MANAGING GAMING PHENOMENA IN THE
TUSCAN PERFORMANCE EVALUATION SYSTEM

Sabina Nuti, Milena Vainieri, Anna Bonini

Working paper n. 01

2008
Working Paper n. 01-2008

Sabina Nuti*, Milena Vainieri*, Anna Bonini*

* Laboratorio Management e Sanità (MeS) - Scuola Superiore Sant'Anna di Pisa

Corresponding Author’s address:
*Milena Vainieri
Laboratorio Management e Sanità - Scuola Superiore Sant'Anna,
Via S. Francesco, 18 56127 PISA – Italy
Tel. (39 +) 050 88 3871/3878/3864
Fax: (39 +) 050 883890
Email: m.vainieri@sssup.it
Web site: http://www.meslab.sssup.it/_sito/

Acknowledgments
The authors want to thank the participants at the 23°PCSI Conference, held in Lisbon 8-12 October 2008, for their helpful comments.

Please quote this way:
MANAGING GAMING PHENOMENA IN THE TUSCAN PERFORMANCE EVALUATION SYSTEM

Sabina Nuti, Milena Vainieri, Anna Bonini

Abstract

Introduction
The problem of gaming in performance evaluation systems is a known phenomenon (Bevan, 2006; Bevan Hood, 2006) that can take place especially when there are economic incentives related to specific targets.

In all Tuscan Local Health Authorities (LHAs) and Teaching Hospitals (THs) a multidimensional performance evaluation system has been adopted since 2005, based on administrative and non-administrative data benchmarking.

The goal of the Tuscany performance measurement system is to give a general outline for the management of the LHAs and THs, both for evaluating performance and for enhancing and promoting the results of the healthcare system. The performance evaluation system consists of 130 indicators classified in six dimensions: Population health assessment; Regional health system; Quality; Patient satisfaction; Staff satisfaction; Efficiency and financial performance.

After three years of the performance evaluation system’s adoption, integrated with an incentive system, improvements were achieved in most of the indicators monitored. The paper reports the effects of the performance evaluation system and how gaming phenomena were managed with a special focus on data manipulation.

Methods
The research team focused the analysis on the larger performance improvements during the years 2005-2007 to evaluate whether gaming phenomena have been taking place. The findings of this analysis were reported and discussed with top managers and professionals through individual interviews and group meetings in order to detect the determinants of the results obtained.
In 2007 LHAs and THs whose indicators clearly show large improvement were highlighted during meetings with the top management asking them how his/her organization reaches its results.

**Results**
Gaming has been found in indicators concerning small and specific areas, such as the rate of hospitalization for heart failure, while indicators that concern larger phenomena, such as the pre-surgical length of stay, were less involved. The request of evidence to explain the improvement obtained has been a deterrent to gaming behaviors.

**Conclusions**
The results suggest that further development is needed to analyse suspicious cases. The public presentation of data and the request for evidence to explain large improvements in peer review meetings ease the accountability process. In addition these can be a deterrent for future activities.

**JEL Classification:** I10, I18
**Keywords:** Gaming phenomena, Health service improvement, reputational level.
Index

Index ............................................................................................................. 5

Introduction ................................................................................................. 6

The tuscany multidimensional performance evaluation system ... 6

The focus: the distortion phenomena ......................................................... 7

How to detect and manage distortion phenomena .................. 9

Results ........................................................................................................... 11

Conclusions ................................................................................................. 13

References ................................................................................................. 15
Introduction

In 2004 Tuscany Region has introduced a multidimensional measurement system in order to assess and monitor its Health Authorities (HAs), that include 12 Local Health Authorities (LHAs) and 4 Teaching Hospitals (THs). (Nuti,2005, Laboratorio Management e Sanità 2007,2008)

After three years of performance evaluation system adoption, improvements were achieved in most of the indicators monitored. But how genuine are these improvements?

When the multidimensional performance system was developed, some assumptions described below were made. After the evidence of some large improvements, the research team became aware that a monitoring system is needed to detect whether gaming phenomena is taking place.

This paper reports the path that the research team has taken in order to discover and manage gaming phenomena.

The structure of the paper consists of a brief description of the multidimensional performance evaluation system adopted in Tuscany Region (Italy); the focus on distortion phenomena; how to detect and how to manage the suspected cases of distortion phenomena; some results and conclusions.

