

# Insights on the effectiveness of reward schemes from 10-year longitudinal case studies in 2 Italian regions.

Milena Vainieri  | Daniel Adrian Lungu  | Sabina Nuti 

Laboratorio Management e Sanità—Institute of Management, Scuola Superiore Sant'Anna of Pisa, Pisa, Italy

## Correspondence

Milena Vainieri, Laboratorio Management e Sanità—Institute of Management, Scuola Superiore Sant'Anna of Pisa, Piazza Martiri della Libertà, 33, 56127 Pisa, Italy.  
Email: milena.vainieri@santannapisa.it

## Summary

**Background:** Pay for performance (P4P) programs have been widely analysed in literature, and the results regarding their impact on performance are mixed. Moreover, in the real-life setting, reward schemes are designed combining multiple elements altogether, yet, it is not clear what happens when they are applied using different combinations.

**Objectives:** To provide insights on how P4P programs are influenced by 5 key elements: *whom*, *what*, *how*, *how many* targets, and *how much* to reward.

**Methods:** A qualitative longitudinal analysis of 10 years of P4P reward schemes adopted by the regional administrations of Tuscany and Lombardy (Italy) was conducted. The effects of the P4P features on performance are discussed considering both overall and specific indicators.

**Results:** Both regions applied financial reward schemes for General Managers by linking the variable pay to performance. While Tuscany maintained a relatively stable financial incentive design and governance tools, Lombardy changed some elements of the design and introduced, in 2012, a P4P program aimed to reward the providers. The main differences between the 2 cases regard the number of targets (*how many*), the type (*what*), and the method applied to set targets (*how*).

**Conclusion:** Considering the overall performance obtained by the 2 regions, it seems that *whom*, *how*, and *how much* to reward are not relevant in the success of P4P programs; instead, the number (*how*

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2018 The Authors. The International Journal of Health Planning and Management published by John Wiley & Sons Ltd.

*many*) and the type (*what*) of targets set may influence the performance improvement processes driven by financial reward schemes.

#### KEYWORDS

incentive, performance, reward

## 1 | INTRODUCTION

Several recent reviews collected studies on pay for performance (P4P) programs and reported mixed results regarding the effects of such programs on performance.<sup>1</sup> The findings showed a significant degree of heterogeneity ranging from positive significant improvement to no impact at all. Moreover, in certain cases, the adoption of P4P has even damaged the prior performance. For instance, the 2011 Cochrane review in primary care found that financial incentives were effective for some outcomes in some settings.<sup>2</sup> Hence, evidence suggests that P4P schemes seem to work under certain circumstances and that both intended and unintended consequences have to be carefully estimated before implementing them on a wide scale, also considering that benefits should always outweigh costs.<sup>3,4</sup>

What makes the success of a P4P program should include several different aspects. A coherent and effective structure should, among others, be able to answer to at least 5 relevant questions: (1) Whom to reward? (2) What to reward? (3) How to reward? (4) How much to reward? (5) How many targets to set?.<sup>1,4-6</sup> In general, scholars reported that the adoption of inappropriate goal-setting procedures is deemed to be a major cause of failure of management control systems.<sup>7</sup>

Although there are suggestions coming from the extant literature, some scholars stated that studies have not investigated which specific design features mainly contributed to obtain the desired effect.<sup>1</sup>

As regards the *whom* to reward, the decision is whether to reward individuals (clinicians, clinical teams or General Managers [GMs]) or organisations. In theory, rewards should be granted to those who are responsible for improved performance, so that it is believed that they are more effective when addressed to individuals. Indeed, most schemes have focused on rewarding groups of practitioners, especially those of the primary care practice. Reviews have found a positive relationship between performance and rewards addressed to individuals.<sup>8-13</sup>

Although individual-level incentives have a high level of accountability, group-level or organisational incentives can be preferred given that the health care sector is characterised by a joint production nature.<sup>3</sup> On the contrary, the main limitation is that they might not translate into private rewards for individuals.<sup>10</sup>

Various studies found that macro level rewards have a positive but poor effect in terms of performance improvement<sup>14-17</sup> or no effect at all.<sup>18</sup> A recent study in Denmark highlighted that institutional level incentives worked when the rewards granted to the department level have been redistributed internally.<sup>5</sup> Hence, individual incentives are more effective, but some targets can be set at a higher level (group or organisation) only. Moreover, institutional targets work better when they are drilled down into the organisations. Therefore, P4P programs addressed to GMs might be effective because they are assigned individual reward schemes, but they may include institutional targets because the GM is the person in charge of the whole health care organisation.

As regards the *what* to reward, there is evidence on the positive effects of P4P when it includes process measures or measures with more room for improvement, rather than solely outcome measures.<sup>1,19</sup> This evidence is related to the so-called "controllability principle" that "is one of the strongest tenets in management accounting, and is considered to be directly relevant to the evaluation of managers' performance."<sup>20</sup> The principle states that indicators used to assess performance, especially in the incentive plans, should be controllable by the workers or the subjects called to achieve the targets, otherwise motivation decreases, and dysfunctional behaviour may occur.<sup>21</sup> The risk of focusing on controllable indicators only is that while individuals or institutions hit the target, the overall health system may miss the point. Hence, management scholars suggest introducing global performance measures in order to coordinate improvement efforts and promote cross-functional cooperation.<sup>22</sup>

Another suggestion derives from the goal setting theory, and it is related to the importance to set specific goals instead of the exhortation to “do one's best.”<sup>23</sup> This statement seems quite obvious, but, in practice, it is not. Indeed, an Italian study on GMs' financial reward schemes highlighted that, in 2011, 36% of the regions mainly applied qualitative indicators which are more difficult to be interpreted and evaluated.<sup>24</sup>

The higher success of process indicators is related to the fact that, if well defined, they are easily taken under control, either by organisations or by individuals. However, indicators should be carefully selected in order to meet the ultimate goal of health care system, which is population (and/or patients) health status.

