C., (1992) Cross-functional, integrative team decision making: Essential for effective QI in health care, *QRB Qual Rev Bull*, 18, 157-163

Farquhar, C. M., Kofa, E. W., & Slutsky, J. R., (2002), Clinicians' attitudes to clinical practice guidelines: a systematic review, *Medical Journal of Australia*, 177(9), 502-506

Fehr, E. & Schmidt, K., (1999), A Theory of Fairness, Competition, and Cooperation, *The Quarterly Journal of Economics*, 114(3): 817-868

France, G., Taroni, F., & Donatini, A., (2005), The Italian health-care sys- tem, *Health economics*, 14(S1), S187-S202

Gabbay, J., & le May, A., (2004), Evidence based guidelines or collectively constructed "mindlines?" Ethnographic study of knowledge management in primary care, *BMJ*, 329(7473):1013

Gachter, S., Nosenzo, D., & Sefton, M., (2013), Peer effects in prosocial behaviour: Social norms or social preferences?, *Journal of the European Eco- nomic Association*, 2013, vol. 11, issue 3, 548-573

Galizzi, M. M., Miraldo, M., & Stavropoulou, C., (2016), In sickness but not in wealth: field evidence on patients' risk preferences in financial and health domains, Medical Decision Making, 36(4), 503-517

Gittell, J.H., (2000), Organizing work to support relational coordination, *International Journal of Human Resource Management*, 11(3), 517-539

Gittell, J., Fairfield, K., Bierbaum, B., Head W., Jackson R., Kelly, M., Laskin, R., Lipson, S., Siliski, J., Thornhill, T., & Zuckerman, J., (2000), The impact of Relational Coordination on Quality of Care, Postoperative Pain and Functioning, and Length of Stay: A Nine-Hospital Study of Surgical Patients, *Medical care*, 807-819

Gneezy, U., & Rustichini, A., (2000), Pay enough or don't pay at all, *The Quarterly Journal of Economics*, 115(3), 791-810

Havighurst, C.C., (1991), Practice guidelines as legal standards governing physician liability, *Law and Contemporary Problems*, 54(2), 87-117

Heinemann, F., Nagel, R., & Ockenfels, P., (2009), Measuring strategic uncertainty in coordination games, A The Review of Economic Studies, 76(1), 181-221

Hermalin, B.E., (1999), 'Leadership and Corporate Culture', in: Gibbons, R., Roberts, J., (2013), *The Handbook of Organizational Economics*, Prince- ton: Princeton University Press

Holt, C. A., & Laury, S. K., (2002), Risk aversion and incentive effects, *American Economic Review*, 92(5), 1644-1655

Institute of Medicine (US), (1990), *Clinical Practice Guidelines: Directions for a New Program*, Committee to Advise the Public Health Service on Clinical Practice Guidelines, Field MJ, Lohr KN, editors, Source Washington (DC): National Academies Press (US)

Jacobs, J.P., Jacobs, M.L., Mavroudis, Maruszewski, B., Tchervenkov, C.I., Lacour-Gayet, F.G., Clarke, D.R., et al., (2007), What is Operative Morbid- ity? Defining Complications in a Surgical Registry Database, *The Annals of Thoracic Surgeons*, 84, 1416-1421

Janssen, J., & Bodemer, B., (2013), Coordinated Computer-Supported Collaborative Learning: Awareness and Awareness Tools, *Educational Psychologist*, 48:1, 40-55, DOI: 10.1080/00461520.2012.749153

Kim, S. E., & Lee, J. W., (2007), Is mission attachment an effective man- agement tool for employee retention? An empirical analysis of a nonprofit human services agency, *Review of Public Personnel Administration*, 27(3), 227-248

Kovacs, R. J., Lagarde, M., & Cairns, J., (2020), Overconfident health workers provide lower quality healthcare. *Journal of Economic Psychology*, 76, 102213

Krupka, E. L., Leider, S., & Jiang, M, (2016), A meeting of the minds: in- formal agreements and social norms, *Management Science*, 63(6), 1708-1729

Krupka, E.L., & Weber, R. A., (2008), *Identifying social norms using co- ordination games: why does dictator game sharing vary?*, (No. 3860), IZA discussion papers

Landrain, E., (2004), *Les réformes de l'assurance maladie en Europe*, Strasbourg: European Union, Assembleé Nationale, Enregistré à la Présidence de l'Assembleé nationale

Lewis, R.D., (2006), *When Cultures Collide: Leading Across Cultures*, Boston: Nicholas Brealey International

Luxembourg: Healthcare & Long-Term Care Systems, (2016), The Joint Re- port on Healthcare and Long-Term Care Systems & Fiscal Sustainability, Paper 37 Volume 2 - *Country Documents*

Mangham, I.L., (1986), *Power and Performance in Organization*, New York: Basil Blackwell Inc

Mapelli, V., & Lucioni C., (2003), Spending on pharmaceuticals in Italy: macro constraints with local autonomy, *Value Health*, 6(S1), S31–S45

McKeganey, N., (2001), To pay or not to pay: respondents' motivation for participating in research, Addiction, 96(9), 1237-1238

Ministero della salute, (2019), *Allattare al seno: un investimento per la vita*, available at www.salute.gov.it

Monfeli, L., (2018), Atto medico e responsabilità erariale, Giustizia Civile.com

National Center for Complementary and Integrative Health, (2017), Na- tional Institutes of Health: *Clinical Practice Guidelines*, available at http:

//nccih.nih.gov/health/providers/clinicalpractice.htm

Neumann, T., & Vogt, B., (2009), Do Players' beliefs or risk attitudes determine the equilibrium selections in 2x2 coordination games?, Univ., FEMM

OECD (2015), OECD Reviews of Health Care Quality: Italy 2014: Raising Standards, *OECD Reviews of Health Care Quality*, OECD Publishing, Paris, available at https://doi.org/10.1787/9789264225428-en

Pikkel, D., Igal, Y. S. P., Sharabi-Nov, A., & Pikkel, J., (2016), Are doctors risk takers?, *Risk management and healthcare policy*, 9, 129

Ratajcak, T., (1998), Approaches to managed care in Germany, *Medicine and Law*, 17(4), 545–51

Romøren, T.G., Torjesen, D.O., & Landmark, B., (2011), Promoting coordination in Norwegian health care, *International Journal of Integrated Care*, 11 (Special 10th Anniversary Edition), e127

Rotenstein, L.S., Sadun, R, Jena, A.B., (2018), Why Doctors Need Leadership Training, *Harward Business Review*

Samuels, M. A., Ropper, A. H., Poses, R. M., & Diaz, J. A, (2005), Clini- cal Experience and Quality of Health Care/IN RESPONSE/IN RESPONSE, *Annals of internal medicine*, 143(1), 84

Shortell, S.M., Zimmerman, J.E., Rousseau, D.M., Gillies, R.R., Wagner, D.P., Draper, E.A., et al., (1994), The performance of intensive care units: Does good management make a difference?, *Medical Care*, 32, 508-525

Simonet, D., (2008), The New Public Management theory and European health-care reforms, *Canadian Public Administration*, 51(4), 617–635

Slof, B., Erkens, G., Kirschner, P. A., Jaspers, J. G., & Janssen, J., (2010), Guiding students' online complex learning-task behavior through representational scripting, Computers in Human Behavior, 26(5), 927-939

Switzer, G. E., Halm, E. A., Chang, C. C. H., Mittman, B. S., Walsh, M. B., & Fine, M. J., (2003), Physician awareness and self-reported use of local and national guidelines for community-acquired pneumonia, *Journal of general internal medicine*, 18(10), 816-823

Tannenbaum, S.J., (1994), Knowing and acting in medical practice: the epis- temological politics of outcomes research, *Journal of Health Politics*, Policy and Law, 19, 27–44

Tedeschi, P., (2010), 'Italy' in: Kringos, D. S., Boerma, W. G., Hutchinson, A., Saltman, R. B. (2015), *Building primary care in a changing Europe*, WHO Regional Office for Europe

The Economics Times, (2019), Hyderabad Allwyn Ltd, in *Budget 2019*, avail-

able at https://economictimes.indiatimes.com/company/hyderabad-allwyn-ltd-/U27109TG1942PLC000607

Viinama¨ki, O.P., (2004), A Theory of Coordination and Its Implication on EU Structural Policy: A Comparative Study of the Challenges for Coordi- nation in Structral Funds in Finland, Ireland and Sweden, *Administrative Science*, Number 132, Vaasa: University of Vaasa/Publication Unit

Webb, A., (1991), Coordination: A problem in public sector management,

Policy Politics, Volume 19, Number 4, 229-242(14)

Appendix A: Instructions

Experimental instructions

Welcome to this experiment.

In this experiment you will be asked to perform one task and you will re- ceive a payoff related to it. During the experiment, we request that you remain quiet and do not attempt to communicate with other participants. Participants not following this request may be asked to leave without receiv- ing payment. If you have any questions, please raise your hand and one of us will come to you.

At the end of the experiment you could receive money based on your choices and the choices of others in the task described below.

There will be one task for all participants to perform. You will not receive feedback on the outcome of the task, and you will not be paid until the end of the experiment.

Introductory questionnaire (taken from Holt and Laury 2002)

Before starting the experiment, please fill in the following questionnaire in all its parts. Your answers will not affect your future earnings.

Choose which of the lotteries you would play between the two lotteries proposed.

Lottery A	**Lottery B**	**Your choice**
2€ with probability 1/10 1,60€ with probability 9/10	3,85€ with probability 1/10 0,10€ with probability 9/10	
2€ with probability 2/10 1,60€with probability 8/10	3,85€with probability 2/10 0,10€ with probability 8/10	
2e with probability 3/10 1,60€ with probability 7/10	3,85€ with probability 3/10 0,10€ with probability 7/10	
2€ with probability 4/10 1,60€ with probability 6/10	3,85€ with probability 4/10 0,10€ with probability 6/10	
2€ with probability 5/10 1,60€ with probability 5/10	3,85€ with probability 5/10 0,10€ with probability 5/10	
2€ with probability 6/10 1,60€ with probability 4/10	3,85€ with probability 6/10 0,10€ with probability 4/10	
2€ with probability 7/10 1,60€ with probability 3/10	3,85€ with probability 7/10 0,10€ with probability 3/10	
2€ with probability 8/10 1,60€ with probability 2/10	3,85€ with probability 8/10 0,10€ with probability 2/10	
2€ with probability 9/10 1,60€ with probability 1/10	3,85€ with probability 9/10 0,10€ with probability 1/10	
2€ with probability 10/10 1,60€ with probability 0/10	3,85€ with probability 10/10 0,10€ with probability 0/10	

Once you have completed the questionnaire, we will start the experiment.

Task one (partially adapted from Barr et al. 2018 and Krupka and Weber 2008)

You will receive a description of three situations. This description corre- sponds to situations in which one person, "Doctor X," must decide how to act. You will be given a description of various possible actions Doctor X can choose to take in response to each situation.

After you receive the description of the situations, you will be asked to eval- uate each of the various possible actions Doctor X can choose to take for each situation. You must indicate, for each of the possible actions, whether taking that action would be "appropriate" or "inappropriate" on a scale of 1 to 4, where 1 means very inappropriate and 4 means very appropriate. We consider an action to be appropriate when you think is the "correct" thing to do in each specific situation.

In each of your responses, we would like you to answer as truthfully as possi- ble, based on your opinions of what constitutes appropriate or inappropriate action. To give you an idea of how the experiment will proceed, we will go through an example situation and show you how you will indicate your responses.

Example Situation

Doctor X is treating a patient who has been admitted in the hospital rehab block due to a compound fracture of his shoulder. Doctor X can choose 4 possible actions to take: referring the patient to a fifteen-minute magneto- therapy per day; referring the patient to a forty-minute magneto-therapy per day, referring the patient to a three-hour magneto-therapy per day; referring the patient to a seven-hour magneto-therapy per night.

The table below presents the list of the possible actions Doctor X can choose. For each of the actions, you would be asked to indicate whether you believe choosing that action is very

inappropriate, somehow inappropriate, somehow appropriate, or very appropriate. To indicate your response, you would put a cross in the row corresponding to your belief about the degree of appropriateness.

	15-minute magneto-therapy per day	40-minute magneto-therapy per day	3-hour magneto-therapy per day	7-hour magneto-therapy per night
1Very inappropriate				
2Somehow inappropriate				
3Somehow appropriate				
4Very appropriate				

If this was the situation for this study, you would consider each of the possible actions above and, for that action, indicate the extent to which you believe taking that action would be “appropriate” or “inappropriate”. Recall that by appropriate we mean action that you think is the “correct” thing to do in each specific situation.

For example, suppose you thought that referring the patient to a three-hour magneto-therapy per day was very inappropriate, referring the patient to a fifteen-minute magneto-therapy per day was somehow inappropriate, referring the patient to a three-hour magneto-therapy per day was somehow appropriate and referring the patient to a seven-hour magneto-therapy per night was very appropriate, then you would indicate your responses as follows:

	15-minute magneto-therapy per day	40-minute magneto-therapy per day	3-hour magneto-therapy per day	7-hour magneto-therapy per night
1 Very inappropriate			X	
2 Somehow inappropriate	X			
3 Somehow appropriate		X		
4 Very appropriate				X

After completing the table, you will be asked to indicate how much certainty do you feel about your choices on a five-point scale, where 5 stands for a great deal of certainty.

If you have any questions about this example situation or about how to indicate your responses, please raise your hand now and we will assist you privately.

You will next be given the description of three situations where Doctor X faces various possible actions. You will be given a paper with the description of the situations and a pen to write down your answer. After you read the description, you must consider the possible actions and indicate on the paper you receive how appropriate these are in a table similar to the one shown above in the example situation.

Once you have completed the task an experimenter will come to collect your paper.

Payment procedure

Once all the papers have been collected, a computer will randomly select one scenario and for that scenario one action Doctor X can choose. Your evaluation of this action will be compared

with the response selected by the most people here today. If your evaluation coincides with the most frequently chosen option, you will receive 10 euros for this task, otherwise you will receive zero. For instance, imagine the example situation above was the actual situation and the possible action “Fifteen-minute magneto-therapy” was selected by the computer. If your evaluation had been “somehow inappropriate” then your task earnings would be 10 euros if this was the response selected by most other people in today's session and zero otherwise.

While the experimenters are calculating your total payoff, we ask you to complete a short, anonymous questionnaire. Please leave the questionnaire on your desk once you have completed it.

Scenarios

Now we present three different scenarios similar to the previous example. For each scenario, we propose four actions that Doctor X can take in response to that. For each of the actions, we ask you to indicate whether you believe choosing that action is very inappropriate, somehow inappropriate, somehow appropriate, or very appropriate. Recall that by appropriate we mean action that you think is the “correct” thing to do in each specific situation.

First scenario

Doctor X deals with a 26-year-old patient who sustained a tear to the anterior cruciate ligament together with a meniscus injury. Doctor X has four alternate options: performing joint aspiration (arthrocentesis), prescribing cryotherapy and magneto-therapy applications, suggesting surgery treatment, recommending ice and rest. Here is the table reporting all the four available actions. For each of the actions you have to indicate whether you believe choosing that action is very inappropriate, somehow inappropriate, somehow appropriate, or very appropriate. Recall that by appropriate we mean action that you think is the “correct” thing to do in each specific situation. To indicate your response, you have to put a cross in the row corresponding to your belief about the degree of appropriateness for each of the actions which Doctor X can take.

Remember that at the end of the experiment your evaluations will be compared with the most common answers provided today. You have a chance to earn additional money by matching the most frequently chosen option.

	Prescribing magneto -therapy and cryotherapy	Recommending ice and rest	Surgery treatment	Performing joint aspiration
1 Very inappropriate				
2 Somehow inappropriate				
3 Somehow appropriate				
4 Very appropriate				

How much certainty do you feel about your choice of optimum treatment? (on a scale 1 to five where 1 stands for no certainty and 5 for a great deal of certainty)

Do you want to know national guidelines content?

- Yes
- No

Do you want to change your previous answers?

- Yes
- No

Second scenario

Doctor X deals with a 28-year-old woman who suffers from breast fissures.

Doctor X has four alternate options: suggesting discontinuing nursing, suggesting discontinuing nursing and in the meanwhile prescribing protective creams, recommending correcting errors in latch, suggesting using pump. Here is the table reporting all the four available actions. For each of the actions you have to indicate whether you believe choosing that action is very inappropriate, somehow inappropriate, somehow appropriate, or very appropriate. Recall that by appropriate we mean action that you think is the ”correct” thing to do in each specific situation. To indicate your response, you have to put a cross in the row corresponding to your belief about the degree of appropriateness for each of the actions which Doctor X can take. Remember that at the end of the experiment your evaluations will be compared with the most common answers provided today. You have a chance to earn additional money by matching the most frequently chosen option.

	Suggesting discontinuing nursing	Suggesting discontinuing nursing and prescribing protective creams	Recommending correcting errors in latch	Suggesting using pump
1 Very inappropriate				
2 Somehow inappropriate				

3 Somehow appropriate				
4 Very appropriate				

How much certainty do you feel about your choice of optimum treatment? (on a scale 1 to five where 1 stands for no certainty and 5 for a great deal of certainty)

Do you want to know national guidelines content?

- Yes
- No

Do you want to change your previous answers?

- Yes
- No

Third scenario

Doctor X deals with a 54-year-old patient who has a cancer to the oral cavity

in the retromolar region which involves the pterygoid muscle. Doctor X has four alternate options: treating it with surgery and radiotherapy, treating it with chemotherapy, treating it with radiotherapy, treating it with surgery. Here is the table reporting all the four available actions. For each of the actions you have to indicate whether you believe choosing that action is very inappropriate, somehow inappropriate, somehow appropriate, or very appropriate. Recall that by appropriate we mean action that you think is the "correct" thing to do in each specific

situation. To indicate your response, you have to put a cross in the row corresponding to your belief about the degree of appropriateness for each of the actions which Doctor X can take. Remember that at the end of the experiment your evaluations will be compared with the most common answers provided today. You have a chance to earn additional money by matching the most frequently chosen option.

	Treating it with surgery and radiotherapy	Treating it with chemotherapy	Treating it with radiotherapy	Treating it with surgery
1 Very inappropriate				
2 Somehow inappropriate				
3 Somehow appropriate				
4 Very appropriate				

How much certainty do you feel about your choice of optimum treatment? (on a scale 1 to five where 1 stands for no certainty and 5 for a great deal of certainty)

Do you want to know national guidelines content?

- o Yes
- o No

Do you want to change your previous answers?

- o Yes
- o No

Questionnaire

_____

Participant ID . . . . . .

Demographic Questionnaire

The following questions ask for some information about you. Please answer each question by placing a mark where appropriate or by writing a brief response.

When you finish a page you may go on to the next one.

In this booklet, unlike the others, if you have a question and you raise your hand, we can assist you privately in deciding what is the best answer for your situation.

1. **What is you age?**.....

2. **What is your gender?**

- Male
- Female

3. **What best describes your race or ethnicity (please select all that apply)?**

- White
- Black/African/America
- Asian or Pacific Islander
- Hispanic
- Multiracial
- Other

4. What hospital/clinic do you work for?.....

5. **How long have you been with that hospital/clinic?.....**

6. **In general, how satisfied are you with the hospital/clinic you work in?**

- Very dissatisfied
- Somehow dissatisfied
- Somehow satisfied
- Very satisfied

7. **Which kind of medical specialty do you have?**.....

8. **How many doctors work in your department?.....**
9. **How often do you ask for colleagues' advices to take decisions?**

   - Never
   - Seldom
   - Often
   - Very often

10. **How much does your colleague's opinion affect you if he agrees with you?**

    - Not at all
    - Slightly
    - Somehow
    - Very much

11. **How much does your colleague's opinion affect you if he does not agree with you?**

    - Not at all
    - Slightly
    - Somehow
    - Very much

12. **How often decisions about a patient are taken as a result of a team's valuation?**

    - Never
    - Seldom

- Often
- Very often

13. How often do you share stuff with other colleagues (e.g. re- sults of diagnostic tests, outcomes of surgical procedures)?

- Never
- Seldom
- Often
- Very often

14. When you work in team

- a. Any decision taken is the result of a shared opinion
- b. There is a member whose opinion weights the most

15. If you have answered a to question 13, please skip to question 16. **What is the role of the team member whose opinion weights the most?**.....

16. Do you have a whatsapp group with your colleagues where you discuss decisions to take?

- Yes
- No

17. How often do you take part in training courses?

- Never
- Seldom
- Often

- Very often

18. How often do you take part in brainstorming sessions?

- Never
- Seldom
- Often
- Very often

19. How often do you consult scientific journals?

- Never
- Seldom
- Often
- Very often

20. Do you carry out and public scientific research?

- Yes
- No

21. How important are national guidelines in your decisions?

- Very unimportant
- Somehow unimportant
- Somehow important
- Very important

22. Does your firm adopt internal guidelines?

- o Yes
- o No

23. If you have answered no to question 22, please skip to question 24.

How important are internal guidelines in your decisions?

