The kindergarten dilemma and the impact of management control tools.
A contribution to a behavioral theory of management

A thesis presented
by
Federico Vola
to
The Class of Social Sciences
for the degree of
Doctor of Philosophy
in the subject of
International PhD in Management: Innovation, Sustainability and Health care

Tutor and Supervisor: Prof.ssa Sabina Nuti

Scuola Superiore Sant'Anna
A.Y. 2014-2015
“Se tutto è sotto controllo, vuol dire che stiamo andando troppo piano”

*Mario Andretti, former Formula 1 pilot*
TABLE OF CONTENTS

TABLE OF CONTENTS............................................................................................................................... 5
EXECUTIVE SUMMARY ............................................................................................................................. 7
INTRODUCTION........................................................................................................................................... 9
  § 1. Overview...............................................................................................................................................11
  § 2. The object of the research project ....................................................................................................11
  § 3. The scientific approach.......................................................................................................................15
  § 4. The scale................................................................................................................................................18
CHAPTER ONE ............................................................................................................................................21
  § 1. Background...........................................................................................................................................21
    § 1.1. The selection of the indicators...................................................................................................23
    § 1.2. Fixing the appropriate targets ....................................................................................................25
  § 2. Research design and research questions ..........................................................................................29
  § 3. Snapshot of the PhD papers..............................................................................................................33
CHAPTER TWO - Priorities and targets: a methodology to support the policy-making process in health care .........................................................................................................................................................37
  Abstract ........................................................................................................................................................38
  Keywords .....................................................................................................................................................38
  § 1. Background...........................................................................................................................................39
  § 2. The method proposed ........................................................................................................................47
  § 3. The use of the methodology...............................................................................................................54
  § 4. Discussion and conclusions...............................................................................................................57
CHAPTER THREE - How to set challenging goals and conduct fair evaluation in regional public health systems. Insights from Valencia and Tuscany Regions .................................................................61
  Abstract ........................................................................................................................................................62
  Keywords .....................................................................................................................................................62
  Acknowledgements .....................................................................................................................................62
  § 1. Introduction .........................................................................................................................................64
  § 2. Background...........................................................................................................................................65
  § 3. The technical framework for setting targets and assessing performance.......................................68
    § 3.1. Method for setting challenging goals ........................................................................................69
    § 3.2. Method for performance evaluation............................................................................................72
  § 4. The two regional cases.......................................................................................................................74
CHAPTER FOUR - Making governance work in the health care sector: evidence from the Italian “natural experiment”

Abstract ........................................................................................................................................................ 86
Keywords ..................................................................................................................................................... 86

§ 1. Introduction ......................................................................................................................................... 87

§ 2. The governance systems adopted by Italian Regions in the health care sector......................... 90
  § 2.1. The network experience ............................................................................................................. 94

§ 3. IRPES utilization as a governance tool .......................................................................................... 101

§ 4. Methodology ...................................................................................................................................... 104

§ 5. Discussion .......................................................................................................................................... 110

§ 6. Conclusions ........................................................................................................................................ 114

Footnotes ................................................................................................................................................... 116

CHAPTER FIVE - Conclusions ................................................................................................................ 119
  Concluding comments ............................................................................................................................. 120

References .................................................................................................................................................. 125

ACKNOWLEDGMENTS ......................................................................................................................... 143
EXECUTIVE SUMMARY

This doctoral research is a compilation of three papers that jointly inquire the design and the behavioral effects of the introduction of specific managerial tools in the Italian health care sector. The Introduction outlines the theoretical framework, by defining the object of the research (management control tools) and the perspective that has been adopted to inquire it (the behavioral approach to management studies). Chapter One defines the research questions and outlines the research design. Chapter Two responds to the first research question, by investigating the potential unexpected consequences brought by the selection of the wrong number of indicators in a performance measurement system and it describes a new methodology to help forestall those risks. Chapter Three deals with the second research question, by identifying the behavioral implications of the goal-setting procedure and by illustrating two case studies, which offer useful insights in designing an algorithm to set the appropriate targets for the organization's agents/units. Chapter Four investigates the association between the adoption of different management control systems and the different performances achieved by the Italian Regions in the health care sector. Final remarks and references are presented in the Conclusions.
INTRODUCTION

There is a paper, recently written by two Israeli scholars – Uri Gneezy and Aldo Rustichini – that opens with these words:

Suppose you are the manager of a day-care center for young children. The center is scheduled to operate every day until four in the afternoon, when the parents are supposed to come and collect their children. Quite frequently, however, parents arrive late, and force you to stay after working hours. You have considered a few alternatives in order to reduce the frequency of this behavior. A natural option is to introduce a fine: every time a parent comes late, she will have to pay a fine. Will that reduce the number of parents who come late?

(Gneezy and Rustichini 2000, 1)

The assumptions behind the idea of *homo economicus* would probably suggest that a fine would induce rational agents to behave in order to avoid it, and literature in psychology agrees on a general finding: when negative consequences are imposed on a behavior, they will produce a reduction of that particular response.

Unfortunately, the parents in Gneezy and Rustichini’s study had not studied classical economics nor psychology and came up violating the prediction, by behaving in the opposite way: the number of them who came late increased. Parents simply started considering the fine as the price of an add-on service (“late pickup”) and felt justified being late.
This study is considered a clear example of the so-called “crowding-out effect”, i.e. one of the responses that explicitly contradict the classical models of human behavior. These unexpected responses are the focus of a strand of contemporary social and cognitive sciences that inquire the complex reactions people might have in front of different kinds of stimuli.

Investigating how unpredictable human behaviors can be is probably one of the most fascinating issues social sciences in general - and management sciences, in particular – are asked to inquire. That is why my research tried to approach this exciting topic, replacing the object beneath the lens: health care organizations instead of kindergartens.
§ 1. Overview

This doctoral research is a compilation of three papers that jointly inquire the design and the effects of the introduction of specific managerial tools in the Italian health care sector. In particular, this thesis summarizes a research path aimed at studying the behavioral implications of selected management control tools in driving the improvement of the Italian regional health care systems.

Any research project is generally the combination of an object of study and a specific approach to inquire it; with particular reference to social sciences, it is moreover necessary to explicit the scale of the research. This introductive chapter is aimed at defining these three elements.

§ 2. The object of the research project

Regarding the first element, the focus of this dissertation is on the responses agents develop in front of the adoption of management tools in health care organizations. In particular, I have inquired the potential impacts of the implementation of a selection of them, characterized by a quantitative nature: the regular collection of key performance measures, the systematic benchmarking of the results, and the assignment of goals to the organization’s agents. Moreover, the research relies on the study of the behavioral implications of the quantitative management control tools to inquire how they can be fine-tuned in order to correctly steer the agents’ conduct.

As is often the case in social sciences, the definition of the central issue might be controversial. The search of the ultimate definition of the object of my research - management control tools - is inevitably inconclusive, as they continuously evolve with the
management practice; therefore, their features cannot be crystallized in a stable definition (Fisher 1995; Malmi and Brown 2008).

For the sake of this research, I have referred to them as those instruments and practices that traditionally constitute the elements of the management control systems (MCSs) (Demartini 2014).

The term “management control” was given many of its current connotations by Robert Newton Anthony in his seminal theoretical study, in which he distinguished management control from both strategic planning and operational control (R. N. Anthony 1965; R. N. Anthony et al. 2014). Management control was defined as “the process by which managers assure that resources are obtained and used effectively and efficiently in the accomplishment of the organization's objectives” (R. N. Anthony 1965, 17). Since then, numerous definitions of “management control” and of “management control systems” were proposed. Among the others, Ernest Anthony Lowe put forward a wider definition:

A system of organizational information seeking and gathering, accountability, and feedback designed to ensure that the enterprise adapts to changes in its substantial environment and that the work behavior of its employees is measured by reference to a set of operational sub-goals (which conform with overall objectives) so that the discrepancy between the two can be reconciled and corrected for. (Lowe 1971, 5; see also Otley 1994; Lowe and Machin 1983)

Michael Hanzlick offers a review of the definitions of “management control systems” (Hanzlick 2015, 176–177; see also Carenys 2012). Three elements are common to almost all the definitions: a) management control systems are a coherent collection of different
techniques/tools/mechanisms/practices/devices, b) to ensure that actions are in accordance with the firm’s strategy, c) by steering employees’ behaviors.

With reference to the first element, control management tools can therefore be defined as “all the devices managers use to ensure that the behaviors and decisions of their employees are consistent with the organization’s objectives and strategies” (Merchant and Van der Stede 2007, 5). Anthony and his colleagues state that typical elements/tools/techniques/devices of the control systems include performance measurement, evaluation and reward, strategic planning, budgeting, resource allocation, responsibility center allocation, transfer pricing (R. N. Anthony et al. 2014). Similarly, Mintzberg recalls the elements of the performance control process: performance measurement, objectives, budgets, operating plans (Mintzberg 1983, chap. 4). Anthony further describes those management control elements by identifying their roles: detecting what is happening in the processes being controlled, assessing the significance of what is actually happening by comparing it with some standards or expectations of what should happen, altering behavior if the assessor indicates the need to do so, communicating and transmitting information, predicting the effects of the changes made by the effector (R. N. Anthony et al. 2014; Chenhall 2003). I will avoid using the

---

1 Management control systems are therefore “the deliberate combination or “package” of tools and techniques by which management controls organizational activities” (R. N. Anthony et al. 2014). See also Malmi and Brown 2008.

2 “Management control systems consist of the various ways in which the organization’s top management team attempt to enhance the organization’s performance in line with strategic objectives” (R. N. Anthony et al. 2014). “The purpose of a unified management control system is to ensure that actions are in accordance with the firm’s plans to achieve its objectives” (Lowe 1971, 1). Robert Simons views MCSs as the means used by senior managers to successfully implement their intended strategies (Simons 1995).

3 Management controls are "behavior-influencing devices" (Merchant and Van der Stede 2007, 12). Anthony maintains that "management control systems are tools to implement strategies. (...) A continuing concern in the design of control systems should be whether the behavior induced by the system is the one called for by the strategy" (R. N. Anthony et al. 2014, 49). Merchant and Van der Stede propose a similar definition: “management control includes all the devices (...) managers use to ensure that the behaviors and decisions of their employees are consistent with the organization’s objectives and strategies” (Merchant and Van der Stede 2007, 5; a similar definition is proposed by Flamholtz, Das, and Tsui 1985, 36). A management control system comprises “a combination of control mechanisms designed and implemented by management to increase the probability that organizational actors will behave in ways consistent with the objectives of the dominant organizational coalition” (Abernethy and Chua 1996, 573).
expression “control mechanisms”, as they mainly refer to the classical William Ouchi’s taxonomy of market mechanisms, bureaucratic mechanisms and social/clan mechanisms (Ouchi 1979), and the expression “coordinating mechanisms”, as Mintzberg uses this expression to refer to something different: mutual adjustment, direct supervision, standardization of work processes, standardization of work outputs, and standardization of worker skills (Mintzberg 1983).

Merchant and Van der Stede distinguish controls based on the results produced (results controls), the actions taken (action controls), or the types of people employed and their shared norms and values (personnel and cultural controls) (Merchant and Van der Stede 2007). The focus of my research is on quantitative results controls. The reason is that management of health care organizations needs to steer the partly autonomous activity of professionals, whose primary loyalty it to the profession, rather than to the organization (Mintzberg 1979a). Departmental managers typically are professionals whose management function is only part-time and whose technical expertise grant them professional autonomy (R. N. Anthony et al. 2014, 680). Results controls are dominant as a means of controlling the behaviors of professional employees, in decentralized forms of organizations with largely autonomous responsibility centers (Mintzberg 1979a; Merchant and Van der Stede 2007).

According to Merchant and Van der Stede, results control requires performance measures and evaluation, and the provision of (monetary and nonmonetary) incentives (Merchant and Van der Stede 2007, xiii): those are the two control tools inquired in Chapter Two and in Chapter Three of this dissertation.

---

4 Michael Hammer explicitly defines professional as “someone who is responsible for achieving a result rather for performing a task” (Hammer 1996).
§ 3. The scientific approach

Regarding the second point of the research project – the scientific approach towards the object of study – I reckon that the twentieth century has witnessed a profound disenchant in many areas of the social sciences. With regards to the epistemological foundations of the social sciences, the radical critique of the concept of rationality introduced by Herbert Alexander Simon destabilized the pillars of the positivistic approach (Simon 1957; Simon 1960; Simon 1961). The confutation of the concept of the “rational agent” as the basic element of scientific research has revolutionized any area of social sciences: political science, psychology, sociology, pedagogy, philosophy, economics... The effects of psychological, social, cognitive, and emotional factors on the choices of human agents have been widely studied by a plethora of scholars. The reflection about how bounded rationality drives people’s behavior still feeds copious, unresolved scientific debates.

With regard to management studies, the concept of management control has been investigated according to two complementary perspectives: a “formal” one and a “psycho/sociological” one (Carenys 2012). The majority of contributions focus on the mechanistic and formal techniques related to control systems, with a strong interest in accounting information systems and more specifically on budgeting procedures. The dominant philosophy underlying the use of management control systems is the notion of managerial cybernetics (Beer 1966; Thompson 1969; Reeves and Woodward 1970; Kloot 1997). It envisages that in the managerial process central is the notion of feedback. This model is clearly based on phenomena in the field of physical and biological sciences, where sciences are kept in control via negative feedback mechanisms (Forrester 1968).

However, as Peter Smith states, “when dealing with human organizations, the system one is trying to control is intelligent. The humans within the organization can anticipate the actions
of the controller and – if it is to their advantage – may take action to frustrate the wishes of the controller” (Smith 1995, 280). All the feedback mechanisms can be flawed if the management control system is not able to direct the actors towards the organization’s aims. The solution relies on the capacity of management control tools to drive the agents’ behavior by reframing the feedback mechanism itself, in order to avoid some of the potential shortcomings. To achieve this result, their design should carefully consider the complexity of human responses to institutional stimuli. For this reason, attention is increasingly devoted to what managerial studies might learn from developments in cognitive studies.

Therefore, since the 1960s the management control concept started being enriched by the incorporation of different approaches (the “psycho/sociological” ones), “in which people’s passive and rational behavior was substituted by a greater consideration of the motivational factors that influence behavior and it began to be accepted that the crucial aspects for the design and implementation of a control system were not limited solely to formal ones” (Carenys 2012, 9). The classical studies of management control systems – mainly focused on accounting techniques – started being complemented by (mostly North American) researches based on psychological, sociological and anthropological theories, focused on the behavioral implications of the control techniques (Macinati 2012, chap. 5; Birnberg 2011; Caplan 1971; Carenys 2012). The inclusion of the “human dimension” within the management studies refers to the focus on the cognitive and behavioral factors that might influence the behavior of the agents of an organization, by leveraging psychological and sociological discoveries to study people’s reactions to managerial choices and to design appropriate instruments (Shields 2009).

As a matter of fact, Herbert Simon’s intuitions explicitly relied upon a managerial text – The Functions of the Executive, by Chester I. Barnard (Barnard 1956) – and since then many management studies have been relying upon behavioristic discoveries. In particular, psychological and sociological theories have been used by the so-called “behavioral
management accounting” discipline to inquire how management and control techniques interact with individual motivation and social interactions. Pioneering studies were conducted by Chris Argyris, Andrew Stedry, Anthony Hopwood, Geert Hofstede, Richard M. Cyert and James G. March (Argyris 1957; Argyris 1964; Stedry 1960; Hopwood 1972; Argyris 1952; Hopwood 1974; Hopwood 1973; Hofstede 1968; Cyert and March 1963). Those studies paved the way for a new interest of managerial scholars towards psychological theories – signally, motivational psychology, cognitive psychology and social psychology (Macinati 2012) – and sociological theories of organizational control - functionalism, general system theory (GST) approach, contingency theory, theory of bureaucracy (Demartini 2014).

Following research about the interaction of psychological/sociological aspects and management control techniques mainly focused on motivational effects and on informative ones (Macinati 2012, 87). Regarding the first, studies have mainly investigated how individual psychological profiles might influence the effectiveness of management control techniques (Frederickson and Waller 2005; Sprinkle, Williamson, and Upton 2008; Rankin, Schwartz, and Young 2008; Riahi-Belkaoui 2002; Luft and Shields 2009; Frederickson 1992; Hannan, Krishnan, and Newman 2008; Towry 2003; Bandiera, Barankay, and Rasul 2005; Stedry 1960; Stedry and Kay 1966; Becker and Green 1962).

Regarding the second issue (informative effects), researchers have studied how managers and agents understand and reinterpret information, according to their cognitive characteristics (Stanovich and West 1999; Stanovich and West 2000; Schwartz et al. 2002; Scott and Bruce 1995; Driver and Mock 1975; Prakash and Rappaport 1977; Macintosh 1981).