As for the *how* to reward, authors identified 2 different approaches: purely positive or competition-based rewards.<sup>25</sup> Purely positive rewards consist into additional financial resources for those actors who meet the criteria requirements, while the competitive approach foresees a reward for the best performers and a punishment for the worst ones. Van Herck et al (2010) conclude that incentives of a purely positive nature seem to have generated more beneficial effects than those based on a competitive approach. Moreover, the UK experience of the QOF suggested that targets should be carefully grouped into categories,<sup>26</sup> while other experiences suggest tailoring target setting and evaluation on the basis of both past and relative performance.<sup>27</sup>

Literature also converges towards the use of frequent rewards and participative goal setting<sup>1,23,28</sup>: when goals are set in a participative manner and rewards are frequent, success is higher.

With reference to *how much* to reward, the determination of the amount of public resources that should be allocated to financial bonuses is an important issue for policy makers.

Size (*how much*) can be expressed both in percentage and in absolute terms. A systematic review by Rosenthal et al (2004) summarised 38 different P4P programs and reported that the magnitude of incentives for organisations generally ranges from approximately 1 to 10 per cent of the annual budget. On the one hand, small rewards are not sufficient to put in place mechanisms able to boost motivation, while on the other hand too big rewards are not effective because of the perverse effects that trigger opportunistic behaviours.<sup>25,29-32</sup> However, reviews found no clear relationship between incentive size and the outcomes of P4P programs.<sup>6,33</sup> It is not possible to generalise how much is enough, indeed, in some cases the size of the reward could be higher than the transaction costs that organisations could avoid.<sup>26</sup>

Another difficult issue to cope with during the target setting phase is to determine *how many* targets have to be included in the evaluation. Both excessive and scarce emphasis on performance indicators can result in a performance paradox which leads to a weak correlation between performance indicators and performance itself.<sup>34</sup> The confusion generated by many targets might disorient the actors of the organisation who may then behave differently from the priority actions.<sup>35</sup> On the other hand, a limited number of targets may induce to the tunnel vision as a consequence of narrowing the managerial attention only to some limited aspects of global performance.<sup>36</sup> To cope with this dilemma, scholars have suggested different solutions, such as introducing subjectivity beside the bonus formula<sup>35</sup> or selecting priorities with different methods,<sup>37</sup> while other indicators could be used for internal benchmarking purposes only.<sup>38</sup>

Finally, some governance mechanisms such as publicly benchmarking outcomes and peer learning can motivate improved performance at both the sub-national and national level but with careful design.<sup>39-41</sup> In particular, improving results requires different approaches when the context is poor or good performing: while naming shaming and punishment may help in the case of poor performance, awards and visibility may help in the case of good performance.<sup>39-41</sup>

This paper aims to contribute to the literature by framing up arguments on the above 5 elements and identify areas for future research on the basis of 2 good performing Italian regions: Tuscany and Lombardy. We analysed the different features of the P4P programs, aimed at improving organisational performance, that the 2 regions have adopted from 2005 to 2015. Finally, we analysed their overall performance trajectory during this time frame.

## 2 | METHODS

A qualitative longitudinal analysis of 2 case studies (Tuscany and Lombardy) was conducted. The analysis was exploratory and was not meant to imply any direct causality between the features of the P4P schemes and the performance.

We chose these cases because they are considered to be 2 of the best performing Italian regions.<sup>42,43</sup> Moreover, they both adopted P4P programs together with performance measurement systems, although making different choices in terms of the scheme design.<sup>43,44</sup> To analyse the above 2 cases, we searched for international and national articles as well as grey literature and Italian documents (Regional acts). The cases were analysed highlighting the choices that the 2 regions made regarding the financial reward schemes during the time frame under investigation. We aimed to seek an answer to the 5 questions previously listed: (1) Whom to reward? (2) What to reward? (3) How to reward? (4) How much to reward? (5) How many targets to set?

For each case, we discuss the strategies adopted and the regional performance results obtained between 2005 and 2015.

As regards the regional performance evaluation, we used the national Essential Levels of Care (ELC) scores that represent the capacity of the regions to provide free of charge (or with a co-payment) care to their citizens.<sup>45-47</sup> Because the Italian GM reward scheme, by law, is set for a 20% of the base pay, we made a focus on a single indicator to better analyse "the size effect" and "how many targets." In particular, we chose the percentage of femur fractures operated within 48 hours indicator because it is included into the ELC evaluation and it is one of the main outcome indicators monitored by the National Outcome Program of the National Agency for health care since its birth in 2010.<sup>48</sup> The data go from 2007 to 2015, respectively, the first and the last year available of ELC score.