- o Very unimportant
- o Somehow unimportant
- o Somehow important
- o Very important

24. Do you think that your colleagues follow guidelines while taking decisions

- o Never
- o Seldom
- o Often
- o Very often

25. What is your average opinion of your colleagues?

- o Bad
- o Mediocre
- o Somehow good
- o Very good

26. What is your average opinion of your colleagues with respect to yourself?

- o Bad
- o Mediocre
- o Somehow good

- Very good

27. Do you have an insurance?

- Yes
- No

28. How do you judge your department in terms of coordination between colleagues?

- Bad
- Mediocre
- Somehow good
- Very good

Appendix B: additional tables and figures

Table 6: Subjects pool

Subject	Male	Age	Hospital	Department
1	1	63	GOM	ORTHOPEDICS
2	1	53	GOM	ORTHOPEDICS
3	1	56	GOM	ORTHOPEDICS
4	1	66	GOM	ORTHOPEDICS
5	0	51	GOM	GYNECOLOGY
6	1	45	GOM	GYNECOLOGY
7	0	44	GOM	GYNECOLOGY
8	1	39	GOM	GYNECOLOGY
18	0	39	GOM	GYNECOLOGY
19	1	55	GOM	HEMATOLOGY
20	0	58	GOM	INTERNAL MEDICINE
9	1	48	GOM	GYNECOLOGY
10	1	59	GOM	GYNECOLOGY
11	0	64	GOM	GYNECOLOGY
12	1	36	GOM	ORTHOPEDICS
13	1	49	GOM	ORTHOPEDICS
15	1	42	GOM	ORTHOPEDICS

16	0	62	GOM	GYNECOLOGY
17	1	41	GOM	RADIODIAGNOSTIC
21	0	54	GOM	GYNECOLOGY
14	1	53	GOM	ORTHOPEDICS
22	1	41	GOM	OTOLARYNGOLOGY
30	1	55	GOM	HEMATOLOGY
23	1	63	GOM	OTOLARYNGOLOGY
24	1	64	GOM	SURGERY
25	0	59	GOM	SURGERY
26	1	62	GOM	SURGERY
27	1	54	GOM	SURGERY
28	1	43	GOM	SURGERY
29	1	63	GOM	NEUROLOGY
31	0	45	GOM	ANESTHESIOLOGY AND INTENSIVE CARE
32	1	55	GOM	ANESTHESIOLOGY AND INTENSIVE CARE
33	0	31	GOM	NEONATOLOGY
34	0	44	GOM	NEONATOLOGY
35	0	56	GOM	NEONATOLOGY
36	0	47	GOM	NEONATOLOGY
37	0	57	GOM	NEONATOLOGY

38	1	34	GOM	ANESTHESIOLOGY AND INTENSIVE CARE
39	0	35	GOM	ANESTHESIOLOGY AND INTENSIVE CARE
40	1	53	GOM	ANESTHESIOLOGY AND INTENSIVE CARE
41	0	58	GOM	ANESTHESIOLOGY AND INTENSIVE CARE
42	1	56	GOM	NEONATOLOGY
43	1	50	GOM	NEONATOLOGY
44	0	46	GOM	ANESTHESIOLOGY AND INTENSIVE CARE
45	0	60	GOM	ANESTHESIOLOGY AND INTENSIVE CARE
46	1	57	GOM	CARDIAC SURGERY
47	1	42	GOM	CARDIAC SURGERY
50	1	52	GOM	TRANSFUSION MEDICINE
52	1	55	GOM	TRANSFUSION MEDICINE
48	0	53	GOM	NEONATOLOGY
49	0	61	GOM	NEONATOLOGY
51	0	50	GOM	TRANSFUSION MEDICINE

Table 7: Variables list

Variables	**Description**
male	=1 if male
age	Age
years of service	Years of employment
specialty	=1 if subject is specialized in the medical field of the vignette
research	=1 if subject carries out scientific research
updating	Frequency of consulting scientific journals
satisfaction	Job satisfaction on a four point scale
colleagues' advice	Frequency of asking for colleagues' advice
negative influence	Influence of colleagues' divergent view
positive influence	Influence of colleagues' concordant view
team	Frequency of taking decisions in team
leader	=1 if team led by a leader
results sharing	Frequency of sharing results with colleagues
training	Frequency of taking part in training sessions
brainstorming	Frequency of taking part in brainstorming sessions
national guidelines	Importance of national guidelines on a four point scale
internal guidelines	Importance of internal guidelines
colleagues' conformity	Assessment of colleagues' adherence to guidelines
colleagues judgment	Valuations of colleagues on a four point scale

risk seeking	=1 if subject risk seeking
valuation	Appropriateness judgment of the action
guidelines	=1 if action prescribed by guidelines
coordination	=1 if subjects coordinate
insurance	=1 if subject is insured
conformity with guidelines	=1 if subjects report guidelines to be ’very appropriate’
department coordination	physicians’ perception of their department coordination on a four-point scale

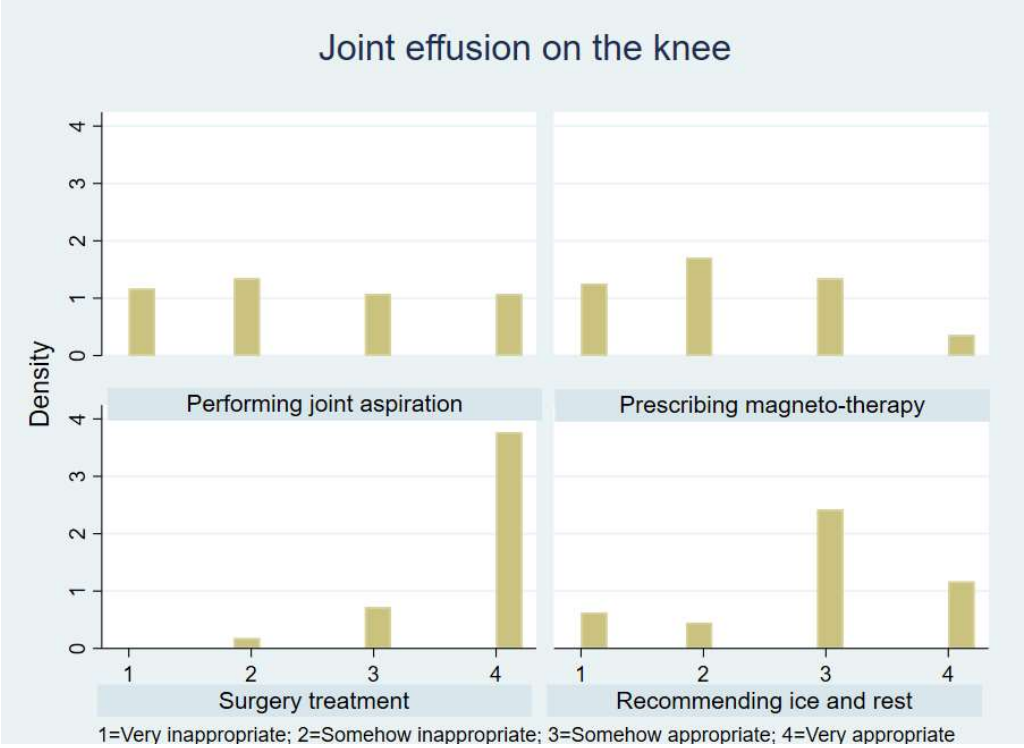


Figure 8: Density of the appropriateness judgment for action for the first vignette

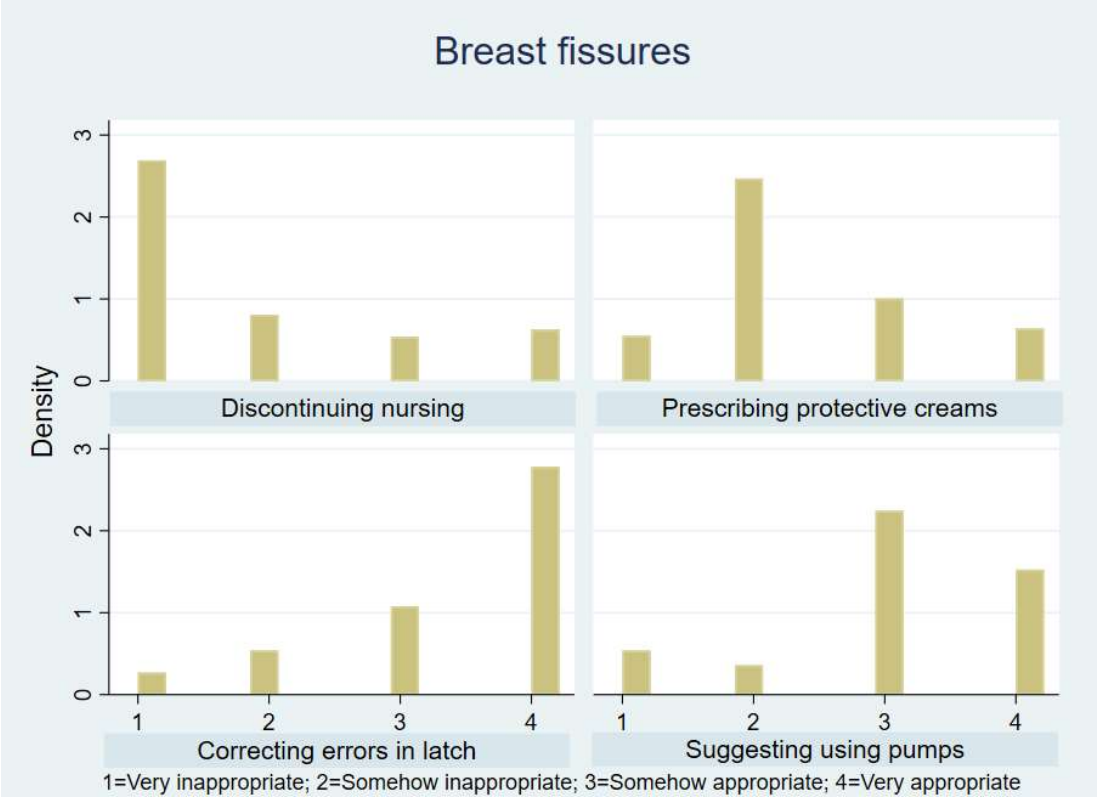


Figure 9: Density of the appropriateness judgment for action for the second vignette

CHAPTER 2

Do National Health Guidelines increase coordination level among physicians? An experimental investigation

Abstract

Practice guidelines are used in the healthcare sector to achieve the hospitals' desired outcomes, improving patient's health and saving resources while providing care. Despite their large promulgation, guidelines continue to generate confusion in interpreting and implementing them, and have hardly changed physicians' behaviour. For all these reasons, it could be useful to introduce a mechanism to elicit guidelines at any level, international, national or local. The purpose of this paper is to adapt the coordination game by Krupka and Weber (2008) in the context of the Italian hospitals to detect local and national guidelines using clinical vignettes taken from standardized cases. In this framed field experiment, physicians were asked to evaluate the appropriateness of each of the possible actions on a scale of one to four in order to match the modal rating. In the second stage of the experiment, physicians could receive information concerning what it should be done, in the specific situation, according to national guidelines (i.e. providing them with the written guidelines) and eventually change decisions. Results show that only 8% of the physicians did not want to know national guidelines content. However, only 23% of the subjects (with a choice frequency of 35%) decided to change their appropriateness judgment after realizing they were in contrast with guidelines in either one or two vignettes. Overconfidence, perceived guidelines ambiguity and default effect could explain such a low value. Finally, the more physicians consult scientific sources (i.e. signal of open-mindedness), the more willing to accept suggestions, coming from guidelines reading, and to change their decisions when they are wrong. Since guidelines dissemination is shown to increase the level of coordination between physicians, hospitals should consider effective programmes to spread their knowledge.

Introduction

Our world is driven by various norms (e.g. social or ethical) which define how people are meant to behave, given their peers' expectations (Opp, 2001 and Bicchieri, 2006). Depending on the culture, these rules give us an idea of what is considered to be appropriate behaviour (Mcleod, 2008). While their impact on individuals' actions has always been recognized by both sociology and psychology (Sherif, 1936), the influence of such norms in economic decisions has been neglected for a long time. However, in recent years many organizations have found shared norms to be the key to their success (Meek et al., 2010). In this regard, practice guidelines, which in fact proxy hospital group 'norms', are used in the healthcare sector to achieve the hospitals' desired outcomes, improving patient's health and saving resources while providing care (Shekelle et al., 1999). Even though they are not actual norms, they can be classified as soft law instruments (Meoli et al., 2018) which assist and uniform practitioners' actions, by identifying recommended courses of actions under certain circumstances (Institute of Medicine, 1990). Practice guidelines, often representing international gold standard reference, result from rigorous clinical research and are largely supported by experts and professionals in the field (Field and Lohr, 1990). This is the reason why they epitomize group norms. The purpose of this paper is to adapt the coordination game by Krupka and Weber (2008) in the context of the Italian hospitals to detect local and national guidelines using clinical vignettes taken from standardized cases (similar to Das and Hammer, 2004). Although this paper builds on the same experiment as in Chapter 1, its further contribution lies in the introduction of national guidelines as a tool to boost the level of coordination among practitioners,.

According to Hermalin (1999), agreed rules, goals and objectives (i.e. prerequisites for guidelines), which constitute each company's (i.e. hospital/health centre in such context) corporate culture, inspire employees' actions. For this reason, it is important to measure norms in order to predict the extent to which they will affect behaviour and ex-post to justify otherwise inexplicable outcomes, especially in the economic context. In the healthcare field, it becomes even necessary to detect practice guidelines for several reasons. First of all, practice guidelines are viewed as a way to help the law governing medical malpractice in setting the benchmarks for the appropriateness and inappropriateness of medical treatments (Gabbay and le May, 2004). In particular, their existence improves the ability of both the tort system and liability

insurers to evaluate the performance of individual physicians, identifying cases of negligence[29] (Havighurst, 1991). Secondly, in recent years some degree of diversity in clinical practice has been documented at national levels (Switzer et al., 2003). Even when considerable research has been done and proven methods have been applied to specific clinical cases, physicians can disagree on what is the most appropriate procedure to be followed (the so called 'macro uncertainty', Baumann et al., 1991) in order to balance patients' benefits with health risks and economic costs, depending on the facility they work in (Field and Lohr, 1990). Consequently, the production of clinical guidelines had an exponential growth so much that physicians, patients and other stakeholders have to juggle several guidelines which vary in quality and are sometimes discordant (Camera civile di Firenze, 2017). In particular, hospitals adopt their own internal guidelines, which do not necessarily overlap (Parker et al., 2019). Potential diversity in practice could arise from heterogeneity in patients' characteristics and differences in delivery system capacity given local resources. More generally, inconsistencies result from variations in the interpretation of scientific evidence and differences in values, interests and experience (Field and Lohr, 1990). The above-mentioned variety, which translate into conflicts in terminology and technique and may contribute to differences in national diseases outcomes (Parker et al., 2019) cannot be ignored. In this regard, public and private organizations are trying to resolve such inconsistencies by clarifying biases, preparing clear recommendations and providing important details about procedures and methods (Shekelle et al., 1999). Despite such efforts, guidelines continue to generate confusion in interpreting and implementing them, given their heterogeneity (Field and Lohr, 1990). Further to this, an increasing number of patients experienced some problems such as receiving conflicting information from different doctors (e.g., conflicting medical prescription) (O'Malley and Reschovsky, 2011), also due to conflicting guidelines (OECD Report, 2015). For all these reasons, it could be useful to introduce a mechanism to elicit guidelines at any level, international, national or local. Although, they may be easily accessible to the public, it is uncertain whether physicians

[29] Further details will be provided in the normative section. In the Italian context, Gelli-Bianco Law (No. 64/2017) ("Provisions on the safety of care and the assisted person, as well as on the professional responsibility of health care professionals") prescribes that when the behaviour of a doctor conforms to guidelines, in the absence of reasons to deviate from them, he cannot be responsible for the damage to the patient, since his behaviour appears to be diligent. However, the court must always check that the physician, who has followed guidelines, has correctly treated the patient (Monfeli, 2018).

perfectly know them and above all whether they share and apply them[30]. To reinforce this contention, Cabana et al. (1999) show that there exist several barriers restricting physicians' adherence to guidelines, for example: lack of awareness, time constraint, patients' characteristics and inertia of previous similar clinical cases. Given the increasing amount of research, it is hard to assume that physicians are familiar with every existing guideline and slavishly follow it**.** An already validated tool used in experimental economics to check whether people belonging to the same organization share common practices is the coordination game introduced in the first chapter (see e.g. Gaechter et al. 2013; Barr et al. 2018; Burks and Krupka 2012).

In this framed field experiment, physicians were asked to evaluate the appropriateness of each of the possible actions on a scale of one to four in order to match the modal rating. In the second stage of the experiment, physicians could receive information concerning what it should be done, in the specific situation, according to national guidelines (i.e. providing them with the written guidelines). Then, in one of the two different treatments, that is the novelty of this chapter, they have the chance to change their appropriateness judgments of the actions given in the first stage, knowing that the probability of getting the reward, whether they match the most frequent appropriateness judgment, could be reduced when they do not judge actions conform with guidelines as the most appropriate ones. Indeed, any participant knows that other colleagues could have decided to ask for the content of national guidelines and, hence, to change their responses in accordance (which obviously can shift the modal answer). Results show that only 8% of the physicians did not want to know national guidelines content. However, only 23% of the subjects (with a choice frequency of 35%) belonging to the treatment group decided to change their appropriateness judgment after realizing they were in contrast with guidelines in either one or two vignettes. Overconfidence, perceived guidelines ambiguity and default effect could explain such a low value. Finally, the more physicians consult scientific sources (i.e. signal of open-mindedness), the more willing to accept suggestions, coming from guidelines reading, and to change their decisions when they are wrong. Since guidelines dissemination is shown to increase the level of coordination between physicians, hospitals should consider effective programmes to spread their knowledge.

[30] Notice that physicians tend to overestimate their adherence to both local and national standards (Bauchner et al., 2001).

Literature review

This paper relates to at least two strands of research. First it integrates the widespread literature on the use of vignettes to learn how people think about a wide range of topics (see e.g Bursztyn. et al., 2018). Despite most used in other settings to measure the existence of personal and social norms (Sorenson and Taylor, 2005) the procedure of submitting to subjects' classical presentations (i.e. vignettes) is spreading in the healthcare literature. According to Glassman et al. (2000) and Das and Hammer (2005), vignettes representing dummy patients with common diseases based on realistic clinical situations are a practical and feasible tool to measure the accuracy of physicians' diagnoses and ultimately assess the quality of care. For example, Das et al. (2008) combined medical vignettes with direct observations of the physician-patient relationship to measure providers' knowledge and find that practitioners' competence and effort are very low in low-income countries. This analysis was taken one step further by Das et al. (2015) who resorted to standardised patients (i.e., real or simulated sent anonymously to hospitals) and medical vignettes to assess providers' competence on tuberculosis in India. Results show that competence in private facilities is higher compared to the public sector. In the same context, Mohanan et al. (2015) test specific knowledge concerning childhood diarrhea and pneumonia, providing evidence for Indian practitioners' poor knowledge of key diagnostic questions. Variations in physicians' diagnoses were instead tested by both Yager et al. (1986) and Gorter et al. (2001). In the first study, vignettes describing patients with psychological problems were shown to physicians including house staff of all levels of training. Data show significant variations in all diagnoses reported by physicians. In the second of the abovementioned studies, Dutch physicians were asked to complete a questionnaire concerning vignettes of patients with foot problems. Most physicians were found to be competent in diagnosing common foot problems. Similarly, variation in sick listing practice was investigated by Englund et al. (2000). Simulated cases were sent to doctors to check for physician gender difference in filling a certificate. Female doctors were found to be more willing to sign certificates than their male colleagues, regardless of other factors.

Alternatively, clinical vignettes can be used to assess physicians' confidence level, since overconfidence is often found to be responsible for diagnostic and treatment errors (see e.g. Berner and Graber, 2008). For example, Friedman et al. (2001) explore the possible connection between confidence and correctness asking participants to give a different diagnosis for each

of the proposed cases. Data indicate a small but positive correlation between correctness and confidence. However, participants show a general tendency for under confidence. Contrasting evidence is put forward by Kovacs et al. (2020) who find that not only most participants underappreciate their likelihood to be wrong, but this attitude is also correlated with lower abilities and worse performances.

While Leonard and Masatu (2005) argue that vignettes are not able to mimic what happens in a real consultation, and thus they should be combined with direct observation for that purpose, a larger strand of literature continues to use them as an instrument to measure physicians' clinical experience. With this in mind, in his study, Sandvik (1995) submitted vignettes depicting incontinent women to physicians, showing that the method was appropriate for group comparisons although responses to vignettes seemed to overestimate real performance. A further review supporting the practice of vignettes can be found in Veloski et al. (2005).

Further to this, this paper contributes to experimental studies on the role of information on individual's behaviour. King (1974) was one of the pioneers of such literature, showing that providing plant managers with artificial reports about job rotation programs improves their productivity. In the same scenario, Griffin (1983) demonstrates that supervisors' advices on tasks attributes positively change employees' approach to duties. Beneficial effects of information are also reported in Duflo and Saez (2003) who find that when proposals for retirement plans are complemented with education programmes the enrolment rate significantly increases. This contention is reinforced in the healthcare sector by Bauchner et al. (2001) who report that medical education in the form of discussions, case studies and local opinion leaders improves physicians' behaviour. More specifically, Braddock et al. (1999) see that providing physicians with patients' information about laboratory tests reduces medical errors and encourages preventive care. Supporting this view, Allery et al. (1997) show that any form of information such as that gleaned from medical journals, scientific conferences, discharge letters, contributes to change and improve physician's behaviour. Further literature on the role of information in changing physicians' behaviour is reviewed in detail in Beilby and Silagy (1997) and Grimhaw et al. (2001).

While different approaches have been already used to investigate physicians' knowledge and mainly to change their behaviour, none of them has combined dissemination strategies of common practices with the incentive structure provided by the coordination game.

Guidelines

According to the Institute of Medicine (1990), guidelines assist practitioners and a wide range of healthcare professionals about diagnostic and treatment modalities for specific clinical circumstances. Typically, their development guarantee that treating patients according to guidelines will bring about the desired outcomes (Shekelle et al., 1999).

Clinical areas for guideline development arise from an assessment of the main causes of mortality and morbidity for a given population, the need to control costs and ensure that funds are optimally spent or evidence that common procedures improve patient outcomes in the long term (Hjelmgren et al., 2001). Topics can be disease-based or procedure-based. When we fall in the first category, the focus is on the appropriate use of all available alternatives for treating a given disease. Alternatively, when guidelines refer to the procedures, they contain all the preconditions for which physicians should resort to particular interventions (Lomas et al., 1991).

Clinical recommendations development is quite expensive in terms of both financial and human resources, requiring the contribution of people with a wide range of skills. In particular, the designing process begins when expert clinicians from different areas are called into question to interpret evidence, eventually coming from small randomised clinical trials or controlled observational studies. Experts' opinions are then used to assess the generalisability of such evidence and finally to extrapolate results which could be extended to the population of interest in the guideline (Shekelle et al., 1999). Once sufficient evidence has been collected and results are gleaned, clinical pathways are designed, specifying their prerequisites for effectiveness and their relevant expected outcomes (Woolf, 1994). In addition, before being enacted, people responsible for guideline development must consider the feasibility of treatments which depends on the resource constraint and the healthcare system ability to implement staff and equipment in accordance with clinical recommendations (Shekelle et al., 1999). However, prior to their adoption, guidelines ought to receive an external review from expert clinicians who

certify their full respondence to validity and clarity requirements (Royal College of General Practitioners, 1995).

Although guidelines were initially aimed to reduce clinically significant variations in services and procedures provided by practitioners, their increasing production is leading to reversed results (Hjelmgren et al., 2001). Guidelines are often inconsistent and since they are not intended to replace any knowledgeable health professional's advice[31], they risk driving physicians to different conclusions, starting from the same evidence, when they do not lead to an unambiguous interpretation (Kahan et al., 1996). In fact, although guidelines should be theoretically exhaustive and straightforward, they are often vague or nonspecific in practice (McDonald and Overhage, 1994). For example, by using vignettes of patients suffering from back pain Shekelle et al. (2000) found that specific guidelines reduce inappropriate orders and increase appropriate orders for electro-diagnostic tests compared to abstract and generic guidelines. In this regard, in the UK, the National Institute for Health and Care Excellence stated that recommendations must be clear and concise, while containing all the information needed to be understood (e.g. description of the clinical problem, patient characteristics and delivery setting) (Feder et al., 1999). However, interests of clarity could sometimes hit the need to not overcome the supporting science. In similar circumstances, ambiguity in the guideline formulation results from the mere adherence to actual evidence. For example, for some screening tests it is not possible to specify any interval of time, given inadequate evidence. Therefore, departing from evidence, coming up with an answer to such lack of specifications would certainly clarify guideline content but might be harmful for both patients and physicians (Woolf et al., 2012).

Another aspect is that developing good guidelines does not always involve their use into practice (Feder et al., 1999). A first explanation comes from overconfidence which leads physicians to deviate from guidelines. As outlined by both Leonard and Masatu (2005) and Kovacs et al. (2020) overconfident providers barely resort to decision support tools in the form of guidelines, feeling that their practice conforms to standards when in fact it is not the case. Additionally, guidelines characteristics could affect compliance. First of all, Cabana et al. (1999) argue that physicians hardly follow guidelines prescribing the elimination of a pre-

[31] Neither research-based medicine nor clinical guideline are supposed to consider the variability of the multifarious considerations which play a role in clinical decisions (Gabbay and Le May, 2016).

existent practice, but they are likely to adhere to guidelines adding a new behaviour to the existing procedure. Secondly, while in some countries guidelines are formalized and thus mandatory, in others their application is still voluntary (Drummond, 1994). Consequently, in spite of their large promulgation, guidelines have hardly changed physicians' behaviour (Hayward, 1997), especially that of family practitioners[32]. In fact, according to Lomas et al. (1989) guidelines are unable to affect physicians' decisions in the absence of incentive mechanisms (i.e. financial interventions) and reminders which should operate at the local level. Nevertheless, disseminating guidelines through their publication in academic journals or making physicians directly aware of them (i.e. proving them with printed educational materials) turned out to be ineffective in improving professionals' behaviour (Freemantle et al., 2000). On the contrary, to maximize physicians' likelihood to conform to clinical guidelines, each organization should identify its own dissemination and implementation strategies, which vary according to available resources and existing barriers to their application (i.e. which are health-care system specific). A good example of educational strategy is given by the implementation programme of sedation guidelines in Italy, informed by 17 seminars all over Italy to spread knowledge of guidelines content, followed by ex-post meeting to monitor their use (Conigliano and Rossi, 2006). In particular, workshops and seminars are successful when barriers affect physicians' knowledge[33], while feedbacks can be useful when clinicians ignore suboptimal practices. Alternatively, marketing strategies together with local consensus mechanisms can fight barriers due to the existing culture and current practices of healthcare professionals (Feder et al., 1999). In conclusion, successful dissemination programs should not only be aimed to increase physicians' knowledge, but also and most importantly to affect their attitudes and behaviour (Lomas et al., 1991).