The Tuscany multidimensional performance evaluation system

The performance evaluation system was developed by using lessons from the many performance measurement systems that have evolved over the last twenty years, in particular from multidimensional systems already used in health care (Pink et al. 2001) as well as the model of the Balanced Scorecard (BSC). (Kaplan,Norton 1996a,b,c) One lesson learned during this process was the importance of developing the performance evaluation
system in close collaboration with health care professionals and managers (Abernethy, Stoelwinder 1995, Jones, Dewing 1997)

The performance evaluation system is based on 50 measures, consisting of more than 130 indicators, classified in six assessment dimensions (the letter is used to indicate each dimension):

(A) Population health.

(B) Regional health system, to guarantee that strategic regional goals are pursued in the time and manner indicated.

(C) Quality, appropriateness, effectiveness, clinical risk management and managing supply to match demand.

(D) Patient satisfaction, the patients' experience and level of satisfaction with health services.

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

(F) Efficiency and financial performance

In order to display each HA performance, a target chart with the six dimensions represented was used. The target chart is also divided into five bands associated with different scores and colors, from dark green, corresponding to excellent performance, to red, corresponding to poor performance. An indicator with a high score is displayed close to the centre (dark green), and one with a low score is displayed far from the centre (red).

**The focus: the distortion phenomena.**

When the research team developed the multidimensional performance evaluation system, as in the UK NHS, two implicit types of assumptions were made: the idea that the selected indicators
represent the overall performance of HAs\(^1\) and the idea that targets change the behaviour of individuals and organizations. (Bevan and Hood, 2006). Phenomena such as “hit the target but miss the point” or “reduce performance where targets do not apply” can take place.

In this paper we have focused our attention on the second type of assumption that is linked with the possibility that HAs manipulate performance measurement data.

Fisher and Downes (2008) classify the manipulation of performance measurement data combining the degree of dishonesty and the four maxims of Grice (see figure 1).

The research team tried to detect what they call “Manipulating information by re-classifying data” according to their classification of deceit involved in data and information manipulation. In the scale of dishonesty it stands at the top level and it breaks the maxim of quality that is “provide true information”.

---

<table>
<thead>
<tr>
<th>Selective presentation</th>
<th>Gaming</th>
<th>Distortion</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Distraction</strong></td>
<td><strong>Economy with the Truth</strong></td>
<td><strong>Implication</strong></td>
</tr>
<tr>
<td>Mislating by hiding the truth amongst a mass of detail</td>
<td>Not telling the whole truth so as to give a false impression</td>
<td>Mislating by anticipating how people will heuristically misinterpret selectively provided information</td>
</tr>
<tr>
<td>Too much</td>
<td>Too little</td>
<td></td>
</tr>
</tbody>
</table>

---

**Fig. 1: Levels of deceit in data and information manipulation**

*Source: Fisher and Downes 2008*

---

\(^1\) It is the idea of synecdoche described by Bevan and Hood: selected indicators are taking a part to stand for whole. But this assumption could lead HAs to focus on the chosen targets leaving uncontrolled the unobserved ones.
**How to detect and manage distortion phenomena.**

In 2007 the improvements recorded by HAs vary from 20% to 68%, with a variable range among HAs (See figure 2).

![Figure 2 Percentage of performance evaluation system indicators improved from 2006 to 2007.](image)

Are all these improvements genuine? In order to uncover whether there is distortion phenomena due to misclassification of data and information, the research team followed two steps:

1. Identification of the pool of suspected cases;
2. Identification of the *real* improvements within the suspected cases

**The first step: Identification of the pool of suspected cases**

The research team analyzed the trend of 130 indicators and selected those indicators that registered notable improvements\(^2\) across the years and the HAs.

---

\(^2\) In the paper we use the term “improvement” both as the growth and the reduction of the indicators performance across the years. In fact, for instance, there could be suspected case of manipulation also when all HAs registered a negative trend while one HA has held steady.
A different source of indicator selection came from professionals who could notify anomalous changes: in the case of motivated warnings, indicators underwent further analysis.

The second step: identification of real improvement

Once the research team selected the suspected indicators, the second step was to report and discuss these improvements with top managers and professionals in order to detect the determinants of the results obtained.

During the year the performance indicators were monitored every three months in individual meetings between the Health Regional Councilor and each HA’s CEO. In these meetings, the research team asked for explanations for those areas with large improvements.

In group meetings with other managers CEOs were called to provide evidence on how they reached these improvements.

Control mechanisms and other contexts influence managers and professionals in manipulating data and information (Fisher Downes 2008). For instance the use of a balanced approach could help in controlling phenomena of deception linked to reduced performance where targets do not apply. Other mechanisms such as the potential for high rewards, associated with a perception of a low probability of being caught, or few damaging consequences if caught, could act as triggers for deception.

