



ELSEVIER

Contents lists available at ScienceDirect

Health Policy

journal homepage: www.elsevier.com/locate/healthpol



How to set challenging goals and conduct fair evaluation in regional public health systems. Insights from Valencia and Tuscany Regions

Milena Vainieri^{a,*}, Federico Vola^a, Gregorio Gomez Soriano^b, Sabina Nuti^a

^a *Laboratory of Management and Healthcare, Institute of Management, Scuola Superiore Sant'Anna, Pisa, Italy*

^b *Conselleria de Sanidad Universal y Salud Pública, Valencia Region, Spain*

ARTICLE INFO

Article history:

Received 11 August 2015

Received in revised form 9 September 2016

Accepted 18 September 2016

Keywords:

Incentives

Targets

Health care sector

Benchmarking

ABSTRACT

The definition of “the right targets” and the way the evaluation of results is performed affect the willingness to commit to new challenges, which is a factor that influences the relationship between goal setting and performance results. Indeed, some authors claim that the choice of an inappropriate goal-setting procedure is a major cause of failure of management control systems. Goal setting theorists found that assigning a 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: (a) what is to be considered a challenging goal and (b) what is a “fair evaluation” in the health care sector. This work suggests that benchmarking is a valid support to solve the previous dilemmas. 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 about their achievement. Although these Regions adopted different governance models, both of them applied very similar techniques, which seem to be associated to an improvement of their performance and a reduction of unwarranted variation.

© 2016 The Authors. Published by Elsevier Ireland Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

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 the 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 the WHO in mid-80s [1], that set targets to member states and renewed them in the mid-1990s with the *Health21 policy framework* [2], flawed in some countries and in some areas [3]. Scholars that analyzed this case [4] stated that some strategies were not met because of: the lack of involvement of key actors at the grass-roots levels; the shift of power and responsibilities from the central to the regional level [5]; the lack of the “right targets” in terms of prioritization, reflecting the specificity of countries and in terms of identification of the correct effort to be required. All these elements are also found in general literature on performance management [6]. Indeed, the adoption

* Corresponding author at: Laboratorio Management e Sanità, Istituto di Management, Scuola Superiore Sant'Anna di Pisa, Piazza Martiri della Libertà, 33, 56127 Pisa.

E-mail address: m.vainieri@sssup.it (M. Vainieri).

<http://dx.doi.org/10.1016/j.healthpol.2016.09.011>

0168-8510/© 2016 The Authors. Published by Elsevier Ireland Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

of an inappropriate goal-setting procedure is deemed to be a major cause of failure of management control systems [7].

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) [8]. Regarding the health care sector, both central and regional levels use targets in their governance models in different ways, getting different results [9–11].

At this purpose, Brown et al. believe 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) [12].

In this scenario, control management studies mainly discussed which indicators should be selected, the criteria to choose them [13–16] and some elements of the process, in particular the importance of feedback and involvement [6,17,18]. 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 [8,19–24]. Yet, few evidence and proposals concern the definition of what is “challenging”, that is an important characteristic goals should have to motivate workers [25].

The definition of “the right targets” and the way the evaluation of results is performed affect the willingness to commit to new challenges, which is a factor that influences the relationship between goal setting and performance results [25]. Seeking to respond to the aforementioned two open issues, it is possible to identify at least four sub-decisions managers and policy makers need to take 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.

Goal setting procedure needs to consider whether a gold standard or a normative target exist (i). When neither the gold standard, nor the normative standard exists, then the definition of the targets often require a subjective decision. This situation can jeopardize the legitimation of goals. Nevertheless once the standard is defined, policy-makers have to decide whether to assign the same target to all units (i.e., health authorities, health departments or professionals) (ii). Homogeneous goals are often assigned to all units. This decision encounters some drawbacks. The first one occurs when the goal is set, for every unit, to the gold standard. The gold standard requires 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 by facilitating improvement because they are seductive, but they can also have a disruptive effect leading to no commitment at all [25,26].

The second drawback is what managerial literature defines the “threshold effect”. This occurs when a minimal 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 [27]. 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 and their starting point (baseline) [6,28]. Performance incentives had the greatest impact on providers whose performance was lower at baseline [29] so that policy-makers could ask more to the worst-performers, considering that the effort should be perceived as challenging but attainable. Indeed, disruptive effects seem to be more frequent in those organizations whose recent performance was low [26].

In laboratory experiments (largely applied in the goal setting theory) challenging goals are usually considered to be those that are fixed at the 90th percentile of the original distribution, while in field experiments “challenging” is what agents perceive as “difficult yet attainable” goals [25]. That implies that the definition of what is challenging is set, most of the time, on subjective basis.