### 3 | THE CASE STUDIES

The Italian National Healthcare System follows a Beveridge model. The responsibility for the organisation and the provision of services is decentralised at a regional level, while the national level has to ensure universal coverage for the whole population.<sup>46</sup> Because regions are autonomous to choose their organisational model, they adopted different governance models and management tools.<sup>49,50</sup> However, more and more the central government has been controlling performance over regions, claiming, since 2010, an increasingly central role when regions were not able to achieve the minimum level of performance set through the ELC.<sup>51</sup>

#### 3.1 | The Tuscany case

The health care system in Tuscany delivers 95% of the services through publicly owned organisations (Local Health Authorities) and spends more than 6.6 billion euros per year in health care services.

Tuscany adopted a mixed governance model that combines hierarchy and targets, public disclosure of performance data, and P4P related to the GM's (*whom to reward*) rewarding schemes.<sup>43</sup> In particular, since 2005, Tuscany's health care system adopted a multidimensional performance evaluation system (PES) that includes 60 composite indicators and more than 600 simple indicators covering multiple performance dimensions that was based on transparent benchmarking and public disclosure of performance results.<sup>52,53</sup>

Before 2006, most of the goals were qualitative, assessed with the "all or none" criterion, and more than 50 per cent of them were based on financial performance. After the integration with the PES, most of the goals became quantitative and related to quality, appropriateness, and process indicators as well as efficiency and patients "satisfaction"<sup>27</sup> (*what to reward*).

Every year, a set of approximately 50 indicators are grouped into categories. In particular, there is a specific category that measures the capacity to improve indicators included into the PES. Hence, the reward scheme is made by a very large set of indicators: those related to the main scheme (around 50) and those included into the PES (around 600) (*how many targets to set*).

Based on the Italian legislative requirements, regions can raise up to 20 per cent of the GMs' compensation by linking it to quality and efficiency performance. In particular, considering the specific indicator chosen to understand the effect of the size, the percentage of femur fracture operated within 2 days, it was incentivized with a very low

weight: less than 0.5 over 100 (*how much to reward*). The regional administration sets the weights both for the categories and for the single indicators, discusses the annual targets with the GMs on December based on performance data of the previous 9 months, and defines specific targets of improvement. In order to set challenging goals, the definition of the targets is done on the basis of the baseline and the relative performance<sup>27</sup> (*how to reward*).

While the overall performance scores linked to the PES were annually presented during the spring months of the next year, the global assessment of the targets achievement was often delayed (up to 3 years). Hence, the financial bonus linked to the reward scheme was not prompt at all. This might have negatively affected the GMs' motivation; indeed, the delay in granting the bonus is a significant demotivation element.<sup>28</sup>

Regional acts show that, during the 9 years under investigation, the average performance-related pay was around 65% to 70% but quite variable within the health authorities, ranging from 50% to 90%.

### 3.2 | The Lombardy case

Lombardy is the sole Italian region which opted for the "choice and competition" governance model, combined together with a P4P scheme and a specific performance measurement system.<sup>43</sup> This model, based on the split between provider and purchaser, where private for profit, private not-for-profit, and public hospitals compete for resources, recently changed in 2016.

This evaluation system, based on benchmarking but without disclosing the names of providers, has been in place until 2014 when Lombardy joined the Italian collaborative network and adopted their shared PES.<sup>43,54,55</sup>

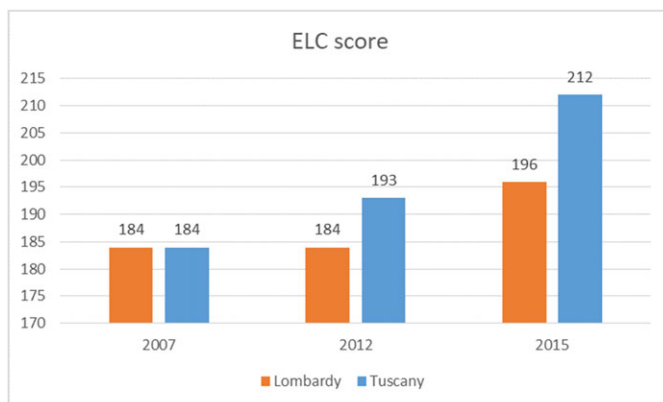
Generally, GMs' financial reward schemes consisted in 2 parts (*what to reward*): one related to their mandate and the other one related to performance results. The first part consisted into a subjective assessment of the GMs' managerial competences. The assessment was done by the regional policy makers, and until 2010 it has been attributed a weight of 40% of the total score. Since 2010, the weight has been reduced to 10%.

The number and the type of targets set by the regional administration changed over the years under investigation (*how many targets to set*). While between 2008 and 2013, GMs' financial reward schemes were based on 30 to 40 objectives expressed in qualitative terms with no clear definition of targets, since 2013, the reward scheme changed. It took into account more than 200 indicators, and despite at the beginning they continued to be mainly qualitative, an increasing number of indicators started to be expressed in quantitative terms and with a clear definition of targets.

