BALANCED SCORECARD HEALTH SYSTEM: A LATENT VARIABLE APPROACH

Pietro Giorgio Lovaglio* **and Giorgio Vittadini****

* *Department of Quantitative Methods for Business Economic Sciences, Faculty of Statistics University Bicocca-Milan, V.le Sarca 202, 20126 Milan piergiorgio.lovaglio@unimib.it*

** *CRISP Interuniversity Center, University Bicocca-Milano*

Abstract

The Balanced Scorecard represents a technique used to translate an organization's mission and strategy into a comprehensive set of performance measures that provide the framework for the implementation of strategic management. The original contribution of this article deals with the exploration of available indicators and statistical methodologies that may facilitate the diffusion of the BSC framework in a hospital setting within the Lombardy Region Health sector.

Specifically, the present paper describes an approach to designing and implementing a Balanced Scorecard system for measuring the performance and productivity of Health structures, suggesting regional stakeholders how the information dimensions of BSC can assist in developing competitive strategic outcomes.

The empirical assessment of causal relationships among four BSC dimensions in Lombardy Region Hospitals was performed by Partial Least Square Path Modelling approach within a multilevel perspective (taking into account the nested structure of Health structures). This approach permits the estimation of two structural models, one for the "between" and one for the "within" model component, demonstrating different causal structures.

The results, based on 167 Lombardy Region Hospitals in 2007, identified robust measurement models but weak significant causal relationships among four BSC dimensions. Particularly, patient satisfaction and quality Hospital levels account for only a moderate quota (30%) of the variance of Hospitals' financial resources.

1. INTRODUCTION

A wide range of systems are available to facilitate Health structures in improving quality, from locally developed systems to those widely recognized by the international health care community, such as Continuous Quality Improvement or Balanced Scorecard (Blumenthal, 1996; Øvretveit and Gustafson, 2003).

In 1992, Kaplan and Norton proposed a managerial model, called Balanced

Scorecard (BSC, Kaplan and Norton, 1992), to support the decision-making process in business, aimed at providing a multidimensional interpretation of economic performance. The BSC is typically proposed to assess which areas best contribute (and to what extent) to the prospective of the creation of monetary value for statistical units (firms, companies, departments, etc). The conceptual scheme of the Balanced Scorecard, using a balanced set of different types of indicators, has gained huge popularity, as it is an effective theoretical tool to support decisions regarding the multidimensionality of evaluation processes, the identification of individual and group targets, and the application of strategies concerning operational and organizational activities.

BSC embodies a set of measures which give top managers a quick but comprehensive vision of key business issues and contains financial measures that demonstrate the consequences of transactions already performed. It complements the financial measures with operational measures on customer satisfaction, internal processes, and the organization's innovation and improvement activities.

More specifically, BSC specifies causal relations between the main dimensions of interest, the so-called key performance areas (KPA), once they are measured by observable indicators (Key performance index, KPI).

Although BSC is typically applied in business disciplines and in the strategic management area for industries, recently it has also been proposed in the health sector, following the guidelines of International Institutions, such as World Health Organization and OECD, which suggest inserting microeconomic efficiency and macroeconomic sustainability as the primary goals to be accomplished by National Health Systems. In their second book on the BSC, Kaplan and Norton introduce cases which demonstrate how the BSC could be implemented in nonprofit, governmental and health care organizations (Kaplan and Norton, 2001).

Baker and Pink (1995) were among the first to argue that the theory and concepts of the balanced scorecard were relevant to hospitals. Since then, the basic principles of the balanced scorecard have been well documented in health care literature (Chow *et al.*, 1998; Zelman *et al.*, 1999; Oliveira, 2001).

The scorecard's measurement and management system provides potential benefits to healthcare organizations such as the alignment of the organization around a more market-oriented, customer-focused strategy, facilitating the monitoring of the internal feedback strategies and the implementation of accountability for performance at all levels of the organization (Inamdar *et al.*, 2002).

Even though the concept of the BSC has had broad application in the international health sector, including Hospitals systems (Zelman *et al.*, 2003, *Sahney, 1998;* Pink *et al.*, 2001), national healthcare systems or organizations (Inamdar *et al.*, 2002; Northcott and France, 2005), moving from concept to practice has often proved difficult (Chow *et al.*, 1998; Ahn, 2001). Furthermore, the implementation of the BSC concept as a strategic management tool has had limited coverage in research (Tuomela, 2005).