Finally, the evaluation of the level of target attainment by each agent (iv) can correspond to 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 [6,8,30].

This paper supports the thesis that the introduction of some benchmarking techniques might be the solution to face the four above-mentioned issues. Indeed, benchmarking techniques have been applied in the public sector since the 1990s [28], becoming the basis for the development of management control systems as dominant form of governance in the health care sector [9–11]. The following paragraphs report the conceptual framework drawn by two European experiences – Valencia in Spain and Tuscany in Italy –, that suggest how benchmarking techniques can be leveraged to set appropriate targets and conduct fair evaluation of their achievement.

2. Methodology

The paper offers a comparison of the methods two regional institutions – Tuscany (Italy) and Valencia (Spain) – independently developed to set appropriate targets to their health care units and to assess their attainment. The study is the result of a longitudinal action research process. The *action research approach* is a research method that aims to simultaneously solve ‘real’ problems in social systems and contribute to the basic knowledge of social

science. The distinctive stronghold of *action research* is that the researchers are involved along the flow of life of the case organization, in close collaboration with its members. This has good potential for producing both practically relevant and theoretically interesting contributions [31–35].

The first output of action research is an idea for change or a design of a solution to the problems faced by the host organization, both of these typically jointly developed with the members of the organization. In practice, this usually means the researcher's participation in a project team in charge of taking care of a change project [36]. Action research also includes the testing of the ideas for change, typically by teaming up with the members of the host organization and by supporting the implementation of new solutions. Hence, organizational change (or at least an attempt to accomplish that) is an important output of this kind of study designs. This close collaboration enables the collection of research material that cannot be usually retrieved by other approaches [31].

The authors of this paper directly cooperate with the regional public agencies that set targets for the health care authorities (for Tuscany) and for the health care professionals (for Valencia). In both cases, the spiral action research process suggested by Berg [37] were applied. Researchers participated in the periodical meetings organized by the institutional bodies in charge of the definition and evaluation of the health care units goals; those meetings were summarized in internal reports and in regional public acts, while single meetings and exchanges were recorded in researchers' notes.

With reference to the Tuscany case, Laboratorio MES of Scuola Superiore Sant'Anna (Pisa) has been actively collaborating with the regional health care administration since 2004. In particular, researchers have been cooperating with the health care regional department in defining a method to set appropriate goals to the regional health authorities, quarterly monitoring the targeted indicators and in helping evaluate the goals' achievement.

Regarding Valencia Region, data collection was made possible by the direct involvement of one of the authors in the daily activity of the Agencia Valenciana de Salud (AVS) as Chief Information Officer.

Thanks to their direct involvement, the authors have been able to collect first hand research material about the procedures the two institutions adopted to set challenging goals, to observe the application of the proposed method and to intervene along the process from the targets definition to the assessment phase. Quali-quantitative data are therefore the main output of the long-time collaboration between the authors and the Tuscany Region/Agencia Valenciana de Salud. In particular, the analyzed period for Tuscany goes from 2007 to date while for Valencia the period runs from 2007 to 2011. Data and information drawn from action research are complemented by regional public acts and regional internal reports for Tuscany; regional strategic plans and internal reports for Valencia.

The epistemological limitations of action research have been thoroughly investigated [38]. We acknowledge that a two case comparison lacking a formalized procedure to collect and interpret data may raise some relevant concerns

about internal and external validity. Nonetheless, the topic this paper inquires remains largely unexplored and we believe that our study can make an important contribution both in terms of scientific novelty and policy implications.

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. The method 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 Ref. [15]), representing the goal and the group of peer units (units with similar mission, such as teaching hospitals or focused hospitals).

The essential ingredients of the method are two: (1) 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 formula.

Considering the target setting phase, literature shows that targets have to be set on the basis of previous or actual results (the baseline), by asking for an inverse effort related to the goal standard or the best performers: [6,28,39] the greater improvement is required to the units with the poorer performance, whilst a less challenging improvement to those that already registered a good performance.