Hypotheses

[32] A few studies find that general practitioners are less likely to implement guidelines than are specialists (see e.g. Lomas et al., 1991 and Wensing et al., 1998). The reason lies in the variety of diseases they face which lead them to decide according to symptoms and patients' problems rather than guidelines.

[33] When organizations do not find published guidelines in line with their priorities, they can adjust their priorities or even develop their own guidelines (Feder et al., 1999).

Based on the model fully described in the previous chapter and on the literature reported above I can derive several hypotheses exclusively concerning guidelines[34] which this experiment is going to test.

First of all, I wonder which factors might affect physicians' likelihood of asking for national guidelines support, when given the opportunity in the second stage of the experiment. According to Tannenbaum (1994) specialists rely on outcome research when the clinical case they face does not fall within their health field and, thus, they cannot draw on their experience. More generally, while more experienced physicians hardly follow guidelines[35] (Meyer et al., 2013), less experienced clinicians who still do not boast acquired skills more easily resort to practice guidelines. Consequently, when physicians are given the opportunity to check guidelines, I expect that less experienced physicians (i.e. measured through their years of service) and physicians who are not familiar with the medical speciality covered by the vignette are more likely to ask for guidelines.

Behavioural hypothesis 2.1: Less experienced physicians are more likely to ask for guidelines than more experienced clinicians.

Behavioural hypothesis 2.2: Physicians who face a vignette which does not fall within their health field are more likely to ask for guidelines compared to their competent colleagues.

A plausible contributing factor is physicians' confidence. As noted by Arkes et al. (1996) overconfident people do not feel the need for assistance, epitomized by clinical guidelines in such a context. To be specific, overconfident providers are unlikely to resort to reference materials to diagnose and treat clinical cases, overstating the correctness of their judgments[36]. For this reason, overconfidence is often listed as one of the many barriers to guideline compliance (Cassam, 2017). Since it can be estimated thanks to physicians' answer to the question 'How much certainty do you feel about your choice of optimum treatment?', I can

[34] As far as coordination hypotheses are concerned, I refer the reader to the previous chapter.

[35] They are said to be contradictory and unable to consider all the factors which play a role in medical decisions (Gabbay and le May, 2004). A more extensive discussion in this regard is provided in the previous chapter.

[36] According to the taxonomy provided by Leonard and Masatu (2005), when overconfidence describes individuals' tendency to overstate the correctness of their judgments, we refer to as overprecision.

investigate the eventually existing correlation between physicians' confidence and their call on guidelines.

Behavioural hypothesis 2.3: Overconfident physicians[37] seldom call on guidelines, compared to underconfident clinicians.

Another factor which may affect physicians' willingness to ask for guidelines is liability. Notably, depending on their risk aversion level, physicians are relatively concerned about their exposure to medical claims. In fact, it is undisputed that adherence to guidelines can reduce the risk for a physician of being sued for malpractice (Elstad et al., 2010). As a result, physicians resort to practice guidelines which provide a shield against malpractice charges (Armed Forces Institute of Pathology U.S., 2008). In fact, in many countries, compliance with guidelines guarantees an absolute defence against medical liability (Hyams et al., 1995). In Italy, for example, recommendations contained in national guidelines are used to assess physicians' conduct in malpractice cases[38]. In the end, since risk aversion is measured at the beginning of the experiment, through the Holt and Laury (2002) questionnaire, physicians' attitude towards risks can be estimated, supporting the following hypothesis:

Behavioural hypothesis 2.4: Risk averse physicians ask more frequently for guidelines content.

Guidelines can represent a solution to lack of coordination. In fact, successful models of care coordination are based on standards of care and decision support tools (World Health Organization, 2018). Physicians agree that not only guidelines improve quality of care, but they could foster coordination (Carrier et al., 2012). In other words, clinical guidelines implementation can solve treatment variation which could result in inadequate health outcomes (Benavidez and Frakt, 2019). According to an informal conversation with some physicians, it is common practice that clinicians resort to guidelines when they do not agree on what is the best thing to do. Consequently, the lower the level of coordination in the department, the more

[37] Once physicians complete the specific vignette, they are asked to indicate how certain they are about their answers on a five-point scale (where 1 stands for no certainty and 5 for a great deal of certainty). Physicians reporting either 4 or 5 are classified as overconfident subjects.

[38] https://www.diritto.it/linee-guida-e-responsabilita-medica-le-risposte-alle-domande-piu-frequenti/

frequent the use of guidelines. Based on the above, I expect that most physicians ask for guidelines in vignettes characterized by lower level of coordination (ex post measured).

Behavioural hypothesis 2.5: The lower the coordination in the vignette, the larger the request for guidelines.

Finally, the last hypothesis refers to the second treatment and specifically to the third stage of the experiment. Physicians have the opportunity to change their previous answer to vignettes, once they know guidelines content. As mentioned in the previous chapter, mindlines, guidelines in the head, acquired over a lifetime, play a crucial role in physicians' decision making (Samuels and Ropper , 2005). Physicians in face of a specific clinical circumstance form their own diagnostic hypothesis based on their own experience (i.e. mindlines). However, when physicians ask for guidelines and notice that they could disagree with their colleagues (i.e. comparing guidelines content with their previous choices), they can change their answers (Gabbay and le May, 2004). In a nutshell, physicians prefer agreeing with their colleagues on what is the best practice to adopt, neglecting their mindlines. As a result, when guidelines prescribe an action which differs from the action ranked as the best possible option, physicians will probably change their answers in accordance with guidelines.

Behavioural hypothesis 2.6: In order to coordinate with others, physicians change their choices when guideline suggests adopting an action not previously chosen as the most appropriate.

Design

The basic framework of the experiment echoes that of a coordination game (see e.g., Gaechter et al. 2013; Barr et al. 2018; Burks and Krupka 2012). Physicians are randomly assigned to either the control group or the treatment group. Both the treatments share the first and the second stage of the experiment. Specifically, physicians face three different vignettes, each referred to a different department, depicting a patient suffering from a particular disease with a given diagnosis. For each vignette, a set of four treatments (i.e. one of them has been set according to national guidelines) in response to the disorder is proposed. In the first stage, physicians have to rank each of the alternatives, on a scale of one to four, according to their

perceived degree of appropriateness, where 4 stands for 'very appropriate', while 1 corresponds to 'very inappropriate'. The instructions clarified that an action is appropriate when a physician believes that it the correct thing to do given the specific situation. Physicians have to judge the degree of appropriateness knowing that they have to match the modal answer in order to get the reward. Before moving to the next stage of the experiment, physicians are asked to indicate how certain they are about the answers given to vignettes, choosing according to a five-point scale (similar to Kovacs et al., 2020 and Baumann et al., 1991).

In the second stage of the experiment, physicians are asked whether they want to know guidelines content, alternative they can move to the next scenario and leave the room once all the scenarios are completed. Since the experiment is run pen and paper, once all participants make their choices and papers are collected, an experimenter reads aloud guidelines content. Differently, in the treatment group, physicians who have decided to ask for guidelines can change their answers given in the first stage of the experiment[39]. Here, modal answers are calculated considering the answers given in the first stage (i.e. for physicians who did not ask for guidelines), together with answers given in the third stage (i.e. for the portion of participants who have changed the answers). At the end of the experiment one of the vignettes and one of the actions are randomly selected and once the experimenter has checked the modal answer, only physicians matching the modal answer are paid with a 10-euro meal ticket.

Procedure

The experiment was conducted at the main hospital of Reggio Calabria, thanks to an agreement signed by the same hospital and the Mediterranean University[40]. 52 physicians took part in the experiment: 24 of them joined the control group while the remaining were assigned to the treatment group.

[39] Notice that they do not know that they have the chance of changing their decisions when they are asked if they want to receive guidelines.

[40] The agreement previews that parties are committed to participating in the research project 'Experiments in health economics' and to sharing results which could be useful to the hospital for management policy.

Since vignettes specifically address three departments (i.e. orthopaedics, paediatrics and oncology), I required the participation of clinicians working there. However, I extended the participation to other departments (e.g. surgery), sometimes inconsistent with the medical case of the vignette, to increase the subject pool. This aspect constitutes the main limitation of the paper: the impossibility of submitting vignettes to only specialized physicians, which could lead to misleading results. In fact, physicians' willingness to know guidelines and then to change decisions could vary with their speciality[41].

Physicians joined the sessions during their coffee-breaks. For this reason, the experiment was conducted in different tranches, and with pool of different sizes. However, physicians know that the modal answer was calculated considering all the answers given in the same day, though through different sessions.

Before starting the experiment, I measured physicians' attitude towards risks. Being, for each of the vignettes, one of the alternative proposed options corresponded to the course of action suggested by national guidelines, physicians could be willing to report such action to be very appropriate, just because they were risk averse. In particular, here the fear of being sued for malpractice may play a crucial role: physicians could try to reduce the probability of a legal claim by complying with guidelines to support a defense of care provided according to standards (Hyams et al., 1995). Since liability and risk aversion are inseparably connected, I need to estimate a dummy for each of the participants whose value depends on their attitude towards risk. To do this, I handed participants a questionnaire first proposed by Holt and Laury (2002), whose details have already been provided in the previous chapter (see e.g. Galizzi et al., 2016).

Once the questionnaire was completed, the experiment started. Participants find written instructions and examples on the way the tasks were to be completed on their desks. Instructions were read aloud.

Payoffs can be calculated at the end of the daily sessions for the above-mentioned reasons. Consequently, participants were finally paid with voucher meals only the day after the

[41] A possible solution could be involving physicians from other hospitals and then other cities, increasing the sample pool, but focusing on specific departments selected in accordance with the concerned vignette.

experiment. The rationale of meal tickets stems from the very high physicians' opportunity cost. Opting for cash payments meant paying physicians almost their opportunity costs, making the whole experiment very costly compared to the disposable funds. Furthermore, since participants are already moved by their intrinsic motivation, I chose to compensate them with a voucher ready to be spent during physicians' lunch break, in order to avoid crowding out effect (Gneezy and Rustichini, 2000).

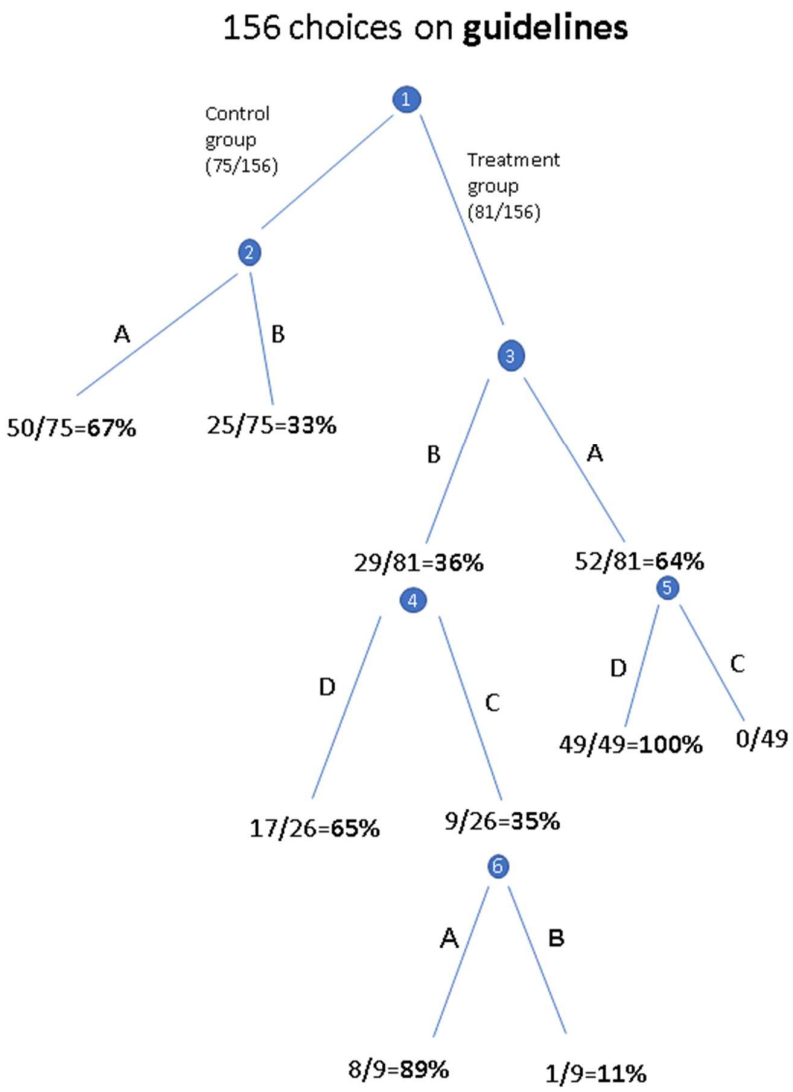
Results

In this section I am going through the experimental results. First, I will provide some descriptive analysis, focusing detailly on physicians' approach towards guidelines, going then to regression models. From then above I will use the concepts of coordination and conformity with guidelines fully described in the previous chapter.

Descriptive analysis and non-parametric tests

Why should physicians ask for guidelines? According to Connelly et al., (1990) accessibility and searchability (i.e. '*easy way to find relevant knowledge in the available source*'), are the two main factors affecting physicians' use of information (Connelly et al., 1990). Reasonably, physicians in this experiment do not have any opportunity cost (e.g. time and energy required to obtain information, Timpka et al., 1989) in knowing guidelines, since they are easily accessible just checking the option 'yes'. In fact, only 8% of the physicians enrolled in the experiment did not want to know national guidelines content, which corresponds to four physicians in the whole, equally divided between the control and the treatment group.

Limiting the analysis to physicians' choices on the actions corresponding to guidelines, the structure of the choices can be represented in figure 2.1.



A= conformity with guidelines
B= nonconformity with guidelines
C=change
D=no change

Figure 2.1: Sequence of the experiment

The starting value, 156, computes all the choices concerning guidelines, that is three choices, one for each vignette, for 52 subjects. All the numbers reported in the decision tree indicate the frequency of choices (i.e. each subject makes several choices which means that frequency does not correspond to the number of subjects). 65% of the times, physicians, regardless of the group, conform with guidelines from the very first stage, assigning the highest appropriateness level to the action reflecting guidelines. This percentage is shared between the two groups as follows: 49% for the control group (i.e. 50 times), corresponding to the left branch of node 2, and 51% for the treatment group (i.e. 52 times) on the right branch of node 3. Going to the right subgame starting from node 3, 36% of the times subjects from the treatment group did not conform with guidelines from the first stage. Of them, 65% of the times physicians knowing guidelines (i.e. checking the option yes) and having the opportunity to go back, decided to maintain their decisions, though realizing they were in contrast with guidelines[42]. On the contrary, 35% of the times physicians decided to change their previous decisions (right branch of node 4), giving the maximum appropriateness judgements to the action fulfilling guidelines 89% of the times (left branch of node 6). For those belonging to the treatment group who did not conform with guidelines at the first stage and either decided not to change their decisions or changed them while not complying with guidelines, guidelines ambiguity and physicians' overconfidence could be the reason. In fact, a negative correlation (though not significant) between physicians' confidence, their guidelines request and, ultimately, their decisions to change their appropriateness judgment is found, confirming Leonard and Masatu (2005). Furthermore, physicians could disagree with guidelines, when they contrast their priorities and expectations (Beaulieu et al., 1999). Additionally, according to Hayward (1997) guidelines rarely change physicians' attitude, despite the incentive mechanism provided in this experiment. Finally, physicians' previously made choice (at the very first stage) may act as a default option and, thus, as a reference point for the subject (Jachimowicz, 2019). Notably, as outlined by Park et al. (2000) the endowment effect (Tversky & Kahneman, 1974) leads decision makers to overvalue the default option, shifting their reference point, when compared with the

[42] Notice that from the 33 choices, the number decreases to 26 in node 4, removing the situations where physicians did not access the third stage, having refused to know guidelines.

alternatives. Similarly, Dinner et al. (2011) show that subjects stay with the pre-selected options to minimize their cognitive effort.

Finally, only 11% of times (right branch of node 6), corresponding to one physician in the oral cavity cancer scenario[43] who was first in contrast with guidelines and decided to change his previous answer did not align his choices with guidelines, maybe due to guidelines ambiguity.

Moving to guidelines effect on coordination, guidelines introduction increases physicians' coordination, as it is shown in figure 2.2.

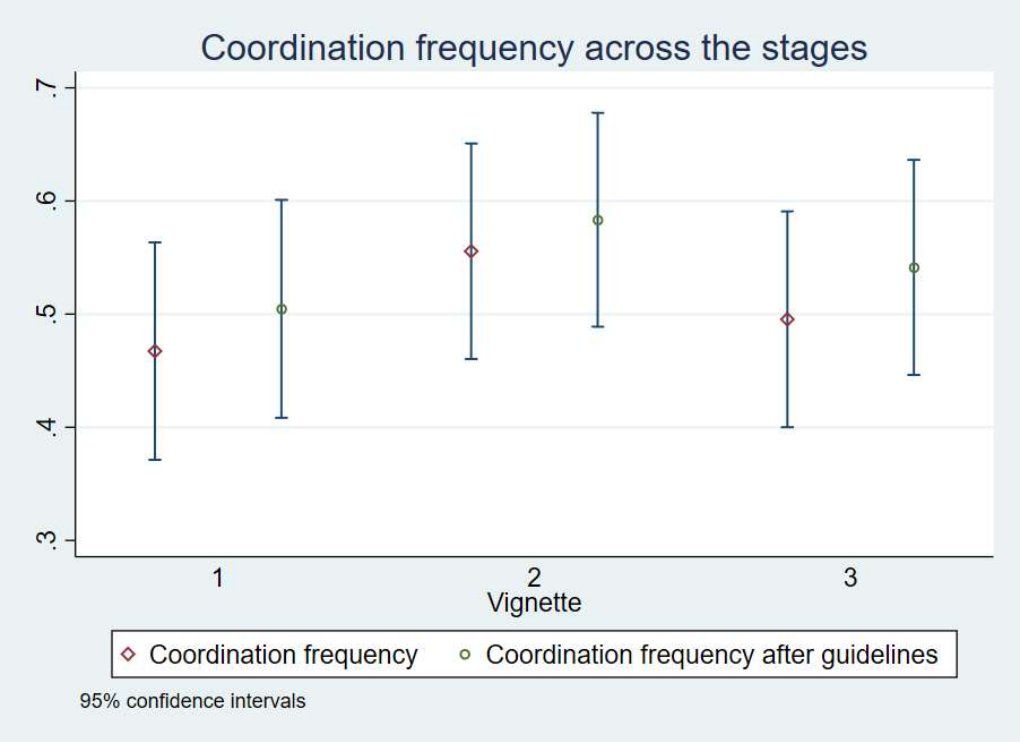


Figure 2.2: Confidence interval for the coordination frequency across the experiment

The average frequency of coordination, for the treatment group, in the first stage of the experiment is 0.51, which rises to 0.54 in the third stage, once physicians have the opportunity to change their decisions. Despite differences are not significant (p-value=0.34), and confidence interval bars seem to overlap, the greatest benefit of guidelines introduction can be seen in the

[43] Appendix C provides figure 2.1 broken down by vignette.

third vignette (oral cavity cancer scenario) even though values are very close from one vignette to another.

As one would expect, conformity with guidelines slightly increases with their introduction, as figure 2.3 demonstrates, despite differences are not significant according to the Wilcoxon test (p-value=0.13).

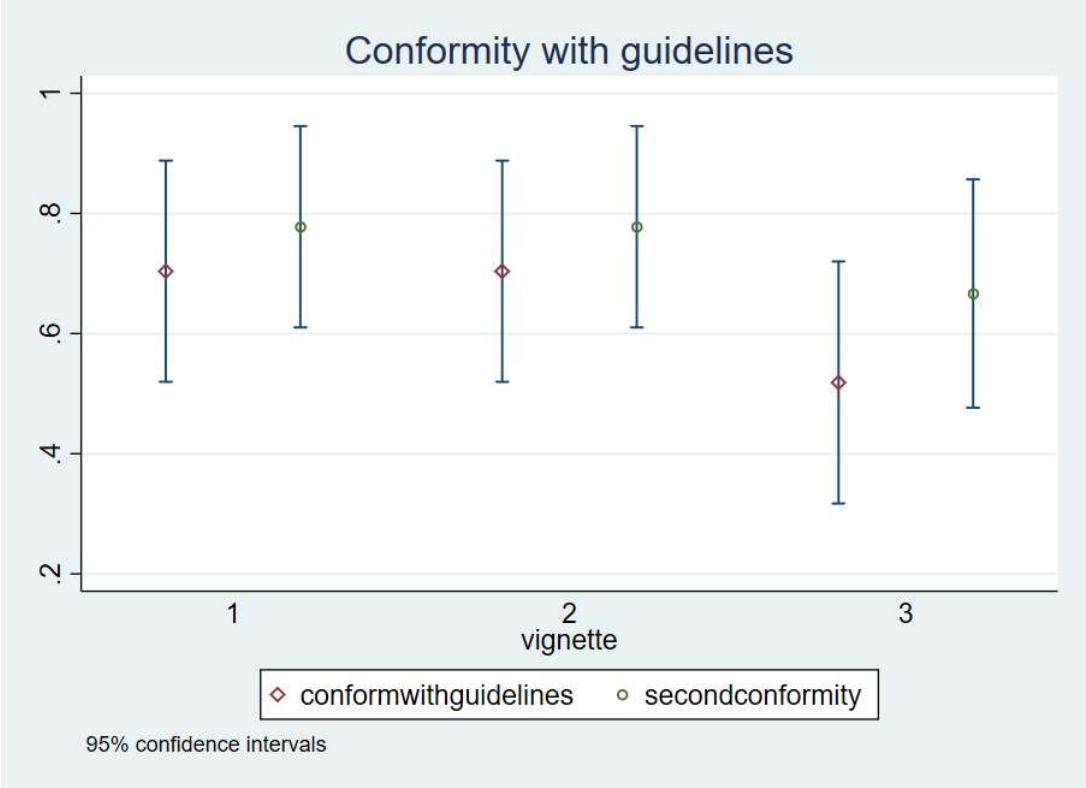


Figure 2.3: Confidence interval for conformity with guidelines frequency across the experiment

Conformity frequency goes from 0.64 (i.e. prior to introducing guidelines) to 0.74(i.e. once physicians have the possibility of changing their judgments) and similar to what happened before, the third vignette seems to benefit the most from guidelines support.

Finally, differences in the appropriateness judgments between the control group and the treatment group are not statistically significant. Specifically, for the treatment group,

differences in the valuations between the first stage and the third stage, where physicians can change their decisions, are not significant according to both the median and the ranksum test.

Regression analysis

For the analyses concerning what affects the probability that physicians conform with guidelines, I refer the reader to the previous chapter. In this section, I first estimated the probability that physicians resort to guidelines for the whole sample, whose output is reported in table 2.1.

Table 2.1: Logit for request guidelines

VARIABLES	(1) First model	(2) Second model	(3) Third model
confidence	-0.383**	-0.223	-0.260
	(0.187)	(0.322)	(0.357)
speciality	0.535	0.280	0.282
	(0.622)	(0.694)	(0.817)
research	1.145	0.374	0.370
	(1.353)	(1.556)	(1.549)
updating	-0.329	0.0404	0.0372
	(0.998)	(1.136)	(1.131)
colleguesadvice		-0.483	-0.477
		(0.457)	(0.455)
yearsofservice		0.250**	0.250**
		(0.104)	(0.104)
vignette1			0.169
			(0.212)
vignette2			0.0979
			(0.262)
Constant	4.507	2.900	2.946
	(3.360)	(4.082)	(4.115)
Observations	604	604	604

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Reasonably, confidence reduces physicians' probability of asking for guidelines. The more physician feels to be right in judging the appropriateness of each of the action, the less his need for guidelines support.

Then, I restricted the dataset to the 28 subjects who joined the treatment group, using robust cluster errors. Given the small number of clusters, bootstrap methods, originating pseudo-samples from the initial pool, are used (Roodman et al., 2019).