A related but different stream of research has investigated how to design an efficient control system, taking into account the different effects and cultural relationships of the organization and its members and considering culture as an internal variable of the organization (Ouchi 1979; J. Kerr and Slocum 1987; Inzerilli and Rosen 1983; Wilkins and Ouchi 1983; Schein 1984; Allaire and Firsirotu 1984). A slightly different approach is centered on analyzing
culture as a broader variable, which goes beyond the organization environment, introducing
the nation’s culture as one of the conditioning elements of control systems (Inzerilli and

Despite those contributions, some scholars still believe that the human dimension of the
management control systems is widely reckoned but not appropriately inquired yet (Macinati
2012, 13) and probably we cannot easily identify yet a “behavioral management theory”,
intended as a cohesive discipline that coherently apply the recent developments in cognitive
and behavioral research to management studies and practices. The thesis I here present
intends to contribute to the reflections about the behavioral reactions that the adoption of
common managerial tools may stimulate, by investigating case studies and performance data
related to the health care sector. The aim of this research is two-fold: it firstly inquires the
reactions agents adopt in responding to the use of specific managerial instruments, whose
goal is exactly steering the agents’ conduct; secondly, it investigates how quantitative
management control tools’ design can be improved according to the discoveries in the
psychological and anthropological disciplines.

§ 4. The scale

The research I here present investigates the design and the effects of the introduction of
specific management control tools by regional administrations (the Italian regions plus a Spanish
case study, Valencia).

The focus on the regional scale is explained by the potential it offers for research on
governance and on managerial studies. During the last 25 years, regions have been at the
intersection of two different but related processes. On the one side, they have experienced
the tendency toward the introduction in the public administration of management control
tools originally developed in the private sector, according to the New Public Management paradigm,\(^5\) at first, and to the Public Governance one, lately (Hood 1991; Hood 1995; Lapsley 1999; Osborne 2009). On the other side, they have been subject to a process of decentralization of powers, from European national states to their regions (Saltman and Bankauskaite 2006; Saltman, Bankauskaite, and Vrangbaek 2007). Since the early 1990s, Italian regions have been transferred administrative, fiscal and managerial autonomy, especially in the health care sector (Toth 2014; Fattore 1999; France, Taroni, and Donatini 2005). During the last 25 years, each Italian region has fine-tuned its governance organization according to the local context and needs (France and Taroni 2005; Pavolini and Vicarelli 2012). With regard to the health care sector, the central level - represented by both the Ministry of Health and the Ministry of Finance - ensures whether the Regions keep their health care expenditures within their budget and guarantee the essential levels of care. Conversely, Regions are in charge of organizing health care services: they define their own Regional health plans, coordinate the strategies of the Regional Health Authorities and allocate the budget within their systems. Since the 2000s, the health care budget has been allocated among the Regions on the basis of a per capita share, partially adjusted by the age distribution of the population. At the same time, the Regions have become more fiscally autonomous and more financially responsible (Ferrario and Zanardi 2011; Ferrè, Cuccurullo, and Lega 2012). Italian Regions now have the political, administrative and financial responsibility for the provision of health care to their residents.

The consequence is that there are now 20 Regional Health Care Systems (RHSs) in Italy with different governance models and management tools (Formez 2007; Censis 2008; Tediosi, Gabriele, and Longo 2009; Carinci et al. 2012; Mapelli 2012).

This implies that the heterogeneity of the Italian regional landscape offers a unique research ground: it can be studied as a natural experiment to test some of the main theories about

---

\(^5\) Indeed, some authors contest the fact that NPM and PG can be defined as “paradigms” and would prefer the expression “set of ideas” (Gow and Dufour 2000; Fattore, Dubois, and Lapenta 2012).
regional governance, policy-making and management. This is what the research group of Laboratorio Management e Sanità (MeS Lab) is involved in and what I contributed to during my PhD.

Therefore, chapter one proposes a view of the literature about the behavioral implications related to the goal-setting process (§1) and defines the research questions (§2). I did not develop a comprehensive and systematic framework for decoding, interpreting and understanding the potential reactions of the political and managerial agents; I rather retrieved and combined two well-developed scientific streams. The first is the so-called “goal-setting theory”, that strongly relies on the contributions of Edwin Locke and Gary Latham (Locke and Latham 2013a; Locke and Latham 2002; Locke et al. 1981; Locke and Latham 1984). The second cannot be defined as a “theory” but is represented by the robust scientific production that – at least since the publication of the paper “Dysfunctional Consequences of Performance Measurements” on the first issue of the Administrative Science Quarterly (Ridgway 1956) – investigates the potential distortions any management control tool may induce.

The three papers of the thesis are illustrated in the following chapters (chapter two, chapter three and chapter four). They jointly articulate a coherent response to the research questions specified in chapter one. They actually inquire the potential behavioral reactions to the goal-setting process by investigating a) the selection of the appropriate number of indicators to be considered in the goal-setting process; b) a quantitative method to rigorously define the appropriate targets to be assigned to the single agents/units; c) the effects the adoption of different governance models has had on the performances of the Italian Regional health care systems.

Concluding remarks and references are finally illustrated in chapter five.
CHAPTER ONE

§ 1. Background

At least since the 1990s, reforms based on New Public Management principles have aimed at making the public sector more efficient, effective and accountable (Hood 1991; Hood 1995; Lapsley 1999). One of the measures to achieve these goals has been importing from the private sector management control tools that – although originally designed for profit-seeking companies – are proven to be suitable to public companies too (Kaplan and Norton 1992; Kaplan and Norton 2000; Kaplan and Norton 2001; Kloot and Martin 2000). Beginning in the late 1980s, a broad range of New Public Management approaches has been adapted to health care too (Saltman and Busse 2002) and measuring performance has become a mantra at all levels of government since the 1990s (Radin 2000; World Health Organization 2003; Chang, Lin, and Northcott 2002; Pink et al. 2001; Ballantine, Brignall, and Modell 1998; McLoughlin et al. 2001).

Governance based on planning and control systems with performance indicators (PIs) and targets is a form of indirect control that is currently widely deemed necessary for the governance of any complex system (Carter, Klein, and Day 1992; Bird et al. 2005; Pollitt 1986).

Initially, performance measurement primarily focused on financial issues and it neglected measures of multiple strategic objectives to drive changes (Ghobadian and Ashworth 1994; Polliit and Bouckaert 1995; Guthrie and English 1997; Lorden, Coustasse, and Singh 2008).
Hence, comprehensive multi-dimensional performance measurement frameworks, such as the balanced scorecard, were progressively introduced (Kloot and Martin 2000; Yang and Tung 2006; Arah et al. 2006; Arah et al. 2003; Smith 2002; Kelley and Hurst 2006). A further development was the introduction of systematic benchmarking of health care performance measurement systems at the international, national and local levels (Johnston 2004; NHS Executive 1999; Pink et al. 2001; Kouzmin et al. 1999; Hibbard, Stockard, and Tusler 2003; Nuti et al. 2015; Bevan and Wilson 2013). Benchmarking can actually help managers learn from best practices (Liebfried and McNair 1992) and be used as a mechanism to detect unwarranted variations and encourage their reduction (Arah et al. 2003).

Performance measurement is the precondition to assign goals and targets to single or collective units of the organization. This is reckoned to be one of the preeminent management control tools in contemporary organizations (Ferreira and Otley 2009). The “management by objectives” (MBO) approach has been widely used by managers and studied by scholars since the 1950s (Drucker 1954; Odiorne 1969; Reddin 1971; Traberg 2011; Kaplan and Norton 1992). In its most basic form, MBO is:

A process whereby the superior and subordinate managers of an organization jointly identify common goals, define each employee’s major areas of responsibilities in terms of the results expected to them, and use these measures as guides for operating the unit and assessing the contribution of each of its members.

(Odiorne 1965, 55–56)

The main theoretical contributions on managing organizations through the definition of challenging goals refer to the so-called “goal-setting theory”, mainly developed by Edwin Locke and Gary Latham at the end of the 80s (Locke and Latham 1990; Locke and Latham
Their main assumption is that goals have pervasive influence on employees’ behaviour and organizational performance. The theory states that a specific high goal leads to higher performance than (a) an easy goal, (b) a general goal or an exhortation to “do one’s best,” or (c) no goal setting (Latham, Borgogni, and Petitta 2008; Locke et al. 1981; Lunenburg 2011; Locke and Latham 2006; Locke and Latham 1990; Latham 2007; Mitchell and Daniels 2003).

The target-setting process implies at least two phases: a) selecting the activities and the related indicators that the organizations deem to deserve primary importance (and therefore to be assigned a target); b) once step “a” is performed, the controlling agents need to define the targets of the controlled employees, on the previously selected indicators. These two subprocesses may potentially involve some relevant behavioral issues and have stimulated some interesting researches over the last 50 years.

§ 1.1. The selection of the indicators

Regarding the first issue, managing by objectives requires defining the appropriate number of indicators that need to be considered as priorities and therefore to be assigned an objective. The failure to design a panoptic performance measurement system causes some potential distortions. The choice of a limited set of indicators to evaluate and manage an organization can actually stimulate undesired behaviours on the part of the agents who are controlled by the system. Evaluated actors may react to measurement processes and to management by objectives by adapting their behaviour merely to achieve single targets, with scarce attention to the quality of the service as a whole.
Studies on the potentially distorting effects of control systems are extensive (Birnberg, Turopolec, and Young 1983; Briers and Hirst 1990; Le Grand and Bartlett 1993) and dates back at least to the first issue of the Administrative Science Quarterly, with Ridgway’s paper on the “dysfunctional consequences of performance measurements” (Ridgway 1956).

The selection of the indicators and the consequent target-setting phase are therefore crucial, since they can potentially generate incoherent behaviours in the system and lead the organizations towards undesired goals (Bubbio 1988; Locke and Latham 2013b; Ferreira and Otley 2009; Smith 1995).

Since the 1950s, scholars have been studying dysfunctional consequences of measurement that may be caused by the selection of too a restricted number of indicators: actors might actually attempt to achieve single targets at the cost of significant but unmeasured aspects of performance (Nove 1958; Miller 1993; Kornai 1994; Smith 1995; Heinrich 2002; Hood 2002; Propper and Wilson 2003; Bird et al. 2005; Bevan and Hood 2006; Ridgway 1956; Van Thiel and Leeuw 2002; Hofstede 1981; Eden and Hyndman 1999; Leeuw 2000). Secondly, some contributions have highlighted that public managers do not always choose priorities according to a rational principle, such as performance gaps or isomorphic rationality, but rather according to other heuristics (Rutherford and Meier 2015).

These elements might suggest managers to enlarge the number of indicators selected as targets. The tendency towards output distortions might be offset by devising a large number of measures. However, specification of more than a handful of objectives is almost certainly meaningless, because it is impossible to devise a reward scheme that satisfactorily reflects achievement in more than three/four sections (Hopwood 1974; Kaplan and Norton 1996). The confusion generated by many targets might disorient the actors of the organization who may then behave differently from the priority actions (Van der Stede 2009). Moreover, the cost and data-collection burden would be prohibitive. Clinical staff could be overwhelmed and efforts could be wasted by trying to identify the indicators that really count (Goddard et al. 2002b).

Therefore, the process of management by objectives needs to solve the following dilemma: whether to rely on a limited number of indicators, in order to clearly communicate the organization’s goals to the controlled actors, or to focus on the containment of potential undesired reactions by enlarging the number of indicators, at the expense of clarity. Van Thiel and Leeuw point out the need to find a balance between an expansion of the number of performance measures on the one hand, and a reduction of the measure pressure on the other. The first issue that this thesis is intended to deal with is that both excessive and scarce emphasis on performance indicators can result in dysfunctional behaviours (Van Thiel and Leeuw 2002; Holmstrom and Milgrom 1991).

§ 1.2. Fixing the appropriate targets

Regarding the second issue, identifying the correct targets is widely deemed to be a crucial element of any management control system (Otley 1999; Ferreira and Otley 2009; Stuart et al. 2009; Stringer 2007; Covaleski et al. 2006). Studies relating to the above-mentioned “goal-
setting theory” demonstrate that specific high goals affect choice, effort and persistence, but too challenging targets may be perceived as unachievable and may induce controlled agents disregarding them. Agents might prefer focusing their efforts on those goals that they deem to be reachable, instead of wasting energy on unachievable targets. The target-setting process implies a trade-off: on the one side, managers need to fix goals that are challenging enough to stimulate the controlled agents; on the other side, they should be careful in not exceeding by setting unachievable targets. Setting the wrong performance targets is deemed to be a major cause of failure of management control systems (Ittner and Larcker 2003).

Defining “the right targets” requires facing at least four sub-issues:

a) choosing whether to set homogeneous targets for all the actors or to set individual goals for each of them;

b) choosing whether to consider the agents’ past performances to set their targets;

c) defining the benchmark the actors are aiming at;

d) defining a rigorous procedure to evaluate the level of target attainment by each agent.

The first point might encounter what managerial literature defines the “threshold effect”. This occurs when a minimum and equal threshold is set for all the controlled actors. On the one side, this mechanism puts some intentional pressure on under-performing agents; on the other side, it instills a perverse incentive for all those agents who are already performing over the threshold, by stimulating a regression toward the threshold level (Bird et al. 2005). The threshold mechanism generally penalizes those actors that perform well but still have single criticalities, while it favors mediocre agents, who systematically perform in the threshold range.

Regarding the second issue, previous studies demonstrate that goals have to be set considering the difference between the units (i.e., Local Health Department or professionals) and their starting point (baseline) (Ferreira and Otley 2009; R. Anthony and Young 1999).
Performance incentives had the greatest impact on providers whose performance was lower at baseline (Greene, Hibbard, and Overton 2015). However, considering the baseline might trigger a second trap, which scholars refer to as the “ratchet effect”. This comes from the tendency by the controlling subjects of defining the targets on the basis of the results the controlled actors obtained in the previous years. This might induce the controlled not overcome the assigned targets, in order to reduce the effort they would be required the following years (Litwack 1993; Goddard, Mannion, and Smith 2000b).

Again, policy makers need to face a trade-off: past performance needs to be taken into account in order to fix individual and achievable targets, but – at the same time – targets need to be partially disconnected by their own heritage, in order to forestall the above-mentioned “ratchet effect”.

Regarding the third issue, in laboratory experiments goals are usually set at the 90th percentile, but in field experiments they are often set on the basis of the perceived judgement of “difficult yet attainable” goals (Locke and Latham 2013a). That implies that the definition of what is challenging is set, most of the time, on qualitative basis. The recourse to subjective judgements may be due to the lack of comparisons. Only few studies (in different contexts, ranging from schools to electricity generating plants) describe procedures to set appropriate targets according to rigorous methods, such as data envelopment analysis (Thanssoulis 1999; Athanassopoulos, Lambroukos, and Seiford 1999).

The fourth issue involves the evaluation of the level of target attainment by each agent. The simple degree of achievement of the set targets might be chosen as the rule to determine it, but other factors need to be necessarily considered. In particular, some contextual variables might have affected the degree of achievement itself. This means that some correctives have to be envisaged (Locke and Latham 1990; Ferreira and Otley 2009; Flamholtz, Das, and Tsui 1985).
The introduction of some benchmarking techniques might help tackle the four above-mentioned issues (Young and Smith 2013). The use of benchmarking, particularly the use of external benchmarks, appears to provide a greater degree of legitimacy for targets (Elnathan, Lin, and Young 1996; Spendolini 1992), as already shown by their use in the health sector (Northcott and Llewellyn 2003).

The procedure to define the right targets and the reflection about how benchmarking can help overcome some of its hurdles are the second issue this thesis is aimed at investigating.
§ 2. Research design and research questions

Given the issues and the scientific contributions we have outlined in the previous paragraphs, we believe some questions still need to be properly explored. In particular, scientific research and managerial practice would benefit from a focus on:

a) the analysis of appropriate methodologies and tools to tackle and minimize the potential distortions that an inappropriate selection of indicators in the target-setting procedure might stimulate;

b) the analysis of rigorous methodologies to individuate the appropriate targets;

c) the demonstration that well-designed management control systems are actually associated to better performances.

This leads to the three research questions of this PhD thesis:

<table>
<thead>
<tr>
<th>Research question 1</th>
<th>How can a rigorous methodology be envisaged to select the appropriate number of indicators to be included in the target-setting procedure, in order to minimize potential opportunistic behaviours of the controlled agents?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research question 2</td>
<td>How can managers and administrators define the correct targets for the agents they control, in order to ask for an appropriate effort (both challenging and achievable)?</td>
</tr>
<tr>
<td>Research question 3</td>
<td>Is the adoption of specific management control systems associated to the achievement of different performances by the Italian Regions in the health care sector?</td>
</tr>
</tbody>
</table>
Regarding the first point, despite some theoretical proposals, literature still lacks contributions that might help build rigorous methodologies to select the indicators to be targeted. Goddard and her co-authors propose to solve the dilemma by selecting some ‘headline’ indicators - focused on priority issues that are published and used to hold organizations to account. Other indicators would be used for internal benchmarking purposes only (Goddard et al. 2002a). Peter Smith suggests his solution to the dilemma by envisaging a sort of dynamic performance evaluation system that could adapt the choice of the “right” indicators to the situation: “Clearly, no package can comprehensively cover the entire domain of a public sector organization’s activity. Indeed, even if such a package could now be envisaged, it might not be adequate in the future. In any case, it is usually necessary to monitor unintended side-effects of new programmes. There is no guarantee that such side-effects will be noted by an existing Performance Indicators system, and so there is a constant need to keep the domain covered by the PI system under review” (Smith 1993, 141; Scriven 1973). Bevan and Hood suggest a solution to limit opportunistic behaviours by introducing some uncertainty in target specification (Bevan and Hood 2004). Meyer and Gupta recommend that organizations adopt the so-called “paradoxical model of performance assessment” with multiple, uncorrelated and varying but comparable performance indicators (Meyer and Gupta 1994).