This can be 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 targets of two units; (3) drawing the line between the two targets; (4) calculating the expected values 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. The two cases we studied consider two options: (a) setting the targets of the best and the worst performers or (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). It is possible to ask the worst performer to improve up to the 25th percentile (or to decrease to the 75th percentile, if the lower the value the better). In this

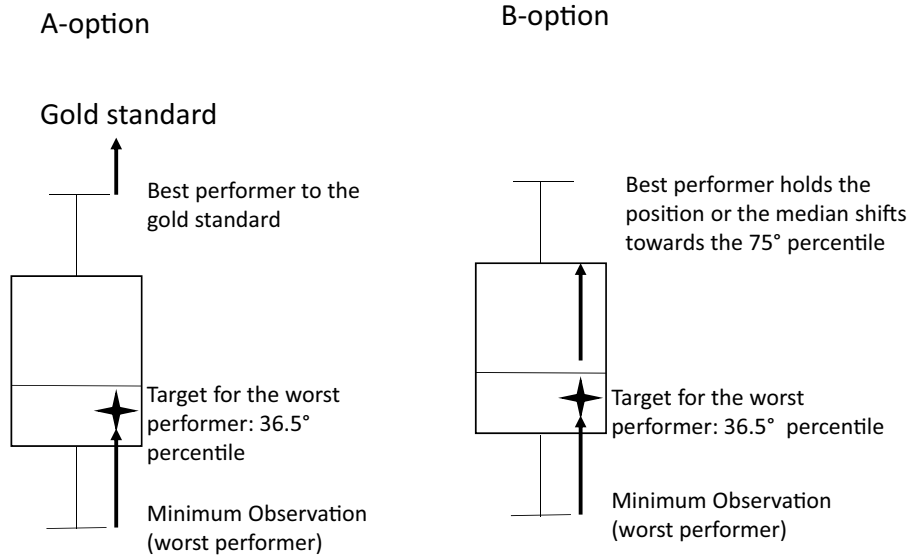


Fig. 1. The reference points, for indicators that have to increase (the higher is the better).

way, it is required to the worst performer to behave as the 1/4 of the units registering a performance lower than the median, as showed in Fig. 1.

If there is no gold standard, then the B-option is preferred. In this case, the two expected values are set to the worst and the median values. In particular, the unit reporting the worst value is assigned a target such as in the A-option: to increase up the 25th percentile or to decrease to the 75th percentile (if the lower the value the better). Regarding the second expected value, the Region (or the central government) could think to shift the median to the 75th percentile (or to the 25th percentile for indicators whose value is expected to decrease), as reported in Fig. 1.

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

In the B-option, it is possible that the expected value for the best performer(s), coming from the equation, tends to worsen the performance, instead of improving it. In this case, the A-option is to be preferred, by asking the best performer(s) to hold the value.

Step 2 determines the range of variability that the Region may consider as 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 anchoring the target definition to ex-ante performance would lead to better results and faster reduction of geographical variation (see for instance Ref. [40]).

Goal achievement evaluation is the last crucial process to be performed. 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 units 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 units: it is clear that some external factors occurred, so that the set target would no more be attainable or the selected indicator would be uncontrollable by the units;
2. The performance of the unit worsened and the relative performance is lower than the median. In this case, performance evaluation is definitely negative;
3. The performance of the unit worsened but the relative performance is higher than the median. In this case, the evaluation is not so negative and an incentive could be given according to its relative position;
4. The performance of the unit improved, but it did not achieve the set target and the relative performance is lower than the median;
5. The performance of the unit improved and the relative performance is higher than the median.

In the fourth and fifth scenarios, it is possible to apply the linear system suggested by Locke [41]: the percentage of the obtained improvement could be credited to the unit. In addition, for the fifth case, policy makers may acknowledge a bonus for the relative performance. In this case, they have to choose: the threshold for providing the bonus (the mean? The median? Another percentile up to the median? The best performance?), and the bonus share between the relative performance and the degree of improvement.

4. The two regional cases

The Spanish Health System's universal coverage is funded by taxes and it predominantly operates within the public sector. The devolution process to the Regions ended up in 2002 even if started before for some Regions [42]. Indeed, Valencia Region has had fully autonomous power in managing and organizing its health care since 1987. The Agencia Valenciana de Salud (AVS) is the public body that provides health care services to approximately 5 million inhabitants, through 24 Local Health Departments.

Since 2004, AVS has adopted a management control system based on targets and their evaluation. Since 2005, this system has also been aligned with the variable salary of all its employees and, since 2007, it has been linked to their professional career (decree 38/2007).

Objectives are declined starting from three domains: responsiveness, health care service provision and financial sustainability.

When goal-setting was first applied, most of the objectives referred to process indicators (such as average length of stay) then, in 2010, outcome and quality indicators were introduced too.

Every year, regional managers selected the indicators to 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.