To be allowed to change their decisions, subjects have to know what national guidelines prescribe. Therefore, not to all the 28 subjects and/or not for all the situations (depending on their choices) it was given the opportunity to change decisions. In the presence of censored observations for changing decisions, when physicians do not access that stage of the game, not requiring guidelines, a two-stage limited dependent variable model is needed. However, a bivariate logit with accounts for sample selection is not convenient for this dataset, since the Likelihood Ratio test for independent equations shows that the null hypothesis that request guidelines and changing decisions are in fact independent choices cannot be rejected. Among all the other options, Tobit test suggests that a Cragg's model is the most suitable for this dataset (Blundell and Meghir, 1987). The Cragg's model consists of two different estimations: a logit for the discrete choice which is the probability that the latent variable takes a nonnegative value (changing decisions can take a value when physicians opt for guidelines) and a truncated regression for the continuous decision of changing decisions. The first estimation is reported in table 2.2 for the variable '*Request guidelines'*, which takes the value 1 if physicians ask for guidelines. '*Research*' is a dummy variable which equals 1 if physicians carry out scientific research. ''*Years of service*' indicates physician's working experience in the metropolitan hospital of Reggio Calabria. 'Colleagues advice' refers to physicians' frequency of asking for peers' suggestions.

Table 2.2: Logit for request guidelines for the treatment group

VARIABLES	(1) First model	(2) Second model
research	-1.626**	-1.626
	(0.768)	(1.034)
yearsofservice	0.407***	0.407**
	(0.146)	(0.171)
colleguesadvice	-0.888	-0.888
	(2.382)	(2.177)
Vignette 1		0
		(0)
Vignette 2		0
		(0)
Constant	3.887	3.887
	(7.160)	(6.382)
Observations	81	81

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Being involved in scientific research decreases the probability of requesting guidelines. According to Katzka (2017), physicians who do clinical research explore a number of possibilities in clinical care, trying to respond to unanswered questions. Researchers sometimes experiment new solutions, never used before, that is the reason why they do not need guidelines. Since they are used to improvise and think critically, in the absence of already known dispositions, they do not perceive the need for guidelines.

Physicians' experience in the hospital increases the probability of asking for guidelines. Intuitively, physicians who have longer been in practice perceive the need to stay abreast of new medicine evidence. In fact, Alper et al. (2004) estimate that virologists for example need 627.5 hours per month to keep current, given the volume of monthly published literature. As a result, the earlier physicians' training, and so the greater their years of services, the more pressing the need for updating. Finally, often asking for colleagues' opinions reduces physicians' willingness to request guidelines. The explanation is straightforward if we assume that peers' consultations are comparable to any other information source (see e.g. the review by Dawes and Sampson, 2003) such as a piece of paper containing guidelines. In a nutshell,

physicians can compensate their lack of knowledge asking for their colleagues' point of views instead of resorting to an external source.

Then, through a truncated regression, which excluded 24 observations, I assessed which variables contribute to physicians' decision of changing appropriateness judgements, once they know guidelines. The dependent variable '*changing decisions*' takes the value 1 whether physicians choose to change their selection in the vignette[44], 0 otherwise. '*Updating*' is a categorical variable referred to physicians' frequency of consulting scientific journals on a four-point scale, where higher values imply greater frequency. '*Riskseeking*' is a dummy variable capturing risk taking subjects. '*Department coordination'* is a categorical variable indicating physicians' feeling about their department level of coordination on a four-point basis. Finally, '*Vignette*' distinguishes the three different vignettes proposed in the coordination game. Table 2.2 shows the output.

Table 2.3: Truncated regression for changing decisions

VARIABLES	(1) First model	(2) First model	(3) Second model	(4) Second model
updating	0.0882		0.124	
	(0.0854)		(0.0821)	
riskseeking	0.223		0.167	
	(0.143)		(0.173)	
speciality	-0.145***		-0.0847	
	(0.0512)		(0.0666)	
age			0.0103	
			(0.00668)	
yearsofservice			-0.0119*	
			(0.00659)	
Vignette 1			-0.127	
			(0.0947)	
Vignette 2			-0.103*	
			(0.0573)	
sigma		0.301***		0.285***
		(0.0422)		(0.0461)
Constant	-0.190		-0.617*	
	(0.272)		(0.342)	
Observations	75	75	75	75

Standard errors in parentheses

[44] Notice that physicians have the chance of changing decisions only if they ask for guidelines and that only 2 physicians over 28 do not require guidelines support.

*** p<0.01, ** p<0.05, * p<0.1

'Updating' increases physicians' probability of changing decisions. Going backwards, according to Haug (1997) physicians try to compensate their lack of knowledge seeking information from different sources (e.g. journals, books). As a confirmation of this, a positive correlation between updating and physicians' request of guidelines is found (p-value<0.001). Such piece of information is then used to solve clinical case (Stinson and Mueller, 1980). As a result, I can reasonably conclude that the more physicians consult scientific sources (i.e. signal of open-mindedness), the more willing to accept suggestions and change their decisions when they are wrong.

Age increases the probability of changing decisions, though younger physicians are more inclined to consult scientific journals than their experienced colleagues (p-value<0.001), confirming Stinson and Mueller (1980). In a nutshell, older physicians, who seem to be less informed of current advances in medicine, when they ask for, welcome new suggestions from guidelines and more easily adopt them.

Speciality, which is positively correlated with physicians' confidence (p-value<0.001), results in practitioners' narrow-mindedness. As a result, physicians specialized in that specific vignette, convinced they are right, are less likely to change their decisions. This result is indirectly related to the positive coefficient for vignette 3 (oral cavity cancer) in the second model. Due to a relocation of the oncology department of the hospital where I conducted the experiment, I was not able to involve cancer experts. As a result, the absence of oncologists explains why the probability that physicians change their appropriateness judgments increases in the third vignette.

Finally, department coordination positively affects physicians' likelihood to change their choices. Logically, physicians' perception of good coordination in their department signals good communication between colleagues and above all broad-mindedness which makes it easy for them to accept opinions and suggestions. As a result, they assume that it is more likely that when their colleagues receive a nudge (i.e. guideline) they implement it to preserve coordination. In summary, each physician knows that his colleagues are likely to change their decisions and does the same to match the modal answer.

Conclusions

This experiment investigates physicians' adherence to guidelines and their potential use as a nudge policy instrument. In the context of the metropolitan hospital of Reggio Calabria, 52 physicians, coming from different departments, were asked to respond to three vignettes representing three medical scenarios. For each of the vignettes, a series of actions, among which the physician could choose to treat the patient, were offered to physicians who have to check their appropriateness level. To receive the final reward, physicians have to select the appropriateness level in order to match the modal appropriateness rating (i.e. coordination game). As a second step, physicians could receive guidelines support for the specific medical case just checking the option yes. Then, 28 of them were allowed to change their previous responses to the vignette. Results show that only 8% of the physicians did not want to know national guidelines content. However, only 23% of the subjects (with a choice frequency of 35%) belonging to the treatment group decided to change their appropriateness judgment after realizing they were in contrast with guidelines in either one or two vignettes. Finally, the more physicians consult scientific sources (i.e. signal of open-mindedness), the more willing to accept suggestions, coming from guidelines reading, and to change their decisions when they are wrong.

Although many physicians comply with guidelines from the very beginning, a fair number of them not only do not recognize them, maybe because of their lack of specialization in the concerned medical field, but also, they are reluctant to change their decisions once they realize they were wrong. This result supports the assumption that guidelines have hardly changed physicians' behaviour (Hayward, 1997), either due to their overconfidence or to guidelines inflexibility and inability to include all the multiple factors which come into play in medical decisions.

However, since guidelines dissemination is shown to increase the level of coordination between physicians, hospitals should consider effective programmes to spread their knowledge. To do this, hospitals should consider physicians' willingness to update, reflected in the high percentage of them eager to know guidelines (92%), and encouraged in this experiment by guidelines ease of accessibility. Introducing a newsletter program and providing an alternative learning option to the standard education courses could be a solution (Strasser, 1978). More

than other instruments, newsletter would have the features of accessibility and searchability required by physicians. Additionally, since physicians' flexibility and open mindedness increases in their frequency of updating which positively correlates with physicians involvement in research, the last should be incentivized. Since 46% of the physicians joining the experiment reported to conduct scientific research, introducing possible incentive mechanisms could allow to increase such percentage. Finally, any measure should consider the subjects' age and training experience, considering that approach to guidelines and to research vary on these subjects' characteristics. In fact, older physicians are more willing to appreciate guidelines suggestions and surprisingly are more involved in scientific research (*p-value*<0.01).

Given the interesting policy implications, it could be useful to resort to the same experimental design with some additional features, to further test physicians' attitude towards guidelines. An example is to allow physicians to choose between different options to get guidelines. In this experiment, physicians just checked the option 'yes' to receive a printed copy of guidelines which were both concise and exhaustive. An alternative is to give enough time to physicians to search for the needed information to solve the vignette, simulating the opportunity cost of seeking guidelines. Furthermore, physicians could be given the possibility of sharing opinions with one or more colleagues, to replicate another preferred information source (e.g. consultation with colleagues, Stinson and Mueller, 1980), before deciding to change their first choices. By adopting a between-subject design, all the above-mentioned options could be employed. Finally, opinions concerned what they read could be used to justify the reason why they do not change their answers despite being in contrast with guidelines.

References

Allery, L.A., Own, P.A., Robling, M.R., (1997), Why general practitioners and consultants change their clinical practice: a critical incident study, BMJ, 314, 870–9

Alper, B. S., Hand, J. A., Elliott, S. G., Kinkade, S., Hauan, M. J., Onion, D. K., & Sklar, B. M., (2004), How much effort is needed to keep up with the literature relevant for primary care?, Journal of the Medical Library Association : JMLA, 92(4), 429–437

Arkes, H. R., Dawes, R. M., & Christensen, C., (1986), Factors influencing the use of a decision rule in a probabilistic task, *Organizational behavior and human decision processes*, 37(1), 93-110

Armed Forces Institute of Pathology (U.S.). Department of Legal Medicine, (2008), Legal Medicine Open File: Analysis of Current Medicolegal Issues, The Department

Barr, A., Lane, T., & Nosenzo D., 2018, On the social appropriateness of discrimination, *Journal of Public Economics*, Volume 164, 153-164

Bauchner, H., Simpson L., Chessare, J., (2001), Changing physician behaviour, Archives of Disease in Childhood, 84 (6), 459-462

Baumann, A. O., Deber, R. B., & Thompson, G. G., (1991), Overconfidence among physicians and nurses: the 'micro-certainty, macro-uncertainty' phenomenon, *Social science & medicine*, 32(2), 167-174

Benavidez, G., & Frakt, A.B., (2019), Fixing Clinical Practice Guidelines, Health Affairs Blog

Berner, E. S., & Graber, M. L., (2008), Overconfidence as a cause of diagnostic error in medicine, *The American journal of medicine*, 121(5), S2-S23

Bicchieri, C. (2006), The grammar of society: The nature and dynamics of social norms, Cambridge University Press

Beilby, J. J., & Silagy, C. A., (1997), Trials of providing costing information to general practitioners: a systematic review, *The Medical journal of Australia*, 167(2), 89-92

Beaulieu, M. D., Hudon, É., Roberge, D., Pineault, R., Forte, D., & Legare, J., (1999), Practice guidelines for clinical prevention: Do patients, physicians and experts share common ground?, Cmaj, 161(5), 519-523

Blundell, R., & Meghir, C., (1987), Bivariate alternatives to the Tobit model, Journal of Econometrics, 34(1-2), 179-200

Braddock III, C. H., Edwards, K. A., Hasenberg, N. M., Laidley, T. L., & Levinson, W., (1999), Informed decision making in outpatient practice: time to get back to basics, *Jama*, 282(24), 2313-2320

Burks, S., & Krupka E., (2012), A Multi-Method Approach to Identifying Norms and Normative Expectations within a Corporate Hierarchy: Evidence from the Financial Services Industry, *MANAGEMENT SCIENCE*, Volume 58, No. 1, 203–217

Bursztyn, L., González, A. L., & Yanagizawa-Drott, D., (2018), Misperceived social norms: Female labor force participation in Saudi Arabia (No. w24736), *National Bureau of Economic Research*

Cabana, M. D., Rand, C. S., Powe, N. R., Wu, A. W., Wilson, M. H., Abboud, P. A. C., & Rubin, H. R., (1999), Why don't physicians follow clinical practice guidelines?: A framework for improvement, *Jama*, 282(15), 1458-1465

Camera civile di Firenze, (2017), Le Corti Fiorentine, Rivista di diritto e procedura civile, n.3/17

Carrier, E., Dowling, M. K., & Pham, H. H., (2012), Care coordination agreements: barriers, facilitators, and lessons learned, *The American journal of managed care*, 18(11), e398-404

Cassam, Q., (2017), Diagnostic error, overconfidence and self-knowledge, Palgrave Communications, 3(1), 1-8

Conigliaro, R., & Rossi, A., (2006), Implementation of sedation guidelines in clinical practice in Italy: results of a prospective longitudinal multicenter study, Endoscopy, 38(11), 1137-1143

Connelly, D. P., Rich, E. C., Curley, S. P., & Kelly, J. T., (1990), Knowledge resource preferences of family physicians, *J Fam Pract*, *30*(3), 353-9

Cragg, J. G., (1971), Some statistical models for limited dependent variables with application to the demand for durable goods, Econometrica: *Journal of the Econometric Society*, 829-844

Das, J., & Hammer, J., (2005), Which doctor? Combining vignettes and item response to measure clinical competence, *Journal of Development Economics*, 78(2), 348-383

Das, J., Hammer, J., & Leonard, K., (2008), The quality of medical advice in low-income countries, *Journal of Economic perspectives*, 22(2), 93-114

Das, J., Kwan, A., Daniels, B., Satyanarayana, S., Subbaraman, R., Bergkvist, S., ... & Pai, M., (2015), Use of standardised patients to assess quality of tuberculosis care: a pilot, cross-sectional study, The Lancet Infectious Diseases, 15(11), 1305-1313

Dawes, M., & Sampson, U., (2003), Knowledge management in clinical practice: a systematic review of information seeking behavior in physicians, *International journal of medical informatics*, 71(1), 9-15

Dinner, I., Johnson, E. J., Goldstein, D. G., & Liu, K., (2011), Partitioning default effects: why people choose not to choose, Journal of Experimental Psychology: Applied, 17(4), 332

Drummond, M. F., (1994), Guidelines for pharmacoeconomic studies, *PharmacoEconomics*, 6(6), 493-497

Duflo, E., & Saez, E., (2003), The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment, *The Quarterly journal of economics*, 118(3), 815-842

Feder, G., Eccles, M., Grol, R., Griffiths, C., & Grimshaw, J., (1999), Using clinical guidelines, *Bmj*, 318(7185), 728-730

Field, M. J., & Lohr, K. N., (1990), Clinical practice guidelines, Directions for a new program. Washington: National Academy Pr

Friedman, C., Gatti, G., Elstein, A., Franz, T., Murphy, G., & Wolf, F., (2001), Are clinicians correct when they believe they are correct? Implications for medical decision support, Studies in health technology and informatics, (1), 454-458

Freemantle, N., Harvey, E. L., Wolf, F., Grimshaw, J. M., Grilli, R., & Bero, L. A., (2000), Printed educational materials: effects on professional practice and health care outcomes, Cochrane Database Syst Rev, 2, CD000172

Gabbay, J., & le May, A., (2004), Evidence based guidelines or collectively constructed "mindlines?" Ethnographic study of knowledge management in primary care, BMJ, 329(7473):1013

Gabbay, J., & le May, A., (2016), Mindlines: making sense of evidence in practice, Royal College of General Practitioners, vol. 66 no. 649, 402-403

Gachter, S., Nosenzo, D., & Sefton, M., (2013), Peer effects in prosocial behaviour: Social norms or social preferences?, *Journal of the European Economic Association*, 2013, vol. 11, issue 3, 548-57

Galizzi, M. M., Miraldo, M., & Stavropoulou, C., (2016), In sickness but not in wealth: field evidence on patients' risk preferences in financial and health domains. Medical Decision Making, 36(4), 503-517

Glassman, P. A., Luck, J., O'Gara, E. M., & Peabody, J. W., (2000), Using standardized patients to measure quality: evidence from the literature and a prospective study, The Joint Commission journal on quality improvement, 26(11), 644-653

Gorter, K., de Poel, S., de Melker, R., & Kuyvenhoven, M., (2001), Variation in diagnosis and management of common foot problems by GPs, Family practice, 18(6), 569-573

Grimshaw, J., Shirran, L., Thomas, R., Mowatt, G., Fraser, C., Bero, L., . . . O'Brien, M., (2001), Changing Provider Behavior: An Overview of Systematic Reviews of Interventions, *Medical Care*, 39(8), II2-II45

Hayward, R.S.A., (1997), Clinical practice guidelines on trial, *CMAJ*, 156, 1725-1727

Haug, J. D., (1997), Physicians' preferences for information sources: a meta-analytic study, Bulletin of the Medical Library Association, 85(3), 223

Havighurst, C.C., (1991), Practice guidelines as legal standards governing physician liability, Law and Contemporary Problems, 54(2), 87-117

Hermalin, B.E., (1999), 'Leadership and Corporate Culture', in: Gibbons, R., Roberts, J., (2013), *The Handbook of Organizational Economics*, Princeton: Princeton University Press

Hyams, A. L., Brandenburg, J. A., Lipsitz, S. R., Shapiro, D. W., & Brennan, T. A., (1995), Practice guidelines and malpractice litigation: a two-way street, *Annals of Internal Medicine*, 122(6), 450-455

Hjelmgren, J., Berggren, F., & Andersson, F., (2001), Health economic guidelines—similarities, differences and some implications, Value in Health, 4(3), 225-250

Institute of Medicine (US), (1990), Clinical Practice Guidelines: Directions

Jachimowicz, J. M., Duncan, S., Weber, E. U., & Johnson, E. J., (2019), When and why defaults influence decisions: A meta-analysis of default effects, Behavioural Public Policy, 3(2), 159-186

Katzka, D. A., (2017), How to balance clinical work and research in the current era of academic medicine, Gastroenterology, 153(5), 1177-1180

Kahan, J.P., Park, R.E., Leape, L.L., Bernstein, S.J., Hilborne, L.H., Parker, L., et al., (1996), Variations by specialty in physician ratings of the appropriateness and necessity of indications for procedures, *Med Care*, 34, 512–523

King, A. S., (1974), Expectation effects in organizational change. Administrative science quarterly, 221-230

Kovacs, R. J., Lagarde, M., & Cairns, J., (2020), Overconfident health workers provide lower quality healthcare, *Journal of Economic Psychology*, 76, 102213

Leonard, K. L., & Masatu, M. C., (2005), The use of direct clinician observation and vignettes for health services quality evaluation in developing countries, *Social science & medicine*, 61(9), 1944-1951

Lomas, J., Anderson, G. M., Domnick-Pierre, K., Vayda, E., Enkin, M. W., & Hannah, W. J., (1989), Do practice guidelines guide practice?, *New England Journal of Medicine*, 321(19), 1306-1311

Lomas, J., (1991), Words without action? The production, dissemination, and impact of consensus recommendations, *Annual review of public health*, 12(1), 41-65

McDonald, C. J., & Overhage, J. M., (1994), Guidelines you can follow and can trust: an ideal and an example. Jama, 271(11), 872-873

McLeod, S. A., (2008), Social roles, retrieved from www.simplypsychology.org/social-roles.html

Meoli, B., Sica, S., & Stanzione, P., (2018), Commentario alla legge 8 Marzo 2017 n.24, "Disposizioni in materia di sicurezza delle cure e della persona assistita, nonché in materia di responsabilità professionale degli esercenti le professioni sanitarie", Edizioni Scientifiche Italiane, 52-59

Meyer, A. N., Payne, V. L., Meeks, D. W., Rao, R., & Singh, H., (2013), Physicians' diagnostic accuracy, confidence, and resource requests: a vignette study, *JAMA internal medicine*, 173(21), 1952-1958

Mohanan, M., Vera-Hernandez, M., Das, V., et al., (2015), The know-do gap in quality of health care for childhood diarrhea and pneumonia in rural India, *JAMA Pediatr*, 169, 349–57

OECD (2015), OECD Reviews of Health Care Quality: Italy 2014: Raising Standards, OECD Reviews of Health Care Quality, OECD Publishing, Paris, available at https://doi.org/10.1787/9789264225428-en

O'Malley, A.S., & Reschovsky, J.D., (2011), Referral and consultation communication between primary care and specialist physicians: finding common ground, *Arch Intern Med*, 171(1), 56-65

Opp, K.D., (2001), How do norms emerge? An outline of a theory, Mind Society 3(2):101–128

Park, C. W., Jun, S. Y., & MacInnis, D. J., (2000), Choosing what I want versus rejecting what I do not want: An application of decision framing to product option choice decisions, Journal of Marketing Research, 37,

187–202

Parker, C. N., Van Netten, J. J., Parker, T. J., Jia, L., Corcoran, H., Garrett, M., ... & Wraight, P, (2019), Differences between national and international guidelines for the management of diabetic foot disease, Diabetes/metabolism research and reviews, 35(2), e3101

Roodman, D., Nielsen, M. Ø., MacKinnon, J. G., & Webb, M. D., (2019), Fast and wild: Bootstrap inference in Stata using boottest. The Stata Journal, 19(1), 4-60

Royal College of General Practitioners, (1995). The development and implementation of clinical guidelines: report of the Clinical Guidelines Working Group. Report from Practice, 26

Samuels, M. A., Ropper, A. H., Poses, R. M., & Diaz, J. A, (2005), Clini- cal Experience and Quality of Health Care/IN RESPONSE/IN RESPONSE, Annals of internal medicine, 143(1), 84

Sandvik, H., (1995), Criterion validity of responses to patient vignettes: an analysis based on management of female urinary incontinence, Family Medicine, 27(6):388-392

Shekelle, P. G., Woolf, S. H., Eccles, M., & Grimshaw, J., (1999), Developing guidelines, *Bmj*, 318(7183), 593-596

Shekelle, P. G., Kravitz, R. L., Beart, J., Marger, M., Wang, M., & Lee, M., (2000), Are nonspecific practice guidelines potentially harmful? A randomized comparison of the effect of nonspecific versus specific guidelines on physician decision making, *Health services research*, 34(7), 1429

Sorenson, S. B., & Taylor, C. A., (2005), Female aggression toward male intimate partners: An examination of social norms in a community-based sample, *Psychology of Women Quarterly*, 29(1), 78-96

Stinson, E. R., & Mueller, D. A., (1980), Survey of health professionals' information habits and needs: conducted through personal interviews, *Jama*, *243*(2), 140-143

Strasser, T. C., (1978), The information needs of practicing physicians in northeastern, New York State, *Bulletin of the Medical Library Association*, *66*(2), 200

Timpka, T., Ekström, M., & Bjurulf, P., (1989), Information needs and information seeking behaviour in primary health care, *Scandinavian journal of primary health care*, 7(2), 105-109

Tversky, A., & Kahneman, D., (1974), Judgment under uncertainty: Heuristics and biases. Science, 185, 1124–1131

Veloski, J., Tai, S., Evans, A. S., & Nash, D. B., (2005), Clinical vignette-based surveys: a tool for assessing physician practice variation, *American Journal of Medical Quality*, 20(3), 151-157

Wensing, M., van der Weijden, T. R. U. D. Y., & Grol, R., (1998), Implementing guidelines and innovations in general practice: which interventions are effective?, Br *J Gen Pract*, 48(427), 991-997

Woolf, S.H., (1994), An organized analytic framework for practice guideline development: using the analytic logic as a guide for reviewing evidence, developing recommendations, and explaining the rationale, In: McCormick KA, Moore SR, Siegel RA, editors. Methodology perspectives, Washington, DC: US Department of Health and Human Services, *Agency for Health Care Policy and Research*, 105–113

Woolf, S., Schünemann, H. J., Eccles, M. P., Grimshaw, J. M., & Shekelle, P., (2012), Developing clinical practice guidelines: types of evidence and outcomes; values and economics, synthesis, grading, and presentation and deriving recommendations, *Implementation Science*, 7(1), 61

World Health Organization, (2018), Continuity and coordination of care: a practice brief to support implementation of the WHO Framework on integrated people-centred health services, In Continuity and coordination of care: a practice brief to support implementation of the WHO Framework on integrated people-centred health services

Yager, J., Linn, L. S., Leake, B., Gastaldo, G., & Palkowski, C., (1986), Initial clinical judgments by internists, family physicians, and psychiatrists in response to patient vignettes: I.