Rewards for GMs were attributed approximately 18 months after the end of the evaluated period (*how to reward*), with a quite fast process considering the lag time related to the collection and the evaluation of data. As already seen for the Tuscany case, also in Lombardy the GMs (*whom to reward*) could receive up to 20% of their base pay (*how much to reward*) as a bonus for performance results. Considering the specific indicator of femur fracture operated in 48 hours, it was part of national requirements as the agreement between Ministry of Health and Regions since 2010 so they are included into the mandate goals for years between 2010 and 2015. In this case, no bonus is linked but a possible penalty related to the risk of losing the role (*how much*).

GMs' final assessment was often around 90% for almost all the health care organisations. Variation within the region was very low across time which may lead us to argue that it is due to the high number of qualitative indicators and to unclear definition of targets (until 2014 it often happened that indicators were defined but no target was assigned to them).<sup>24</sup>

In 2012, Lombardy introduced another reward mechanism in an experimental way to some providers (*whom to reward*). Regional managers defined the hospitals' annual budgets based on their historical expenditure and rewarded their outcomes (*what to reward*) monitored throughout the 5 indicators of the PES.<sup>56</sup> The scheme assumed a competitive approach (*how to reward*): from a penalty of -2 per cent for the worst performer to a bonus of +2 per cent for the best performer (*how much to reward*), while intermediate actors receive a percentage that is proportional to the distance between their result and the best practice. Authors who studied this P4P program found a positive correlation between the introduction of the program and 2 of the 5 indicators: the readmissions for the same condition in the Major Diagnostic Categories within 12 months from the discharge date and the transfers to a different hospital.<sup>56</sup>



**FIGURE 1** Essential level of care score 2007 to 2012 and 2015

## 4 | DISCUSSION

Both regions adopted GMs reward schemes that link pay to organisational performance. Moreover, they both registered improvements, although with different magnitudes. However, there are significant differences in the design phase of such schemes that need to be further discussed.

A first difference consists in *what* is the object of P4P schemes. While Lombardy seems to reward mainly general and qualitative goals (at least until 2014), since 2006 Tuscany focuses mainly on quantitative indicators related to quality and appropriateness.

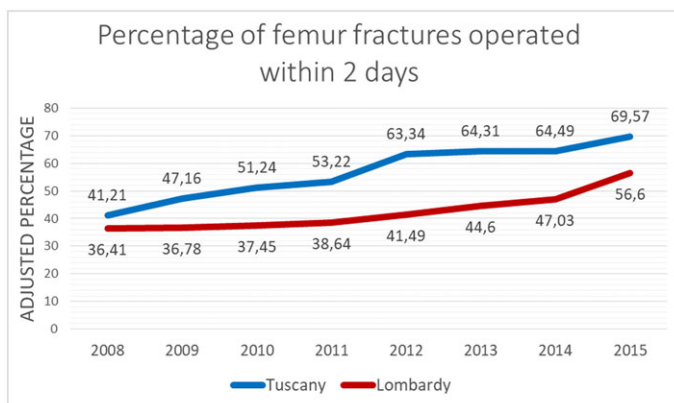
A second difference stands in the target-setting process (*how to reward and how many targets to set*): Tuscany sets a high number of goals (around 50 and the PES) based on quantitative measures and uses relative and past performance to assess them. Lombardy used to set around 40 goals that were mainly qualitative and instrumental until 2013; then, it started to introduce more and more indicators (around 200) enhancing also the number of quantitative targets.

Finally, although Tuscany shows the results promptly (around 4 months after the end of the year), it provides the rewards with a very large delay (around 3 years). Instead, Lombardy pays the rewards after 18 months. These differences could have had some effects on the capacity of policy makers to evaluate the final achievement: while Tuscan GMs achievement varied between 50% and 90%, in Lombardy most of GMs achieved 90%. The argument that Lombardy is one of the best performing regions in Italy seems to be not enough to justify the difference between the percentage of achievement given to GMs with respect to Tuscany.

Considering the ELC scores in 2007, 2012, and 2015,<sup>1</sup> it can be observed in Figure 1 that both regions started with the same score (184 points out of 225). While Tuscany steadily improved and achieved the maximum score in 2015, Lombardy slightly improved its performance only after 2012. On the basis of this evidence, we may argue that the introduction of a large set of indicators (*how many targets*) related to clinical performance (*what*), into the GMs' reward schemes has potentially helped regions to improve their performance and achieve higher scores in the ELC evaluation. The delay in paying the rewards (case of Tuscany) seems to have had no such detrimental effect on performance or it is possible that it was smoothed by the presence of prompt feedback on performance results and the adoption of a clear process of setting challenging goals (*how*).

As regards the specific performance indicator of the percentage of femur fractures operated within 2 days, we can observe from Figure 2 that both regions have experienced a constant and significant improvement over the years.

<sup>1</sup>We used these years on the basis of the availability of data and because the 2007 is the first year considered, the 2012 because it is the year before the changes introduced into the Lombardy reward schemes and 2015 because it is the last available score.



**FIGURE 2** Results obtained for the percentage of femur fracture operated within 2 days—source: National Outcome Program

Considering the case of Tuscany, we may argue that the incentive size seems to be not so relevant: despite the limited weight of this indicator (less than 0.5 on 100 of the bonus), Tuscany steadily improved from 41% to almost 70% in 2015. Considering Lombardy, we may argue that the introduction of this indicator into the GMs' mandate in 2010 (even if it was not directly mentioned), with a lag time of 2 years, led Lombardy to increase the percentage from 41% to 56%.