The decree 38/2007 of the Department of Health reports the basic principles and actors involved in the definition of the variable pay of professionals (see yearly plans from 2007 to 2012, and the periodic strategic regional plans). Results and other information about algorithms can be derived from internal AVS reports and other Spanish documents [43]. Appendix A shows the last goals available. In 2013, the law no. 7181/2013 determined the closure of the AVS. Its functions and personnel were transferred to the Department of Health, that still monitors and assesses health units and professionals. With this change, in 2013, the new regional management decided to go backward to the traditional goal-setting procedures. The technique described in this paper is still used to manage private health care institutions. However, it seems that the Department of Health is currently reconsidering the application of the method for public institutions too.

The Italian health care system ensures universal coverage and, after the devolution process of the 90s, each Region is responsible for organizing and delivering health services [44–47]. Tuscany's health care system covers approximately 3.7 million inhabitants, delivers 95% of its services through its publicly-owned organizations, and

spends more than 6.6 billion euros in health care services per year.

In 2005, Tuscany's health care system adopted a Performance Evaluation System (PES) that consists of more than 100 quantitative indicators, publicly disclosed [48]. In 2006, the region administration decided to link the PES to the Chief Executive Officers' (CEOs) reward system. Before 2006, most of the 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 became quantitative and the weight of the financial assessment goals was reduced [49]. Above mentioned information can be found in the Regional acts that assign the goals to the health care authorities CEOs and internal documents and reports.

Every year, regional managers select the indicators that are included in the reward system, on the basis of the strategic plan and the regional priorities [16], the contextual environment and the list of indicators used at the national level to assess Italian Regions. Since the introduction of the National outcome program [50], some indicators have been relating to the health care outcomes too (such as 30 days mortality rate for AMI). Appendix A shows the list of performance goals of the last year available.

Both Valencia and Tuscany apply similar goal-setting methods, although some differences persist.

To sum up, it is possible to group the main differences in two aspects: (1) the process of communication and (2) the level of implementation. As regards the communication, Tuscany publicly discloses all the information – in order to leverage the professionals' reputation –, while Valencia carefully selects some information to be publicly disseminated and some other information to be communicated only to peers. Concerning the level of implementation, Tuscany assigns health care goals to the health care authorities (formally represented by their CEOs), whilst Valencia has a centralized and pervasive system, which sets targets and assesses them not only at the macro level (health departments), but also at the micro level (professionals).

Moreover, there are differences in the technical methods the two Regions adopted. For the target-setting phase, Tuscany uses a global performance goal and an overall improvement goal, with the aim to motivate health authorities to pay attention to all the indicators, in order to reduce potential output distortions [9,51]. Regarding 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 accepts a small variation to hold the same position). In particular, the three components that are applied to the evaluation phase are those of Section 2 [43]: (1) the degree of achievement of the set target (linear component); (2) the performance improvement or worsening (asymptotic component) and (3) the relative performance, compared to similar Local Health Departments' one (exponential component). The performance improvement is corrected by a factor ranging from 0.5 and 1.5 from an arctangent function, with the aim of reducing or increasing the value of the objective by 50%, depending on whether

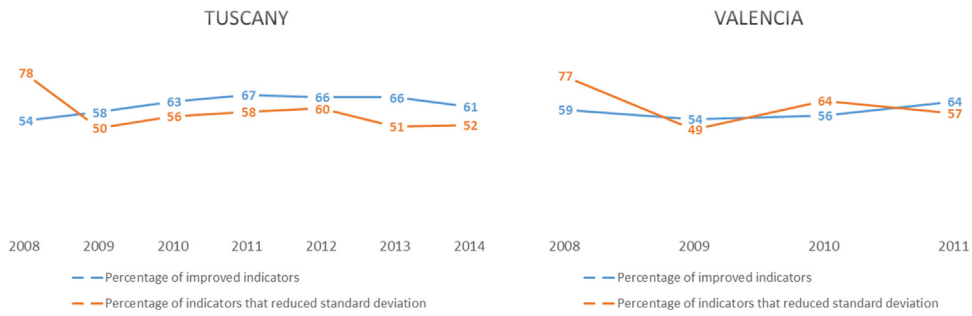


Fig. 2. The yearly percentage of performance improvement and reduction of variation in Tuscany and in Valencia.