Assessment of problems and diagnostic possibilities, *General hospital psychiatry*, 8(3), 145-151

Appendix C

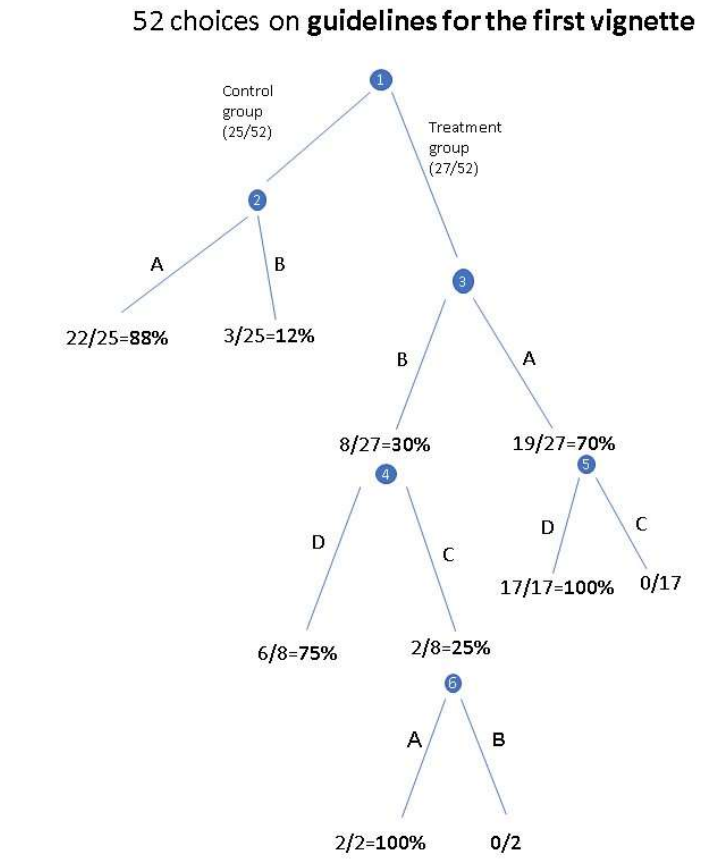


Figure 2.1.1: Structure of the experiment for the first vignette

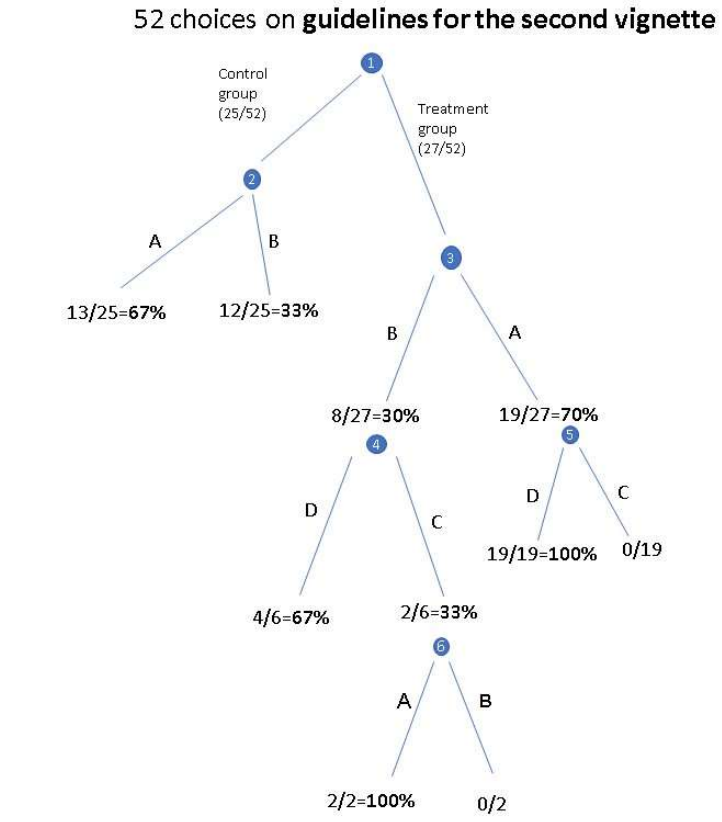


Figure 2.1.2: Structure of the experiment for the second vignette

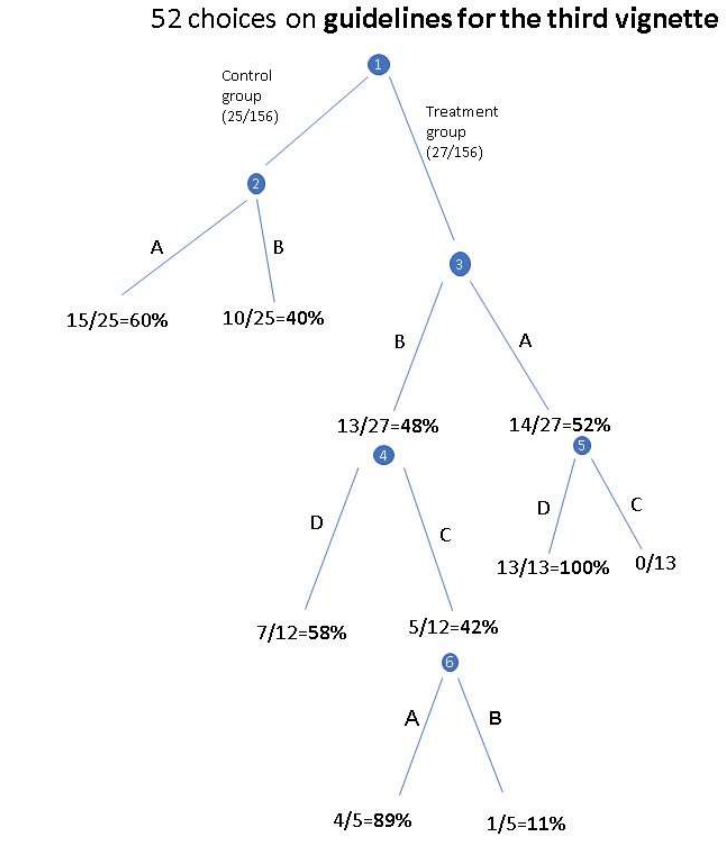


Figure 2.1.3: Structure of the experiment for the third vignette

CHAPTER 3

An experimental analysis of patient dumping under different payment systems

Abstract

Physicians behave differently depending on the payment systems, giving rise to several problems such as patient dumping, when patients are refused because of economic or liability reasons. While different studies have focused on the impact of remuneration scheme on the quantity of medical services provided, little research has been dedicated to patient dumping. This paper aims to test whether and to which extent the adoption of fee-for-service (FFS) or salary system can induce physicians to practice patient dumping. For, we will use an artefactual field experiment where physicians facing the possibility of being sued for medical malpractice, decide whether or not to provide medical services for patients with different state of health. Also, we check whether the introduction of the risk of being sued for a physician for having practiced dumping can have effect on his behaviour. Results show that dumping is more often observed under Salary than under FFS. However, physicians seem to be insensitive to the introduction of dumping liability under the same incentive mechanism, though it seems to trigger a higher amount of services provided. Policy strategies concerning the incentive scheme may vary on the hospital purposes.

Introduction

It is commonly acknowledged that financial incentives from payment systems have a crucial impact on the behaviour of healthcare providers (Clemens and Gottlieb, 2014). Not only physicians choose the quality and the quantity of medical services to provide but, even more fundamentally, they could decide whether to treat patients depending on the remuneration scheme (Treiger, 1986; Barnett and Mayer, 1992). To support governments in their attempt to optimize healthcare delivery, a considerable theoretical literature analyses pros and cons of the alternative remuneration schemes (see e.g. Simoens and Giuffrida, 2004). In his review, Robinson (2001) argues that, albeit most commonly used, fee for service (FFS) and capitation (CAP) present many drawbacks. First of all, a remuneration which gives physicians a fixed fee for each service provided (FFS), incentivizes the provision of inappropriate services, inflating healthcare costs without any effect on health itself (Donaldson and Gerard, 1989). Even worse, a lump-sum payment for a treated patient (CAP)[45], which does not take account of the quantity and the costs of services, could lead some practitioners to curtail consultation time, harming patients' health (Maynard *et al.* 1986). On the one hand, under such scheme physicians have an incentive to cream-skim patients, selecting only consumers who are in a good health status, given their reduced treatment costs (Matsaganis and Glennerster, 1994). On the other hand, this system rewards the dumping of the chronically ill patients owing to treatments expensiveness compared to physician's reimbursement (Preston et al., 1997). Similarly, an alternate prospective payment such as diagnosis related group (DRGs) promotes low risk patients' treatment at the expense of high-risk patients (Takahara, 2016). Under DRGs, indeed, hospitals receive a lump-sum payment for each patient treated, given a certain diagnosis (Siciliani, 2006). Consequently, hospitals which decide to treat expensive patients incur losses whenever costs are greater than the corresponding DRG payments. In a nutshell, prospective payment systems induce hospitals to aim for efficiency and profit maximization and most importantly to refuse treatment on patients, a phenomenon known as *patient dumping* (Eze and Wolfe, 1992).

[45] Notice that salary, largely used by hospitals, produces outcomes largely in line with CAP (Blomqvist and Busby, 2012).

Despite initially limited to USA, the concept of patient dumping progressively spread on multiple continents, including Europe, given the introduction of case payment mechanisms (Busse et al., 2006). DRG systems were originally thought to provide a financial incentive to reduce health care expenditures. To this end, the patient was classified into one of the 468 diagnostic indices and the hospital was reimbursed according to a formula derived from the average costs of treating patients with that diagnosis (Preston et al., 1997). As a result, hospitals were induced to save costs and to seek increased efficiency (Horwitz, 1988). In particular, if the cost for treating the patient was lower than the fixed sum, the hospital could realise profits, otherwise the hospital absorbed the additional cost (Dranove, 1987). Therefore, hospitals were likely to dump charity patients who could not pay for the needed care (Mullner et al., 1986). Based on the above, one of the possible reasons for the observed dumping is that hospitals pursuing profit maximation strategies try to reduce the number of their high-cost admissions (Eze and Wolfe, 1992). Under such assumption, many theorists explored alternative payment structures such us a risk adjustment prospective system to prevent cream skimming and, mainly, patient dumping (see e.g. Sappington *et al.,* 1999; Barros, 2003). Similarly, hospitals could be insured for some fraction of the additional costs incurred to treat expensive patients (Chalkley and Khalil, 2005). Nevertheless, the reason behind dumping could be not so straightforward. Specifically, patient dumping and medical malpractice liability are strictly related. Since physicians constantly face the risk of being sued for malpractice, they could turn away a priori patients to protect themselves in the event that they are unable to provide patients with the best care in fearing to be sued (Zibulewsky, 2001). However, regardless the underlying reason, it is undeniable that patient dumping carries significant social costs. First, patients can face fatal treatment delays (Busse et al., 2006). Further to this, dumping makes patients converge to particular hospitals, especially public hospitals, triggering further treatment delays (Newhouse, 1983). Contrarily, another strand of literature supports dumping policy said to be welfare-improving to some extent. As outlined by Busse et al. (2006), information asymmetry (i.e. between patient and hospital) concerning patient type is mitigated, given the endogenous change in the distribution of patients across the hospitals following dumping. For this reason, reduction in information rent allows hospitals to save costs for treating high-severity patients.

Since there is little consensus on the causes and the effects of patient dumping, the phenomenon requires further investigation. The assessment of the hospital's perspective is beyond the scope of this paper, however, we can reasonably assume that physicians' choices, being all employed

at public hospitals, reflect employers directives. Therefore, our main focus is to investigate the occurrence of patient dumping under two different payment schemes, FFS and salary[46], building on a framed field experiment. Although experimental evidence on the impact of payment systems on the quantity of medical services provided is fast growing (see e.g. Green, 2014; Brosig et al., 2016; Lagarde and Blaauwn, 2017), the same cannot be said about patient dumping. This paper can be seen as a first attempt to fill this gap in the literature, assessing how the introduction of medical liability and the risk of being sued for dumping may affect physicians' practice of patient dumping under different payment systems. In this experiment, each participant playing the role of a physician decides whether to take charge of the patient and, then, to what extent to treat him, given a certain payment structure. Under both the payment schemes, participants always face the possibility of being sued for malpractice and eventually an additional risk of being sued for practicing dumping. While the design closely mirrors Hennig-Schmidt et al. (2011) and Finocchiaro Castro et al. (2019), the novelty lies in the interplay of three issues heavily affecting medical care decisions: patient dumping, medical liability and the probability of being sued for dumping. Results show that dumping is more often observed under Salary than under Ffs. However, physicians seem to be insensitive to the introduction of dumping liability under the same incentive mechanism, though it seems to trigger a higher amount of services provided. Older physicians are more likely to take charge of the patient than their younger colleagues. Policy strategies concerning the incentive scheme may vary on the hospital purposes.

Patient dumping

Before the Consolidated Omnibus Reconciliation Act legislation (COBRA) was enacted in USA by Congress in 1985, hospitals could refuse treatment to any patient. Cost containment efforts by both government and private sector and the increase in the number of uninsured were the main reasons for patient dumping (Treiger, 1986). Most of the uninsured delayed seeking healthcare until their illness had become serious, due to their economic status. For this reason,

[46] Choosing salary as payment method is plausible being the most common compensation scheme employed by hospitals. Moreover, similarly to CAP and DRG, it does not provide physician with any financial incentive to work harder than the minimum (Hewak and Kovacs-Litman, 2015).

as the severity of their medical conditions increased, the cost of providing treatment for these patients increased in the same way, giving hospitals an economic incentive to dump them (Rice *et al.,* 1988). This was partly due to the prospective payment system introduced in the USA by the Medicare Program[47], the afore-mentioned DRG (Siciliani, 2006).

As patient dumping was becoming more and more frequent, in 1985 the U.S. Congress decided to enact COBRA to prohibit it. According to this law, hospitals, who received Medicare funds, were required to screen and stabilise all patients in need of emergency care and women in active labour before transferring them if necessary, regardless of their economic status (Saks, 2004). A hospital could transfer a patient in an emergency condition if and only if the patient requested a transfer or the physician certified that another facility could provide better treatment[48]. All transfer decisions needed the appraisal of a qualified staff member and the receiving hospital had to be informed of the pending transfer and had to consent to it. Moreover, appropriate transportation and medical services during the transfer had to be guaranteed to the patient.

Hospitals and physicians who knowingly violated COBRA statute were subject to monetary fines and risked losing participation in Medicare. Furthermore, patients and the receiving hospitals could sue a referring hospital which had transgressed the provisions of COBRA, obtaining damages for personal harm or financial loss (Kusserow, 1988). In addition to COBRA, many states had enacted statutes to guarantee access to emergency care to uninsured and indigent (see e.g. Dallek and Waxman, 1985; Treiger, 1986; Volp and Siegel, 1993)[49].

However, it is undisputed that COBRA legislation and the following statutes fail to fully prevent dumping (Kellermann and Hackman, 1990). To confirm this, Struik (2015) makes

[47] Medicare cares for those who are 65 years of age and older by guaranteeing access to hospital care and physician services, whereas Medicaid addresses to the problems of the poor and near poor (Rice *et al.*, 1988).

[48] The physician had to demonstrate and certify that the benefits of the transfer outweighed the risks because the transferring hospital could provide highly specialized technical teams and needed equipment (COBRA, 1985).

[49] Texas and Alameda regulations required that each hospital had its written transfer policy establishing that the patient had to be transported with qualified personnel and medical equipment (Dallek and Waxman, 1985). Wisconsin's law prohibited delay in treatment due to financial reasons (Treiger, 1986). New Jersey established the Uncompensated Care Trust Fund, an 'indigent-pool' which reimbursed hospitals for all their bad debt and charity care to eliminate financial incentive to practice patient dumping (Volpp and Siegel, 1993).

several examples of patient dumping episodes in Los Angeles, Las Vegas, Chicago, Denver and New York which prove that this practice persists. The United States Commission on Civil Rights (September 2014) stated that the Centres for Medicare and Medicaid Services (CMS) received approximately an average of 500 EMTALA[50] complaints annually in the period between 2006 and 2012. More recent episodes have been reported by the Office of Inspector General (U.S. Department of Health and Human Service, 2019[51]). To these is added the story of a 17 year old boy from California suffering from COVID-19 who was denied medical treatment since he lacked insurance and died during the transportation from the dumping hospital to a public hospital (Parris, 2020[52]).

There may be several reasons explaining why this problem still exists. First of all, since hospitals act like businessmen, they try to cut their losses. Treating the poor who cannot afford any medical insurance is the main reason for their losses (New York Times, 1989). These losses could result in their closing, thus jeopardising the availability of medical services for everyone. Second, it may be an issue of enforcing[53], or rewriting the law to avoid ambiguity (e.g. on what we mean for 'emergency') (New York Times, 1991). Notably, monetary fines do not seem enough to prevent physicians and hospitals from practicing patient dumping.

As far as our context is concerned, since the nineties most European countries have begun to reimburse their hospitals according to a DRG system, which is supposed to cover all costs, based on U.S.A. Medicare program (Busse et al., 2011). In Italy, in particular, health care is publicly financed[54] and DRG tariffs are set equal at a national level, though regions have the opportunity to modulate charges according to various criteria such as organisational features (e.g. presence of emergency room), or activities performed (e.g. research activities) (Cavalieri et al., 2013). Additionally, Italian healthcare system (*Sistema Sanitario Nazionale*) governance

[50] Emergency Medical treatment and active Labor Act (EMTALA) is an act belonging to COBRA legislation.

[51] https://oig.hhs.gov/fraud/enforcement/cmp/cmp-ae.asp

[52] News taken from the speech of the Lancaster major for COVID-19 Daily Update, available at https://www.youtube.com/watch?time_continue=165&v=NkhnAF7E5xw&feature=emb_title

[53] E.g. the Statute should define appropriate circumstances for the revocation of Medicare contract, thus a hospital's contract with Medicare should be suspended for repeated violations.

[54] According to the Italian Constitution, the central government distributes tax revenues for publicly financing health care. Regions are in charge of the organization and the delivery of health services through local health units (Donatini,2015).

was recently reorganized, transferring partial financial responsibility to the Regions in order to contain increasing health expenditure (France, 2006). Thus, since health protection is a competence shared between the State and the Regions, healthcare services in terms of both quantity and quality are not uniformly distributed throughout the national territory (Gabriele, 2019). However, despite such inequalities, the Italian National Health service has ensured access to free or low-cost healthcare, such as treatment at public hospitals, to all citizens and residents since its establishment (Cantù et al., 2011). Consequently, the emergency department or emergency room[55] cannot refuse emergency patients who need to be diagnosed and stabilized (Cass. sentenza n. 40753/16 del 29.09.16). Then, if and only if there is another medical facility equipped with specialists who can provide a better treatment according to the medical diagnosis, the patient can be transferred (Agenzia di Sanità pubblica, 2002). Nevertheless, the initial medical facility should always find a place for the patient in its structure or in other ones, and in the latter case it should act as a bridge between the patient and the new structure (D'Athis et al., 1992).

Although patient dumping is very mitigated in Italy and should not need for a particular regulation, there has been a recent judgement by the court of Cassation (i.e. the court of last resort in Italy[56]) in this regard. According to the ruling n. 45844 (5 November 2014), the doctor on call who refuses to treat an emergency patient is criminally responsible for dereliction of duty[57] (art. 328 Code of Criminal Procedure). Such conduct constitutes a crime if the admission is urgent and cannot be delayed without harming the patient's health. The intervention by the court of Cassation followed an episode of patient dumping occurred in 2011, where a E.R. doctor[58] refused to admit a patient in serious conditions as a result of an accident (Asprone, 2015). Generally, similar events are not as frequent in Italy as in U.S.A., because of the public

[55] By Emergency department we refer to **t**he department of a hospital responsible for the provision of medical and surgical care to patients arriving at the hospital in need of immediate care. The emergency department is also called the emergency room or ER (www.medicinenet.com).

[56] The Supreme Court of Cassation is the Italian highest court, whose decisions cannot be subject to further review by any other court (Traina, 2009).

[57] It happens in response to a request or an order but also when there is a substantial emergency for which an action is needed. In this way, public official's inertia constitutes a guilty refusal of the act.

[58] See footnote 58.

nature of Italian healthcare provision. However, the recent need to reduce healthcare spending, among others, reducing beds, could trigger treatments rationing (Gabriele, 2019) which in the worst scenario results in dumping. In such a scenario, it becomes necessary to investigate the reasons behind patient dumping which does not always have a financial nature and may be very insightful in a policy perspective. An emblematic episode has been recently recorded in Abruzi where a 70-year old man died in front of the hospital waiting to be admitted, after a previous rejection by another medical facility[59].

Literature Review

This paper contributes to different streams of literature. First, it deals with the role of medical liability in affecting medical decisions, adding to previous works by Nashed et al. (2012) and Kessler and McClellan (2002). Typically, insurance covers physicians from the financial risks of malpractice litigation, however there are several non-insurable costs, such as psychological and reputational costs, (Currie and Macleod, 2008). This means, albeit being insured, physicians are still concerned about legal liability[60]. Nonetheless, not all physicians suffer from the same liability risks since medical specialists are more exposed to litigations compared to their peers (Jena et al., 2011). As a result, surgeons prefer performing elective procedures which do not really affect the quality of care provided to patients, in the attempt to evade litigation[61] (Agarwal, 2018). Alternatively, in response to liability, physicians can choose to treat only less risky patients (i.e. cream skimming) to decrease the probability of negative outcomes. Both practices known as defensive medicine have been further reviewed by Danzon, (2000), Kessler (2011) and Bertoli and Grembi (2018).

[59] https://www.fanpage.it/attualita/avezzano-enzo-muore-in-auto-in-attesa-del-ricovero-tra-le-urla-della-moglie-fateci-e-entrare/

[60] In Italy, we record the highest number in Europe of physicians sued for medical malpractice. For this reason, the national health system is obliged to pay high insurance premiums when it succeeds in finding an insurance company ready to bear risk of monetary claims due to medical malpractice. (Fusciani, 2004). A brief overview of the Italian regulation of medical malpractice can be found in the Appendix.

[61] Notice that this practice (i.e. excessively using treatments and diagnostic tools) does nothing but increasing healthcare expenditures (Traina, 2009).

The second contribution of the paper relates to the growing experimental literature devoted to finding how different payment structures affect medical service provision. The initial studies carried out in this field can be attributed to Henning Schmidt *et al.* (2011) who test how medical students respond to fee-for-service and capitation. They observe that patients are over-treated under FFS and under-treated under CAP and that healthier patients are better off under CAP, while sickest patients benefit otherwise. These results echo those of the artefactual field experiment by Brosig *et al.* (2015). However, their analysis goes one step forward showing that students are more influenced in their decisions by financial incentives compared to physicians who are more patient-regarding (similarly to Schmidt et al., 2014[62]). Building on the same experimental design, Brosig *et al.* (2016) analyse the possibility of introducing mixed payment systems as an alternative to pure FFS and CAP. They find that such schemes mitigate both over-provision and under-provision, thus benefitting patients' health. Moreover, in line with the findings of the aforementioned experiments, participants show a certain degree of altruism towards patients' health. In a similar setting, but allowing for uncertainty of the patient health outcome, Martinsson and Persson (2019) show that physicians' willingness to altruism varies across patients with different medical needs and that between all the possible principles for provision of medical care (i.e., severity of illness; capacity to benefit; ex post health equality), severity of illness drives such variation. Drawing from Brosig et al. (2015), in their laboratory experiment Finocchiaro Castro et al. (2019) introduce the risk for physicians of being sued for malpractice under FFS and CAP, and test how it impacts on medical service provision. Results show that physicians provide a higher number of medical services when malpractice liability pressure comes into play. However, while introducing the probability of being sued for malpractice reduces physicians' tendency to under-treat patients under CAP, it exacerbates overprovision under FFS. Other studies in the field employing similar experiments assess physicians' behaviour under alternative payment systems, such as performance pay, (Brosig et al., 2019), report cards (Green, 2014), and salary (Lagarde and Blaauw, 2017), showing that quality of care improves under prospective payment systems. Finally, Godager et al. (2016) depart from the abovementioned evidence and test whether disclosing physicians' performance information to their peers can have an impact on their medical decisions under FFS. When

[62] Notice that they compared medical students and nonmedical students' behaviour.

performance information is displayed, subjects choose more frequently to maximise patient benefit only, or the sum of their profit and patient's benefit.