The first paper that composes this thesis (“Priorities and targets: a methodology to support the policy-making process in health care”) contributes to the debate by describing an innovative methodology MeS Lab has recently developed and some Italian Regions adopted in order to detect the priorities of the regional health care systems and select an appropriate number of indicators in the target-setting process accordingly. The paper investigates the principal behavioral distortions that the choice of an incorrect number of indicators might provoke, examines a methodology to prevent them and investigates how several Italian Regions have differently adopted this methodology. It finally discusses how the
characteristics of the methodology help avoid the majority of the potential unintended behaviors outlined before.

Regarding the second point - the analysis of rigorous methodologies to individuate the appropriate targets - we already reported that only very few studies describe some examples of rigorous procedures to set challenging but achievable targets. The second paper we present in this thesis (“Setting challenging and fair targets in regional public health systems. Insights from Valencia and Tuscany Regions”) aims at contributing filling the gap. Once policy-makers have selected an appropriate number of important indicators, the target-setting process requires defining the correct goals for each of them: this second phase is the core of the paper illustrated in chapter three. An impressive number of publications proves the efficacy of setting quantitative and challenging targets to stimulate a desired behavior by the employees, but very few studies investigate the appropriate effort to be required and a methodology to identify correct targets accordingly. By focusing on two case studies – Tuscany and Valencia Regions – the paper explores how these two advanced administrations yearly identifies the targets for their local health authorities. Therefore, the paper offers valuable insights to the literature gap, by critically discussing the solution these two Regions have developed to define appropriate targets through a rigorous procedure.

We finally move to the third issue: the demonstration that well-designed management control systems are actually associated to better performances. According to the different management control systems an administration may adopt, scholars identify five ideal typical governance models (Cromwell et al. 2011; A. Brown et al. 2012; P. Brown et al. 2012; Bevan and Fasolo 2013; Bevan and Wilson 2013):

1. the “trust and altruism” model relies on the perspective that all public servants behave like knights. It was the traditional model applied by the NHS and it does not focus on success and failure; on the contrary, it can lead to rewarding failure and ignoring success;
2. the “choice and competition” model is based on the quasi-market system where patients can choose and the money follows the patients. This model introduces external incentives and patients (or insurance companies) can choose providers on the basis of information on quality;

3. The “hierarchy and targets” model, also known as “command and control”, is based on recourse to external incentives and the strong role of performance management (generally by the central government). It has side effects like high monitoring costs and low acceptance by professionals;

4. the “transparent public ranking” model is based on the lever of reputation. This model has been applied and is known in England as the “naming and shaming” model;

5. the “pay for performance” (P4P) model draws upon economic incentives to direct the managers' behavior.

The third paper (“Making governance work in the health care sector: evidence from the Italian «natural experiments»”), that we expound in chapter four, logically follows the previous ones: once we have structured a management control system, can we test if it actually proves to improve the performance of the contexts where it is applied? Relying on the natural experiment the Italian regional health care systems offer, the paper compares the different performances associated to different governance mechanisms across them. The work demonstrates that the combination of different governance methods (pay for performance, transparent public ranking, hierarchy and targets) is associated to more balanced and sustained improvements. Secondly, it supports the hypothesis that benchmarking is the precondition to grant continuous improvements, especially for already highly performing systems.
§ 3. Snapshot of the PhD papers

In the following table, details of the three papers are reported in a snapshot, by summarizing:

- the research questions of each paper;

- a brief summary of the background/theory the papers are based on;

- data/cases used;

- the methodology;

- the contribution to the literature.
## Priorities and targets: a methodology to support the policy-making process in health care

Setting challenging and fair targets in regional public health systems. Insights from Valencia and Tuscany Regions

### Authors


### Journal

- **Public Money & Management**
- **Health Policy**
- **Health Economics, Policy and Law** (I.F. 1.696)

### Status

- Under review
- Submitted
- In press (published online)

### Research question

- How can managers and administrators select the correct indicators to be included in the target-setting process?
- How can managers and administrators define the correct targets of the agents they control, in order to ask for an appropriate effort (both challenging and achievable)?
- Are specific governance models associated to different performances in the health care sector?

### Literature background

- Researches about performance paradox and performance systems distortions (Ridgway 1956; Van Thiel and Leeuw 2002; Meyer and Gupta 1994)
- Goal-setting theory (Locke and Latham 2006; Locke 1968; Locke and Latham 1990; Locke and Latham 2002; Locke et al. 1981; Locke and Latham 2013a; Locke and Latham 2013b)
- Literature inquiring behavioral responses associated to different models of governance (Bevan and Fasolo 2013; Bevan and Hood 2006; Mannion and Davies 2008)
<table>
<thead>
<tr>
<th>Data/cases</th>
<th>Methodology</th>
<th>Contribution to literature</th>
</tr>
</thead>
<tbody>
<tr>
<td>The paper studies the algorithm MeS Lab designed to assist regional policy-makers in the health care sector.</td>
<td>Literature analysis and throughout description of the methodology and of its adoption by a few regional administrations</td>
<td>The paper contributes to the current debate about the potential adverse reactions management control systems may stimulate. It presents a methodology to limit the potential distortions implied by the selection of an inappropriate number of indicators to be included in the target-setting process.</td>
</tr>
<tr>
<td>The paper compares two cases: how Tuscany and Valencia Regions administrations define targets for their Health Authorities.</td>
<td>Literature analysis and comparison of the case studies</td>
<td>The paper proposes two original solutions to the unresolved issue of setting the correct (both challenging and achievable) targets to the agents of any organization. The two cases the study reports offer innovative insights about rigorous and quantitative approaches to the target-setting procedure.</td>
</tr>
<tr>
<td>The paper relies on a set of 15 indicators, measured in 2007 and 2012. Those indicators were available for all the Italian Regions.</td>
<td>Literature analysis and qualitative comments about the association of different regional governance models to different performances in the health care sector</td>
<td>The administrative heterogeneity of the regional Italian health care systems offered the opportunity to test the association between different governance models and different performances. The paper demonstrated that mixed models are associated to higher performances and improvements, especially if combined with benchmarking tools.</td>
</tr>
</tbody>
</table>
CHAPTER TWO - Priorities and targets: a methodology to support the policy-making process in health care
Abstract

Management by objectives requires selecting the appropriate number of indicators to measure objectives and, among these, defining the high priority ones. Failing to address the two issues may result in the so-called performance paradox. This paper describes the algorithm applied in the health sector, in most Italian regions that have adopted the same performance evaluation system, to detect priority indicators for the target-setting phase, with the aim of overcoming the performance paradox.

Keywords

Performance paradox, management by objectives, priority setting, performance evaluation
§ 1. Background

Governance based on planning and control systems with performance indicators (PIs) and targets is a form of indirect control that is widely deemed necessary for the governance of any complex system (Beer 1966; Carter, Klein, and Day 1992; Bird et al. 2005; Pollitt 1986). However, managing by objectives requires tackling at least two issues: defining the appropriate number of indicators that work as objectives and choosing a rigorous principle to define which indicators should be considered as high priorities.

Regarding the first point, the process of governance by PIs and targets relies on a set of measures that should represent social complex phenomena. This translation process – from social phenomena to quantitative measures – requires careful design.

Firstly, only a part of the phenomena can be represented by quantitative figures, because indicators portray the “measurable” part of the observed object (for example, “speed” represents the measurable element of the more complex process of “driving”).

Secondly, the measurement process has different degrees of accuracy. Inaccuracies in the measurement process may bring about false positives and false negatives (in the previous example, an inaccurate speedometer may result in wrong or in missed warnings).

The process of governance by PIs and targets relies on the assumption that available indicators are both representative and accurate.

The so-called “synecdoche” principle assumes that the measurable part is able to satisfy both conditions (representativeness and accuracy), by accurately representing the whole object (Carter, Klein, and Day 1992). However, measurement systems almost never comply with the “synecdoche” principle, because the set of selected indicators is rarely able to portray the complexity of the evaluated processes (Bevan and Hood 2006).
The failure to design a panoptic performance measurement system causes some potential distortions. The choice of a limited set of indicators to evaluate and manage an organization can actually stimulate undesired behaviours on the part of the agents who are controlled by the system. Evaluated actors may react to measurement processes and to management by objectives by adapting their behaviour merely to achieve single targets, with scarce attention to the quality of the service as a whole.

Studies on of the potentially distorting effects of control systems are extensive (Birnberg, Turopoliec, and Young 1983; Briers and Hirst 1990; Le Grand and Bartlett 1993). The basic point is that, in order to make the planning and control scheme meaningful, the measures used should be consistent with the performance targets – which should, in turn, be representations of organizational objectives – so that the behavioural responses of the agents are likely to be matched to the requirements of the principal (Smith 1995). The selection of the indicators and the consequent target-setting phase are therefore crucial, since they can potentially generate incoherent behaviours in the system and lead the organizations towards undesired goals (Bubbio 1988; Locke and Latham 2013b; Ferreira and Otley 2009).

Since the 1950s, scholars have been studying dysfunctional consequences of measurement; in particular, many of them have inquired the so-called “output distortions”, i.e. the attempts to achieve single targets at the cost of significant but unmeasured aspects of performance (Nove 1958; Miller 1993; Kornai 1994; Smith 1995; Heinrich 2002; Hood 2002; Propper and Wilson 2003; Bird et al. 2005; Bevan and Hood 2006; Ridgway 1956).

In 1991, Bouckaert and Balk wrote about “13 diseases” of public productivity measurement, which resulted from wrong assumptions underlying measurement, measurement errors, and problems concerning the content, position and amount of measures (Bouckaert and Balk 1991). Similarly, Smith wrote about eight potential and unintended behavioural consequences of performance data use (Smith 1995). Indeed, they are all result of some violations of the “synecdoche” principle by the performance evaluation system, which causes an inconsistency
between the goals of the agents and the actual goals of the principal: “tunnel vision”, “suboptimization”, “myopia”, “measure fixation”, “misrepresentation”, “misinterpretation”, “gaming”, “ossification” (Goddard et al. 2002a; NHS Executive 1999; Goddard, Mannion, and Smith 2000a; NHS Executive 2000; Smith 1993; Smith 1995; Hopwood 1973). In particular, “tunnel vision” happens when management emphasizes phenomena quantified in the performance measurement scheme, at the expense of unquantified or not considered aspects of the performance. “Suboptimization” occurs when key organizational objectives are compromised by the lack of congruence with targets for individual benefit offices (Dearden 1960; Dearden 1961; Henderson and Dearden 1966; Beyer 1963; Jasinski 1956; Ridgway 1956). “Measure fixation” occurs when the pursuit of success as measured, rather than as intended, becomes the main focus (S. Kerr 1975). Smith defines it as “the emphasis on [single] measures of success rather than [on] the underlying objective” (Smith 1995, 290). “Misrepresentation” occurs when managers deliberately manipulate data under their control to show their organization’s performance in the most advantageous light. This risk is more common when indicators are not measured in benchmarking and the outputs measured are self-referred (Bardsley, Coles, and Jenkins 1987; Flynn 1986).

Since the 2000s, many scholars have started investigating the reasons behind the potential failures of results-based management, with a particular focus on the public sector (Van Thiel and Leeuw 2002; Hofstede 1981; Eden and Hyndman 1999). They have confirmed the hypothesis that reliance on quantitative measures to monitor social phenomena can lead to several unintended consequences that may not only invalidate conclusions on public sector performance, but also negatively influence performance (Leeuw 2000; Organisation for Economic Co-operation and Development 1996). Smith highlights that the problem of objectives that cannot be quantified is particularly acute in the public sector for three reasons (Smith 1995):
a) stakeholders usually hold a large number of diverse objectives with respect to a public sector organization;

b) many important objectives in the public sector simply defy adequate quantification;

c) it is a common feature of public sector services that the ramifications of their activities extend well beyond the immediate target of service delivery.

So far, efforts to meet the old-style NHS efficiency targets have often been directed at increasing measured activity within the hospital setting, at the expense of improvements that might have been achieved by increasing care outside the hospital (Goddard et al. 2002a).

We previously mentioned that managing by objectives requires tackling two issues: first, defining the appropriate number of indicators that work as objectives; second, choosing a correct principle to define which indicators should be considered as high priorities. Regarding the second point, very recent studies have found that public managers prioritize competing performance goals by political control mechanisms (Rutherford and Meier 2015). This means that the determinants of managerial priorities are not always a rational principle, such as performance gaps or isomorphic rationality, but rather bureaucratic control/hierarchical pressure. With the latter expression, Rutherford and Meier indicate that public managers (in their case, university presidents) choose priorities according to the perceived priorities of their principals and not according to a rigorous and rational algorithm.

The interaction of the above-mentioned limitations gives rise to the so-called “performance paradox”, which refers to a weak correlation between performance indicators and performance itself (Meyer and Gupta 1994; Meyer and O'Shaughnessy 1993). A performance paradox is the result of a discrepancy between the policy objectives set by politicians and the goals of executive agents (Smith 1995). Certain characteristics of the public sector can, indeed, increase the chance of a performance paradox occurring (Fountain 2003; Torenvlied 2000; Le Grand 1991). The risk is that the performance measurement system could degenerate into a formal ceremony that has little impact on the behaviour of
managers and does not act to improve the efficiency, effectiveness and accountability of an organization (Sharifi and Bovaird 1995).

In countries that have adopted a universal health care coverage and that are usually financed by general taxation on a capitation basis (Morton and Cornwell 2009) profit is not significant while the concept of “value for money” (i.e., what results are obtained with the available resources) becomes much more relevant than in the other industries. Therefore, the performance evaluation systems require identifying a wide range of indicators that can encompass the complexity and diversity of the results. Indicators should focus not only on the outcomes but also on the determiners of the outcomes. Thus, process, efficiency and productivity indicators should be included in the performance evaluation systems (Arah et al. 2003; Arah et al. 2006; Murray and Frenk 2000). Therefore, because of the high complexity of the health care sector, a key issue is the choice of the appropriate number of indicators to use in the planning and control system. Specification of more than a handful of objectives is almost certainly meaningless, because it is impossible to devise a managerial reward scheme that satisfactorily reflects achievement in more than three/four sections (Hopwood 1974). The confusion generated by many targets might disorient the actors of the organization who may then behave differently from the priority actions (Van der Stede 2009).

On the other hand, the fewer the number of performance indicators, the more difficult it becomes to obtain an accurate performance report (Meyer and Gupta 1994). Selecting only a few indicators for a small section of total performance can cause the occurrence of a performance paradox. This effect is reinforced when indicators do not change over time (Van Thiel and Leeuw 2002, 276).

Increasing the number of PIs may therefore be an efficient means to tackle the synecdoche problem and its distortions. For instance, the NHS performance assessment framework seeks to manage potential distortions by covering six areas, which can give a balanced view of health care performance. However, there is clearly a long way to go: the English
Government’s consultation document suggested 470 possible indicators (Goddard et al. 2002a). The tendency towards output distortions might be offset by devising a large number of measures. However, the cost and data-collection burden would be prohibitive. Clinical staff could be overwhelmed and efforts could be wasted by trying to identify the indicators that really count (Goddard et al. 2002b).

Therefore, the process of management by objectives needs to solve the following dilemma: whether to rely on a limited number of indicators, in order to clearly communicate the organization’s goals to the controlled actors, or to focus on the containment of the paradox problem by enlarging the number of indicators, at the expense of clarity. Van Thiel and Leeuw point out the need to find a balance between an expansion of the number of performance measures on the one hand, and a reduction of the measure pressure on the other. Both excessive and scarce emphasis on performance indicators can result in a performance paradox (Van Thiel and Leeuw 2002).

The target-setting phase should, therefore, include a number of targets large enough to limit the synecdoche and “output distortion” problems and, at the same time, sufficiently small to keep the controlled actors focused on the priority issues.

Goddard proposes to solve the dilemma by selecting some ‘headline’ indicators - focused on priority issues that are published and used to hold organizations to account. Other indicators would be used for internal benchmarking purposes only (Goddard et al. 2002a).

Peter Smith suggests his solution to the dilemma by envisaging a sort of dynamic performance evaluation system that could adapt the choice of the “right” indicators to the situation: “Clearly, no package can comprehensively cover the entire domain of a public sector organization’s activity. Indeed, even if such a package could now be envisaged, it might not be adequate in the future. In any case, it is usually necessary to monitor unintended side-effects of new programmes. There is no guarantee that such side-effects will be noted by an
existing Performance Indicators system, and so there is a constant need to keep the domain covered by the PI system under review” (Smith 1993, 141; Scriven 1973).

The methodology we describe in the following paragraphs reflects both proposals, by combining the completeness of a wide PES with a rigorous and iterative selection of a few headline indicators. The algorithm we suggest aims at detecting regional priorities and selecting a subset of indicators for the target-setting phase, so that they can represent the complexity of health care systems and, at the same time, not disorient the actors of the systems with too many indicators?