The experience of Lombardy with the introduction in 2012 of the P4P program at the provider level confirmed what we found in literature: that rewards on outcome are not always effective. In particular, the 2 indicators that improved could be considered intermediary outcomes (*what to reward*).

## 5 | CONCLUSIONS

A multitude of studies investigated the impact of P4P programs, obtaining mixed results. This opened the debate over whether health care systems should continue to pursue performance improvement by the means of P4P programs or not. The current research contributes to the P4P knowledge by bringing up the experience of 2 Italian regions, Tuscany and Lombardy, who both decided to reward health care performance by financial means. Although the pathways of the 2 regions were quite different in terms of both implementation strategy and context, they obtained significant beneficial effects.

Both regions managed to obtain positive effects from the adoption of individual P4P programs addressed to GMs when they included clinical performance indicators. Integrative suggestions come from the Lombardy experience on P4P at the provider level: the program seems to be more effective for intermediary outcomes than for pure outcome indicators. These findings may lead us to argue that the level of implementation is not the main determinant factor in the success of such programs (*whom to reward*) while the content of the reward is relevant (*what to reward*). On the basis of this evidence, we may argue that regions should keep on designing individual P4P addressed towards the GMs. This strategy could be more feasible in terms of costs with respect to provider incentives, although it assumes that GMs redistribute goals within the organisations throughout budget and evaluation processes advocating the positive return in terms of organisational performance.<sup>5</sup> The analyses of these 2 case studies provide evidence on the fact that, when correctly designed, GMs reward schemes seem to lead to performance improvement, answering to some doubts cast by other authors.<sup>57,58</sup>

These experiences led us conclude that the real mover across the 5 factors is *what to reward*. We believe that in order to make the organisation to perform better it is essential to obtain the engagement of the people who work into it. The adoption of indicators capable to measure the essence of their work is a key strategy to enable this

improvement process (*what to reward*). This may depend on the fact that indicators which are closer to professionals' work enable their engagement.<sup>59</sup> On this type of goal, clinicians and managers can converge, and they can both achieve the objectives they are more interested in. Indeed, some authors believe that when quality of care and appropriateness is pursued, in the long run, also the financial sustainability of the health care systems will be ensured.<sup>60-62</sup> In this perspective, the inclusion of a large set of indicators seems to be a good solution (*how many targets to set*), as suggested by the shift of Lombardy's reward scheme: when the region started to link performance of GMs to a larger set of health care goals (although not always expressed in quantitative terms and clear targets) the overall performance, measured through the ELC score, improved. Although single individuals may control a few activities, it is also true that GMs are in charge of the overall performance of their organisation, so that the goal achievement is related not so much to their direct activity but to the capacity to delegate and cascade goals and targets to the right units. This is particularly true in complex organisations where there are multiple targets and services to provide. In this sense, the controllability principle for GMs can be substituted by the capacity to communicate to their staff the goals to be pursued.<sup>63</sup> The communication can be eased if all the actors are familiar with the indicators and thus, a high number of indicators seems to be a good strategy to increase overall performance of the health care system.

Interestingly, this comparison highlighted that while literature converges on the fact that rewards should be given timely to individuals, the case of Tuscany, where GMs obtained the bonuses with 3 years of delay, frequency of the rewards seems to be not so relevant. A possible explanation relies on the fact that GMs received periodic feedback (quarterly) on performance trend of their organisations.

This opens the debate about the fact that it is not easy to disentangle improvement when it is obtained by combining several elements. For instance, in the case of Tuscany, systematic use of clinical indicators, target-setting process, benchmarking, and transparent public disclosure of results may have played an important role. Indeed, when all these governance tools are put in place, the incentive size seems not important (*how much to reward*) as the results obtained for the percentage of femur fractures operated within 2 days suggest. When GMs' goals are aligned to those of organisations and focus on quality and appropriateness measures that closer to professionals activities (*what*), it is possible to enable the mechanism of identity above all throughout reputation lever.<sup>41</sup>

This study has a number of limitations and opens to further studies. A first limitation pertains the transferability of our findings, because the study is based on 2 Italian regional cases. Both regions were considered best performing and both of them already systematically used PESs; hence, they already shared a culture of measure not so common in other regions.<sup>43,50</sup> Further qualitative analyses may be welcome to confirm our preliminary arguments. Because performance improvement is the consequence of multiple factors, the environment and the context where the reward scheme is settled may cloud some of our considerations. On the one hand, because the context may exert a significant influence on the effectiveness of performance management systems.<sup>64</sup> Indeed, because reward schemes play different roles in regions with recovery plans, it would be interesting to understand whether the analysis of such regions leads to similar conclusions. On the other hand, qualitative studies may also compare whether different results occur with regard to different governance models adopted (for a classification of governance model in health care see Bevan and Wilson<sup>65</sup>; for the application to the Italian case see Nuti et al<sup>66</sup>; and for the effect of recovery plans in Italy see Ferrè et al<sup>67</sup>). Therefore, further evidence from other Italian regions or countries would contribute to the P4P body of knowledge.

A second limitation lies on the fact that even though the arguments proposed are based on quantitative elements, our study is qualitative. Every final proposition on the 5 factors: *whom*, *what*, *how*, *how many* targets and *how much* to reward, requires to be tested on empirical basis.