Although building on the above-mentioned works, the present paper does not assume that physicians always choose to treat patients, thus allowing for the possibility of patient dumping, which previous studies fail to consider.

Theoretical Model

To derive the theoretical equilibrium as level of medical care provided by physicians under different payment systems, we employ a simple model theorised by Ellis and Mcguire (1986) and then readjusted by Finocchiaro Castro M. et al. (2019) to include medical malpractice liability. However, since this model is aimed to test whether the risk of being sued for both medical malpractice and dumping could affect physicians' behavior, we are taking an important step forward for our experimental purposes. Starting from the previous models, we add another variable which could capture the potential punishment for practicing dumping.

First of all, we imagine considering a physician who cares for both his profit and the benefits to patients. Physician's profit can be represented as follows

$$\Pi(q) = R(q) - C(q) \tag{3.1}$$

where $R(q)$ is physician's revenue which does vary according to the payment system. Specifically, under FFS physicians receive a fee, based on a national fee schedule[63], for each

[63] The fee schedule assigns a fixed relative value to each health care service, recognising that goods and services can have different production costs. Following Schmidt *et al.* (2011), we refer to the German scale of charges and fees for physician services. In Germany, physicians working in outpatient care are enrolled in their respective regional Association of the Statutory Health Insurance Physicians which pays them according to the "Uniform Evaluation Scale". This payment system is based on the mix of services provided, the number of patients served and finally a fixed budget distribution scheme (An IGES Group company Assessment in Medicine, Reimbursement of Medical Devices in Germany, 2018).

unit of medical services provided; thus the revenue function is: $R_{FFS} = pq$. On the contrary, according to salary, subjects are paid a monthly sum, which does not vary with the services provided; thus, the revenue function is: $R_{Salary} = L$. Going back to equation 1, $C(q)$ is the total cost for providing medical care. We can state that $R'(q) \geq 0$ and $R''(q) = 0$ in coherence with the standard payment systems. In the same way the total cost is set increasing and convex, $c'(q) > 0$ and $c''(q) > 0$. As mentioned before, the physician is also interested in the patient's benefit, thus his utility function includes *B(q)* which is the patient's expected health benefit sent increasing and concave, that is $B'(q) > 0, B''(q) < 0$). Patient's benefit resulting from medical treatment depends on $B(q) + \varepsilon$ where ε refers to a random component depending on the unavoidable uncertainty surrounding the provision of medical care, which is independent from the quantity of treatment given, that is $E[\varepsilon|q] = E[\varepsilon] = 0$. The total benefit function is assumed to increase to a point, that is a certain quantity of treatment, after which point it starts to fall (Ellis and Mcguire, 1986).

Therefore, physician i's expected utility function when he does not face the risk of being sued can be represented as follows:

$$E[U(q)] = R(q) - C(q) + \alpha B(q) \tag{3.2}$$

where $0 \leq \alpha \leq 1$ is defined as the rate at which the physician is willing to give up one euro of profit for one euro of patient benefit (Ellis and Mcguire, 1986).

A rational physician *i* wants to maximize his utility function. The first order condition for the optimal quantity of medical service, q^*, can be written as:

$$R'(q^*) + \alpha B'(q^*) = C'(q^*) \tag{3.3}$$

If we consider the risk for a physician *i* to be sued, we need to add another term to the previous condition to include the expected disutility caused by a complaint, $p(q)$ (Finocchiaro Castro *et al.,* 2019). This term is assumed to decrease in the quantity of medical services provided to the patient, $p'(q) < 0$, consistently with the idea of defensive medicine. Although it may be difficult for patients to judge whether medical provision is appropriate (Kershbamer and Sutter, 2017), physicians use overtreatment to persuade that the patient's poor health is not due to malpractice but rather to the uncertainty surrounding the provision of medical care (Finocchiaro Castro et al., 2019). In fact, according to Studdern et al. (2005), physicians are willing to practice positive defensive medicine, that means providing additional unnecessary care, so as to discourage

patients from suing them or, even worse to document diligence and prudence (Kessler and McClellan, 2002) persuading the legal system that the patient has been treated according to the standard care. However, for the purposes of our study, we need to include another variable in the model so far described to account for a punishing mechanism for dumping physicians. We refer to this variable as *d(j),* which represents the possibility for a physician of being sued for dumping, where *j* is the degree of patient's illness. It is reasonable to assume that this probability increases with the severity of the patient's disease ($d'(j) > 0$) and that it is independent from the quantity of medical services provided, $E[d|q] = 0$. In fact, reports for patient dumping are likely to come from patients with serious health conditions, since mainly in these cases the refuse to treat them could cause irreparable damage to their health, leading patients themselves or their relatives to instigate a prosecution (Rice et al., 1988). Moreover, it is implausible that a physician who provides at least a minimum treatment to patients could be sued for dumping and this explains the independence between the quantity of services and the probability of being sued for dumping. For the same reason, if a physician is sued for dumping, he cannot be sued for malpractice, and vice-versa (i.e. one event precludes the other), that is $E[d|p] = 0$. In fact, malpractice can only arise if the physician takes care of the patient. On the contrary, when the physician decides to reject the patient a priori, he can only be accused of dumping.

Based on the above, the physician's expected utility which includes both medical liability and dumping becomes as follows is the following:

$$E[U(q)] = R(q) - C(q) + \alpha B(q) - p(q)H - d(j)D \tag{3.4}$$

Where H is the disutility coming from any potential complaint including all the monetary and nonmonetary costs incurred to undertake a legal defensive action (Finocchiaro Castro et al., 2019); similarly, D is the disutility coming from complaint for patient dumping[64].

Here, the first order condition to obtain the optimal quantity of medical services to be provided is defined as follows:

$$R'(q^*) + \alpha B'(q^*) - p'(q^*)H = C'(q^*) \tag{3.5}$$

[64] Notice that although according to the COBRA legislation mentioned in the previous sections, only hospitals can be directly sued for EMTALA violations, physicians responsible for such repeated violations could face disciplinary actions and could be subject to civil monetary penalties (Fabrikant et al., 2006 and Zibulewsky, 2001).

The first order condition, used to find the optimal quantity of medical services provided does not depend on d(j), since it does not vary according to q. However, physician's profits are inversely proportional to $d(j)$:

$$\frac{d\Pi}{\partial d(j)} < 0 \tag{3.6}$$

which could significantly affect his choice. As a result, the introduction of the risk of a complaint for dumping could discourage physicians from practicing it.

Design and Hypotheses

Hypotheses

Our first hypothesis deals with the impact that the two payment systems have on the practice of patient dumping. According to the Cook Count Study[65] conducted in 1986 and referred to the period 1980-1983 eighty-seven percent of the hospitals which decided to reject or transfer patients to other medical faculties cited the lack of insurance as the sole reason for such a conduct. Moreover, Harvard Medical school's study[66] indicated that when there was no medical reason to transfer the patient, some patients were transferred due to financial interests of hospitals and physicians (Treiger, 1986).

Evidence shows that patient dumping is still practiced 30 years after EMTALA and most of the time it is due to financial reasons (see Zaubi et al., 2016). Indeed, when physicians are paid through a prospective system, such as salary, they are induced to control for the costs of patient care[67] (Ellis and Mcguire, 1986). With this in mind, suppliers could decide to screen patients, to separate those who are in a good health status (more profitable patients) from those who have an undiagnosed disease or defect (less profitable patients), thus dumping high-risk patients (Matsaganis and Glennerster, 1994). Therefore, under such a payment scheme, we reasonably

[65] A Prospective Study of 467 Patients, 314 New Eng. J. Med. 552-556 (1986)

[66] It was conducted at Highland Hospital in Oakland (Treiger, 1986).

[67] Physicians act as agents of the hospital where they are employed.

expect physicians to refuse to treat patients whose needs are costly to fulfil (Sappington et al., 1999). On the contrary, a retrospective payment system, such as FFS, gives physicians an incentive to provide as many health care services as possible (Donaldson and Gerard, 1989). In other words, providers have no reason to hold down their costs, because the amount paid is determined by the actual realised costs, computed only after that services had been rendered to beneficiaries. If we assume that dumping is only due to financial reason, FFS physicians should be less willing to dump patients because they are covered for treatment expenditures. In particular, based on our model, when the risk of being sued for dumping does not come into play, physicians should be more eager to dump patients when they are paid according to salary than when they are paid according to FFS.

Hypothesis 3.1: *The level of patient dumping achieved under salary is higher than the one reached under FFS, when physicians do not face the risk of being sued for dumping.*

Our second hypothesis builds on both Green (2014) and Finocchiaro Castro et al. (2019)'s experimental results. Generally, collected data show that physicians paid by retrospective payment systems (e.g. FFS) provided more total services when compared with prospective payment systems (e.g. salary). In fact, according to Faloon (2012), under salary physicians are willing to minimise their level of effort because extra services are not properly remunerated. However, such difference could be partly mitigated by the constant presence of medical malpractice liability. As confirmed by Finocchiaro Castro et al. (2019), when the probability of being sued for medical malpractice is introduced, the quantity of medical services provided increases under both CAP and FFS, despite such payment systems should provide physicians with conflicting interests (i.e. in the absence of medical liability).

Hypothesis 3.2: *When physicians do not dump patients, the level of medical services provided is higher under the FFS than under Salary.*

Then, we check whether the introduction of the possibility for a physician of being sued for dumping affects physicians' attitude towards the patient. According to equation 3.4, physician's expected utility decreases in $d(j)$. However, when the physician does not refuse to treat the

patient, he does not face the risk of a report for dumping. As a result, we expect this variable to be a deterrent against the practice of dumping.

Hypothesis 3.3*: When physicians do face a risk of being sued for patient dumping, the level of dumping decreases, regardless of the payment systems.*

Finally, we investigate on whether the introduction of the possibility of being sued for dumping affects physician's choices on the quantity of services provided. The first-order condition does not change when we introduce $d(j)$, because it does not figurate in it, since $d(j)$ does not depend on the quantity of services provided (equation 3.5).

Hypothesis 3.4: W*hen physicians do not practice dumping, given a certain payment structure, they provide the same quantity of services, in the presence or in the absence of the risk of being sued for dumping.*

Design

In this experiment, each participant plays the role of a physician and decides whether to take charge of and eventually to what extent to treat the patient (e.g. the amount of medical services to provide), given a certain payment structure. In fact, at the beginning of the experiment, participants are randomly assigned to different payment conditions, either salary or FFS. Under both the payment schemes, participants always face the possibility of being sued for malpractice. Then, participants face an additional risk of being sued for practicing dumping, regardless of the payment structure. This potential sanction is the only deterrent we can use against the incentive to practice patient dumping[68]. In fact, the lack of effective means to prove violations jeopardises the capacity of monitoring the problem and, then, punishing

[68] A parallel may be drawn here with the traditional dumping theory, in the form of price-discrimination between national markets (Salvatore, 1989). In this model, the foreign exporter who wants to maximize his profits from foreign sales must consider a variable cost of exporting coming from the likelihood of antidumping action. This cost is increasing with additional quantities exported, due to a higher probability of getting sued for dumping behaviour. In fact, in case domestic firms file an antidumping petition, the foreign exporter has to undertake a defensive action which has a cost.

infringements through appropriate fines (Kusserow, 1988). Thus, hospitals can often find a way to subvert the legislation, because there is no entity checking discharge plans which are often falsified due to the lack of any monitoring system (Struik, 2015). To make an example of what we mean by subversion, medical facilities could cite the inability to properly treat the patient as the sole reason for refusing him, when this is not the case. Alternatively, physicians could give the patient a minimum level of treatment persuading him he is receiving appropriate care, but really dumping him (Treiger, 1986). Based on the above, the only risk for a physician or a hospital practicing dumping is being sued by the damaged patient. Subjects who have suffered dumping should refer matters for civil and criminal prosecution to the relevant authorities, putting out the investigation. The physician and the hospital held responsible for the fact will incur reputation damage together with all the legal consequences referred to in the previous sections (Meoli et al., 2018). For all these reasons, we believe that introducing potential liability in our model can be appropriate to reproduce the above-mentioned systems of incentives in the lab.

Table 3.1 reports the 2x2 design which allows for both a within-subject and a between-subject analysis, similar to Brosig et al. (2000). At a within level, participants under the same payment structure act both in the absence and in the presence of the possibility of being sued for dumping[69]. Additionally, the comparison between the two different payment conditions allows for between-groups tests.

[69] Notice that the order of the treatments varies across subjects to avoid carry-over bias (Charness et al., 2012).

Table 3.1: Experimental structure

	DUMPING		
PAYMENT SCHEME		YES	NO
	FFS	T1	T2
	SALARY	T3	T4

In all treatments, each physician *i* faces 9 different patients, pooled into three different groups according to the degree of their illness (*j*=1,2,3). For each patient, physicians observe the generic diagnosis[70] referred to a common disease and, hence, easy understandable to any physician regardless of his medical specialty. Patients are presented in a random order, which varies across treatments, to avoid carryover effects (Charness et al., 2012). The three types of patients belong to an heterogenous population. Patient types reflect the patients' different states of health: good(j=1), intermediate(j=2) and bad(j=3).

Diagnoses are provided in table 3.2.

Table 3.2: Diagnoses

Pathologies	**Severity of illness**
hypertension	1
measles	1
fever and a cough	1
cholecystitis	2
femur Fracture	2
respiratory distress	2
hepatic coma	3
intestinal obstruction	3
stroke	3

[70] Diagnoses have been provided by a general practitioner and then classified according to the severity of illness.

Once the participant has observed their diagnosis[71], he must decide if he wants to take charge of the patient or not. Then, if he has previously decided to treat the patient, he chooses the quantity of medical services q to give ($0 \leq q \leq 10$), knowing that, under the FFS system, both his profit Π and the patient's health benefit $B(q)$ are affected by that choice.

When it comes to another remuneration scheme such as salary, our model is simplified because subjects receive regular time-based payments, which means a fixed euro amount per specific time period. In contrast to FFS, physician *i'* profit is independent from the quantity of medical services provided but varies with the actual costs. Physician's revenue can be represented as follows:

$$\Pi(q) = \begin{cases} pq - cq^2 \; under \; FFS \\ L - cq^2 \; under \; salary \end{cases}$$

We set p=2 which is the fee for unit of service provided, c=0.1 which represents the marginal cost for the provision of medical services (Finocchiaro Castro *et al.,* 2019) and L=10 which refers to a fixed hourly amount received, similar to Lagarde and Blaauwn (2017). When participants are sued for malpractice their total profit is reset to zero (i.e. for that period, which means for that patient) (Finocchiaro Castro *et al.,* 2019).

As far as patients are concerned, their expected benefit used which is the same as in Ellis and Mcguire model (1986) can be derived as follows:

$$B^j(q) = \begin{cases} B_0^j + q \;\; if q \leq q^* \\ B_1^j - q \; if \; q \geq q^* \end{cases}$$

With $B_0^{j=1} = 7$, $B_0^{j=2}$=5, $B_0^{j=3}$=3 and B_1^j=B_0^s+2q^* ∀ j.

Based on the above, Finocchiaro Castro et al. (2019) have derived the optimal quantity to provide for patients, in both the payment structures, which depends on the degree of their illness: $q_1^* = 3(\text{patient with a good health status}), q_2^* = 5(\text{patient with an intermediate health status}), q_3^* = 7$ (patient with a bad health status).

[71] Notice that we do not report patients' sensitive data to protect privacy

Moreover, they fixed the value of *p(q)*, so as that $p^{j=1}(q) < p^{j=2}(q) < p^{j=3}(q)$ since it increases with the severity of illness and decreases with the quantity of services provided. In particular, they set:

$p^j\ (q) = \lambda^j\ \left(1 - \frac{q}{10}\right)$ where $\lambda^{j=1} = 0.3, \lambda^{j=2} = 0.4$ and $\lambda^{j=3} = 0.5$.

Finally, we set the variable which we add to this model *d(j),* which increases with the degree of illness, but does not vary with the quantity of services provided. We impose: *d(1)=0.1, d(2)=0.15 and d(3)=0.2*. Such low probabilities are supported by the arguments presented at the beginning of this section. In summary, since hospitals and then physicians can always subvert the antidumping law, their probability of facing a complaint for dumping is conceivably low.

Patient are assumed to be passive, accepting each level of medical services provided by physicians.

Having completed these decision tasks for all the treatments (i.e. two for each participant), participants answer a questionnaire about their social-economic status and the motivations which have driven their choices.

Although real patients are not actually present in our experiment, participants are conscious that their choices affect real patients outside the lab, because *B(q)* is converted into money which will be given to the 'Per Mano ONLUS', a local voluntary association which monitors and assists people affected by Duchenne Muscular Dystrophy (similar to Henning Schmidt et al., 2011; Finocchiaro Castro et al., 2019) .

Procedure

The experiment was conducted at the main hospital of Reggio Calabria, thanks to an agreement signed by the same hospital and the Mediterranean University. Local physicians who chose to contribute to this research took part in the experiment during their coffee-breaks. For this

reason, the experiment was conducted in different tranches, and with pool of different sizes[72]. 35 physicians joined the experimental sessions: 18 of them were assigned to the ffs treatment, while the remaining 17 were incentivized through salary. 49% of the subjects were emergency room doctors, to preserve the external validity of the experiment. In fact, ER physicians are those who more frequently decide to take charge of the patient, and then hospitalize him, or screen the patient and leave him go home (Iannello et al., 2015). The rest of physicians were cardiologists or oncologists.

Upon arrival, subjects were randomly allocated to the given seats, where they completed their task in full anonymity, using Z-tree (Fischbacher, 2007). Before starting the experiment, we measured physicians' attitude towards risks, since physicians make their choices facing liability, either for malpractice or dumping. Notably, liability condition can change physicians' attitude towards risks, therefore it is appropriate to assess it ex ante. To do this, we handed participants a questionnaire first proposed by Holt and Laury (2002), but with hypothetical rewards (see e.g. Galizzi et al., 2016). Such questionnaire included 10 hypothetical choices between a safer lottery and a riskier lottery, respectively A and B. Payoff and probabilities distributions allowed to estimate a subject' attitude towards risk by the number of times he had chosen lottery A. Typically, as the probability associated to the high payoff outcome increases, subjects ought to shift from option A to option B, as outlined by Holt and Laury (2002). Results show that 46% of the subjects opted for lottery A for the first three or four choices and then definitely moved to lottery B. 13% of the subjects preferred option B for all the strings. Finally, three subjects can be classified as very risk averse, choosing lottery A for all but the last option, while three subject showed irrational preferences, randomly moving from A to B.

Instructions were read aloud, and all the participants' doubts were clarified before the starting session. At the end of the experiment, one of the periods was randomly selected and physicians were paid with voucher meals whose value corresponded to their profit in that period[73]. We

[72] Notice that since the design provided for no participants' interaction, we did not require a specific amount of people in the room to start the experiment.

[73] According to McKeganey (2001), food voucher could replace cash payment for research participation, being

chose to use meal tickets to preserve the salience of the incentive mechanism (i.e. they can soon spend them at their cafeteria, during their lunch break), though being very low compared to physicians' opportunity cost. In fact, physicians should have intrinsic motivation on their own, knowing that they are contributing to research, and paying them something corresponding to their opportunity cost could reduce the quality of information they provide during the experiment (Gneezy and Rustichini, 2000)[74]. The whole experiment lasted approximately half an hour and the average reward was 9.03€ per participant.

Results

In this section I will provide the preliminary results of this experiment. The main variables of interest are respectively physicians' choice concerning whether to take charge of the patient and then the level of services provided. Regression analysis will be preceded by the descriptive analysis and nonparametric tests, given the small sample.

Descriptive analysis and nonparametric tests

Before starting with the descriptive analysis, it is important to specify what variable 'choice' stands for. '*Choice*' refers to the first decision which a physician has to make in this experiment, choosing whether to take charge of the patient, in that case the variable is equal to 1, or leave him and moving to the next patient, in that case the variable is equal to 0. If and only if the physician decides to treat the patient, he moves to the next stage and decides on the level of services to provide.

Table 3.3 and Table 3.4 report the summary statics for the four different treatments.

appropriate to the category of the participant group.

[74] The above-mentioned incentive is reasonably salient for two reasons. First, cafeteria is the only hospital internal alternative available to physicians. Although there are some external cafés, walking distance from the hospital, their opportunity cost may be high (physicians would have to push out and walk for 15 minutes). Additionally, according to the regulation, the internal cafeteria must charge discounted rates (20% less) to the hospital's employees.

As it was predicted in the hypotheses, the average frequency for choosing to treat the patient is higher under ffs (0.86) than under salary (0.65), and differences are significant at the 1% level according to the Wilcoxon ranksum, the median test and the chi-squared test (p-value<0.001). However, physicians seem to be insensitive to the introduction of dumping liability under the same incentive mechanism (p-value=0.571).

Table 3.3: Summary statics Salary

	Salary			Salary with dumping		
VARIABLES	mean	sd	obs	mean	Sd	obs
Choice	.653594	.4773	153	.6928105	.4628	153
Quantity of services provided	4.2	2.025	100	4.647619	1.951	100

Table 3.4: Summary statics Fee for service

	Fee for service			Fee for services with dumping		
VARIABLES	mean	sd	obs	mean	Sd	obs
Choice	.8580247	.3501	162	.8580247	.3501	162
Quantity of services provided	4.76259	2.175	139	5.05036	2.161	139

As far as the quantity of services provided is concerned, moving from salary to ffs raises the quantity of services provided, as shown in literature (e.g. Brosig et al., 2015 and Finocchiaro

Castro et al., 2019), and such increase is significant according to both the ranksum and median test at the 1% level. However, as in the previous case, physicians seem to be indifferent to dumping liability regardless of the payment structure (p-value>0.1). Table 3.5 provided below summarizes all the nonparametric tests conducted to compare different treatments.

Table 3.5: Nonparametric tests (p-value)

VARIABLES	**CHOICE**	**QUANTITY OF SERVICES PROVIDED**	
	Ranksum	Ranksum	Median
FFS vs SALARY	0.000	0.0008	0.000
FFS without d. vs SALARY without d.	0.000	0.0088	0.001
FFS with d. vs SALARY with d.	0.0004	0.02	0.000
FFS without d. vs FFS with d.	1.000	0.19	0.140
SALARY without d. vs SALARY with d.	0.46	0.04	0.246

Caption: d.=dumping liability

Based on the above, under malpractice liability the introduction of an additional liability for dumping does not affect physicians' decisions concerning both the choice of treating the patient and the amount of services to provide. Nevertheless, what is relevant for doctors' decisions is the payment structure. In fact, the frequency of dumping is higher under Salary than under FFS. More specifically, dumping frequency mean distinguished by incentive scheme and broken down by patient's severity of illness is shown in Table 3.6.

Table 3.6: Choice frequency mean by treatment and severity

VARIABLES	FFS	FFS w.d.l.	SALARY	SALARY w.d.l.
Severity of illness=1	0.67	0.63	0.1	0.16
Severity of illness=2	0.96	0.96	0.88	0.96
Severity of illness=3	0.94	0.98	0.98	0.96

Caption: w.d.l= with dumping liability

The greatest difference can be seen for the patients with the lowest level of illness, where the choice frequency is 0.1 under Salary and 0.67 under FFS. This result indicates that if physicians have to choose between different patients, they prioritize seriously ill patients.

Going to the quantity of services provided, showed in table 3.7, as mentioned before, the largest difference is observed for the patient whose degree of illness is low. Moreover, dumping liability introduction results in a slightly larger amount of services provided regardless of the severity of illness.

Table 3.7: Quantity of services mean by treatment and severity

VARIABLES	FFS	FFS w.d.l.	SALARY	SALARY w.d.l.
Severity of illness=1	3.17	3.41	2.2	2.86
Severity of illness=2	4.71	5.07	4.04	4.59
Severity of illness=3	5.94	6.07	4.54	4.95

Caption: w.d.l= with dumping liability

Regression analysis

Table 3.8 briefly describes all the variables being used in the following regressions.