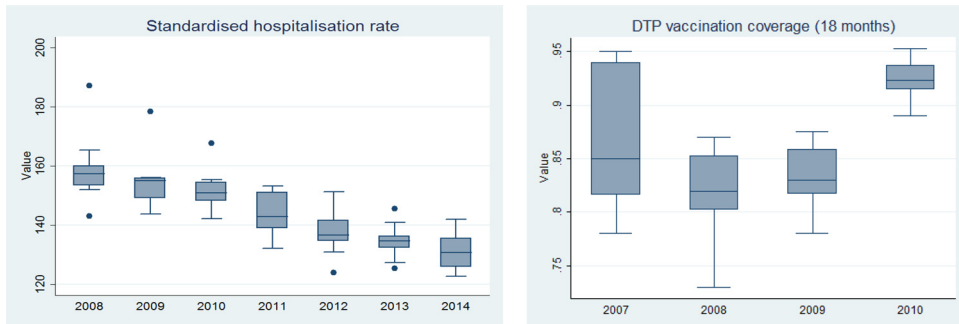


Fig. 3. Standardised hospitalization rate in Tuscany; DTP vaccination coverage in Valencia.

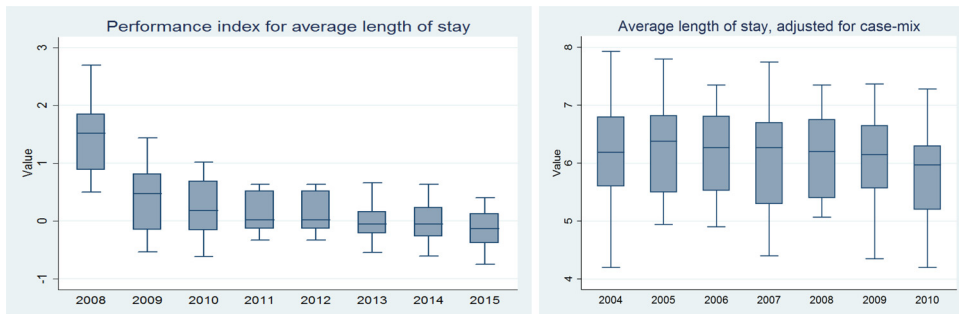


Fig. 4. Hospital length of stay.

a deterioration or an improvement occurs. 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 rescale all the other scores.

Results obtained by the two Regions seem rather encouraging. On the one side, the proposed method requires some effort and stimulates technical competences by the Region/supporting bodies; on the other side, it eases the communication process between evaluators and evaluated. Since this technique was applied, the claims about goal unfairness (both from the worse and from the best performing units) has heavily reduced.

Both in Valencia and in Tuscany, this technical framework helped make the mechanism of financial incentives for CEOs (Tuscany) or employees (Valencia) credible and acceptable. It led not only to continuous improvement

but also to strategy alignment between the Region and the Health Departments. Finally, most of the indicators reported a reduction of variability in the performance of the units analyzed [48,52]. The following descriptive statistics can offer some insights about the implications of the adoption of a coherent managerial strategy (that includes the goal-setting process proposed in this paper) on performance and variation: the percentage of improved indicators has varied across years between 54% and 67% in Tuscany; 54% and 64% in Valencia (see Fig. 2). In addition, more than 50% of indicators reduced their variability every year.

The list of indicators of the last year available are reported in Appendix A, while some examples are shown below. In particular, Fig. 3 reports the results obtained by the 12 Tuscan Local Health Authorities in the standardized hospitalization rate from 2008 to 2014 and results obtained by the 24 centers that have to monitor whether

0–6 months babies have completed all the requested exams in outpatient primary care visits (2008–2010). For the Tuscan case, the normative standard for hospitalization rate was 120×1000 residents (DGRT 1235/2012), while for the Valencian case (DTP vaccination coverage) the gold standard is 100% (Fig. 3).

As Fig. 4 shows the Tuscan Local Health Authorities have steadily reduced their hospitalization, while the reduction of variation is unstable, because of the different speed in improving performance; however, in the long run, every LHA is aligning to the best performers, as 2013 results show. In the three years analyzed for Valencia (2008–2010), the majority of units improved, reducing differences across them.

Fig. 4 reports another example, for indicators that do not have a standard: the average length of stay. It is calculated differently by Valencia and Tuscany: Valencia considers the average DRG weight of the hospital, while Tuscany considers the average number of days above or below the regional average length of stay (per DRG).

In the Tuscan case, the average length of stay has reduced on average of one day: the 16 units (12 Local Health Authorities and the 4 teaching hospitals) moved from a 2008 median of up 1.5 days to less than 0 in 2015. The reduction of variation in this case is more evident. In the Valencian case, the average length of stay reduced of 0.3 days on average from 2006 to 2010. In this case, the reduction of variability is less evident.

5. Discussions and conclusions

The two cases this paper describes aim to help solve the dilemma of “how challenging is challenging” and “how to conduct a fair assessment of the target achievement”. The methodologies adopted by the two Regions share at least two components: performance benchmarking (*relative* performance) and performance baseline (*past* performance). They both integrate a *cross-sectional* perspective (benchmarking) with a *longitudinal* one (past performance).