This paper presents a methodology developed to support Italian regional policy makers in the target-setting phase. This methodology is being experimented by the majority of the Italian regions (the Autonomous province of Bolzano, the Autonomous province of Trento, Basilicata, Calabria, Emilia Romagna, Friuli-Venezia Giulia, Liguria, Marche, Lazio, Lombardy, Sardinia, Toscana, Umbria and Veneto) that have adopted the same Performance Evaluation System (PES). This system provides the regions with about 200 indicators that mirror the multiple facets of the health care system (see for details Nuti et al. 2015). The Laboratorio MeS developed the performance evaluation framework as an independent research unit and coordinates and manages information sharing and data acquisition. The 14 regions in the network have agreed on the criteria to calculate the indicators. Each region is responsible for processing its own data in order to increase the awareness and expertise of regional managers and their staff. The PES measures the results with a multidimensional approach in quantitative terms and then assesses the performance of 100 indicators out of 200: excellent, good, sufficient, poor, or very poor. These five evaluation tiers are associated with different colours, from dark green (excellent performance), to red (poor performance). Regions use the same reference standards for their evaluation, based on scientific literature, national standards or, when these are lacking, on the median of the 150 HAs.
At the start of the project, the regional network agreed on transparency for public accountability. An annual performance report is published and the web platform where the data are stored is public (performance.sssup.it/netval). The report includes all the regions and their local performance (HAs).
§ 2. The method proposed

The method we propose identifies regional priorities by jointly evaluating three different issues that should be relevant in the strategies of Regions working in a universal coverage context: 1. The performance, 2. The improvement capacity and 3. The reduction of geographical disparities. The latter point is tightly connected with the concept of equity: when unwarranted variation occurs, unexplained by population needs, the distribution of fair access to health services across the population is unmet (Vainieri and Nuti 2015; Wennberg, Fisher, and Skinner 2003; Barsanti and Nuti 2014).

Hence, for each indicator included in the adopted PES, the method proposed evaluates:

- the regional performance, measured in benchmarking against the other regions;
- the trend, measured in comparison to the other regions’ trends;
- the internal variability, i.e. the measurement of the differences among regional Local Health Authorities.

More in detail, the algorithm first selects – among the 200 indicators – those whose performance, trend and variability are jointly available. It then excludes the indicators that adhering regions consider to be “observational”: they are indicators that have only been recently adopted (and need some time to be validated), indicators whose accuracy is controversial and indicators that are not relevant by themselves but are important to understand the dynamics of other indicators. This process assures that the remaining indicators are sufficiently relevant and accurate.

Second, the procedure excludes the indicators that depict the epidemiological status of the regions (e.g., the infant mortality or the suicide mortality): these indicators portray the health care need of each region. Each local health authority and every regional health care system aims at their improvement; however, these are measured with a two-year delay and the
phenomena they monitor cannot be modified – in the short term – by the regional health care governance on its own.

Once these exclusions are made, about 85 indicators (with small fluctuations among regions) are left: these indicators provide the quantitative basis for detecting regional priorities. In order to do this, the regional performance of each indicator, its trend and its geographical (intra-regional) variability are computed as it follows.

First of all, each indicator is evaluated against the performance of the other regions or – when available – against a national or an international standard. Each region is therefore assigned a discrete score for each indicator, ranging from “1” (“poor performance”) to “5” (“excellent performance”). The score is associated to a colour: red for “1”, orange for “2”, then yellow (score equal to “3”), green (“4”) and dark green (“5”). The same colours are associated to the trend and the variability score, described below. The chromatic reference is very effective because it immediately depicts how the indicators are performing.

Secondly, the algorithm identifies regional priorities by considering the annual trend each indicator shows. The annual trend of each indicator is computed for each region and compared to the trend of all the other regions. More precisely, the difference between each regional indicator trend and the median regional trend is computed; this difference is then divided by the inter-regional standard deviation of the trends, according to the following formula:

\[
\frac{t_r - t_{\text{median}}}{\text{stddev}(t_R)}
\]

The resulting coefficient measures “how much” each regional indicator’s trend deviates from the median tendency of the network. Each indicator is therefore assigned a second score for
each region (the "trend score"), ranging again from “1” ("strong worsening") to “5” ("strong improvement"). This score reflects each regional trend compared to the trends of the other regions.

Tab. 1 – The trend score

<table>
<thead>
<tr>
<th>Trend</th>
<th>Coefficient</th>
<th>Score</th>
<th>Color</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong improvement</td>
<td>&gt;1</td>
<td>5</td>
<td>Dark green</td>
</tr>
<tr>
<td>Improving trend</td>
<td>From 0.5 to 1</td>
<td>4</td>
<td>Green</td>
</tr>
<tr>
<td>Average trend</td>
<td>From -0.5 to 0.5</td>
<td>3</td>
<td>Yellow</td>
</tr>
<tr>
<td>Worsening trend</td>
<td>From -1 to -0.5</td>
<td>2</td>
<td>Orange</td>
</tr>
<tr>
<td>Strong worsening</td>
<td>&lt; -1</td>
<td>1</td>
<td>Red</td>
</tr>
</tbody>
</table>

The following example can be used to explain the above-mentioned procedure: the percentage of femur fractures operated within two days after hospital admission differently changed from 2012 to 2013 among the regions that adhere to the network. The percentage increased by nearly 26% in the autonomous province of Trento, while it did not change very much in Tuscany; therefore, considering the median and the standard deviation, the autonomous province of Trento is given a score equal to “5” and Tuscany equal to “2”.
Finally, the geographical performance variability among Local Health Authorities (within each region) is the third criterion the algorithm uses to identify regional criticalities. The inter-LHAs coefficients of variation are computed for each region and each indicator. The quintile distribution is subsequently calculated for each region. Mono-specialist hospitals (i.e., hospitals that treat specific types of patients, such as paediatric hospitals) are excluded from the calculation of the coefficients of variation, as these show peculiar dynamics and this would distort comparisons with other hospitals. Equally, the coefficients of variation are not included for those indicators that refer to only one LHA in the region. A “variability score” that ranges from 1 to 5 is attributed to each indicator, according to the quintile occupied by the indicator.
Table 2 – The variability score

As an example, the coefficients of variation of Tuscan can be computed for 88 indicators in 2013. A variability score, based on quintile, assigns “5” to the first quintile, “4” to the second and so on.

![Figure 2 – Tuscany: coefficients of variation - 2013](image)

<table>
<thead>
<tr>
<th>Color</th>
<th>Number</th>
<th>Variability</th>
<th>Variation coefficient (quintile occupied)</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>1</td>
<td>Very high</td>
<td></td>
<td>V</td>
</tr>
<tr>
<td>Orange</td>
<td>2</td>
<td>High</td>
<td></td>
<td>V</td>
</tr>
<tr>
<td>Yellow</td>
<td>3</td>
<td>Average</td>
<td></td>
<td>V</td>
</tr>
<tr>
<td>Green</td>
<td>4</td>
<td>Low</td>
<td></td>
<td>I</td>
</tr>
<tr>
<td>Dark Green</td>
<td>5</td>
<td>Very low</td>
<td></td>
<td>I</td>
</tr>
</tbody>
</table>

Table 2 – The variability score
Finally, three different scores are available for each indicator and for each region: performance, trend and variability scores, each of them ranging from 1 to 5. It is therefore possible to sort the indicators for each region: first based on the performance score, second on the trend score, and finally, on the variability score.

Moreover, we deemed necessary to add a fourth column to give policy makers information on the financial impact each indicator might have in the short-medium term. The fourth column has a simple tick next to the indicators whose improvement would result in financial savings within a year. In Region $\alpha$, for example, a reduction in the pharmaceutical expenditure would translate into immediate savings for the public funder. The fourth element has been added to the methodology to provide information on the relevance each indicator has in terms of resource re-allocation strategies (Nuti, Vainieri, and Bonini 2010).

The final output is exemplified in Table 3.
Table 3 – An excerpt of the final output for Region α

<table>
<thead>
<tr>
<th>Code</th>
<th>Region</th>
<th>Indicator</th>
<th>Performance score</th>
<th>Trend score</th>
<th>Variability score</th>
<th>Short-term savings</th>
</tr>
</thead>
<tbody>
<tr>
<td>C11a.3.1</td>
<td>Region α</td>
<td>Hospitalization rate for COPD per 100,000 residents (50-74 years)</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>C18.6</td>
<td>Region α</td>
<td>Hospitalization rate for varicose vein stripping per 100,000 residents</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>D9</td>
<td>Region α</td>
<td>Percentage of people leaving the ED without being treated</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>C16.2</td>
<td>Region α</td>
<td>Percentage of green code patients visited within 1 hour</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>C4.7</td>
<td>Region α</td>
<td>Surgical ELC DRGs: percentage of hospital admissions in Day Surgery (Healthcare Agreement 2010)</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>F10</td>
<td>Region α</td>
<td>Pharmaceutical Expenditure</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>✓</td>
</tr>
<tr>
<td>C1.1</td>
<td>Region α</td>
<td>Standardized hospitalization rate per 1,000 residents</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>C14.2</td>
<td>Region α</td>
<td>Percentage of medical outpatient admissions for diagnostic purposes (Healthcare Agreement 2010)</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>C18.1</td>
<td>Region α</td>
<td>Hospitalization rate for Tonsillectomies</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>✓</td>
</tr>
<tr>
<td>C8a.19.2</td>
<td>Region α</td>
<td>Paediatric hospitalization rate for gastroenteritis per 100,000 residents aged ≤ 17 years</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
§ 3. The use of the methodology

Table 3 provides an intuitive graphic tool to detect regional criticalities, by jointly considering the performance of each region (benchmarked against the other regions), the trend of each indicator and intra-regional variability.

Each region should start its analysis by examining the first column. The benchmarked performance actually represents the first element to be considered. In the above mentioned case, Region α should first take into consideration indicators C11a.3.1 and C18.6, which show a very poor performance, and subsequently move to indicators with a poor performance (highlighted in orange). Each indicator deserves even more attention if low performances are associated to worsening trends. If the criticalities worsened in the previous year, they are highlighted in red in the second column, as they register a trend that is significantly worse than the other regions’. For example, Region α policy-makers should focus on people leaving the ED without being treated, not only because the indicator (D9) shows a critical performance, but also because it strongly worsened from 2012 to 2013.

The inclusion of the “variability score” provides policy-makers with valuable information to design different strategies that are tailored to single criticalities. If the third column warns about a strong intra-regional variability, each LHA should be assigned a specific target, based on the LHA’s previous situation. Under-performing LHAs should be required to show more improvement. On the contrary, if internal variability is low, the criticality refers to the region as a whole and it is reasonable to design a structural answer, through a transversal intervention that involves all the actors of the health care system with uniform targets among the providers. Referring again to Region α, the standardized hospitalization rate (indicator C1.1) shows a critical regional performance and the low variability suggests a structural
problem. Conversely, recourse to the day hospital for diagnostic purposes (indicator C14.2) exhibits a poor regional performance and a high intra-regional variability. This second case might suggest that the organizational problems involve specific providers. Moreover, the introduction of variability in an explicit way is an attempt to respond to the search for a managerial tool that can cope with disparities, which are still largely unattained (Barsanti & Nuti 2013)

Outputs similar to the above-mentioned table are provided for all the regions in the Network and they provide valuable information to set the 2015 targets of the regional health authorities. This information is used according to the managerial skills of the policy-makers’ staff and according to the presence of an annual planning and control system. The method we propose has been used by the regions with existing planning and control systems (Nuti et al. 2015; Vainieri and Nuti 2015) to steer the selection of the indicators to be included in the regional decree that sets the Health Authorities’ targets. The inclusion of some information about each indicator’s variability has also been useful in selecting the indicators that deserve a target for each HA.

In particular, the tool we describe has triggered a discussion among Tuscan regional managers on the priorities to be considered in the definition of the HAs’ targets. In some other cases, the tool is used to define the HA managers’ “mandate goals”, i.e. the results the CEOs are expected to meet in a 3-year period. In other cases, the tool is used as an opportunity to open a discussion among the regional policy makers on the procedures they usually adopt to set the HAs’ targets. Moreover, the tool can be applied not only at the inter-regional level but also at the intra-regional level to assess the HAs’ performance, especially where each HA manages more than one district and/or hospital.

In any case, the tool for priority detection relies on information made available by the network performance evaluation system and it does not consider regional criticalities that are not monitored by the evaluation system. Therefore, it is up to the regional policy makers to
decide how to weigh the three aspects (performance, trend and variability) in the priority-setting process and how to integrate the regional governance mechanisms with the evidence provided by the evaluation system.
§ 4. Discussion and conclusions

The dominant philosophy underlying the use of performance indicators is the notion of managerial cybernetics (Beer 1966). It envisages that in the managerial process central is the notion of feedback. This model is clearly based on phenomena in the field of physical and biological sciences where sciences are kept in control via negative feedback mechanisms (Forrester 1968). As Smith states, “when dealing with human organizations, the system one is trying to control is intelligent. The humans within the organization can anticipate the actions of the controller and – if it is to their advantage – may take action to frustrate the wishes of the controller” (Smith 1995, 280). All the feedback mechanisms can be flawed if the performance indicators are not able to direct the actors towards the organization’s aims. Indeed, some authors have highlighted the limits of the management by performance indicators.

The solution relies on the capacity of measurements and targets to drive the agents’ behaviour by reframing the feedback mechanism itself in order to avoid some of the shortcomings. In line with the practices of the corporate sector, scholars and managers who embrace this approach find it desirable to quantify each objective, though crudely, and use the measurement scheme as a pivotal control tool (Kaplan and Norton 1992).

None of the studies we mentioned in Section 1 explicitly challenges the desirability of using performance indicators per se. Rather, a few papers emphasize that their use can introduce risks and side effects that may not be anticipated. Our work provides a methodology that can minimize some adverse repercussions. In particular, the procedure we describe can help manage the limits associated with the “synecdoche principle” and therefore help contain potential opportunistic behaviours (“output distortions”).
Regarding the synecdoche effect, the above-mentioned selection of the indicators (about 85 out of 200) first assures their *accuracy* and their *relevance* (we recall that accuracy and relevance are the preconditions of the synecdoche principle). Secondly, we are aware that neither the adoption of a limited set of indicators, nor the use of a wide range of measures automatically leads to distortion-free results *per se* (Holmstrom and Milgrom 1991). However, the algorithm helps policy-makers select an appropriate number of indicators, which does not neglect any health care areas and – at the same time – does not disorient the actors.

On the one hand, the procedure systematically analyses *all* the available indicators, including *all* the indicators monitored by the performance evaluation system in the criticalities detection process. On the other hand, the selection of the problematic indicators gives the policy-makers the opportunity to set up an incentive system on a *limited number* of indicators, which keeps the controlled actors focused on the priorities. The mechanism we describe helps administrators set targets on “a part” of the indicators, selected because of their relevance to “the whole” system. Therefore, it responds to the objections concerning the excessive number of indicators monitored by the system.

Regarding the output distortions, the iterative nature of the procedure – we note that the selection of the priorities is repeated each year – limits some of the opportunistic behaviours of the controlled subjects. In particular, the mechanism would prevent professionals from neglecting the indicators that are not incentivized (“suboptimization”). It would detect the worsening indicators (or their increase in variability) and they would be automatically selected as priorities in the following year. Bevan and Hood suggest a solution to limit opportunistic behaviours by introducing some *uncertainty* in target specification. Their thesis is that excessive predictability could undermine the control itself, by facilitating opportunistic reactions (Bevan and Hood 2004).

The methodology we describe reverses their hypothesis: opportunistic behaviours are limited not by the inclusion of a degree of uncertainty but by the *certainty* that they will be detected
the following year, when the algorithm includes all the indicators and not only the incentivized ones. The algorithm is designed to monitor a large number of indicators – and to account for the different viewpoints and goals stakeholders might have –, but at the same time, it focuses the attention of the managers and the policy-makers on the indicators that are more critical in terms of performance, trends and variability.

To prevent the performance paradox, Meyer and Gupta (1994) recommend that organizations adopt the so-called “paradoxical model of performance assessment” with multiple, uncorrelated and varying but comparable performance indicators. The method we describe comes up with the same results (“multiple, uncorrelated and varying but comparable performance indicators”) through a “multi-criteria model of performance assessment”.
CHAPTER THREE - How to set challenging goals and conduct fair evaluation in regional public health systems. Insights from Valencia and Tuscany Regions
Abstract
The definition of “the right targets” and the way in which the evaluation of results is performed affect the willingness to commit to new challenges, which are factors that influence the relationship between goal setting and performance results. Indeed, some authors discussed that wrong performance targets is deemed to be a major cause of failure of management control systems. Goal setting theory scholars found that specific and challenging goal leads to higher performance than (a) an easy goal, (b) a general goal or (c) no goal setting. Despite this evidence, yet, few proposals concern the definition of what is challenging. This paper focuses on two issues: what is challenging and what is fair in health care sector? Benchmarking might be a valid tool to overcome the previous dilemmas. Moreover, the benchmarking techniques in public sector have been wide spreading and they have been the basis for the development of management control systems as dominant form of governance in the health care. Relying on two Regional European advanced experiences, Valencia in Spain and Tuscany in Italy, this paper aims to provide conceptual methods that can help managers define challenging goals and conduct fair evaluation. Although these Regions adopted different governance models, both of them applied very similar techniques improving performance and reducing variability.

Keywords
Incentives, targets, health care sector, benchmarking

Acknowledgements
A previous version of this manuscript was presented at the 2014 EHPG autumn meeting. The authors want to thank EHPG members for their helpful comments and in particular Barbara Bini who helped us in collecting some materials.
How to set challenging goals and conduct fair evaluation in regional public health systems.