Table 3.8: Variable list

VARIABLES	mean	sd	min	max	obs
Age	46.27	11.16	24	68	630
Emergency department	0.486	0.500	0	1	630
Male	0.400	0.490	0	1	630
Degree of illness	2	0.817	1	3	630
FFS	0.514	0.500	0	1	630
Dumping	0.500	0.500	0	1	630
Period	5	2.584	1	9	630
Choice	0.768	0.422	0	1	630
Quantity	4.704	2.108	0	10	630
Risk seeking	0.371	0.484	0	1	630
Gender*riskseeking	0.114	0.318	0	1	630
Ffs*dumping	0.257	0.437	0	1	630

'*Emergency department*' is a dummy variable capturing emergency department doctors. '*FFS*' is a dummy variable equal to 1 if physicians are paid according to FFS and 0 otherwise. 'Degree of illness' is a categorical variable which can assume the value 1,2 or 3 depending on the patient's severity of illness. '*Dumping*' is a dummy variable referred to the absence or the

presence of dumping liability in the treatment. '*FFS*dumping*' is an interaction variable catching FFS with dumping liability treatment. '*Male*' is the gender variable, '*Risk seeking*' indicates risk loving subjects, while '*Gender*riskseeking*' is the corresponding interaction between the former and the latter.

Since practitioners make 9 different decisions, standard errors are clustered at the subject level (Cameron et al., 2008). However, given the small number of clusters, bootstrap methods, originating pseudo-samples from the initial pool, are used (Roodman et al., 2019).

Prior to estimating the effect of the above-mentioned variables on the quantity of services, I need to account for some censored observations (when q does not take a value). In fact, given the structure of this experiment, the quantity of services takes a value only if physician's choice equals 1 at the very first stage, triggering the so-called selection problem (Sartori, 2003), since the sample only includes observations for not dumped patients. In similar cases, Heckmann and more generally selection models are the alternative solution to the standard OLS, since they not only consider the selection bias but also allow for possible dependence between two decisions, in our case choice and quantity. These models include two separate regressions: the selection equation which, in this case estimates the effect of some variables on the probability of choosing to treat the patient; and the outcome equation which considers the quantity of services provided as a dependent variable. However, the lamda parameter from the Heckman model for this dataset is not significant, which means that although a selection bias might exist, results converge to an OLS. A possible alternative to Heckman model is using a limited dependent variable model[75] such as the Cragg's model (1971) which similarly consists of a two-step estimation: a logit model for the discrete decision of whether the quantity of services takes a nonnegative value (i.e. choice of practicing dumping) and a truncated regression for the continuous decision (i.e. amount of services to provide). Before opting for it, I run the Tobit Test for the Tobit one-step model[76] versus the Cragg's two-step model (Lin and Schmidt, 1984), which rejected the first one in favour of the latter ($\lambda > \chi$ critical value).

[75] The actual value for the dependent variable, quantity of services to provide, is observed if the latent variable is above the limit (which I put -1 whenever patients are dumped).

[76] Tobit model assumes that the discrete decision (choice>0 and then quantity of services≥0) and the continuous decision $E[q|q \geq o]$ are the same. Here, such assumption is quite implausible (dumping decisions differ from choices concerning the amount of services to provide).

As far as the decision on whether to practice dumping is concerned (i.e. discrete decision), table 3.9 reports the output of the logit where 'choice' is the dependent variable. Numbers reported in the tables correspond to coefficients.

Table 3.9: Logit for choice

VARIABLES	First model	Second model	Third model
2.degreeofillness	3.718***	3.940***	3.975***
	(0.601)	(0.614)	(0.549)
3.degreeofillness	4.298***	4.532***	4.567***
	(0.790)	(0.892)	(0.905)
ffs	1.927***	2.583	2.973
	(0.631)	(3.714)	(3.220)
dumping	0.205	0.222	0.420**
	(0.172)	(0.200)	(0.190)
male		0.797	0.266
		(0.685)	(0.642)
age		-0.0624**	-0.0642**
		(0.0268)	(0.0309)
riskseeking		0.647	0.170
		(0.551)	(0.763)
emergency department		-0.714	-0.896
		(3.928)	(3.024)
genderriskseeking			1.973**
			(0.803)
ffsdumping			-0.420
			(0.466)
Constant	-1.627***	0.678	0.871
	(0.383)	(1.140)	(1.173)
Observations	630	630	630

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The probability that physicians decide to take charge of the patient increases with their severity of illness. In fact, given patient complexity there is not always enough time to meet all the patients' needs (Grant et al., 2013). As a result, physicians compare patients and prioritize the most serious ill (Chan et al., 2018). In this experiment, physicians decide to treat the patient depending on how urgently he requires physician's attention.

FFS payment scheme incentivizes patients' treatment. As it was anticipated in the hypotheses, the more physicians are paid for the services they provide, the less dumping is observed, the more overtreatment is boosted (Green, 2014). On the other hand, receiving a fixed payment such as salary results in physicians' low level of effort (Faloon, 2012). Consequently, if physicians decide to dump patients, they prefer doing it under Salary than under FFS.

Physicians' age decreases the probability of taking charge of the patient. According to Iannello et al. (2015), younger physicians make more decisions than their older colleagues. However, their decisions involve less responsibility than that taken by their older peers. As a result, if we assume that admittance decisions involve greater responsibility and that they are mainly managed by older physicians, we conclude that age contributes to physicians' unwillingness to take risks which would result in taking charge of the patient. In fact, if they refuse the patient, the probability of being sued for dumping, which is present in 2 over 4 treatments, is very low compared to the probability of malpractice claims which is not at play for dumped patients.

Men are more likely to take charge of the patients. This result is explained by introducing the interaction variable which links risk taking to gender. As it is often found in literature, men are more risk takers than women (Fogel and Nehmad, 2009; Pikkel et al., 2016). This contributes to men's willingness to accept patients, assuming medical malpractice claim risks.

Moving to the continuous decision, concerning the amount of services to provide table 3.10 provides the truncated regression, where 147 observations were truncated because of dumping.

Table 3.10: Truncated regression for the quantity of services provided

VARIABLES	First model	/	Second model	
2.degreeofillness	1.845***		1.769***	
	(0.337)		(0.280)	
3.degreeofillness	2.612***		2.542***	
	(0.353)		(0.266)	
ffs	1.096*		1.060	
	(0.570)		(0.873)	
dumping	0.358***		0.531***	
	(0.103)		(0.108)	
age	0.0261		0.0215	
	(0.0344)		(0.0243)	
prontosoccorso	-0.588		-0.276	
	(0.617)		(0.794)	
male	0.503		1.239*	
	(0.719)		(0.649)	
riskseeking	-0.335		0.383	
	(0.636)		(0.614)	
genderriskseeking			-1.888	
			(1.395)	
ffsdumping			-0.284	
			(0.208)	
sigma		1.846***		1.794***
		(0.181)		(0.140)
Constant	1.043		0.841	
	(1.380)		(1.029)	
Observations	483	483	483	483

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Older physicians provide greater services than their younger colleagues. Reasonably, younger physicians coming from a recent training are more concerned about services cost-effectiveness than their older colleagues. In fact, in recent years, education programmes addressed to medical students integrates economics to incentivize cost-conscious care[77] (Stammen et al., 2015). As a

[77] https://www.renalandurologynews.com/home/departments/practice-management/integrating-economics-into-medical-training/

result, younger physicians are more affected by treatments cost assessment in their decisions and try to reduce the amount of unnecessary treatments whenever possible.

FFS and the presence of dumping liability increases the amount of services to provide. As I mentioned in the introduction, on the one hand, a fixed fee for each unit of service (FFS) encourages the provision of services. On the other hand, a reverse incentive is provided under a prospective payment system such as salary (Donaldson and Gerard, 1989). As far as liability is concerned, although this it is not actually linked to the quantity of services provided, it affects the extent to which physicians treat patients. Similar to Finocchiaro Castro et al. (2019), the introduction of liability increases medical services by 0.525.

Men provide more services than women doctors. Many studies did not find any difference in medical expenditures (e.g. drug prescriptions, office visits) between men and women doctors (see e.g. Jerant et al., 2013). However, a stream of literature revealed that women are more likely to follow protocols prescribed by guidelines than men (see e.g. Baumhäkel et al., 2009; Kim et al., 2015). Therefore, if men doctors go beyond practice patterns, they could exceed in the quantity of services provided.

Finally, severity of illness increases the amount of services provided, as in Finocchiaro Castro et al. (2019), since more serious patients generally require more treatments than others.

Conclusions

This artefactual field experiment conducted in the context of the metropolitan hospital of Reggio Calabria saw the participation of 35 physicians, either emergency department doctors, cardiologists and oncologists. Drawing from standard procedures (e.g. Henning Schmidt *et al.,* 2011: Finocchiaro Castro et al., 2019) and adding patients' diagnoses, I tested whether and to which extent the adoption of fee-for-service or salary system can induce physicians to practice patient dumping. Facing the possibility of being sued for medical malpractice, physicians decide whether to take charge of patients. Also, I checked whether the introduction of the risk of being sued for a physician for having practiced dumping can have effect on his behaviour. Results show that dumping is more often observed under Salary than under Ffs. However, physicians seem to be insensitive to the introduction of dumping liability under the same incentive mechanism, though it seems to trigger a higher amount of services provided. Older

physicians are more likely to take charge of the patient than their younger colleagues. Finally, men physicians are found to more frequently take charge of the patients, due to their risk taking behaviour, and to increase the amount of services to provide.

Though more relevant in the American context, this experiment shows that dumping can be viewed with varying perspectives (e.g. hospital physicians believe that patient's pathology falls under the competence of a family practitioner; physicians tend to prioritize serious ill patient if they have to decide which patient to take charge of, and so on). Since older physicians are found to take charge of patients more than their younger colleagues and to significantly increase the amount of services to provide, introducing economics training among more experienced physicians to incentivize cost-conscious care could be a solution. In fact, according to Cohen et al. (1982) assessing the level of physicians' training, and, we add, their years of experience, is the first step to design effective medical education programmes. As far as dumping is concerned, policy strategies may vary on the hospital purposes. If the hospital wants to avoid unnecessary treatments and hospitalizations, reducing expenditures, salary must be adopted, since it incentivizes physicians to refuse not seriously ill patients. If hospital opts for reducing the rate of refused patients, to avoid further legal complications, FFS is the preferred alternative. On the contrary, introducing liability only affects the amount of services provided.

The above-mentioned measures, should be more specifically addressed to doctors who always decide on patients' admission (e.g. emergency department doctors).

To provide an incentive against patient dumping we introduced a probability of being sued whether physicians decide not to take charge of the patient. Though very low, such probability increases in patient's severity of illness and just depends on it. It does not take into account the circumstances which lead physician to dump the patient, such as the desire to prioritize more serious patients. To overcome this limit, the experimental design could provide physicians with the possibility of choosing between different patients (e.g. cream skim them), making dumping probability vary on such decision. Another possible extension of the experiment could be making the random order of patients' diagnoses depend on physicians' decision of taking charge of patients[78]. This extension could allow to investigate how physicians arrange priority

[78] For example, if the first patient which physician faces is affected by fewer, and physician refuses to treat him, this could result in a high/low probability of receiving a serious ill patient as next patient (i.e. depending on the severity of illness of the patient refused).

lists. Furthermore, physicians could be given an additional detail about their work shift to check if dumping decision can be attributed to that. For example, Spetz et al. 2001 found that in the Californian context physicians are more likely to perform caesarean sections in the evening hours since they may be affected by leisure incentive. In our context, telling a physician the hour of the day and its match with his work shift could significantly vary his decisions.

References

Agenzia di Sanità pubblica, (2002), Criteri clinico-organizzativi per il trasferimento del paziente critico nella rete dell'emergenza

Ansell, D.A., & Schiff, R.L., (1987), Patient Dumping: Status, Implications, and Policy Recommendations. *JAMA*, 257(11):1500–1502

Asprone, E., (2015)**,** Diritto Penale. Il rifiuto da parte del medico del ricovero ospedaliero del malato in condizioni d'urgenza integra il reato di rifiuto d'atti d'ufficio, *Gazzetta Forense*

Barros, P.P., (2003), Cream-skimming, incentives for efficiency and payment system, *Journal of Health Economics*, 22, 419-443

Barnett, A. E., & Mayer, G. G., (Eds.), (1992), Ambulatory care management and practice, Jones & Bartlett Learning

Bayer, R. C., Renner, E., & Sausgruber, R., (2013), Confusion and learning in the voluntary contributions game, *Experimental Economics*, 16(4), 478-496

Baumhäkel, M., Müller, U., & Böhm, M., (2009), Influence of gender of physicians and patients on guideline‐recommended treatment of chronic heart failure in a cross‐sectional study, *European journal of heart failure*, 11(3), 299-303

Bertoli, P., & Grembi, V., (2018), Medical Malpractice: How Legal Liability Affects Medical Decisions, *Health Econometrics*, 235-261, Emerald Publishing Limited

Blomqvist, Å., Busby, C., (2012), How to pay family doctors: why "pay per patient" is better than fee for service, CD Howe Institute

Brosig-Koch, J., Henning Schmidt, H., Kairies-Schwartz, N., & Wiesen, D., (2016), Using artefactual and lab experiments to investigate how fee-for-service and capitation affect medical service provision, *Journal of Economic Behavior & Organization*, 131, 17-23

Brosig-Koch J., Henning Schmidt H., Kairies-Schwarz N. & Wiesen D., (2017), The effect of introducing mixed payment systems for physician: experimental evidence, *Health Economics*, 26, 243-262.

Brosig-Koch, J., Hennig‐Schmidt, H., Kairies-Schwarz, N., Kokot, J., & Wiesen, D., (2019), Physician performance pay: Experimental evidence, Available at SSRN 3467583

Busse, R., Schreyögg, J., & Smith, P. C., (2006), Hospital case payment systems in Europe, *Health care management science*, 9(3), 211-213

Busse, R., Geissler, A., Aaviksoo, A., Cots, F., Häkkinen, U., Kobel, C., ... & Street, A., (2013), Diagnosis related groups in Europe: moving towards transparency, efficiency, and quality in hospitals?, *Bmj*, 346, f3197

Cameron, A. C., Gelbach, J. B., & Miller, D. L., (2008), Bootstrap-based improvements for inference with clustered errors, The Review of Economics and Statistics, 90(3), 414-427

Cantù, E., Carbone, C., & Anessi-Pessina, E., (2011), 21, Do Italian regions effectively use DRG funding to steer provider behaviour?, Policy, performance and management in governance and intergovernmental relations: transatlantic perspectives, 341

Cavaliere, A., (2017), Responsabilità medica alla luce della Riforma, avaible at www.diritto.it

Cavalieri, M., Gitto, L., & Guccio, C., (2013), Reimbursement systems and quality of hospital care: an empirical analysis for Italy, *Health policy*, 111(3), 273-289

Chalkley, M., & Malcomson, J.M., (1988), Contracting for health services when patient demand does not reflect quality, *Journal of Health Economics*, 17(1), 1-19

Chan, T. M., Mercuri, M., Van Dewark, K., Sherbino, J., Schwartz, A., Norman, G., & Lineberry, M., (2018), Managing multiplicity: Conceptualizing physician cognition in multipatient environments, *Academic Medicine*, 93(5), 786-793

Charness, G., Gneezy, U., & Kuhn, M.A., (2012), Experimental methods: Between-subject and within-subject design, *Journal of Economic Behavior and Organization*, 81, 1-8

Clemens, J., & Gottlieb, J. D., (2014), Do physicians' financial incentives affect medical treatment and patient health?, *American Economic Review*, 104(4), 1320-49

Cohen, S. J., Weinberger, M., Mazzuca, S. A., & McDonald, C. J., (1982), Perceived influence of different information sources on the decision-making of internal medicine house staff and faculty, *Social Science & Medicine*, 16(14), 1361-1364

Consolidated Omnibus Budget Reconciliation Act (COBRA) of 1985, PL 99-272, 42 U.S.C. §§ 300 et seq

Corte di Cassazione, sez. VI Penale, sentenza 5 luglio – 29 settembre 2016, n. 40753

Cragg, J. G., (1971), Some statistical models for limited dependent variables with application to the demand for durable goods, Econometrica: Journal of the Econometric Society, 829-844

Cupelli, C. (2017). Lo statuto penale della colpa medica e le incerte novità della legge Gelli-Bianco. *Diritto Penale Contemporaneo*, (4), 200-216

Currie, J., & MacLeod, W. B. (2008). First do no harm? Tort reform and birth outcomes, *The Quarterly Journal of Economics*, 123(2), 795-830

Dallek, G., & Waxman, J, (1985), National Health Law Program. Clearinghouse Rev., 19, 1413

Donaldson, C., Gerard, K., (1989), Paying general practitioners: shedding light on the review of health services, *Journal of the Royal College of General Practitioners*

Danzon, P. M., (2000), Liability for Medical Malpractice, Handbook of Health Economics, In A. J. Culyer and J. P. Newhouse (ed.), *Handbook of Health Economics*(Chapter 26,pp. 1339-1404)

D'Athis, P. F., Barriot, P., Buffat, P. J., & Dubouloz, M., (1992), Recommandations concernant la surveillance des patients au cours des transferts interhospitaliers médicalisés, Sociètè Francaise d'Anesthèsie et de Rèanimation

Dranove, D., (1987), Rate-setting by diagnosis related groups and hospital specialization, *The RAND Journal of Economics*, 417-427

Eberlin, K.R., Hartzell, T.L., Kuo, P., Winograd, J., & Day, C., (2013), Patients Transferred for Emergency Upper Extremity Evaluation. Plastic and Reconstructive Surgery 131:3, 593-600

Ethier, W.J., (1982), Dumping, *Journal of Political Economy*, 90, 3, 487-506

Eze, P., & Wolfe, B., (1993), Is dumping socially inefficient?: An analysis of the effect of Medicare's prospective payment system on the utilization of Veterans Affairs inpatient services, *Journal of Public Economics*, 52(3), 329-344

Fabrikant, R., Kalb, P. E., Bucy, P. H., Hopson, M. D., & Stansel, J. C., (2018), Health care fraud: enforcement and compliance, Law Journal Press

Faloon, T., (2012), Physician Remuneration Options, Physicians services CMA companies.

Fattorini, P., Peretti, A., Bergamini, P., & Valentini, R., (2007), Casistica ortopedica dell'osservatorio GISDI. Medical malpractice daily*, Supplemento di Medicina Legale Quaderni Camerti, reg. Trib. di Camerino,* avaible at

http://www.malpracticedaily.org/index.php?option=com_content&task=view&id=19&Itemid=2.

Finocchiaro Castro, M., Ferrara, P., Guccio, C., & Lisi, D., (2019), Medical Malpractice Liability and Physicians' Behavior: Experimental Evidence, *Journal of Economic Behavior & Organization*, 166, 646-666

Fischbacher, U., (2007), z-Tree: Zurich toolbox for ready-made economic experiments, Experimental economics, 10(2), 171-178

Fogel, J., & Nehmad, E., (2009), Internet social network communities: Risk taking, trust, and privacy concerns. Computers in human behavior, 25(1), 153-160

France, G. (2006), "Seeking a better balance: Developments in intergovernmental relations in Italian health care", Eurohealth, vol. 13, n. 3

Fusciani M., (2004), Rischio tecnologico e responsabilità legale in sanità, At Convegno ANIA, Milano

Gabriele, S., (2019), Lo stato della sanità in Italia, Focus n.6, Ufficio parlamentare di bilancio

Galizzi, M. M., Miraldo, M., & Stavropoulou, C., (2016), In sickness but not in wealth: field evidence on patients' risk preferences in financial and health domains, Medical Decision Making, 36(4), 503-517

Gneezy, U., & Rustichini, A., (2000), Pay enough or don't pay at all, *The Quarterly Journal of Economics*, 115(3), 791-810

Godager, G., Henning Schmidt, H., & Iversen, T., (2016), Does performance disclosure influence physicians' medical decision? An experimental study, *Journal of Economic Behavior & Organization*, 131, 36-46

Grant, R. W., Adams, A. S., Bayliss, E. A., & Heisler, M., (2013), Establishing visit priorities for complex patients: A summary of the literature and conceptual model to guide innovative interventions, *Healthcare (Amsterdam, Netherlands)*, 1(3-4), 117–122

Green, P.E., (2014), Payment systems in the healthcare industry: An experimental study of physicians' incentives, *Journal of Economic Behavior & Organization*, 106, 367-378

Henning Schmidt, H., Selten, R., & Wiesen, D., (2011), How payment systems affect physicians' provision behaviour- An experimental investigation, *Journal of Health Economics*, 30, 637-646

Henning Schmidt, & H., Wiesen, D., (2014), Other-regarding behaviour and motivation in health care provision: An experiment with medical and non-medical students, *Social Science & Medicine*, 106, 156-165

Iannello, P., Perucca, V., Riva, S., Antonietti, A., & Pravettoni, G., (2015), What do physicians believe about the way decisions are made? A pilot study on metacognitive knowledge in the medical context, *Europe's journal of psychology*, *11*(4), 691

Jena, A.B., Seabury, S., Lakdawalla, D., et al., (2011), Malpractice risk according to physician specialty, *N Engl J Med*, 365, 629–36

Jena, A. B., Schoemaker, L., Bhattacharya, J., & Seabury, S. A., (2015), Physician spending and subsequent risk of malpractice claims: observational study, *BMJ*, 351, h5516

Jerant, A., Bertakis, K. D., Fenton, J. J., & Franks, P., (2013), Gender of physician as the usual source of care and patient health care utilization and mortality, *The Journal of the American Board of Family Medicine*, 26(2), 138-148

Kellermann, A. L., & Hackman, B. B, (1990), Patient 'dumping' post-COBRA, *American journal of public health*, 80(7), 864-867

Kerschbamer, R., & Sutter, M., (2017), The economics of credence goods–a survey of recent lab and field experiments, *CESifo Econ. Stud*. 63 (1), 1–23

Kessler, D. P., (2011), Evaluating the Medical Malpractice Systems and Options for Re-form, *Journal of Economic Perspectives*, 25(2), 93-110

Kessler, D. P., & McClellan, M. B., (2002), How liability law affects medical productivity, *Journal of health economics*, 21(6), 931-955

Kim, C., McEwen, L. N., Gerzoff, R. B., Marrero, D. G., Mangione, C. M., Selby, J. V., & Herman, W. H., (2005), Is physician gender associated with the quality of diabetes care?, Diabetes care, 28(7), 1594-1598

Kusserow, P.R., (1988), Patient dumping after COBRA, OAI-12-88-00830

Lagarde, M., & Blaauw, D., (2017), Physicians respond to financial and social incentives: A medically framed real effort experiment, *Social Science & Medicine*, 179, 147-159

Lin, T. F., & Schmidt, P., (1984), A test of the Tobit specification against an alternative suggested by Cragg, The Review of Economics and Statistics, 174-177

Matsaganis, M., & Glennerster, H. (1994), The threat of 'cream skimming'in the post-reform NHS, *Journal of health economics*, 13(1), 31-60

Maynard, A., Marinker, M. & Gray, P.D., (1986), The doctor, the patient, and their contract, *British Medical Journal*, 292, 1438-1440.