On the one side, 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.

On the other side, following the assumption that incremental goals can be more motivating than radical changes, the technical framework (Section 3) proposes to set targets by asking for an inverse effort on the basis of past performance. The expected consequence is a high 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, as both of them register performance improvement and variability reduction. The process we described is aimed not only at guaranteeing high-quality services (defined by the chosen targets), but also at reducing the distance between best and worst performers: it is therefore designed to support Regions in coping with potential inequalities in services.

Moreover, reports of the Ministry of Health on the performance of the Italian Regions to guarantee essential level

of care from 2007 to 2014 highlight that Tuscany steadily improved its performance results (compared to the other Italian Regions) becoming the best performing Region in 2013 and in 2014, also registering the highest level of improvement.

Results obtained by the two Regions (both in terms of performance and in terms of variability) might have significant implications, if we consider that Tuscany and Valencia have adopted very different governance models. If we take up Bevan and Wilson's classification [11], Valencia can be defined to have adopted the centrally driven “hierarchy and targets” model, with some characteristics that are linked to the “choice and competition” model. Actually, the public-private partnership that was introduced into the system enables quasi-market mechanisms. Tuscany, instead, combines the “hierarchy and targets” model with public ranking [9]. Despite different governance models, the technical solutions 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 seems to be useful and adaptable to different contexts and seems to be applicable to different levels (CEOs, heads of departments, individuals).

The two case studies we described differ in a relevant characteristic, that might deserve some further consideration: as mentioned above, the goal setting procedure is addressed to the CEOs of Local Health Authorities and Teaching Hospitals in Tuscany, while it focuses on the health departments and on professionals in Valencia. Hence, the 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 shared performance measurement system already in use.

It is beyond the scope of this paper to inquire whether the application of the goal setting techniques we discussed prove to be differently effective according to the different agents it is addressed to (CEOs, heads of departments, professionals). They are only a part of the Tuscan and Valencian managerial strategies and further research is needed to understand if (and how much of) the positive aforementioned preliminary results can be attributed to the methods applied in the two Regions. However, we want to emphasize that the underlying assumptions that some of the recommendations this paper infers – the importance of considering the baseline in the goal-setting procedure, for instance – can be generalized across the different levels. Moreover, the preliminary evidence of the effectiveness of this method are influenced also by the application of broader governance tools.

We deem it necessary to dedicate some conclusive remarks to the reorganization that affected the AVS in 2013. This specific event gives us the opportunity to highlight the need of systematically combining the scientific accuracy of the management tools with their transparent disclosure: allegedly, the main reason behind the choice of dismissing the goal-setting procedure described in the paper and to opt for a simpler one was the need to adopt a methodology that could be easily understood by the new management of the department of health, that took in charge the definition of goals and its assessment in

2013. The policy-makers mainly embraced transparency to leverage the professionals' and the managers' reputation in improving performances, but it also could have a (positive) side-effect: once granted, transparency cannot be easily withdrawn. This means that it anchors the policy-makers themselves to an irrevocable public accountability logic. Maybe a different choice of Valencia on public disclosure could have acted as a deterrent to go backward to the traditional way of setting targets.

Indeed, public disclosure of data contributes to fully exploit the opportunities a performance evaluation tool (target setting, performance measurement system or evaluation) might offer. Once all the stakeholders have become familiar with the regular use of the evaluation tool, the same endorsement by policy-makers may become less relevant, as the legitimization would come from the stakeholders' expectations themselves. It therefore turns into a "common language" that the various agents adopt to be accountable with an evidence-based approach.

In conclusion, the transparent process, reinforced by public disclosure of data, can convert the technique from a governance tool in the policy-makers' hands into an "open asset", at the service of all the stakeholders that jointly constitute the health care system.

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 Barbara Bini, for her precious contribution in collecting some materials.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.healthpol.2016.09.011>.