For instance, scholars may analyse if there are any differences across the type of outcome or intermediary outcome included into the reward schemes (eg, whether indicators of integrated care are more difficult to achieve with respect to quality and appropriateness indicators).

A third stream of research may lie on what happens within the region: our study analysed the acts and the performance at a regional level, but much more can be done to understand the regional variations. Do all the GMs transfer the regional goals into their internal performance management systems? Do they apply the same methodology? Are there any performance variations between the organisations within a region? Indeed, these conclusions seek to



shed new light on how scholars could formulate and test alternative hypotheses on the complex and debated topic of how to set effective reward schemes in health care.

## ACKNOWLEDGEMENT

The authors wish to thank the participants at the LSE International Health Policy Conference 2017 for their fruitful comments.

## ORCID

Milena Vainieri  <http://orcid.org/0000-0002-0914-4487>

Daniel Adrian Lungu  <http://orcid.org/0000-0002-8612-8384>

Sabina Nuti  <http://orcid.org/0000-0002-2093-1828>

## REFERENCES

1. Eijkenaar F, Emmert M, Scheppach M, Schoffski O. Effects of pay for performance in health care: a systematic review of systematic reviews. *Health Policy (New York)*. 2013;110(2-3):115-130. <https://doi.org/10.1016/j.healthpol.2013.01.008>
2. Scott A, Sivey P, Ait Ouakrim D, et al. The effect of financial incentives on the quality of health care provided by primary care physicians. *Cochrane Database Syst Rev*. 2011;9:
3. Mannion R, Davies HTO. Payment for performance in health care. *BMJ*. 2008;336(7639):306-308. <https://doi.org/10.1136/bmj.39463.454815.94>
4. Glasziou PP, Buchan H, Del MC, et al. When financial incentives do more good than harm: a their premature or inappropriate implementation. *BMJ*. 2012;5047(August):1-5. <https://doi.org/10.1136/bmj.e5047>.
5. Kristensen SR, Bech M, Lauridsen JT. Who to pay for performance? The choice of organisational level for hospital performance incentives. *Eur J Health Econ*. 2016;17(4):435-442. <https://doi.org/10.1007/s10198-015-0690-0>.
6. Van Herck P, De Smedt D, Annemans L, Remmen R, Rosenthal MB, Sermeus W. Systematic review: effects, design choices, and context of pay-for-performance in health care. *BMC Health Serv Res*. 2010;10(1):247. <https://doi.org/10.1186/1472-6963-10-247>
7. Ittner CD, Larcker DF. Coming up short on nonfinancial performance measurement. *Harv Bus Rev*. 2003;81(11):
8. Basinga P, Gertler PJ, Binagwaho A, Soucat AL, Sturdy J, Vermeersch CM. Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: an impact evaluation. *Lancet*. 2011;377(9775):1421-1428. [https://doi.org/10.1016/S0140-6736\(11\)60177-3](https://doi.org/10.1016/S0140-6736(11)60177-3)
9. Gertler P, Vermeersch C. Using performance incentives to improve medical care productivity and health outcomes. *Natl Bur Econ Res Work Pap Ser*. 2013;No. 19046. doi:<https://doi.org/10.3386/w19046>.
10. Miller G, Babiarz KS. Pay-for-performance incentives in low- and middle-income country health programs. *Encycl Heal Econ*. 2014;457-466. <https://doi.org/10.1016/B978-0-12-375678-7.00126-7>
11. Regalia F, Castro L. Nicaragua: combining demand- and supply-side incentives. In: *Perform Incent Glob Heal*; 2009:215-236.
12. Rusa L, Schneidman M, Fritsche G, Musango L. Rwanda: performance-based financing in the public sector. In: *Perform Incent Glob Heal potential pitfalls*; 2009:189-214.
13. Singh P. Performance pay and information: reducing child undernutrition in India. *J Econ Behav Organ*. 2015;112:141-163. <https://doi.org/10.1016/j.jebo.2015.01.008>
14. Pearson SD, Schneider EC, Kleinman KP, Coltin KL, Singer JA. The impact of pay-for-performance on health care quality in Massachusetts, 2001-2003. *Health Aff*. 2008;27(4):1167-1176. <https://doi.org/10.1377/hlthaff.27.4.1167>
15. Roski J, Jeddeloh R, An L, et al. The impact of financial incentives and a patient registry on preventive care quality: increasing provider adherence to evidence-based smoking cessation practice guidelines. *Prev Med (Baltim)*. 2003;36(3):291-299. [https://doi.org/10.1016/S0091-7435\(02\)00052-X](https://doi.org/10.1016/S0091-7435(02)00052-X).
16. Christensen DB, Neil N, Fassett WE, Smith DH, Holmes G, Stergachis A. Frequency and characteristics of cognitive services provided in response to a financial incentive. *J Am Pharm Assoc*. 2000;40(5):609-617.
17. Glickman SW, Ou F-S, DeLong ER, et al. Pay for performance, quality of care, and outcomes in acute myocardial infarction. *JAMA*. 2007;297(21):2373-2380. <https://doi.org/10.1001/jama.297.21.2373>
18. Petersen LA, Woodard LD, Urech T, Daw C, Sookanan S. Does pay-for-performance improve the quality of health care? *Ann Intern Med*. 2006;145(4):265-272. <https://doi.org/10.7326/0003-4819-145-4-200608150-00006>