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

Meoli, B., Sica, S., Stanzione, P., (2018), Commentario alla legge 8 marzo 2017 n.24, Disposizioni in materia di sicurezza delle cure e della persona assistita, nonché in materia di responsabilità degli esercenti le professioni sanitarie, Edizioni Scientifiche Italiane

Mcguire T.G., (2000), Physician Agency, *Health Economics,* 1, Part A, 461-536

McKeganey, N., (2001), To pay or not to pay: respondents' motivation for participating in research, Addiction, 96(9), 1237-1238

Monfeli, L., (2018), Atto medico e responsabilità erariale, Giustizia Civile.com

Mullner, R. M., McNeil, D. L., & Andes, S., (1986), Hospital closure: Who will be at risk in the upcoming decade?, *Healthcare financial management: journal of the Healthcare Financial Management Association*, *40*(1), 42-48

Nahed, B. V., Babu, M. A., Smith, T. R., & Heary, R. F., (2012), Malpractice liability and defensive medicine: a national survey of neurosurgeons, PloS one, 7(6)

Newhouse, J.P., (1983), Two prospective difficulties with prospective payment of hospitals, or, it's better to be a resident than a patient with an complex problem, *Journal of Health Economics*, 2(3), 269–74

New York Times, December 6, (1989), A Nurse Is Sentenced for Denying Treatment

New York Times, March 23, (1991), U. S. Law on Hospital Care of Poor Faces Test

Pikkel, D., Pikkel Igal, Y. S., Sharabi-Nov, A., & Pikkel, J., (2016), Are doctors risk takers?, *Risk management and healthcare policy*, 9, 129–133

Preston, A. M., Chua, W. F., & Neu, D, (1997), The diagnosis-related group-prospective payment system and the problem of the government of rationing health care to the elderly, *Accounting, Organizations and Society*, *22*(2), 147-164

Rice, M., Jones, W., & Rouge, B., (1988), The Uninsured and patient dumping: recent policy responses in indigent care, *Journal of the national medical association*, Vol. 83, NO. 10.

Robinson, J. C., (2001), Theory and practice in the design of physician payment incentives, *The Milbank Quarterly*, 79(2), 149-177

Roodman, D., Nielsen, M. Ø., MacKinnon, J. G., & Webb, M. D., (2019), Fast and wild: Bootstrap inference in Stata using boottest. The Stata Journal, 19(1), 4-60

Saks, S. J., (2004), Call 911: psychiatry and the new Emergency Medical Treatment and Active Labor Act (EMTALA) regulations, *The Journal of psychiatry & law*, 32(4), 483-512

Salvatore, D., (1989), A model of dumping and protectionism in the United States, *Weltwirtschaftliches Archiv*, Bd. 125, H. 4, 763-781

Santovito, D., Iorio, M., Mastro, F., Conforti, M., & Rispoli, P., (2007), Symptomatic carotid stenosis. Diagnosis and treatment guidelines. Juridical and medico-legal approach. Minerva Cardioangiol, 55:167–198

Sappington, D.E.M., & Lewis, T.R., (1999), Using subjective risk adjusting to prevent patient dumping in the health care industry, *Journal of Economics & Management Strategy*, 3, 351-382.

Sartori, A. E. (2003), An estimator for some binary-outcome selection models without exclusion restrictions, Political Analysis, 11(2), 111-138

Simoens, S., & Giuffrida, A., (2004), The impact of physician payment methods on raising the efficiency of the healthcare system, *Applied health economics and health policy*, 3(1), 39-46

Spetz, J., Smith, M. W., & Ennis, S. F., (2001), Physician incentives and the timing of cesarean sections: evidence from California, *Medical care*, 536-550

Stammen, L. A., Stalmeijer, R. E., Paternotte, E., Pool, A. O., Driessen, E. W., Scheele, F., & Stassen, L. P., (2015), Training physicians to provide high-value, cost-conscious care: a systematic review, *Jama*, 314(22), 2384-2400

Struyk, K., (2015), Cause and Effect of "Patient Dumping": A Dark Side of Health Care in America, www.academia.edu.it

Studdert, D.M, Mello, M. M., Sage, W., DesRoches, W.M., Peugh, C.M., Zapert, K., & Brennan, T.A., (2005), Defensive Medicine among high-risk specialist physicians in a volatile malpractice environment, *Jama*, 293(21), 2609-2617

Takahara, T., (2016), Patient dumping, outlier payments, and optimal healthcare payment policy under asymmetric information, *Health economics review*, 6(1), 57

Traina, F., 2009, Medical Malpractice: The Experience in Italy, *Clinical Orthopaedics and Related Research*, 467(2), 434–442

Treiger, K.I., (1986), Preventing Patient Dumping: Sharpening the Cobra's Fangs, 61 N.Y.U. L. Rev. 1186

Volpp, K. G., & Siegel, B., (1993), New Jersey: Long-Term Experience with All-Payer State Rate Setting. Health Affairs, 12(2), 59-65

Zibulewsky, J., (2001), The Emergency Medical Treatment and Active Labor Act (EMTALA): what it is and what it means for physicians, Proceedings (Baylor University. Medical Center), 14(4), 339–346

Zuabi, N., Weiss, L. D., & Langdorf, M. I., (2016), Emergency Medical Treatment and Labor Act (EMTALA) 2002-15: Review of Office of Inspector General Patient Dumping Settlements, The western journal of emergency medicine, 17(3), 245–251

Appendix D

Italian regulation of medical malpractice

From a legal point of view, for malpractice cases, a physician is both liable to prosecution (in a criminal court) and civil action (in a civil court) in our country (Traina, 2009).

According to decree Balduzzi (law 8 November 2012, n. 189) any physician incurred criminal liability for malice, gross negligence and even slight fault unless he succeeded in demonstrating that he slavishly followed the international guidelines, contained in the same law, or alternatively the good clinical care practices (Cavaliere, 2017).

Data show that healthcare costs are drastically increasing (i.e. partly due to medical malpractice) and that the practice of defensive medicine is becoming even more common than in the past (Traina, 2009). This was due to the previous Italian legislation on medical malpractice which was said to be excessively favourable to injured patients and their attorneys but too disadvantageous to physicians (Monfeli, 2018). For this reason, with the aim of protecting personal and professional dignity of medical professionals, Gelli-Bianco Law (No. 64/2017[79]) replaced the above-mentioned law (Cupelli, 2017). From a criminal point of view, this law has overcome the previous distinction between gross and slight fault and has introduced a cause for exemption from liability: in the event of personal injuries to the patient or manslaughter caused by the physician's unskillfulness[80] , criminality is excluded as long as the physician had followed the international guidelines which are considered appropriate[81] for that case (art. 590-sexies, penal code).

From a civil point of view, the new rules provide for a clear separation between the responsibility of the hospital and that of the physician (Meoli et al., 2018). According to the Court of Cassation "the acceptance of the patient in the hospital for admission or for a clinical

[79] "Provisions on the safety of care and the assisted person, as well as on the professional responsibility of health care professionals".
[80] 'Unskillfulness' consists of a poor attitude in those activities which do need special technical knowledge and implies a deficiency of culture, practice, intuition and capacity of observation (Santovito et al., 2007).
[81] Notice that the specificity of the case could require that the physician deviates from the guidelines (art.5 Gelli-Bianco Law).

control, involves the conclusion of a contract[82]" , and for this reason the medical facility guilty of willfully or voluntary misconduct is contractual liable towards the patient (art. 7 Gelli-Bianco Law). However, the healthcare practitioner is answerable to the non-contractual liability towards the patient. This important demarcation between the hospital and the physician has a direct impact on the burden of proof as well as on the terms of prescription. In fact, when it comes to contractual liability (non-contractual), the burden of proof is on the defendant (the accuser) and the statute of limitation is 10 years (5 years).

In conclusion, the new legislation is trying to minimise the criticisms of the physician's guilt in order not to mortify the professional's actions. Indeed, the new legislation has been enacted so as to restore physician's operating calm and prevent him from practising defensive medicine (Cavaliere, 2017).

[82] 28. Italian Court of Cassation, United Sections, Judgment n.577, 11 January 2008.

Appendix E

Instructions (largely adapted from Finocchiaro Castro et al., 2019)

Welcome to our experiment

You are going to join an experiment on individual decision-making. Instructions are straightforward and, if you pay close attention, you may gain a monetary amount that will be paid to you in corresponding meal tickets at the end of the experiment. The amount of cash you may win depends only on your decisions and will not be affected by other participants' decisions. Your monetary gains, measured in Experimental Crown (EC), will be converted into Euro at the following exchange rate 1 EC = 0.45 Euro. For example, if, at the end of the experiment, you achieve 40 EC, you will receive a 18 Euro meal ticket.

Experimental design

The experiment lasts approximately 30 min and is divided into two stages. You will receive detailed instructions at the beginning of each stage. Please, remind that the decisions taken in one stage of the experiment do not have effects on the decisions that you will have to take in the following stage of the experiment.

Stage I

Please, read carefully the following instructions regarding stage I. If anything in the instructions is not clear please raise your hand and one of the experimenters will approach you. From this moment onward, you cannot communicate with any other participant. If you fail to do so, you will be asked to leave the room

Stage I lasts for nine periods. In each period, you will play in the role of a physician and you will have to decide whether to take charge of an already diagnosed patient. In each period you will face a patient with a different diagnosis. Each diagnosis is associated with a different level

of severity of illness (low, medium, high). If you decide to treat the patient you then have to decide how many medical prescriptions to provide to patients. In other words, you have to decide on the level of medical care (in terms of drugs, diagnostic exams, …) to provide to patients according to their severity of illness. Thus, you will face nine patients. When taking the decision on patient's medical care, you can choose among 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 prescriptions per patient.

If you decide not to treat the patient you will skip to the following period. If you decide otherwise to treat the patient, after the decision on the level of medical prescriptions to provide, the patient could sue you for medical malpractice with probability *Pr*, which depends on the level of medical prescriptions already provided.

The following table show the relationship between patient's severity of illness and your profit, if you decide not to treat the patient.

Severity of illness	Your profit
Low (1)	0
Medium (2)	0
High (3)	0

The other tables we will provide before taking your decision, show the relationship between provided prescriptions and the probability of being sued.

Earnings

In each period of stage I, you will be paid according to the FFS payment system. Your earnings increase together with the number of medical prescriptions that you provide to patients. Moreover, you bear a cost due to the level of effort devoted to visiting each patient that depends on how many medical prescriptions you provide to patients. If you get sued by a patient, you will incur a fixed monetary loss equal to the profits earned in the same period you are sued. Hence, your profit in each period is computed as the payment you receive from the FFS system

minus the cost due to the provision of medical services minus, if sued, the monetary loss due to being sued by the patient.

Each level of medical prescription provided accrues a certain level of benefit to patient according to her/his severity of illness. Therefore, your choice on the quantity of medical prescriptions to provide determines both your profits and the patients' benefits.

In each period, you will see on the screen (see below) all the information regarding the patient you currently face: his diagnosis, the associated severity of illness, your earning according to the payment system in use, the related costs, the probability of being sued for each possible level of medical prescriptions, the monetary loss due to being sued, your profits and the corresponding patient's benefits.

Stage II

Please, read carefully the following instructions regarding stage I. If anything in the instructions is not clear please raise your hand and one of the experimenters will approach you. From this moment onward, you cannot communicate with any other participant. If you fail to do so, you will be asked to leave the room.

Stage I lasts for nine periods. In each period, you will play in the role of a physician and you will have to decide whether to take charge of an already diagnosed patient. In each period you will face a patient with a different diagnosis. Each diagnosis is associated with a different level of severity of illness (low, medium, high). If you decide to treat the patient you then have to decide how many medical prescriptions to provide to patients. In other words, you have to decide on the level of medical care (in terms of drugs, diagnostic exams, …) to provide to patients according to their severity of illness. Thus, you will face nine patients. When taking the decision on patient's medical care, you can choose among 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 prescriptions per patient.

If you decide not to treat the patient, before skipping to the following period, you may be sued with probability *d*, which depends on the patient's severity of illness. If you decide otherwise to treat the patient, after the decision on the level of medical prescriptions to provide, the patient could sue you for medical malpractice with probability Pr, which depends on the level of medical prescriptions already provided.

The following table show the relationship between the probability of being sued for not treating the patient, *d*, and the patient's severity of illness.

Severity of illness	Probability *d*	Your profit
Low (1)	10%	-10
Medium (2)	15%	-10
High (3)	20%	-10

Earnings

In each period of stage II, you will be paid according to the FFS payment system. Your earnings increase together with the number of medical prescriptions that you provide to patients. Moreover, you bear a cost due to the level of effort devoted to visiting each patient that depends on how many medical prescriptions you provide to patients. If you decide not to treat the patient and you get sued for that, you will incur a loss as shown in table and your profit will be simply equal to it. If you decide to treat the patient and you get sued by a patient for malpractice, you will incur a fixed monetary loss equal to the profits earned in the same period you are sued. Hence, if you treat the patient, your profit in each period is computed as the payment you receive from the FFS system minus the cost due to the provision of medical services minus, if sued, the monetary loss due to being sued by the patient.

Each level of medical prescription provided accrues a certain level of benefit to patient according to her/his severity of illness. Therefore, your choice on the quantity of medical prescriptions to provide determines both your profits and the patients' benefits.

In each period, you will see on the screen (see below) all the information regarding the patient you currently face: his diagnosis, the associated severity of illness, your earning according to the payment system in use, the related costs, the probability of being sued for each possible level of medical prescriptions, the monetary loss due to being sued, your profits and the corresponding patient's benefits.

Stage I (for a different pool)

Please, read carefully the following instructions regarding stage I. If anything in the instructions is not clear please raise your hand and one of the experimenters will approach you. From this moment onward, you cannot communicate with any other participant. If you fail to do so, you will be asked to leave the room.

Stage I lasts for nine periods. In each period, you will play in the role of a physician and you will have to decide whether to take charge of an already diagnosed patient. In each period you will face a patient with a different diagnosis. Each diagnosis is associated with a different level of severity of illness (low, medium, high). If you decide to treat the patient you then have to decide how many medical prescriptions to provide to patients. In other words, you have to decide on the level of medical care (in terms of drugs, diagnostic exams, …) to provide to patients according to their severity of illness. Thus, you will face nine patients. When taking the decision on patient's medical care, you can choose among 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 prescriptions per patient.

If you decide not to treat the patient you will skip to the following period. If you decide otherwise to treat the patient, after the decision on the level of medical prescriptions to provide, the patient could sue you for medical malpractice with probability *Pr*, which depends on the level of medical prescriptions already provided.

The following table show the relationship between patient's severity of illness and your profit, if you decide not to treat the patient.

Severity of illness	Your profit
Low (1)	10
Medium (2)	10
High (3)	10

The other tables we will provide before taking your decision, show the relationship between provided prescriptions and the probability of being sued.

Earnings

In each period of Stage I, you will be given a fixed salary. Your remuneration does not vary with the quantity of medical services provided. Your profit in each period is computed as your fixed salary equal to 10, minus the cost due to the provision of medical services if you treat the patient, minus, if sued, the monetary loss due to being sued by the patient.

Each level of medical prescription provided accrues a certain level of benefit to patient according to her/his severity of illness. Therefore, your choice on the quantity of medical prescriptions to provide determines both your profits and the patients' benefits.

In each period, you will see on the screen (see below) all the information regarding the patient you currently face: his diagnosis, the associated severity of illness, your earning according to the payment system in use, the related costs, the probability of being sued for each possible level of medical prescriptions, the monetary loss due to being sued, your profits and the corresponding patient's benefits.

Stage II (for a different pool)

Please, read carefully the following instructions regarding stage I. If anything in the instructions is not clear please raise your hand and one of the experimenters will approach you. From this moment onward, you cannot communicate with any other participant. If you fail to do so, you will be asked to leave the room.

Stage I lasts for nine periods. In each period, you will play in the role of a physician and you will have to decide whether to take charge of an already diagnosed patient. In each period you will face a patient with a different diagnosis. Each diagnosis is associated with a different level of severity of illness (low, medium, high). If you decide to treat the patient you then have to decide how many medical prescriptions to provide to patients. In other words, you have to decide on the level of medical care (in terms of drugs, diagnostic exams, …) to provide to patients according to their severity of illness. Thus, you will face nine patients. When taking the decision on patient's medical care, you can choose among 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 prescriptions per patient.

If you decide not to treat the patient, before skipping to the following period, you may be sued with probability *d*, which depends on the patient's severity of illness. If you decide otherwise to treat the patient, after the decision on the level of medical prescriptions to provide, the patient could sue you for medical malpractice with probability Pr, which depends on the level of medical prescriptions already provided.

The following table show the relationship between the probability of being sued for not treating the patient, *d*, and the patient's severity of illness.

Severity of illness	Probability *d*	Profit
Low (1)	10%	0
Medium (2)	15%	0
High (3)	20%	0

The other tables we will provide before taking your decision, show the relationship between provided prescriptions and the probability of being sued.

Earnings

In each period of Stage II, you will be given a fixed salary. Your remuneration does not vary with the quantity of medical services provided. If you decide not to treat the patient and you get sued for that, you will incur a loss as shown in table and your profit will be simply equal to it. Otherwise, your profit in each period is computed as your fixed salary equal to 10, minus the cost due to the provision of medical services if you treat the patient, minus, if sued, the monetary loss due to being sued by the patient.

Each level of medical prescription provided accrues a certain level of benefit to patient according to her/his severity of illness. Therefore, your choice on the quantity of medical prescriptions to provide determines both your profits and the patients' benefits.

In each period, you will see on the screen (see below) all the information regarding the patient you currently face: his diagnosis, the associated severity of illness, your earning according to the payment system in use, the related costs, the probability of being sued for each possible level of medical prescriptions, the monetary loss due to being sued, your profits and the corresponding patient's benefits.

Payment

At the end of the experiment, one of the nine periods of stage I will be randomly drawn. The profit achieved in that period will be paid to you in corresponding meal tickets. While you in this stage have decided in the role of physician on service provision for hypothetical patients, real patients' health outside the lab is affected by your choices. The overall benefits accruing to patients will be converted into Euro and donated to the charity 'Per Mano onlus', https://permanoonlus.wixsite.com/per-mano-onlus. To verify that the monetary amount corresponding to the sum of the patients' benefits in a session is actually transferred, one of the subjects will be randomly chosen to be a monitor. When the experiment is over, the monitor will verify that one of the experimenters will actually transfer the monetary amount through credit card payment on the Per Mano ONLUS website. The money will support the charity assisting people affected by Duchenne Muscular Dystrophy.

Questionnaire

Before starting the experiment, we kindly ask you to answer some simple questions aiming at checking your comprehension of the design of stage I and of the profit generation mechanism.

If you have any question regarding the questionnaire, please raise your hand and one of the experimenters will come to your seat. Stage I will start only when all the participants answer to all questions correctly.

Conclusions

The three experiments saw the participation of 87 physicians in the whole: 52 for the first two experiments and 35 for the last one.

The first experiment investigates the level of coordination of the Italian healthcare sector using a simple, portable and incentive-compatible tool such as the coordination game. Results show that the overall average level of coordination across the experiment is 52%. The regression analyses evidence that coordination increases when physicians do often exchange opinions and share positive feedbacks with colleagues. In addition, the presence of a leader in the medical department, either a primary doctor or a professor outside of the team, facilitates coordination. Furthermore, when physicians have been working longer for the same hospital, the probability that physicians share guidelines and judge them as the most appropriate actions to adopt in the specific clinical case increases. Even though, in recent years Italy distinguished itself for the initiatives aimed at promoting integrated care through the creation of networks and cooperatives between different health professionals (European Union Report, 2017), there is still a lot to do. More generally, collaborations between general practitioners and specialists together with more formal integrated path among professionals belonging to different levels of care organizations should be encouraged to fight lack of coordination. Furthermore, since sharing ideas and feedback is said to increase the level of coordination, teamwork should be incentivized through brainstorming and planning sessions. In this regard, leaders must be trained to be able to manage a team, facilitating collaboration and communication between different members. Moreover, since actions prescribed by national guidelines are judged as very appropriate in 62% of the cases, training courses specifically addressed to the dissemination of guidelines should be introduced with more frequency in order to increase such percentage.

The second experiment, built on the same experiment as in Chapter 1, investigates physicians' adherence to guidelines and their potential use as a nudge policy instrument. Results show that only 8% of the physicians did not want to know national guidelines content. However, only 23% of the subjects decided to change their appropriateness judgment after realizing they were in contrast with guidelines in either one or two vignettes. Overconfidence, perceived guidelines ambiguity and default effect could explain such a low value. Finally, the more physicians consult scientific sources (i.e. signal of open-mindedness), the more willing to accept

suggestions, coming from guidelines reading, and to change their decisions when they are wrong.

Although many physicians comply with guidelines from the very beginning, a fair number of them not only do not recognize them, maybe because of their lack of specialization in the concerned medical field, but also they are reluctant to change their decisions once they realize they were wrong. This result supports the assumption that guidelines have hardly changed physicians' behaviour (Hayward, 1997), either due to their overconfidence or to guidelines inflexibility and inability to include all the multiple factors which come into play in medical decisions.

However, since guidelines dissemination is shown to increase the level of coordination between physicians, hospitals should consider effective programmes to spread their knowledge. To do this, hospitals should consider physicians' willingness to update, encouraged in this experiment by guidelines ease of accessibility. Introducing a newsletter program and providing an alternative learning option to the standard education courses could be a solution (Strasser, 1978).

Finally, in the last experiment I test whether and to which extent the adoption of fee-for-service or salary system can induce physicians to practice patient dumping.. Results show that dumping is more often observed under Salary than under Ffs. However, physicians seem to be insensitive to the introduction of dumping liability under the same incentive mechanism, though it seems to trigger a higher amount of services provided. Older physicians are more likely to take charge of the patient than their younger colleagues. Finally, men physicians are found to more frequently take charge of the patients, due to their risk taking behaviour, and to increase the amount of services to provide. Though more relevant in the American context, this experiment shows that dumping can be viewed with varying perspectives (e.g. hospital physicians believe that patient's pathology falls under the competence of a family practitioner; physicians tend to prioritize serious ill patient if they have to decide which patient to take charge of, and so on). Since older physicians are found to take charge of patients more than their younger colleagues and to significantly increase the amount of services to provide, introducing economics training among more experienced physicians to incentivize cost-conscious care could be a solution. As far as dumping is concerned, policy strategies may vary on the hospital purposes. If the hospital

wants to avoid unnecessary treatments and hospitalizations, reducing expenditures, salary must be adopted, since it incentivizes physicians to refuse not seriously ill patients. If hospital opts for reducing the rate of refused patients, to avoid further legal complications, FFS is the preferred alternative. On the contrary, introducing liability only affects the amount of services provided.

Acknowledgements

Nine years ago my love story with Mediterranean University started, and I hope that this dissertation can represent its crowning touch. It has been a long and often uphill journey, in which obstacles were not lacking. It is specifically in these phases that the presence of my mentor, Professor Finocchiaro Castro, was fundamental since guided me, like only a father, more than a teacher, could do. I extend my largest gratitude to him, not only for in-field training, for his patience and committment, but above all for pushing me to look forward, without looking back, when this was necessary. Two years ago he drove me to the hardest decision of my life, at times I thought I hated him, however now I realize that without my temporary move to England, I would not have ended this journey endowed with the luggage gathered from this experience. It was a very difficult period, sometimes unsustainable for my excessive bond with my family and friends, physically distant in that moments, but made colourful by my guardian angels, Theresa (my guardian angel par excellence), Paola and Alexandra, coming from different parts of the world but united by the desire to see me winning in every circumstance.

Going back to my background at home, I cannot neglect the constant presence of my friends, who fortunately, are a lot, but worth to be individually mentioned because of their essential contribution.

I thank Luana, the sister I never had, my missing part, who accompanies me step by step in all the moments of my life, drying every tear and photographing every smile.

I thank Martina, who full of tireless patience gives me her time, by listening in silence to all my complaints.

I am grateful to Katia, the mirror where I reflect all my doubts, who day by day is able to curb my instinct driving me to the fairer choices.

I thank Ketty, the embodiment of the strength and perseverance, who follows her dreams against the world and encourages me to do the same.

I thank Giuditta, indefatigable and tenacious, for her contagious ambition.

I thank Raffaella for the joy she gives everyday thanks to her bright smile.

I thank Chicca, because I can think of no one more delightful than her.

I thank Giovanna, my lifelong friend, who in her authenticity helps me to see the good sight of things.

I thank Emanuela, selfless and generous as few.

Finally, I am grateful to my family, the bedrock of my life. I thank my mother who in sickness gave me strength and with prayer conveyed to me faith. I thank my father who taught me that the family makes you better and that to protect it you have to be willing to do whatever.

Last but not the least, I thank Leo to whom this work is dedicated, for following me in each part of it, for not letting me miss his support, also when my choices seemed to go against the tide, but above all, for being a constant presence in my life. It is thanks to him that I have succeeded in finishing this work, for believing in me more than I believe in myself.