References

- [1] World Health Organization. Ottawa charter of health promotion. World Health Organization; 1986.
- [2] World Health Organization. The health for all policy framework for the WHO European Region. Health21; 1999.
- [3] Wismar M, McKee M, Ernst K, Srivastana D, Busse R. Health targets in Europe. *Observatory Studies Series* 2008:13.
- [4] Busse R, Wismar M. Health target programmes and health care services—any link? A conceptual and comparative study (part 1). *Health Policy* 2002;59:209–21.
- [5] Decentralization in health care. Saltman RB, Bankauskaite V, Vrangbaek K, editors. *Strategies and outcomes*. London: Open University Press; 2007.
- [6] Ferreira A, Otley D. The design and use of performance management systems: an extended framework for analysis. *Mana* 2009;20:263–82, <http://dx.doi.org/10.1016/j.mar.2009.07.003>.
- [7] Ittner CD, Larcker DF. Coming up short on nonfinancial performance measurement. *Harvard Business Review* 2003:81.
- [8] Locke EA, Latham GP. *A theory of goal setting and task performance*. Englewood Cliffs, NJ: Prentice Hall; 1990.
- [9] Nuti S, Vola F, Bonini A, Vainieri M. Making governance work in the health care sector: evidence from a natural experiment in Italy. *Health Economics, Policy and Law* 2015;11:17–38, <http://dx.doi.org/10.1017/S1744133115000067>.

- [10] Le Grand J. *Motivation, agency and public policy: of knights and knaves, pawns and queens*. Oxford, UK: Oxford University Press; 2003.
- [11] Bevan G, Wilson D. Does naming and shaming work for schools and hospitals? Lessons from natural experiments following devolution in England and Wales. *Public Money & Management* 2013;33:245–52, <http://dx.doi.org/10.1080/09540962.2013.799801>.
- [12] Brown A, Closson T, Sullivan T, Baker GR. The journey toward high performance and excellent quality. *Healthcare Quarterly* 2012;15:6–9.
- [13] Pencheon D. *The good indicators guide: understanding how to use and choose indicators*. London: NHS Institute for Innovation and Improvement; 2007.
- [14] Gagliardi AR, Fung MKF, Langer B, Stern H, Brown AD. Development of ovarian cancer surgery quality indicators using a modified Delphi approach. *Gynecologic Oncology* 2005;97:446–56, <http://dx.doi.org/10.1016/j.ygyno.2004.12.059>.
- [15] Jacobs R, Goddard M, Smith PC. Public services: are composite measures a robust reflection of performance in the public sector. *Centre for Health Economics Paper* n.16: 2006.
- [16] Nuti S, Vainieri M, Vola F. Priorities and targets: a methodology to support the policy-making process in health care. *Public Money & Management*, 2017; February.
- [17] Murante AM, Vainieri M, Rojas D, Nuti S. Does feedback influence patient–professional communication? Empirical evidence from Italy. *Health Policy (New York)* 2014;116:273–80.
- [18] Vandewalle D. A goal orientation model of feedback-seeking behavior. *Human Resource Management Review* 2003;13:581–604, <http://dx.doi.org/10.1016/j.hrmr.2003.11.004>.
- [19] Latham GP, Borgogni L, Petitta L. Goal setting and performance management in the public sector. *International Public Management Journal* 2008;11:385–403, <http://dx.doi.org/10.1080/10967490802491087>.
- [20] Locke EA, Shaw KN, Saari LM, Latham GP. Goal setting and task performance: 1969–1980. *Psychological Bulletin* 1981;90:125–52.
- [21] Lunenburg FC. Goal-setting theory of motivation. *International Journal of Management Science Business and Administration* 2011;15:1–6.
- [22] Locke EA, Latham GP. New directions in goal-setting theory. *Current Directions in Psychological Science* 2006;15:265–8, <http://dx.doi.org/10.1111/j.1467-8721.2006.00449.x>.
- [23] Latham GP. *Work motivation: theory, research and practice*. Thousand Oaks, CA: Sage; 2007.
- [24] Mitchell TR, Daniels D. Motivation. In: Borman WC, Ilgen DR, Klimosky RJ, Weiner IB, editors. *Compr. Handb. Psychol. Ind. Organ. Psychol.* New York: Wiley & Sons; 2003. p. 225–54.
- [25] Locke EA, Latham GP. *New developments in goal setting and task performance*. New York and London: Routledge Taylor & Francis Group; 2013.
- [26] Sitkin S, See K, Miller C, Lawless M, Carton A. The paradox of stretch goals: organizations in pursuit of the seemingly impossible. *Academy of Management Review* 2011;36:544–66, <http://dx.doi.org/10.5465/AMR.2011.61031811>.
- [27] Bird S, Cox D, Farewell V, Goldstein H. Performance indicators: good, bad, and ugly. *Journal of the Royal Statistical Society* 2005;168:1–27.
- [28] Anthony R, Young D. *Management control in nonprofit organizations*, VI ed. Boston, MA: McGraw-Hill; 1999.
- [29] Greene J, Hibbard JH, Overton V. Large performance incentives had greatest impact on providers whose quality metrics were lowest at baseline. *Health Affairs* 2015;34:673–80.
- [30] Flamholtz EG, Das TK, Tsui AS. *Toward an integrative framework of organizational control*. Accounting, Organizations and Society 1985;10.
- [31] Jonsson S, Lukka K. There and back again: doing interventionist research in management accounting. In: Chapman CS, Hopwood AG, Shields MD, editors. *Handbooks of management accounting research*, vol. 1. Elsevier; 2006. p. 373–97, [http://dx.doi.org/10.1016/S1751-3243\(06\)01015-7](http://dx.doi.org/10.1016/S1751-3243(06)01015-7).
- [32] Susman GI, Evered RD. An assessment of the scientific merits of action research. *Administrative Science Quarterly* 1978;23:582–603, <http://dx.doi.org/10.2307/2392581>.
- [33] Lewin K. Action research and minority problems. *Journal of Social Issues* 1946;2:34–46, <http://dx.doi.org/10.1111/j.1540-4560.1946.tb02295.x>.
- [34] Reason P, Bradbury H, editors. *Handbook of action research*. London–Thousand Oaks–New Delhi: SAGE Publications; 2006.
- [35] Rapoport RN. Three dilemmas in action research: with special reference to the Tavistock expe-