The primary aim of this paper is the description of approaches implementing a balanced scorecard system for measuring performance and productivity in a hospital setting within the Lombardy Region Health sector. We propose a conceptualization of the BSC scheme to describe the mechanism producing creation of (monetary) value for health structures, utilizing a suitable statistical approach taking into account the unobservable nature of Key areas and the hierarchical structure of data.

METHODOLOGICAL FRAMEWORK

For operational purposes, the Balanced Scorecard must evolve into a measurement system using explicit, objective formulas that define causal relationships among the areas analyzed and prescribes the weights to be attached to each Key performance area (Ittner and Larcker, 1998).

Following Kaplan and Norton (2001), our conceptualization of BSC hypothesizes that the economic performance (Economy KPA), referring to the monetary-physical output of healthcare structures and their productivity depends on three macro-areas: Human Capital (referring to the quality of human resources in Health structures), Process (inherent to the quality and efficiency of healthcare processes) and Patient Satisfaction. Figure 1 specifies the theoretical framework adopted for implementing BSC in the Lombardy Region Health sector.

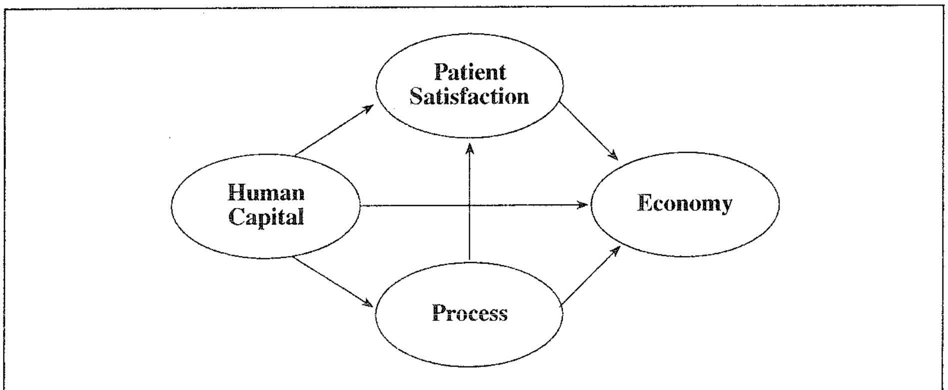

Fig. 1: Conceptual BSC model for Lombardy Region Health structures.

Statistically speaking, the BSC is a system of causal relationships among composite indicators (Key performance areas), which integrates large amounts of information (Key performance Indicators) into an easily understood single metric.

The basic methods to construct composite indicators in a single block of observed variables are suggested (Cox et al., 1992), putting emphasis on weighting and aggregation.

However, the perspective of measuring composite indicators within a statistical (causal) model framework, as occurs in the development of a BSC managerial model, moves away from traditional weighting methodologies as they do not take into account the causal links between the specified dimensions.

Thus, the primary question arising in this new perspective is linked to the choice of a proper methodological approach, simultaneously resolving the twofold problem of KPA construction and causal model estimation.

Recently, leading accounting journals have placed great emphasis on the future directions of management accounting research, including the structures of theoretical modeling and data analysis that can be used (Atkinson *et al.* 1997). To conceptualize Health Quality Systems, the use of structural equation models (systems of simultaneous causal equations) may overcome some of the limitations presented by the more traditional statistical techniques generally used by management accounting (Hulland, 1999).

In fact, since KPAs are unobservable multidimensional constructs (hypothetical or latent constructs) measurable with errors by means of blocks of observed indicators, the most natural methodological context for BSC conceptualization and estimation deals with structural equation models with latent variables (LVs).

Statistical literature has proposed two different LV definitions.

The first definition, sharing the Factor Analysis approach, is supplied by the Structural Equation Models or *LI*near *S*tructural *REL*ations (from here on, LISREL model, following the approach of Jöreskog, 1981), where latent variables are "true LVs" in the sense that the equations cannot be manipulated so as to express themselves as a function of observed indicators only (Bentler, 1982).

Despite its widespread use, the LISREL model presents many drawbacks. Firstly, under general conditions, the LISREL model is not identified (Vittadini, 1989). Because of the complexity of the structural models it supports, large numbers of alternative but statistically equivalent, models can be supported by the same data: for example, reversing the direction of any causation path or replacing it with a correlation path will produce an equivalent model with the same fit indices (Stelzl, 1986). Furthermore, even if the parameters of the LISREL Model are perfectly identified, the scores of the LVs are not unique, i.e. there are infinite sets of latent scores for the same identified model (Bentler, 1982; Vittadini, 1989).

Finally, real data often does not meet the stringent LISREL assumptions referring to the multivariate normality of observed variables and strict requirements as a high sample size and a complex structure of errors' independency.