19. Van Herck P, De Smedt D, Annemans L, Remmen R, Rosenthal MB, Sermeus W. Systematic review: effects, design choices, and context of pay-for-performance in health care. *BMC Health Serv Res*. 2010;10(247):
20. Burkert M, Fischer FM, Schaffer U. Application of the controllability principle and managerial performance: the role of role perceptions. *Manag Account Res*. 2011;22(3):143-159. <https://doi.org/10.1016/j.mar.2011.03.002>
21. Merchant K. How and why firms disregard the controllability principle. *Account Manag*. 1987;
22. Ittner CD, Larcker DF. Determinants of performance measure choices in worker incentive plans. *J Labor Econ*. 2002;20(s2):S58-S90. <https://doi.org/10.1086/338674>
23. Locke EA, Latham GP. In: Locke EA, Latham GP, eds. *New Developments in Goal Setting and Task Performance*. New York and London: Routledge Taylor & Francis Group; 2013.
24. Vainieri M, Nuti S. *Governance E Meccanismi Di Incentivazione Nei Sistemi Sanitari Regionali*. Bologna: Il Mulino; 2015.
25. Rosenthal MB, Fernandopulle R, Song HR, Landon B. Paying for quality: providers' incentives for quality improvement. *Health Aff*. 2004;23(2):127-141. <https://doi.org/10.1377/HLTHAFF.23.2.127>
26. Christianson J, Leatherman S, Sutherland K. Financial incentives, healthcare providers and quality improvements a review of the evidence.; 2007.
27. Vainieri M, Vola F, Gomez G, Nuti S. How to set challenging goals and conduct fair evaluation in regional public health systems. Insights from Valencia and Tuscany regions. *Health Policy (New York)*. 2016;120(11):1270-1278. <https://doi.org/10.1016/j.healthpol.2016.09.011>
28. Gomez-Mejia LR, Berrone P, Franco-Santos M. *Compensation and Organizational Performance*. New York: M.E. Sharpe; 2010.
29. Hillman AL, Ripley K, Goldfarb N, Nuamah I, Weiner J, Lusk E. Physician financial incentives and feedback: failure to increase cancer screening in Medicaid managed care. *Am J Public Health*. 1998;88(11):1699-1701. <https://doi.org/10.2105/AJPH.88.11.1699>
30. Hillman AL, Ripley K, Goldfarb N, Weiner J, Nuamah I, Lusk E. The use of physician financial incentives and feedback to improve pediatric preventive care in Medicaid managed care. *Pediatrics*. 1999;104(4 Pt 1):931-935. <https://doi.org/10.1542/peds.104.4.931>.
31. Kouides RW, Bennett NM, Lewis B, Cappuccio JD, Barker WH, LaForce FM. Performance-based physician reimbursement and influenza immunization rates in the elderly. *Am J Prev Med*. 1998;14(2):89-95. [https://doi.org/10.1016/S0749-3797\(97\)00028-7](https://doi.org/10.1016/S0749-3797(97)00028-7)
32. Grady KE, Lemkau JP, Lee NR, Caddell C. Enhancing mammography referral in primary care. *Prev Med (Baltim)*. 1997;26(6):791-800. <https://doi.org/10.1006/pmed.1997.0219>
33. Dudley R, Miller RH, Korenbrot TY, Luft HS. The impact of financial incentives on quality of health care. *Milbank Q*. 1998;76(4):649-686. 511. doi:<https://doi.org/10.1111/1468-0009.00109>
34. Van TS, Leeuw FL. The performance paradox in the public sector. *Public Perform Manag Rev*. 2002;25(3):267. <https://doi.org/10.2307/3381236>
35. Van der Stede W. Designing effective reward systems. *Financ Manag*. 2009;44(170):
36. Wadmann S, Johansen S, Lind a, Okkels Birk H, Hoeyer K. Analytical perspectives on performance-based management: an outline of theoretical assumptions in the existing literature. *Health Econ Policy Law*. 2013;8(2013):511-527. <https://doi.org/10.1017/S174413311300011X>  
<https://doi.org/10.1017/S174413311300011X> [pii]
37. Nuti S, Vainieri M, Vola F. Priorities and targets: a methodology to support the policy-making process in health care. *Public Money Manag*. 2017;7(4):277-284. <https://doi.org/10.1080/09540962.2017.1295728>
38. Goddard M, Davies HTO, Dawson D, Mannion R, McInnes F. Clinical performance measurement: part 2—avoiding the pitfalls. *J R Soc Med*. 2002;95(November):549-551.
39. Oliver A. Incentivising improvements in health care. *mics, Policy Law*. 2015;10:
40. Oliver A. Do unto others: on the importance of reciprocity in public administration. (in press). *Am Rev Public Adm*. 2017; 027507401668682
41. Bevan G, Evans A, Nuti S. Reputations count: why benchmarking performance is improving health care across the world. *Health Econ Policy Law*.
42. Ricciardi W, Favaretti C, Bellantone R. A healthy lesson, Italian-style ? *ITALIAN JOURNAL OF PUBLIC HEALTH*. 2009; 6(4):323-329.
43. Nuti S, Vola F, Bonini A, Vainieri M. Making governance work in the health care sector: evidence from a "natural experiment" in Italy. *Health Econ Policy Law*. 2015;11(1):17-38. <https://doi.org/10.1017/S1744133115000067>
44. Carinci F, Caracci G, Di Stanislao F, Moirano F. Performance measurement in response to the Tallinn Charter: experiences from the decentralized Italian framework. *Health Policy*. 2012;108(1):60-66. <https://doi.org/10.1016/j.healthpol.2012.07.010>