Please cite this article in press as: Vainieri M, et al. How to set challenging goals and conduct fair evaluation in regional public health systems. Insights from Valencia and Tuscany Regions. *Health Policy* (2016), <http://dx.doi.org/10.1016/j.healthpol.2016.09.011>

- rience. *Human Relations* 1970;23:499–513, <http://dx.doi.org/10.1177/001872677002300601>.
- [36] Eero K, Kari L, Arto S. *Journal of Management Accounting Research* 1993;5:243–64.
- [37] Berg BL. *Qualitative methods for the social scientist*. Boston: Pearson Education; 2004.
- [38] Mctaggart R. *Participatory action research: issues in theory and practice*. *Educational Action Research* 1994;2:313–37.
- [39] Drucker P. *The practice of management*. New York: Harper & Row; 1954.
- [40] Nuti S, Seghieri C. Is variation management included in regional healthcare governance systems? Some proposals from Italy. *Health Policy (New York)* 2014;114:71–8, <http://dx.doi.org/10.1016/j.healthpol.2013.08.003>.
- [41] Locke EA. *Linking goals to monetary incentives*. *The Academy of Management Executive* 2004;18.
- [42] García-armesto S, Abadía-taira MB, Hernández-quevedo C, Bernal-delgado E. *Health Systems Review* 2010;12.
- [43] Gómez G, Carrillo E. *Evaluación del cumplimiento de objetivos en los sistemas de incentivación del desempeño*. Barcelona: 2009.
- [44] Ferré F, Giulio A, Valerio L, Longhi S, Lazzari A, Fattore G, et al. *Health system review: Italy*. *Health Systems in Transition* 2014;16:1–168.
- [45] Fattore G. Clarifying the scope of Italian NHS coverage. Is it feasible? Is it desirable? *Health Policy (New York)* 1999;50:123–42, [http://dx.doi.org/10.1016/S0168-8510\(99\)00068-8](http://dx.doi.org/10.1016/S0168-8510(99)00068-8).
- [46] France G, Taroni F, Donatini A. The Italian health-care system. *Health Economics* 2005;14:187–202, <http://dx.doi.org/10.1002/hec.1035>.
- [47] Tediosi F, Gabriele S, Longo F. Governing decentralization in health care under tough budget constraint: what can we learn from the Italian experience? *Health Policy (New York)* 2009;90:303–12, <http://dx.doi.org/10.1016/j.healthpol.2008.10.012>.
- [48] Nuti S, Seghieri C, Vainieri M. Assessing the effectiveness of a performance evaluation system in the public health care sector: some novel evidence from the Tuscany region experience. *Journal of Management & Governance* 2012, <http://dx.doi.org/10.1007/s10997-012-9218-5>.
- [49] Nuti S. *La valutazione della performance in sanità*. Bologna: Il Mulino; 2008.
- [50] Agabiti N, Davoli M, Fusco D, Stafoggia M, Perucci CA. *Valutazione comparativa di esito degli interventi sanitari*. *Epidemiologia e Prevenzione* 2011;(Suppl. 1):1–80.
- [51] Bevan G, Hood C. What's measured is what matters: targets and gaming in the English public health care system. *Public Administration* 2006;84:517–38.
- [52] *Revista de Medicina de Familia y Atención Primaria*. Vol. 15(1 Suppl);50 p. ISSN: 1989-6832.