Insights from Valencia and Tuscany Regions.
§ 1. Introduction

Goals have pervasive influence on employees’ behavior and in turn on organizational performance. This basic assumption of goal setting theory developed by Locke and Latham at the end of 80s for the individual level, has been analyzed for the organization and system levels by control management scholars. Literature and experience on goal setting showed that assigning targets is not sufficient. For instance the experience of Health for All program, launched by WHO in mid-80s (World Health Organization 1986) that set targets to member states renewing them in the mid-1990s with the Health21 policy framework (World Health Organization 1999), flawed in some countries and in some areas (Wismar et al. 2008).

Scholars that analysed the case of Health for all stated that some strategies were not met because of (Busse and Wismar 2002): the lack of involvement of key actors at the grass-roots levels; the shifted of power and responsibilities from the central to the regional level (Saltman, Bankauskaite, and Vrangbaek 2007); the lack of the “right targets” in terms of prioritization reflecting specificity of countries and in terms of identification of the correct effort to be required. All these elements are also found in the general literature on performance management (Ferreira and Otley 2009), indeed, setting the wrong performance targets is deemed to be a major cause of failure of management control systems (Ittner and Larcker 2003). Scholars of goal-setting theory stated that effective goals should be assigned considering the content (what have to be sought) and the intensity (how to attain the goal) (Locke and Latham 1990).

Remaining in the health care sector, both central and regional levels used targets in their governance models in different ways, getting different results (Nuti et al. 2015; Le Grand 2003; Bevan and Wilson 2013).

At this purpose Brown and his colleagues (A. Brown et al. 2012) believed that successful health care systems have: a public, specific statement of goals with a plan for reaching these
goals; a public report of improvement results and strong physician and clinical leadership in improvement efforts, aligned to improvement goals, again supported by useful data.

In this scenario, control management studies mainly discussed what indicators should be selected, the criteria to be followed to choose them (Pencheon 2008; Gagliardi et al. 2005; Jacobs, Goddard, and Smith 2006; Nuti and Vola 2015) and some elements of the process, in particular the importance of feedback and involvement (Ferreira and Otley 2009; Murante et al. 2014; VandeWalle 2003). When goal is specific and challenging, it leads to higher performance than (a) an easy goal, (b) a general goal or an exhortation to “do one’s best,” or (c) no goal setting (Locke and Latham 1990; Latham, Borgogni, and Petitta 2008; Locke et al. 1981; Lunenburg 2011; Locke and Latham 2006; Latham 2007; Mitchell and Daniels 2003). Yet few evidence and proposals concern the definition of what is challenging, that is an important characteristic goals should have to motivate workers (Locke and Latham 2013a).

Given these premises, this paper aims to contribute to the debate on the one hand on how targets should be set to be challenging, and, on the other hand, on how evaluation of results should be conducted to be perceived as fair. In particular, relying on two European advanced experiences, this paper aims at providing conceptual methods that can help managers define challenging goals and conduct fair evaluation.

§ 2. Background

The definition of “the right targets” and the way in which the evaluation of results is performed affect the willingness to commit to new challenges, which are factors that influence the relationship between goal setting and performance results (Locke and Latham 2013b). Seeking to respond to the aforementioned two open issues, it is possible to identify
at least four sub-decisions managers and policy makers follow when they set and evaluate targets:

i. whether to define the benchmark the actors are aiming at;

ii. whether to set homogeneous targets for all the actors;

iii. whether to consider the agents’ past and relative performances to set targets;

iv. whether to adjust results on the basis of environmental factors.

When policy-makers assign targets they usually consider whether a gold standard or a normative target exist (i). When the gold standard neither the normative standard exist then the definition of the targets requires more thinking and often a subjective decision. This situation may decrease the level of legitimation of goal. Once the standard was defined, policy-makers have to decide whether to assign the same target to all units (ii). Often policy-makers set homogeneous goals to all units. This decision might encounter some drawbacks. The first one occurs when the goal was set, for every unit, to the gold standard. The gold standard could require extreme effort for some agents so that it can be perceived as unattainable. Seemingly impossible goals can have two opposite effects known as “the paradox of stretch goals”: stretch goals could influence organizational learning and performance in a positive way facilitating improvement because they are seductive or they can they can have a disruptive effect leading to no commitment at all (Locke and Latham 2013b; Sitkin et al. 2011).

Another drawback is what managerial literature defines the “threshold effect”. This occurs when a minimum and equal threshold is set for all the controlled actors. On the one side, this mechanism puts some intentional pressure on under-performing agents; on the other side, it instills a perverse incentive for all those agents who are already performing over the threshold, by stimulating a regression toward the threshold level (Bird et al. 2005). The threshold mechanism generally penalizes those actors that perform well but still have single
criticalities, while it favors mediocre agents, who systematically perform in the threshold range. To overcome these problems, individual goals can be preferred.

When policy-makers have to set individual goals or they do not have the gold standard, a way to set targets is considering the past and relative performance of agents (iii). Indeed, previous studies demonstrate that goals have to be set considering the difference between the units (i.e., Local Health Department or professionals) and their starting point (baseline) (Ferreira and Otley 2009; R. Anthony and Young 1999). Performance incentives had the greatest impact on providers whose performance was lower at baseline (Greene, Hibbard, and Overton 2015) so that policy-makers could ask more to lower performer remembering that the effort should be perceived challenging but attainable. Indeed disruptive effects seem to be more frequent in organization whose recent performance was low (Sitkin et al. 2011). Then how challenging is challenging? In laboratory experiments (largely applied in the goal setting theory) challenging goals are usually considered those at the 90th percentile while in field experiments challenging is what agents perceived “difficult yet attainable” goals (Locke and Latham 2013b). That implies that the definition of what is challenging is set, most of the time, on qualitative basis. The recourse to subjective judgements may be due to the lack of comparisons.

The evaluation of the level of target attained by each agent (iv) can follow the simple degree of achievement of the set targets, but other factors need to be considered. In particular, some contextual variables might have affected the degree of achievement itself. This means that some correctives have to be envisaged (Ferreira and Otley 2009; Locke and Latham 1990; Flamholtz, Das, and Tsui 1985).

The introduction of benchmarking techniques might be the solution to face the four above-mentioned. While in the private sector is difficult to get data because of competition, in the public sector information are available so that benchmarking can help define challenging goals and fair evaluation. Indeed the benchmarking techniques in public sector
have been applying since the 1990s (R. Anthony and Young 1999), becoming the basis for
the development of management control systems as dominant form of governance in the
health care sector (Le Grand 2003; Bevan and Wilson 2013; Nuti et al. 2015). The following
paragraphs report the conceptual framework drawn by two European experiences.

§ 3. The technical framework for setting targets and assessing performance

Before investigating the specific experiences of the two case studies, we describe the
technical framework those experiences need to be contextualized in. The two case studies
actually share the same methodology to set challenging goals and assess their achievement.
The process is jointly based on systematic comparison (benchmarking) and on information
on past performance. The method follows a reasonable heuristic rather than statistical
methods (such as data envelopment analysis), in order to ease the process and to streamline
the communication toward stakeholders. It can be divided in two phases: 1) setting goals and
2) assessing performances.

For both phases, regional policy makers and managers have to identify the appropriate key
performance measures (as suggested for instance by Jacobs et al, 2006), representing the goal
and the group of peer units (units with similar mission, such as teaching hospital or focused
hospitals).

The essential ingredients of the method are two: 1) the past performance measures which
represent the baselines for improvement; 2) the relative performance. The method works
with goals expressed in quantitative terms. Indicators should be easily measurable with an
explicit and clear method of calculation. However, to be effective, the targets and goals
should follow the general rules (Ferreira and Otley 2009; R. Anthony and Young 1999;
Drucker 1954):
• Goals have to be coherent and aligned with the strategic objectives of the Region;
• Goals have to be set considering the difference between the units (i.e. Local Health
  Department or professionals) and their starting point (baseline);
• Goals have to be communicated in a clear and prompt way.

These rules are taken for granted in the method presented in this paper because it mainly
focuses on how to define what is challenging or what is fair.

§ 3.1. Method for setting challenging goals

Targets have to be set on the basis of the baseline for improvement, asking for an inverse
effort related to the goal standard or the best performer: greater improvement is required to
the units with poor performance whilst a lower improvement to those that already registered
a good performance. This method allows to set targets in a fair way, because units with the
worst performance are not rewarded more than those who perform already well.

This is easily done by executing, for each indicator, the following steps:

1. ordering the comparable units on the basis of their baseline (past or actual
  performance);

2. setting the target to two units;

3. drawing the line between the two targets;

4. calculating the expected value on the basis of the line;

5. proposing the targets to the general managers of the units, to fine-tune the target,
  according to local peculiarities.
Step 2 is a crucial phase. There are different combinations that policy makers could use. The two cases we studied consider two options:

a. setting the targets of the best and the worst performers;

b. setting the targets of the worst performers and the median value.

A-option is used when there is a gold standard that the best performer must achieve or there is the intention to ask the best performer for holding its position. B-option is adopted when there is no gold standard.

For the A-option, the gold standard (derived from literature or national/regional plans, such as the vaccination coverage) can be used as a reference for the best performer(s). To the worst performer it is possible to ask for going up the 25 percentile (or down the 75 percentile, if the lower value is the better value). In this way it is required to the worst performer to behave as the \( \frac{1}{4} \) of the units registering a performance lower than the median as showed in the figure 1.

Figure 1 – the references in the case of the existence of a gold standard
If there is no gold standard (such as efficiency indicators) then the B-option is preferred. In this case, the two expected values are set to the worst and median values. In particular, the worst value is set as in the A-option: to go up the 25 percentile or down the 75 percentile, if the lower value is the better value. As the second expected value, the region (or the central government) could think to shift the median to the 75 percentile (or 25 percentile for indicators that has to decrease) as reported in figure 2.

Once the two expected values are set, two dots have been identified: for increasing indicator, the actual performance of best practice and the gold standard (or the actual median and the 25 percentile) is one dot, the other dot is the actual performance of the worst practice and the 36.5 percentile. The expected values of the other units can be estimated using the linear equation that comes from the connection of the two dots.

In the B-option, it is possible that the expected value, coming from the equation, for the best performer(s) tend to worsen the performance instead of improve it. In this case, the A-option is to be executed asking the best performer(s) to hold the value.
Step 2 determines the range of variability that Region considers acceptable. Indeed, it could sound odd to accept a certain degree of unwarranted variation. The underpinning of this choice is linked to the empirical evidence provided by the goal setting theory that challenging goals lead to better performance. Although it is desirable that all units achieve the performance of the best practice or the gold standard, assigning to every unit, for the same goal, the same expected target may be perceived as unfair, thus reducing motivation to achieve it. Indeed, incremental goals can be more motivating than radical changes, which can be perceived as stretch goals. Hence, the assumption is that the \textit{ex ante} range set, jointly with a fair evaluation, would lead to better performance and faster reduction of variability.

Once targets are defined with the above-mentioned method, they are communicated and discussed with the units, which receive them in a proper timeframe. This phase is implemented in order to involve units (hence employees) into the process.

\textit{§ 3.2. Method for performance evaluation}

If the set target is reached, then the achievement is 100% and no further evaluation has to be done.

If the set target is partially attained, then, in order to give a fair evaluation, it is important to compare performance with the other comparable and the baseline.

Indeed, the relative performance and the baseline can help understand if the target was stretch or if environmental factors occurred.

How to assess in a fair way target partially attained? It depends upon circumstances. It is possible to identify five scenarios:

1. the performance of the unit worsened as well as all the other comparable;
2. the performance of the unit worsened and the relative performance is lower than the median;

3. the performance of the unit worsened but the relative performance is upper than the median;

4. the performance of the unit improved although it doesn’t achieved the set target and the relative performance is lower than the median;

5. the performance of the unit improved and the relative performance is upper than the median.

In the first scenario, it is clear that some external factors happened so that the set target was no more attainable or the selected indicator was uncontrollable by units.

In the second scenario, performance evaluation is bad so that if monetary incentives are linked to this evaluation no incentive has to be given.

In the third scenario, whilst the performance worsened, the unit is positioned better than the half of other units. In this case, performance evaluation is not so bad so that regional policy makers could provide a premium for the relative performance.

In the fourth and fifth scenario it is possible to apply the linear system suggested by Locke (Locke 2004). In addiction for the fifth case, policy makers may acknowledge a bonus for the relative performance.

Table 2 reports a synthesis of the previous cases if the goal is partially attained.

<table>
<thead>
<tr>
<th>Case</th>
<th>Position</th>
<th>Trend</th>
<th>Evaluation linked to incentives</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.</td>
<td>-</td>
<td>-</td>
<td>No incentive at all</td>
</tr>
<tr>
<td>3.</td>
<td>= or +</td>
<td>-</td>
<td>It can be given an incentive to the position. If the share between trend and position is 50%-50% then the incentive would be less than 50%</td>
</tr>
</tbody>
</table>
4. It can be assigned a linear percentage of the result obtained but only to the share reserved to the trend.

5. It can be assigned a linear percentage of the result obtained and an incentive to position.

In order to apply this evaluation, policy makers have to decide if they want to provide a bonus for the relative performance (the position). If the answer is positive then they have to choose:

- the threshold of providing the bonus: the mean? the median? Another percentile up to the median? Only the best performer?
- the bonus share between the relative performance and the degree of improvement.

Although it could be preferred to recognize a bonus for the relative performance, the size share and the threshold may be a matter of political decisions.

§ 4. The two regional cases

§ 4.1 The Valencia Region

The Spanish Health System’s universal coverage is funded by taxes and it predominantly operates within the public sector. Health competences were totally devolved to the regional level at the end of 2002 even if the devolution process started before 2002 for some Regions, because of their self-governing tradition and strong regional identity (García-Armesto et al. 2010). In fact, the Valencia Region has had fully autonomous power in managing and organizing its health care since 1987. The Agencia Valenciana de Salud is the public body responsible for providing health care services to its approximately 5 million of inhabitants.
through its 24 Local Health Departments. Local Health Departments are geographically-based organizations responsible for the provision of primary health care services, acute services and community services.

Since the beginning of 2000, Valencia has introduced public-private joint ventures through the administrative concessions to corporations or temporary union of enterprises for the provision of care (García-Armesto et al. 2010; The NHS European Office 2011; Trescoli Serrano, Marin Ferrer, and Torner de Rosa 2009). Agencia Valenciana controls the whole system but public providers are still predominant: in 2008 the percentage of publicly-owned hospital beds was around 80% in Valencia and the percentage of the population that was using this model was around 16% (García-Armesto et al. 2010).

Since 2004, Valencia’s Regional health care system has adopted a management control system based on targets and their evaluation. The assessment of each organizational level (represented by the health departments) and each employee is based on the annual Management Agreement Acts (Acuerdos de Gestión) of each health department. The target setting and evaluation system is characterized by a clear framework of objectives. A pivotal role is played by the Agencia Valenciana de Salud, which is the governmental authority in charge of the full process of target setting and assessment. Since 2005, this system has also been aligned with the variable salary of all its employees and since 2007, it has been linked with their professional career (Decree 38/2007).

About 20% of the objectives of each health department relates to specific initiatives for the area/territory, while the remaining 80% regards the strategic objectives defined by the Agencia Valenciana de Salud, although some of them can be structured in different ways among the health departments.

Objectives are declined starting from the three domains: responsiveness, health care service provision and financial sustainability.
When this method was first applied, most of the objectives referred to process indicators (such as hospitalization rates, average length of stay, etc.). Recently, outcome and quality indicators have also been introduced, given the importance of measuring the full health care path and its impact on patients’ health status.

§ 4.2 Tuscany Region

The Italian health care system ensures universal coverage and, after the devolution process of the 90s, it makes all the regions responsible for organizing and delivering health services. Tuscany’s health care system covers approximately 3.7 million inhabitants, delivers 95% of its services through its public organizations, and spends more than 6.6 billion euro in health care services per year.

Since 2005, Tuscany’s health care system has adopted a Performance Evaluation System (PES) that consists of more than 100 indicators grouped into six categories (Nuti, Seghieri, and Vainieri 2013; Nuti et al. 2009): population health status; capacity to pursue regional strategies; clinical performance; patient satisfaction; staff satisfaction; efficiency and financial performance. The PES gives a performance overview of the 12 local health authorities and 5 teaching hospitals in the region through a benchmarking process. Indicators for the PES were selected by Tuscany’s health authorities and regional administrative professionals by investigating performance indicators in other countries (e.g., Ontario Health System, English NHS). In general, this system shows performance evaluation in five tiers: excellent, good, average, bad and very bad performances. The system set up a public website in 2007. In 2006, the Tuscany Region decided to link the PES system to the CEO reward system. Before 2006, most CEO goals were qualitative and assessed following the “all or none” criterion. They were mainly based (more than 50%) on financial performance and the average achievement level reached up to 90%, with low variability. After integration with the PES,
more than 50% of the goals have been quantitative, the weight of the financial assessment goals has been reduced and the degree of the global achievement of goals has varied between 45% and 70% (Nuti 2008).

Every year, the regional managers select the indicators that are be included in the reward system on the basis of the strategic plan, the contextual environment and the list of indicators used at the national level to assess Regions. Goals are usually common across units, but there are also specific indicators for units with different missions (such as teaching hospitals) or some Regions want their health authorities to focus their attention on particular issues.

After the introduction of the national outcome program (Agabiti et al. 2011), some indicators relates also to the outcomes (such as 30 days mortality rate for AMI).