45. Toth F. In: Fierbeck K, Palley H, eds. *Italy, Comparing Health Care Federalism*. Routledge Taylor & Francis Group; 2015.
46. Ferré F, Giulio A, Valerio L, et al. Health system review: Italy. *Health Syst Transit*. 2014;16(4):1-168.
47. Torbica A, Fattore G. The “essential levels of care” in Italy: when being explicit serves the devolution of powers. *Eur J Health Econ*. 2005;6(SUPPL. 1):46-52. <https://doi.org/10.1007/s10198-005-0318-x>
48. Expert Group on Health Systems Performance Assessment. *So what? Strategies across Europe to assess quality of care*. Health and Food Safety; 2016. doi:<https://doi.org/10.2875/03667>.
49. 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(2-3):303-312. <https://doi.org/10.1016/j.healthpol.2008.10.012>
50. Vainieri M, Nuti S. Performance measurement features of the Italian regional healthcare systems: differences and similarities. In: Smigorski K, ed. *Health Management-Different Approaches and Solutions*. In Tech; 2011.
51. Mauro M, Maresso A, Guglielmo A. Health decentralization at a dead-end: towards new recovery plans for Italian hospitals. *Health Policy* (New York). 2017;121(6):582-587. <https://doi.org/10.1016/j.healthpol.2017.04.003>
52. 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. *J Manag Gov*. 2012;17(1):59-69. <https://doi.org/10.1007/s10997-012-9218-5>
53. Nuti S, Vainieri M. Strategies and tools to manage variation in regional governance systems. In: Sobolev B, ed. *Handbook of Health Services Research*. Springer; 2014.
54. Berta P, Seghieri C, Vittadini G. Comparing health outcomes among hospitals: the experience of the Lombardy region. *Health Care Manag Sci*. 2013;16(3):245-257. <https://doi.org/10.1007/s10729-013-9227-1>
55. Martini G, Berta P, Mullahy J, Vittadini G. The effectiveness-efficiency trade-off in health care: the case of hospitals in Lombardy, Italy. *Reg Sci Urban Econ*. 2014;49:217-231. <https://doi.org/10.1016/j.regsciurbeco.2014.02.003>
56. Peluso A, Berta P, Vinciotti V. Do pay-for-performance incentives lead to a better health outcome ? *arXiv Prepr arXiv170305103*. 2017;44(0):0-19.
57. Caldarelli A, Fiondella C, Marco M, Spanò R, Aria M. CEO performance evaluation systems: empirical findings from the Italian health service. *Public Money Manag*. 2013;33(5):369-376.
58. Lega F. The rise and fall(acy) of clinical directorates in Italy. Health policy. *Health Policy* (New York). 2008;85(2):252-262.
59. Lee TH, Cosgrove T. Engaging doctors in the health care revolution. *Harv Bus Rev*. 2014;92(6):
60. Nuti S, Vainieri M, Frey M. Healthcare resources and expenditure in financial crisis: scenarios and managerial strategies. *J Matern Neonatal Med*. 2012;25(S4):40-43. <https://doi.org/10.3109/14767058.2012.715022>
61. Berwick DM, Nolan TW, Whittington J. The triple aim: care, health, and cost. *Health Aff*. 2008;27(3):759-769. <https://doi.org/10.1377/hlthaff.27.3.759>
62. Brown A, Closson T, Sullivan T, Baker GR. The journey toward high performance and excellent quality. *Healthc Q*. 2012;15(Special Issue CP-PB):6-9.
63. Vainieri M, Ferré F, Giacomelli G, Nuti S. Explaining performance in health care: how and when top management competencies make the difference. *Health Care Manage Rev*. 2017;44(1). <https://doi.org/10.1097/HMR.000000000000164>
64. Ferreira A, Otley D. The design and use of performance management systems: an extended framework for analysis. *Man*. 2009;20(4):263-282. <https://doi.org/10.1016/j.mar.2009.07.003>
65. 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 Manag*. 2013;33(4):245-252. <https://doi.org/10.1080/09540962.2013.799801>
66. Nuti S, Vola F, Bonini A, Vainieri M. Making governance work in the health care sector: evidence from a “natural experiment” in Italy. *Health Econ Policy Law*. 2014;11(1). <https://doi.org/10.1017/S1744133115000067>
67. Ferré F, Cuccurullo C, Lega F. The challenge and the future of health care turnaround plans : Evidence from the Italian experience. 2012;106(October 2001):3-9. <https://doi.org/10.1016/j.healthpol.2012.03.007>

**How to cite this article:** Vainieri M, Lungu DA, Nuti S. Insights on the effectiveness of reward schemes from 10-year longitudinal case studies in 2 Italian regions. *Int J Health Plann Mgmt*. 2018;33:e474–e484. <https://doi.org/10.1002/hpm.2496>