According to some authors (Hibbard et al 2003; Bevan 2006) the driver of performance change in the health care sector is the risk of reputation damage³. If reputation drives changes in performance it could also drive to deception in order to avoid hassle and scrutiny mechanisms by professionals.

---

³ In order to have an impact on health services providers, the performance evaluation system should satisfy 4 characteristics: a ranking system, published and widely disseminated, easily understood by the public; and followed up by future reports.
Reputation is a strong driver, above all in the context of professional bureaucracies. It is considered a fundamental lever for successfully introducing multidimensional performance systems (Hibbard 2003). The hypothesis of the research team is exactly that of using reputation lever in order to tackle some problems linked to the gaming phenomena.

The research team assumed that the public investigation could discourage data manipulation: the request for evidence to explain large improvements in peer review meetings ease the accountability process and can be a deterrent for the future.

**Results**

For the first time in 2007 the distortion phenomenon was detected and managed.

The research team focused its attention on the ten indicators that show impressive improvements by one HA among the others. The research team noticed that during the interviews with CEOs whose HA reported notable improvements, when CEOs were not only able to describe the actions put in place to increase the performance it was due to distortion phenomena. In fact in some cases, after the inquiry, they admit that data has been manipulated. On the contrary when the improvement was the result of a real organizational change, CEO was eager to demonstrate how his HA achieved it.

Here it is reported one example of real improvements driven by organizational changes and one example of deceptive improvement.

Figure 3 shows the percentage of hip fractures operated timely (within 2 days from the admission): LHA 12 increases its result from 9% in 2006 to 51% in 2007. This impressive improvement was due to the hiring of a new professional that completely reorganized the
clinical pathway in the orthopaedic unit. In particular he has introduced a dedicated path for the urgent cases.

Figure 4 reports an example of data manipulation. The reduction of the 100,000 inhabitant hospitalization rate for heart failure in LHA 4 from 2.802 in 2006 to 1.362 in 2007 was due to the re-codification of ICD9-CM diagnosis.

Fig. 3

Fig. 4
The 2008 partial results confirm which the real and deceptive improvements are. In fact, in the 2008 first six-month period LHA 12 has increased even more the percentage of timely hip fracture surgeries, to 60%. At the opposite, the hospitalization rate for heart failure in LHA 4 has come back to a value of 2.033, regressing to nearly the 2006 levels.

These examples support the investigation results, and confirm the hypothesis that public explanation is a deterrent to considering the manipulation route.

Gaming has been found more in indicators concerning small and specific areas, such as the rate of hospitalization for heart failure, while indicators that concern larger phenomena, such as the pre-surgical length of stay, were less involved. This effect could be linked to the difficulty for the HA to manipulate a large number of medical records spread out in several Departments, or, in a different view, to identify if there are recodifications and in particular to calculate the effect of recodifications on indicators that cover large phenomena.

**Conclusions**

Problems linked to gaming phenomena are widespread. The experience reported in this paper shows that the path is at its early stages and necessary developments should be applied. In particular further developments are planned in order to streamline the identification process of suspected cases and to better analyze suspicious cases more in depth.

In order to cope with the gaming phenomena Bevan and Hood suggested two types of remedies: the introduction of on-site inspections, through randomized controls, and the introduction of uncertainty in to targets and rating systems. In this manner Bevan and Hood (2006) introduce into the regulation an high degree of unpredictability because it makes it difficult for managers to ascertain
in advance how they will be assessed. Too much predictability in the name of transparency weakens controls by the gaming responses it invites.

Following the first remedy, in 2008 the Health Regional Council, a technical consultant commission of Tuscany Region, has been asked for a check control. After preliminary selections made by the research team, the Health Regional Council should investigate on a sample of clinical histories. Moreover it should identify, ex-ante, a pool of indicators used in the performance evaluation system that could be more involved in the manipulation phenomena.

The contribution of this paper is linked to the role that the lever of reputation can play in the gaming phenomena through peer reviews meetings and CEO public presentations as an accountability process. In fact publishing the performance results on one side and on the other, the agreement with the HAs CEOs to argue how the outlier improvement were achieved by professionals, worked as deterrent to gaming behaviours. Moreover the habit of sharing the strategies adopted to achieve better results has been a learning tool.
References

Abernethy MA, Stoelwinder JU (2005), The role of professional control in the management of complex organizations. Accounting, Organizations and Society, 20 (1): 1-17