§ 4.3 A comparative view of the two case studies

Both Valencia and Tuscany seem to apply the same methods, although some differences persist. Table 1 sums up the principal characteristics of the two systems.

<table>
<thead>
<tr>
<th>A. Components of the method</th>
<th>Tuscany</th>
<th>Valencia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criteria of Past and Relative performance</td>
<td>The target is set considering past performance and the relative performance asking for more effort from units performing worse.</td>
<td></td>
</tr>
</tbody>
</table>
There are two indicators which sum up the global relative performance and the global ability to improve performance.

**Communication**

By each December, targets are discussed on the basis of the previous last nine months. Then targets are issued in a regional act which includes all the information, publicly accessible via web. Targets can be adjusted once the total of yearly data is available.

By each December, targets are discussed. Each health department and professional is aware of their targets. The targets can be adjusted during the year (usually in April or May), but there are usually not many changes.

**Level of implementation**

*Who is in charge of the process?*

Chief Executive Officers of Local Health Authorities and Teaching Hospitals

Tuscany Region supported by an independent body, public university (Laboratorio Management & Sanità)

Health departments and professionals

Governmental body (Agencia Valenciana de Salud)
### B. Components of the method

<table>
<thead>
<tr>
<th></th>
<th>Tuscany</th>
<th>Valencia</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Criteria of Past performance</strong></td>
<td>Accounts for 50%</td>
<td>Accounts for 50% plus corrections with the arctangent function</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Criteria of Relative performance</strong></td>
<td>Accounts for 50%. Correction occurs in case of “hold your performance” goal.</td>
<td>Accounts for 50% plus a premium if it is the best performer</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Communication</strong></td>
<td>By the month of May, the results of the quantitative indicators are disclosed. Final evaluation is registered after 2 years. Once final assessment is done, after a discussion phase, the ranking and assessment of each goal per unit are in a public act publicly available online for citizens.</td>
<td>By each May, the results of the quantitative indicators are presented. The Ranking and assessment of each goal are shown to each unit. Only the results of the top ten units are publicly disclosed.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Level of implementation (who is assessed?)</strong></td>
<td>Chief Executive Officers of Local Health Authorities and Teaching Hospitals</td>
<td>Health departments and professionals</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Who is in charge of the process</strong></td>
<td>Tuscany Region supported by an independent body, public university</td>
<td>Governmental body (Agencia Valenciana de Salud)</td>
</tr>
</tbody>
</table>
Table 1 – A comparison of the Tuscan and Valencia method for setting targets (A) and their assessment (B)

To sum up the results presented in the table, it is possible to group the main differences into two aspects: 1) the process of communication and 2) the level of implementation. As regards the process, Tuscany, which bases its performance management system on reputation, publicly discloses all its information while Valencia carefully chooses what has to be shown in public and among peers. Concerning the level of implementation, Tuscany applies the system on the health care authorities (formally represented by the Chief Executive Officers) at the macro level, whilst Valencia has a very centralized system, which sets targets and assesses them not only at the macro level (health departments) but also professionals.

There are differences in the technical methods adopted. For the setting target phase, Tuscany uses a global performance goal and an overall improvement goal to limit the synecdoche problem. Overall performance is calculated as the percentage of good and excellent performance indicators minus the bad and very bad performance indicators over the total. In addition, there is also the percentage of indicators improved. These two indicators sum up the relative performance and the trend in one shot. The aim is to motivate health authorities to pay attention to all the indicators, in order to reduce potential output distortions (Nuti et al. 2015; Bevan and Hood 2006).

For the assessment phase, Valencia uses the median (Tuscany adopted the mean), the arctangent function to adjust for the past performance effect and the premium for best performers (Tuscany accepted a small variation to hold the same position). In particular, the
three components applied to the evaluation phase are those of section 2 (Gòmez and Carrillo 2009): 1. the degree of achievement of the set target (linear component); 2. the performance improvement or worsening (asymptotic component) and 3. the relative performance with regard to similar Local Health Departments (exponential component). The performance improvement compared to the previous year is corrected by a factor ranging from 0.5 and 1.5 from an arctangent function with the aim to reduce or increase by 50% the value of the objective, depending on whether is a deterioration or an improvement compared to the previous year. An additional and final correction is applied in the Valencian method to ensure that there is always someone that achieves 100% of the target; if p is the percentage of achievement of the best performer and this value is less than 100, a correction factor of 100/p is applied to the scale all the other scores.

Results obtained by the two Regions are quite positive. In particular, in Valencia this technical framework has helped make the mechanism of financial incentives for employees credible and acceptable. It has led not only to continuous improvement but also to strategy alignment between Regions and Health Departments. Finally, most of the indicators have reported also a reduction of variability in the performance of health departments (Gòmez 2008; Gòmez 2011).

In Tuscany, the overall governance systems (PES monitoring, CEOs reward system, and other PES tools) has registered very positive results, both in terms of improvement and variation reduction. With the integrated management system, the Region of Tuscany was able to improve its performance and align its strategies and objectives (Nuti, Seghieri, and Vainieri 2013). A recent study shows that this region registered a high performance in 2007 and was still offering good general assistance in 2012 discussing on the positive role played by PES to drive the health authorities toward a sustained improvement path (Nuti et al. 2015).
§ 5. Discussion and conclusions

The technical framework proposed aims to help solve the dilemma of how challenging is challenging also supporting a fair assessment of the goal attainment. The method is essentially based on two components: performance benchmarking (relative performance) and performance baseline (past performance). Both of them play a pivotal role in setting challenging goals and fairly assessing their attainment.

Benchmarking introduces a yardstick competition among the actors of the health systems, by helping Regions set difficult, yet attainable targets and avoid problems linked to uncertainty.

Following the assumption that incremental goals can be more motivating than radical changes (which can be perceived as stretch goals), the technical framework proposes to set targets by asking for an inverse effort on the basis of past performance. The expected consequence is a higher degree of success which leads to an overall regional improvement and a reduction of geographic variation as well. The results obtained by the two Regions seem promising, both of them register improvement and performance alignment. This is a surprising result considering that the two Regions have adopted a very different governance model on the basis of Bevan and Wilson (Bevan and Wilson 2013) classification. Valencia has adopted the centrally driven hierarchy and target model, with some characteristics that are linked to the choice and competition model, because of the public-private partnership introduced into the system which enables quasi-market mechanisms (based on price and quality in contracts with providers). Tuscany, instead, combines the hierarchy and target model with public rankings (Nuti et al. 2015). Despite different governance models, the technical steps adopted by the two Regions are very similar. Both of them use past performance and relative performance to set challenging goals and to fairly assess their
achievement. Hence, the operational framework proposed seems to be useful and adaptable to different contexts and it can be also applied to individuals, as in the case of Valencia.

Moreover, if setting challenging targets and conducting fair evaluations lead to better performance, then this process, which reduces the distance between best and worst performers, will help Regions to cope with unwarranted variations: the proposed method will not only help Regions guarantee high-quality services (defined by the targets chosen), but also guarantee equity by reducing unwarranted variations.

Hence, this method is mainly proposed at the macro level to regional administrations, but it can also be applied at the micro level, as long as units can be compared and there is a performance measurement system already in use. Indeed, some choices can be generalized, on the basis of the principles sustained by the goal setting theory and benchmarking process, while other choices strictly depend on the strategies Regions decide to follow, such as the communication process. Further studies are needed to better understand if different applications lead to different results.
CHAPTER FOUR - Making governance work in the health care sector: evidence from the Italian “natural experiment”
Abstract

The Italian Health Care System is a public health system that provides universal coverage for comprehensive health services and it is mainly financed through general taxation. Since the early 1990s, a strong decentralization policy has taken place in Italy and the State has gradually ceded its jurisdiction to its 20 Italian Regions. They now have political, administrative, fiscal and organizational responsibility for the provision of health care.

This paper systematizes the different governance models Italian Regions have adopted and investigates the performance evaluation systems (PES) associated to them, with a particular focus on the experience of a network of ten Regions that shares the same PES.

The article draws on the wide range of governance models and PES to design a “natural experiment”. Thus, the research assesses - through the analysis of 14 indicators measured in 2007 and in 2012 for all the Italian Regions - how different performance evaluation models are associated to different health care regional performances and if the network-shared PES has made any difference in the results achieved by the adhering Regions.

First results support the idea that systematic benchmarking and public disclosure of data prove to be powerful tools to guarantee balanced and sustained improvement of the health care systems, but only if integrated with the regional governance mechanisms.

Keywords

Management control systems, governance models, benchmarking, health care, Italy
§ 1. Introduction

Following the wave of the international New Public Management movement (Hood 1991), many health care systems underwent reforms in the 1990s to shift control from the national to the local level and hence raise the scope for flexibility in local governance (Fattore 1999; Saltman, Bankauskaite, and Vrangbaek 2007). Since then, reforms based on New Public Management principles have aimed at making the public sector more efficient, effective and accountable (Hood 1995; Lapsley 1999; Saltman, Bankauskaite, and Vrangbaek 2007). Some countries have introduced quasi-market mechanisms to enable competition; others have focused on measuring performance which has become a mantra at all levels of government since the 1990s (Radin 2000). As a consequence, health systems and institutions have adopted different strategies and governance models with a particular interest in measurement tools and techniques. Initially, performance measurement primarily focused on financial issues and it neglected measures of multiple strategic objectives to drive changes (Ghobadian and Ashworth 1994; Pollitt and Bouckaert 1995; Guthrie and English 1997; Lorden, Coustasse, and Singh 2008). Hence, comprehensive multi-dimensional performance measurement frameworks, such as the balanced scorecard, were introduced (Kloot and Martin 2000; Yang and Tung 2006). Another development was the benchmarking of health performance measurement systems at the international, national and local levels (Johnston 2004; NHS Executive 1999; Pink et al. 2001; Vainieri and Nuti 2011). Benchmarking can help managers learn from best practices (Liebfried and McNair 1992) and be used as a mechanism to detect unwarranted variations and encourage their reduction (Arah et al. 2003).

We argue here that to drive health care system improvement at national or local levels, the performance evaluation system should be designed to align with the national (or sub-national for local governments) strategy, mission and vision in order to provide coherent messages for those running units and their employee (Ferreira and Otley 2009).
Relying on previous studies (Cromwell et al. 2011; P. Brown et al. 2012; Bevan and Fasolo 2013; Bevan and Wilson 2013), we identify five governance models:

1. The “trust and altruism” model relies on the perspective that all public servants behave like knights. It was the traditional model applied by the NHS and it does not focus on success and failure; on the contrary, it can lead to rewarding failure and ignoring success;

2. The “choice and competition” model is based on the quasi-market system where patients can choose and the money follows the patients. This model introduces external incentives and patients (or insurance companies) can choose providers on the basis of information on quality.

3. The “hierarchy and targets” model, also known as “command and control”, is based on recourse to external incentives and the strong role of performance management (generally by the central government). It has side effects like high monitoring costs and low acceptance by professionals;

4. The “transparent public ranking” model is based on the lever of reputation; this model has been applied and is known in England as the “naming and shaming” model.

5. The “pay for performance” (P4P) model draws upon economic incentives to direct the managers' behavior. Regarding the specific use of the expression “pay for performance” in this paper, Regions that adopt the “P4P” model link the rewarding scheme of their Health Authorities’ CEOs to the performance they achieve. This model is based on the assumption that financial payments can motivate people to achieve performance targets. It aims to improve the quality and efficiency by paying more for obtained results or actions such as evidence-based preventive care services or denying payment for preventable complications.
These governance models can be adopted at the macro level by state, at the meso level by Regions (or counties and provinces, depending on the country organization) and at the micro level by local institutions (municipalities, health care authorities, hospitals…). The basic ingredients of the five “ideal typical” models can be mixed. For instance, Bevan and Fasolo (2013) have described the “star rating” model, which was applied in the English NHS from 2000 to 2005, as a combination of the third and fourth model.

This paper discusses which governance models have been adopted by the Italian Regions (the meso level) and their impact on performance. The paper firstly classifies the Italian Regions using the five governance models described above. Secondly, it describes the Inter-Regional Performance Evaluation System (IRPES), which is an evaluation tool currently adopted by a network of 10 Italian Regions and how this was used in terms of a governance model during the 2006-2012 period. Thirdly, this paper examines how the performance of Italy’s Regions changed between 2007 and 2012. The paper concludes by discussing the outcomes of the different governance models adopted by the Italian Regions and, in particular, how IRPES has, or has not, driven improvement in the network.
§ 2. The governance systems adopted by Italian Regions in the health care sector

The Italian National Health Care System (NHS), which follows the Beveridge model, is a public health system that provides universal coverage for comprehensive and essential health services through general taxation. Since the early 1990s, a strong decentralization policy has taken place in Italy and the state has gradually ceded its jurisdiction to its 20 Italian Regions (France and Taroni 2005).\(^1\) The central level - represented by both the Ministry of Health and the Ministry of Finance - ensures whether the Regions keep their health care expenditures within their budget and guarantee the essential levels of care. Conversely, Regions are in charge of organizing health care services: they define their own Regional health plans, coordinate the strategies of the Regional Health Authorities and allocate the budget within their systems. Since the 2000s, the health care budget has been allocated among the Regions on the basis of a per capita share, partially adjusted by the age distribution of the population. At the same time, the Regions have become more fiscally autonomous and more financially responsible (Ferrario and Zanardi 2011; Ferrè, Cuccurullo, and Lega 2012).\(^2\)

Italian Regions now have the political, administrative and financial responsibility for the provision of health care to their residents.

De Vries (2000) argued that the results of the decentralization process depend on the cultural and political context, on the administrative capabilities of the actors involved and on how the process is promoted (De Vries 2000; Putnam 1993). The consequence is that there are now 20 Regional Health Care Systems (RHSs) in Italy with different governance models and management tools (Formez 2007; Censis 2008; Tediosi, Gabriele, and Longo 2009; Vainieri and Nuti 2011; Carinci et al. 2012; Mapelli 2012). France et al. showed the North-South performance disparities in mortality, expenditure and equity up until 2002 (France, Taroni, and Donatini 2005). About ten years later, Federico Toth reviewed the first decade of Italian
decentralization (1999-2009) and concluded that the shift of power from the central to the Regional level had accentuated the North-South divide, in terms of expenditure and perceived quality of health care services (Toth 2014).

The high degree of geographic variation in various measures of performance demonstrates that these general conditions and quality/volume standards are not equally achieved among the Italian Regions. Such variation is common in systems of health care (Wennberg, Fisher, and Skinner 2003; Wennberg 1999; Wennberg and Gittelsohn 1973; Appleby et al. 2011).

During the first years few of devolution (2001-2005), the central government bailed out the past health care deficits of the Regions. In order to prevent growing deficit, the Italian government approved legislation that introduced a new recovery process to reduce the financial deficit of the Regional health care systems (Bordignon and Turati 2009; Ferrè, Cuccurullo, and Lega 2012). The Financial Stability Law L. 311 (30 December 2004) and the Financial Stability Law L. 296 (27 December 2006) regulated the design and the adoption of the recovery plans.

The laws decree that if the Regions are in deficit - even with extra finance from Regional taxes - they have the right to access a bail out fund, financed by national taxation. To access this fund, the Regions are required by the central government to produce a recovery plan, which should identify strategic actions to address the structural determinants of costs to achieve financial balance (Ferrè, Cuccurullo, and Lega 2012). These recovery plans are subject to approval by the National Ministry of Health and by the Ministry of Economy and Finance. If the plans are deemed inadequate, the President of the Region is formally replaced by an “ad acta commissioner”, that the law states to be the president himself, and Regional taxes have to be automatically increased up to a predefined threshold. Since 2007, 10 out of 20 RHSs have carried out a recovery plan: Abruzzo, Molise, Apulia (since 2010), Campania, Calabria (since 2009), Sicily, Lazio, Piedmont (since 2010), Sardinia and Liguria. In five of them, the central government has nominated a commissioner in charge of local
implementation. So far, the Liguria Region and Sardinia have successfully implemented their recovery plans with a balanced budget (these Regions have succeeded by allocating financial resources from other public sectors to health care).

We now describe the models of governance adopted by each Italian Region, illustrating how Italian Regions differently declined the five above-mentioned “ideal typical” models in the 2007-2012 period. We identified four groups of Regions, according to how they mixed the five governance models.

First, Lombardy is the only Region that has opted for the “choice and competition” model by splitting purchasers and providers (including private institutions) in order to stress the role of patients’ choice to boost competition (Lombardy Region 1997). The principal tools adopted by Lombardy to manage its services are represented by a) tariffs; b) the adoption of the Joint Commission International Accreditation Program for Hospital Care (Joint Commission International 2014); c) hospital care outcomes and patients’ satisfaction (Vittadini 2012). General managers of the LHAs are rewarded according to the achievement of targets negotiated with the Regional administration. Lombardy, therefore, combines some elements of the “choice and competition” model (tariffs and patients’ choice) with “pay for performance” ones (the managers’ rewarding system). However, despite the link between CEO rewards and performance results, the variability of managers’ results and the related economic incentives is low, thus weakening the P4P strategy as a governance tool (Vainieri et al. 2013). Finally, it is noteworthy noting that even if Lombardy measures outcomes in benchmark, it does not fully disclose results either to the hospitals, or to the patients: Lombardy hospitals only have the opportunity to know their own results, without understanding how they compare with each other; therefore, they cannot identify and learn from the best practices (Vittadini 2011; Berta, Seghieri, and Vittadini 2013). Moreover, despite the alleged stress on patients’ choice, Lombardy does not use transparent public ranking and it does not publicly disclose results.
Second, the “hierarchy and targets” (or “command and control”) model has been applied by the State for the following eight Regions that are still subject to recovery plans: Abruzzo, Molise, Apulia, Campania, Calabria, Sicily, Lazio and Piedmont. Although the central government mainly specifies financial targets for all of them, the response differs according to their previous governance models and the managerial skills of their staff. Regardless of these differences, none of these eight Regions nor the state systematically benchmark clinical results between Regions and publicly disclose data.

Recent studies on top management evaluation systems have highlighted that the setting targets phase in these Regions does not follow objective and rational processes (Caldarelli et al. 2013; Vainieri et al. 2013). Indeed, they often do not take into consideration past performance and largely depend on qualitative targets, which are usually vague and can be interpreted in different ways (Vainieri et al. 2013). Moreover, these Regions lack a reliable supervision and monitoring system (Ferrè, Cuccurullo, and Lega 2012). Sanctions for failure are therefore not clearly applied and the “hierarchy and targets” (or “command and control”) model appears to be applied loosely: thus, the difference with the “trust and altruism” model (which is not formally adopted by any Region) is blurred.

Third, since 2006 Tuscany and an increasing number of Regions (ten Regions in 2014) have adopted a mixed governance model that combines “hierarchy and targets” elements with “transparent public ranking” (in the form of public disclosure of performance data) and “pay for performance” (limited to the CEOs’ rewarding schemes) ones. This mixed model has been adopted at different times between 2007 and 2014 by Tuscany, Liguria, Umbria, Basilicata, Trento, Bolzano, Marche, Veneto, Emilia Romagna and Friuli.

Moreover, another two Regions - the Aosta Valley and Piedmont - joined the Tuscan Performance Evaluation System for 3 years, from 2008 to 2010. As previously mentioned, Piedmont partially lost its autonomy when it shifted to the command and control model, with the strong role of the central government, while the Aosta Valley went back to its
Regional model of hierarchy and targets mainly focused on epidemiological issues. The Aosta Valley has disclosed performance results to internal users rather than to the public and they are not always translated into policy decisions (Carinci et al. 2012).

Fourth, Emilia Romagna, Veneto and Friuli have only recently opted for the governance model suggested by Tuscany. These Regions, before 2014, applied mixed governance models differently (Carinci et al. 2012; Vainieri and Nuti 2011). In the 2007-2012 period, all of them used some performance evaluation mechanisms on several dimensions. For instance, Friuli linked health databases, delivering detailed reports and regular publications for internal users; Veneto conducted patient surveys and Emilia Romagna conducted self-evaluation cycles, involving health professionals (Carinci et al. 2012; Vainieri and Nuti 2011). However, these tools were not systematically used in the Regional decision making process and their performance was not benchmarked against the other Regions’ one.

In conclusion, no Region can be considered to have adopted one single clear-cut governance model but rather a combination of them.

In this context, the advanced experience of the Italian Regions that adopt the same Performance Evaluation System is worth examining.

§ 2.1. The network experience

Since 2008, a growing number of Italian Regions has adopted the same Inter-Regional Performance Evaluation System (IRPES), which was designed and implemented for the first time in 2005 in all of the Tuscany’s Local Health Authorities by the Laboratorio Management e Sanità (MeS) of the Scuola Superiore Sant’Anna to measure and monitor indicators of quality, efficiency, appropriateness, continuity of care, patient satisfaction and staff satisfaction (Nuti and Bonini 2013; Nuti and Bonini 2014; Nuti, Seghieri, and Vainieri 2013).
In 2014, there were ten Regions in the network: Basilicata, Liguria, Marche, the Autonomous Province of Bolzano, the Autonomous Province of Trento, Toscana, Umbria, Veneto, Emilia Romagna and Friuli Venezia Giulia. The Regions have joined the network in different years, as it is reported in figure 1.

**Figure 1 – Regional adhesion to IRPES over the years**

<table>
<thead>
<tr>
<th>IRPES-adhering Regions</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aosta Valley</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basilicata</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bolzano</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emilia Romagna</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friuli Venezia Giulia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liguria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marche</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Piedmont</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trento</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tuscany</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Umbria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Veneto</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The Laboratorio MeS develops the performance evaluation framework and brings objectivity to the benchmarking processes as an independent research unit. It coordinates and manages information sharing and data acquisition. The ten Regions in the network agree on the indicators to be selected for benchmarking and on all the details about how they should be calculated. Each Region is responsible for processing its own data, in order to increase the awareness and the expertise of the Regional managers and their staff.

The aim of the IRPES is to assess and monitor health system performance at a Regional and local level: the results are shown by Region and by Health Authorities (HA) (both Local Health Authorities (LHAs) and Teaching Hospitals (THs)). In 2014, IRPES monitored the performance of 99 HAs.
The Regional Network integrates a *longitudinal* perspective (the trend) with a *cross-sectional* one, based on the benchmarking process. It offers Regions valuable information to define priorities and to fix appropriate targets, considering the results in benchmarking. Moreover, thanks to the same PES, the Regions may evaluate, share and spread best practices.

Indicators are defined by endorsing a “managerial” perspective designed to drive the change to organizational improvement (Mannion and Davies 2008). The rationale behind the selection of each indicator is the informational contribution it can offer the managers and policy makers. Indicators are chosen not only because they appropriately represent the epidemiological situation of single Regions/Local Authorities, but because they also detect best (organisational) practices or, on the contrary, flawed clinical processes.²

Indicators are defined in regular meetings with Regional representatives that include both managers and clinicians. For an evaluation system to be able to influence and change behaviours, it must actually win support from clinicians on rules and criteria their performance is measured on (Locke and Latham 2013a).

PES encompasses a large set of indicators that are up-to-date because they are calculated and disseminated in a six-month period. The indicators are grouped into 60 indexes and classified in six dimensions (a letter is used to indicate each dimension):

(A) **Population health.**

(B) **Regional strategy compliance**, to guarantee that strategic Regional goals are pursued in the time and manner indicated.

(C) **Quality**, appropriateness, continuity of care, patient safety and managing supply to match demand.
(D) **Patient satisfaction**, the patients’ experience and level of satisfaction with health services.

(E) **Staff satisfaction**, results of surveys on the satisfaction level of staff with their working conditions and management.

(F) **Efficiency and financial performance**

PES measures results in quantitative terms and then assesses performance for 100 of the 160 indicators: excellent, good, sufficient, poor or very poor. These five evaluation tiers are associated with different colours, from dark green - corresponding to excellent performance -, to red - corresponding to poor performance. Regions use the same reference standards for evaluation, based on scientific literature, national standards or, in their absence, on the median of the 99 HAs. Figure 2, as an example, displays the indicator of femur fractures operated within two days.
In order to show the performance of each Region or HA, a dartboard chart with the six dimensions has been used (see figure 3). The dartboard chart is also divided into five evaluation bands, associated with different scores and colours, from dark green, corresponding to excellent performance, to red, corresponding to poor performance. Each indicator is positioned on the dartboard and there is no overall unique ranking for Regions/HAs. When the result has a high score, it is displayed close to the centre (dark green), and when the score is low, it is displayed far from the centre (red).
The number of indicators varies by Region because each Region chooses which ones to include, with reference to local context and strategies, but there is a core group of indicators that all the Regions consider mandatory for the main pillars of the health care system. Indeed, the majority of indicators are common to all the Regions because the main objectives are the same at the national level. The IRPES structure also allows Regions to choose different indicators to reflect the different Regional strategies. The inclusion of a specific indicator within IRPES signals the strategic relevance the indicator is deemed to have, for all the Regions or for a subset of them.

From the beginning, the Regional Network has agreed on transparency for public accountability. An annual performance report is published and the web platform is public (http://performance.sssup.it/network). The report includes all the Regions and local performance (HAs) is showed as well.
There are regular meetings between the Regional representatives to share the results of the assessment system, identify best practices and compare outcomes of different Regional strategies. The systematic reporting of comparisons of performance that IRPES provides may result in some elements of competition among the Regions. Working groups are established as issues arise to discuss the different impacts of policies and to develop new indicators.
§ 3. IRPES utilization as a governance tool

Different governance models can rely on the informational contribution offered by IRPES (P. Brown et al. 2012; Nuti, Seghieri, and Vainieri 2013). With reference to the five above-mentioned “ideal typical” models:

1. IRPES can be linked to strategic planning and health authorities’ goal setting so that it is integral to political accountability. The IRPES provides a basis for Regions to identify priorities and to set challenging targets: it can therefore be used as a tool to sanction managers according to their performance (“hierarchy and target” governance model);

2. IRPES can be linked to the CEOs’ financial reward system. Indeed, it is largely acknowledged that reward schemes reinforce orientation and directions. Hence, performance indicators monitored and assessed by IRPES can be included in CEO schemes in order to better align CEOs’ objectives with those of the institution and of the health care system in general (“pay for performance” governance model);

3. Regions can use IRPES information to leverage reputation as an improvement tool, by publicly disclosing data to all the stakeholders of the Regional health system (“transparent public ranking” governance model). Regions can disseminate results through public events, such as press conferences, meetings and internal periodic monitoring. To wholly enable peer review mechanisms, the performance results can be recalled in different occasions, in order to stimulate professionals who are the basic operators of change. The professionals’ involvement can be stimulated by encouraging feedback from them. An effective way to achieve this is to include the PES in managerial training activities for top and middle management.

4. IRPES can be finally used as a tool to align the three above-mentioned governance mechanisms (mainly addressed to managers) to the operative units (signally, clinical
units) of the Regional health care systems. IRPES results can be integrated within Health Authorities’ organizational mechanisms, such as the budgeting process.

The integration and the adoption of all these strategies boost the opportunity to improve performance and allow more than the application of a single strategy, as demonstrated by the comparison of Lazio and Tuscany regarding hip fractures operated within 2 days (Pinnarelli et al. 2012).

We now describe how IRPES-adhering Regions have differently integrated the PES with their internal governance mechanisms.

There is a group of Regions, such as Tuscany and Basilicata, that are currently using all four strategies, and the new entries - Veneto, Emilia Romagna and Friuli – seem to be on the same track. Piedmont has also applied all four strategies but only when it participated in the network.

There is a second group, represented by Trento, Liguria and Umbria, which has adopted three of the four strategies. In particular, Trento has linked IRPES to other governance tools (strategy 1), CEO reward schemes (strategy 2) and both internal and public events (strategy 3). Full integration in the local budget process is still missing. Liguria has also applied the first three strategies but in a light and non-systematic way. Finally, Umbria has adopted the first two strategies and partially introduced IRPES in its managerial training programs.

Finally, there is a third group of Regions - Bolzano, Marche and the Aosta Valley – that have not endorsed any of the aforementioned four strategies. Table 1 summarizes the different governance models adopted by the Italian Regions in the 2007-2012 period.
Table 1 – The regional governance models in the 2007-2012 period

<table>
<thead>
<tr>
<th>Region</th>
<th>Trust and Altruism</th>
<th>Choice and Competition</th>
<th>Hierarchy</th>
<th>Pay for performance</th>
<th>Transparent public ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abruzzo</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aosta Valley</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apulia</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autonomous Province of Bolzano</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autonomous Province of Trento</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Basilicata</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Calabria</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Campania</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Emilia-Romagna</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friuli-Venezia Giulia</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lazio</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liguria</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lombardy</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marche</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Molise</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Piedmont</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sardinia</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sicily</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tuscany</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Umbria</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Veneto</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
§ 4. Methodology

To compare the results achieved by the Italian Regions with different governance models, we chose 14 performance indicators measured in 2007 and in 2012 (see Table 2). These specific indicators were chosen because they had already been validated by the pilot study the McS Laboratory coordinated in 2009, on behalf of the Italian Ministry of Health (Nuti et al. 2012). Most of the indicators had been derived from the framework already developed by the Tuscany Region or from international studies (Canadian Institute for Health Information 2001; Organisation for Economic Co-operation and Development 2003; World Health Organization 2003; Department of Health 2008; Agency for Healthcare Research and Quality 2006) and they had been selected on the basis of the following criteria (Kelley and Hurst 2006):

- relevance in terms of policy;
- scientific soundness of the indicators in terms of their validity and reliability;
- feasibility of obtaining nationally comparable data;
- ability of providing a comprehensive overview of the different levels the Italian health care system is composed of (hospital, primary and preventive care).

The national hospital discharges database for the years 2007 and 2012 was used for all the measurements on hospital and primary care dimensions; the 2007 and 2012 OsMed reports were used for the indicators on pharmaceutical care (OsMed 2007; OsMed 2012); the 2007 and 2012 national screening reports were used for the measurements on prevention (National Screening Observatory 2014). Avoidable hospitalization rates for chronic conditions from inpatients data were used as a proxy of primary care performance because of the lack of national comparable sources on territorial services (Ricketts et al. 2001).
Indicators from hospital inpatient data, when possible, were standardized according to sex and age, using the Italian residents of the year 2001 as a standard population.

Table 2 - Set of selected indicators

<table>
<thead>
<tr>
<th>Indicator Code</th>
<th>Indicator Label</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HOSPITAL CARE (H)</strong></td>
<td></td>
</tr>
<tr>
<td>H1</td>
<td>Ordinary hospitalization rate</td>
</tr>
<tr>
<td>H3</td>
<td>Percentage of medical DRG from surgical departments</td>
</tr>
<tr>
<td>H4</td>
<td>Percentage of laparoscopic cholecystectomies in day surgery or 0-1 day admissions</td>
</tr>
<tr>
<td>H5</td>
<td>Surgical essential levels of health services DRG - Standard percentage achieved</td>
</tr>
<tr>
<td>H9</td>
<td>Percentage of cesarean births</td>
</tr>
<tr>
<td>H11</td>
<td>Percentage of femur fractures operated within 2 days</td>
</tr>
<tr>
<td>H13</td>
<td>Preoperative average hospital stay</td>
</tr>
<tr>
<td>H14</td>
<td>Percentage of short medical hospitalizations</td>
</tr>
<tr>
<td><strong>PRIMARY CARE (T)</strong></td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td>Hospitalization rate for heart failure (50-74 years old)</td>
</tr>
<tr>
<td>T3</td>
<td>Hospitalization rate for diabetes (20-74 years old)</td>
</tr>
<tr>
<td>T4</td>
<td>Hospitalization rate for COPD (50-74 years old)</td>
</tr>
<tr>
<td>AF5</td>
<td>Per capita net pharmaceutical expenditure</td>
</tr>
<tr>
<td><strong>PREVENTIVE CARE (P)</strong></td>
<td></td>
</tr>
<tr>
<td>P3</td>
<td>Mammography screening extension</td>
</tr>
<tr>
<td>P4</td>
<td>Compliance with mammography screening</td>
</tr>
</tbody>
</table>

The indicators refer to two years - 2007 and 2012 - and they provide information for a pre-post comparison: IRPES was actually first developed in 2008.

All the selected indicators were considered by the IRPES Regions to have the same relevance in measuring the performance of the Regional health care system. Therefore, this set of indicators offers a preliminary overview of the differences across Regional health care performances and how they shifted in the 2007-2012 period.
In order to provide a synthetic assessment of the different Regional performances in 2007 and 2012, the 14 indicators were composed into a single indicator according to the following methodology:

- we ranked each indicator for each year (2007 and 2012) and we assigned the quintile each Region occupied for each specific indicator;
- different coefficients were then assigned: 0,2 to worst performing indicators, 0,4 to badly performing indicators, 0,6 to average performing indicators, 0,8 to well performing indicators and 1 to best performing indicators. For each Region, the weighted indicators were first summed and then divided by 14 (the total number of indicators), obtaining a synthetic performance score that hypothetically ranged from 0,2 (all the 14 indicators in the worst quintile) to 1 (all the 14 indicators in the best quintile). This procedure was applied both to the 2007 and to the 2012 indicators.

The synthetic performance score is the mean of the 14 (ranked and weighted) indicators. Despite the potential controversy of this final operation, which is based on only 14 indicators, the choice was supported by the decision of all the IRPES Regions to consider the indicators equally important and relevant to offer an overview of performance of the Regional health care systems. Moreover, according to the national legislative framework, the three health care levels – hospital, primary and preventive care – should be financed according to fixed shares (respectively: 44%, 51%, 5%), which mirror their respective importance (State-Regional Conference 2009; Presidency of the Repubblic 2011): the proportion of the indicators selected approximately reflects this balance (even if slightly overestimating hospital care relevance). Finally, it should be noted that the synthetic indicator is not conceived as a tool to rank the Regions, but as an explanatory expedient to offer an overview of their performances in the 2007-2012 period.
First, the method allows cross-Regional comparisons, regardless of the scale of each indicator, and offers a synthetic view of the Regional health care systems performances; second, it allows longitudinal comparisons (2012 vs 2007) that are not affected by different Regional starting points in 2007 and by national spontaneous dynamics, as trends are assessed in relative terms, in relation with all the ones of the Regions.

Figures 4 and figure 5 offer a graphical representation of the Regional performances in 2007 and in 2012. Figure 4 provides a detailed overview of each Region’s performance in 2007 and 2012, by listing the number of indicators according to the quintile they occupied.

The dynamics of each Region’s relative performance is displayed by figure 5: the blue line shows the 2007 score; the green line and the red line respectively portray each Region’s improvement and worsening in the 2007-2012 period.

Figure 4 – The regional performances in 2007 and in 2012
Figure 5 – The regional performance histogram

The above-mentioned methodology and its graphic representation offer some preliminary insights for comments on the impact of the different governance models adopted by the Italian Regions on performance and on the relevance of the IRPES tool. The variability of the governance models adopted by the Italian Regions – that follows the above-mentioned decentralization process – actually provides a “natural experiment” to study the association of Regional health care performances to different governance models.

Other Regional variables might be associated to health care performances and might represent a confounding factor for our analysis. Actually, the Italian Regions have historically been heterogeneous in terms of size, population, economic development, civic culture and
institutional performance, showing a sharp cleavage between the North and the South of the country (Putnam 1993; Pavolini and Vicarelli 2012; Toth 2014; Cotta and Verzichelli 2007). From an economic point of view, despite the substantial difference between the northern Regions (with a per capita income of 27,500 euros in 2012) and the southern ones (18,200 euros), such a disparity at the economic level is not reflected in their public health spending (Istituto Nazionale di Statistica 2012): as above-mentioned, since the 2000s the health care budget has been allocated among the Regions on the basis of a per capita share, partially adjusted by the age distribution of the population. Therefore, all Regions are roughly guaranteed the same per capita resources for health care (Toth 2014).
§ 5. Discussion

In the second paragraph we grouped the Italian Regions in four clusters, according to how they differently mixed the five “ideal typical” governance models we had previously outlined. We now discuss about the performance each group achieved in the 2007-2012 period, according to the methodology explained in the fourth paragraph.

As already mentioned, Lombardy is the only Region that adopted a "choice and competition" governance model. According to the 14 indicators we considered, this governance system does not seem to be associated with outstanding 2012 performances, or with exceptional improvement capabilities. Lombardy actually shows a 2012 performance that is slightly better than the other Italian Regions but proves to have (relatively) worsened its performance in comparison to 2007. In particular, hospital-related performance seems to be detrimentally affected by this governance model, both regarding appropriateness indicators (H3, H4, H5) and quality indicators (H9 and H11).

Both the Regions that had a “trust and altruism”/"hierarchy & targets” governance model (debt-rescheduling plan) and the Regions that chose to adhere to a "hierarchy & targets"/"transparent public ranking"/"pay for performance” (IRPES) show different performances, suggesting that governance models can be differently applied and interact with endogenous characteristics.

Regions with a recovery plan (group 2) generally show a poor performance in 2007 and some degree of improvement from 2007 to 2012. Hence, it seems that the strict commitment of the central government to setting targets and controlling their achievement has pushed Regions towards performance improvement. However, there are doubts on the real effectiveness of Regional recovery plans, which are more oriented towards financial expenditure and hospital performance rather than on the quality of services. Sicily and Piedmont represent two interesting cases.
Sicily registered one of the worst performances in 2007 but the Regional system was able to achieve a significant improvement from 2007 to 2012. However, the analysis of single indicators shows that improvements almost uniquely refer to hospital care indicators. Primary and preventive care, which were poor in 2007, did not significantly improve in 2012 and the same goes for pharmaceutical expenditure.

On the one hand, Sicily’s improvement may be due to the introduction of a clause in top manager contracts which required the achievement of specific performance targets linked to the national outcome evaluation program (PNE) run by AGENAS (the National Agency for Regional Health Services). Target achievement is one of the conditions needed to have appointments confirmed: the commitment to a strict “hierarchy & targets” models seems therefore to be associated to a relevant performance improvement. The Sicilian case might suggest that “hierarchy & targets” models prove to be more effective in dealing with hospital care re-organization, where structural reforms require strong political commitment, while it might be more difficult to deal with primary and preventive care.

On the other hand, these results seem to confirm the findings on the above-mentioned evaluation of Regional recovery plans made by Ferrè et al. (2012): the pluralistic nature of complex systems might require the hierarchy model to be integrated with different governance models (signally, “transparent public ranking” and “pay for performance”) that help align the different powerful players’ goals with the Regional ones.

Piedmont is a Northern Region that has generally shown high quality performances. It adhered to the IRPES network in 2008 and left it in 2010, when it entered the debt-rescheduling plan. It seems that, independently from the recovery plan, the Piedmont health care system was able to ensure increasing quality performances. Indeed, the Region proved to be able to improve hospital performances (see indicators H4, H11, H13) confirming, at the same time, its excellent primary care.

IRPES-adopting Regions (group 3) showed different internal patterns. They were, in
general, characterized by higher performances (both in 2007 and in 2012) than the Regions with recovery plans, but they showed relevant variability, especially in their dynamics. The two Regions which improved more than the others were Basilicata and Tuscany.

Regarding Basilicata, single indicators point out more balanced dynamics than the above-mentioned Sicilian Region: improvements are registered for the three assistance levels (hospital, primary and preventive care), even if a couple of hospital care indicators – referring to appropriateness – worsened (H3 and H5).

The second interesting case is Tuscany. This Region registered a high performance in 2007 and was still offering a good general assistance in 2012, even improving some hospital and primary care processes (H3, H11, H13, T2).

These two Regions (Basilicata and Tuscany) are the ones that, more than the others, have integrated the IRPES with their governance tools, combining some elements of the “hierarchy and targets” model with some of the “transparent public ranking” and “pay for performance” ones. Information provided by IRPES has been used to set HAs’ targets and define priorities, linking them with the CEO reward management systems. Basilicata and Tuscany publicly disclose their results and disseminate them at the local level through meetings and training programs for professionals.

Marche had a similar performance to Tuscany in 2007, but it shows an opposite 2007-2012 trend. It worsened its hospital care offer, regarding appropriateness and quality (H4, H5, H11). This is probably due to a continuous reorganization process carried out at the local level and it is equally associated to a different approach toward performance evaluation.

The comparison between the autonomous provinces of Trento and Bolzano probably provides the most interesting suggestions. As figures 4 and 5 show, they present a similar successful 2007 performance, but - again - with opposite trends, despite similar geographic conditions. They embraced a rather different approach toward performance evaluation: only
Trento systematically disclosed and shared data through public meetings while Bolzano only started in 2014.

The different models adopted seem to have affected hospital care appropriateness/efficiency and primary care. Bolzano seems to have relatively worsened its ability to efficiently manage its hospital processes and to divert demand towards the primary care setting. Trento jointly improved its hospital, primary care and prevention performance. Again, it might be suggested that a combination of "hierarchy & targets"/"transparent public ranking"/"pay for performance" governance models can be associated with a balanced improvement path.

Umbria and Aosta Valley did not disseminate their IRPES results and they poorly linked the system with other mechanisms, as reported in par. 2. Indeed, they registered worsening performances.

Liguria did not use the IRPES in a systematic way; the Region only slightly improved its 2007 performance.

Finally, the group of Regions which adopted a mixed model of governance (group 4) did not benchmark their results against the other Regions’ and only partially disclosed their results. Emilia Romagna, Veneto and Friuli (all of them joined the network after 2012) registered a very high performance in 2007 but they showed worsening in 2012 (with the exception of Veneto, that held its starting point position). This may suggest that the mixed model of “hierarchy and targets”/”Pay for performance” alone is not enough to keep high performing Regions improving. External benchmarking and public disclosure of data may represent a valid incentive to activate peer review processes, reputation pressure and emulation of best practices.
§ 6. Conclusions

This research draws upon the organizational autonomy Italian Regions have been granted since 2001 to evaluate if different governance models are systematically associated to different performances in the health care sector. There is no Region that endorsed a single clear-cut governance model – most of them combined the five ideal typical models laid out in paragraph 2 – but the analysis of how they combined them by differently using the Regional performance management tools (signally, the IRPES) and of the related performance results can suggest some interesting conclusions.

First, the only Region that quite clearly endorsed the “choice and competition” governance model – Lombardy – has a 2012 performance that is above the national average but proves to have (relatively) worsened its performance in comparison to 2007. The “choice and competition” governance model by itself does not seem to be associated to a sustained performance improvement.

Second, it might be suggested that – regardless of the chosen mix of governance models – external benchmarking represents a precondition to sustained improvement. In order to avoid a self-referential attitude, systematic comparison to other providers offers a powerful tool to systematically detect best practices and organizational flaws. In particular, it seems that especially high performing Regions – such as Lombardy, Emilia Romagna and Friuli Venezia Giulia – might benefit from more external reference points. IRPES can be considered a starting point in the performance evaluation process as it provides information that individual Regions cannot gather by themselves: internal benchmarking is important but it may be not enough to improve Regional performance.

Third, despite the fact that no Region has uniquely adopted a “transparent public ranking” governance model, the analysis of the different Regions suggests that public disclosure of data can be a powerful tool to drive the improvement of the health care system. This can be
explained by the specific lever that public disclosure activates: reputation. The latter can pave the way to the systematic involvement of clinicians in the improvement process by supporting best practice detection and peer review mechanisms.

Fourth, the improvement achieved by two southern Regions – Sicilia and Basilicata – proves that the coherent adoption of appropriate governance models might reduce the Italy’s geographical divide.

Finally, further research is needed to understand and analyze if and how the adoption of different governance models affects Regional health care performance, by updating available data and by examining the impacts of the IRPES on newly-adhering Regions (signally, Emilia Romagna and Friuli Venezia Giulia).
Footnotes


2. Decree Law 56/2000 abolished the national health fund and directly attributed taxes to the Regions. A “national equalization fund” was set up to counterbalance differences in Regional GDP. Regions are accountable for covering their deficit with their own resources, which include Regional taxes and co-payments for health care services.

3. The Financial Stability Law L. 296 (27 December 2006) enforced the pact (“Patto per la Salute”) that the Italian government and the Regional governments had signed on October 5th, 2006. The health care funding for the period 2007-2009 was slightly increased and extra funds were allocated to cover 2006 deficits. A specific fund was created for those Regions with high deficits (greater than 7% of the funding) (Tediosi, Gabriele, and Longo 2009; Ferrè, Cucurullo, and Lega 2012).

4. In particular, the Lombardy Regional Health Care Directorate, in collaboration with the Interuniversity Research Centre on Public Services (CRISP), started developing in 2002 a set of performance measurements to systematically evaluate the performance of health care providers in terms of the quality of care. This set of indicators comprises: (1) intra-hospital mortality, (2) mortality within 30 days after discharge, (3) overall mortality (intra-hospital plus within-30-day mortality), (4) voluntary hospital discharges, (5) readmission to an operating room, (6) inter-hospital transfer of patients, and (7) readmission for the same Major
Diagnostic Categories (MDC) (Formez 2007; Vittadini 2011; Vainieri and Nuti 2011; Berta, Seghieri, and Vittadini 2013).

5. As an example, the timeliness of the surgical treatment for femur fractured patients can be monitored in different ways, including or excluding specific segments of the population. The highest incidence of fractures refers to people older than 65 and the epidemiological perspective would focus on this cohort, measuring the percentage of femur fractures operated within two days after the hospital admission among patients older than 65. When used as a performance indicator, though, it would overlook patients younger than 65 and would potentially trigger opportunistic behaviours among clinicians, who might be incentivized to postpone interventions for younger patients in order to achieve the target. By endorsing a “managerial” perspective, IRPES chose to include all the patients in the computation of the indicator, in order to avoid this kind of “gaming” side effects.

6. The replication data set can be accessed via the electronic version of *Health Economics, Policy and Law*. 
CHAPTER FIVE - Conclusions
Concluding comments

One of my professors used to repeat us that “the role of intellectuals is more than ever seeding doubts, instead of harvesting certainties” (Bobbio 1955, 15). The main contribution of my research to the current debate is probably the critical collection of those works that reflect upon the unexpected responses to management control tools. Implications are self-evident: literature advises managers to be cautious when dealing with the design, the adoption and the implementation of management control tools – such as targets and rewards –, because the agents they are in charge directing might react in unexpected and undesired manners.

The thesis I have presented says however something more: the third paper I illustrated (“Making governance work in the health care sector: evidence from the Italian “natural experiment””) actually demonstrates at least three things:

1) management matters. Different governance models prove to be associated to very different performances. This implies that good managers have the opportunity to make a real difference;

2) despite the potential undesired consequences triggered by management tools, management control systems can be well designed, in order to support the systems’ improvements;

3) a key element to design good managerial systems is the introduction of some element of benchmarking techniques. Unhinging the self-referential attitude that any professional group potentially develops is not the panacea to any managerial dilemma, but surely helps legitimate the stimuli – unless unpopular – management gives to the organization’s agents.
Papers presented in chapter two and three (“Priorities and targets: a methodology to support the policy-making process in health care” and “How to set challenging goals and conduct fair evaluation in regional public health systems. Insights from Valencia and Tuscany Regions”, respectively) rest on the three above-mentioned assumptions: both of them are the result of the firm belief that management can make the difference. They give a contribution in designing appropriate management tools, that could minimize behavioral side effects. Both of them introduce systematic comparison as a tool to support the quantitative and rigorous methodologies they describe.

Surely, this research suffers from relevant limitations. Regarding the scale of the research, the focus on a specific sector (the health care one) and on a single administrative level (the regional one) may ingenerate some generalizability issues.

Health care organizations show very peculiar characteristics and any study relating to the health care sector may not be valid on other institutional and professional areas (Bini 2015). In particular, management control of the health care industry needs to recognize the peculiar characteristics of the agents involved, who work as “professionals” (see the Introduction): discoveries on their behavior may be not generalizable to other sectors. Moreover, this thesis focused on management controls based on the results, because results controls are dominant as a means of controlling the behaviors of professional employees: this does not mean that the behavioral implications of action controls and personnel/cultural controls are not equally if not more interesting, and might suggest the topic for further research.

With reference to the regional scale, it has already been underlined (see the Introduction) the interest that the regional level deserves for research on managerial and governance patterns, because of the different situations Italian regions show in terms of governance and managerial models. Regarding the two trajectories I drew in the Introduction – the decentralization process and the potential transition from the Old Public Administration logics to different ones (New Public Management and/or Public Governance) –, the research
I presented portrays how differently the Italian Regions positioned themselves along the two ideal typical axis (“Decentralization” and “Governance Logics”) over the last 15 years. The three studies share a common interest on the opportunities brought by the regionalization process and agree on the conclusion that these opportunities can be fully seized only by those regional administrations that are capable of structuring a coherent management control system, in order to exert the governance of their complex systems. The third paper I discussed explored the association between regional institutional and managerial characteristics and the performance they achieved in the health care sector, but further research might cast light on new elements, such as the association between the regional governance setting and the performance in other sectors, or the impact on performance of the two variables (decentralization and governance models) separately, for instance.

With reference to the regional scale, there is a second observation that emerges from this research. Some scholars deem that the role of the Italian regions in the health care sector should be compared to an “holding” one (Mapelli 2012): regions should coordinate the semi-autonomous conduct of the Health Authorities, aiming at attaining the goals set by the central level. Regardless of the different ways single regions have interpreted this role, this research seems to confirm the distinctiveness of the regional level within the Italian administrative landscape. Despite its relatively narrow viewpoint (managerial studies focused on the heath care sector), this research shows how most of the Italian regions seem to have taken up the challenge of partially abandoning the Old Public Administration model. Relying on the results of the three papers presented, it can be hypothesized – as previous studies already did (Fattore, Dubois, and Lapenta 2012) – that the Italian regional administration is currently experiencing the hybrid coexistence of some elements of the “Old Public Administration” logic (the bureaucratic reliance on a set of rules and regulations stemming from public law), the New Public Management one (the introduction of management control tools we illustrated in our papers), and the Public Governance one (the systematic insistence
on the paradigm of the network, that especially permeates the first and third paper presented).

This unique and dynamic situation of the regional level encourages further research. The study of the different combinations of the three ideal typical models the Italian Regions have chosen and the association they may have on different levels of performance (not exclusively in the health care sector) could surely be an interesting topic to be investigated.

Regarding the scientific approach I have endorsed, I am aware that more sophisticated expertise on cognitive, sociological and anthropological studies would bring more sound and original results.

These limitations pave the way for future research opportunities: management studies have been the cradle of the most disruptive theories in social sciences and have never been afraid of interdisciplinarity. Surely, they will be able to benefit from the revolution behavioral studies are stimulating and such a dynamic sector as the health care might be the appropriate laboratory to test a reframe of managerial theories and practices.
References


Taylor & Francis Group.


ACKNOWLEDGMENTS

“No Man is an Island”, John Donne declaims in a famous poem. Surely, no researcher is an island.

Each step of my research activity has been the result of a common stroll, walked with my MeS-mates. This is why I want to express my deep gratitude to Alessandro, Andrea, Anna, Anna Cristina, Anna Maria, Antonella, Armando, Barbara & Barbara, Beatrice, Beppe, Bruna, Chiara & Chiara, Cinzia, Cristiana, Domenico, Elena & Elena, Emiliano, Emmanuel, Fabio, Federica & Federica, Francesca & Francesca & Francesca, Francesco, Gaetano, Gianluca, Giorgio, Giovanna, Giulia & Giulia, Laura, Linda & Linda, Luca, Manila, Manuela, Maria Sole, Massimiliano, Milena, Ilnicco, Nicola, Pierluigi, Pietro, Sabina & Sabina, Sara, Serena, Silvia & Silvia & Silvia & Silvia, Valentina and Tommy.

It is all because of you that this thesis sees the light.