To resolve these drawbacks, a second definition arises in literature defining an LV as "Unobservable Component Variable" (Kmenta, 1991), whose scores are estimated as linear combinations of their observed indicators. In this view, observed indicators appear as causes of an LV (formative indicators), whereas in the LISREL model they are effects (reflective indicators).

Within the framework of causal models with latent variables, Partial Least Squares Path Modeling (PLSPM; Wold, 1982) is the counterpart of LISREL model embodying this LV definition.

PLSPM, or soft modeling, avoids the traditional assumptions for which the observations need to be independent and jointly sampled by the same normal distribution and works with only a few observations and numerous variables. It creates optimal linear predictive relationships among variables, aiming at identifying predictive links, rather than causal ones. These relationships are interpreted as the best set of predictions available for a given study, considering all theoretical, measurement, distributional and practical limitations implicit in the data.

PLSPM proceeds in two stages (Wold, 1982): in the first stage, it estimates the latent scores as linear combinations of its own observed indicators whereas in the second stage, the parameters of the structural model (links between LVs) and of the measurement model (links between observed indicators and LV) are achieved through simple and multiple regressions, utilizing latent scores estimated in the previous step. To this end, the structural model becomes a model of Path Analysis between linear combinations of the observed variables. By iterating the steps of this procedure, the convergence of the algorithm presents the final estimation of the weights that defining the LV scores as linear combinations of their observed indicators.

Hence, PLSPM robustly estimates the latent variable scores of different specified models, obtaining consistent estimates and minimum variance predictions, without creating problems of either identification or indeterminacy.

PLSPM provides useful diagnostics, testing model fit, robustness and uncertainty of estimates, according to composite indicators guidelines stressing the usefulness of sensitivity analysis and adequate allowance (plausible ranges of estimates) for uncertainty associated with summary measures. Finally, the assessment of the model in terms of robustness and sensitivity can be obtained by methodologies such as cross-validation and jackknife, whereas in presence of nonnormal data, robust standard errors (and t-statistics) are empirically obtained by bootstrap replications.

The utilization of PLSPM has been examined in literature in the fields of strategic management (see Hulland, 1999 for a complete review), consumer beliefs (Gelper and Croux, 2007) and customer satisfaction (Tenenhaus *et al.*, 2005).

2. PARTIAL LEAST SQUARES PATH MODELING IN A HIERARCHICAL PERSPECTIVE

As explained, BSC is a structural model and thus estimable by PLSPM (Tenenhaus *et al.,* 2005). However, an accurate estimation process must take into account that observations, such as those typically occurring in health context, present hierarchical structure: statistical units are nested in gradually more aggregated levels. The Lombardy Region Health system embodies four different levels of hierarchy: the most aggregated refers to Local Health Agency, collecting Hospitals, Health structures, Ambulatory structures, Centers for mental health and for elderly patients in the same geographical area. The second deals public and private Health Agencies which distribute local Health services. Health Agencies, managed by Local Health Agencies, collects among their Health structures, Presidium of Health Agency and Hospitals (third level). Finally, Operative units (units of care of different areas belonging to the same Hospital or Presidium) constitute the most disaggregated level.

When specifying a structural model, the correlation between micro level units (e.g. patients) belonging to the same group (e.g. Hospital) must be considered. Ignoring the hierarchical structure of data provokes underestimation of the estimates' variability and distortion of fit statistics (Muthen, 1994).

In order to estimate the BSC model, taking into account the hierarchical structure of data, we have considered a recent technique, called *Multilevel Simultaneous Component Analysis* (MSCA; Timmerman, 2006).

The MSCA obtains latent scores underlying formative set of indicators as their Principal Components (PC), allowing the separation of the overall variability (among all first level units) in two components: the contribution due to the groups (variability among second level units) and that due to individual contribution (variability among first level units in each second level unit). In a block of formative indicators, MSCA obtains two sets of latent scores describing the variability within groups (first level units net of second level units effect) and between different groups (second level units effect) that reproduce the maximum part of overall data variability.

In the following discussion, Hospitals are considered first level (micro) units, whereas Local Health Agency (from here, ASL) second level (macro) units.

We start by considering a sample composed of S ($S=\Sigma_k n_k$, $k=1,\ldots,m$) Hospitals nested in one of the m ASL, where n_k describes the number of Hospitals belonging to the k-th ASL.

Let $\mathbf{Y}_k$ (n_k,p_k) be the matrix of (grand mean) centered p_k observed formative indicators of n_k Hospitals in the k-th ASL. Stacking vertically each $\mathbf{Y}_k$ block in the matrix $\mathbf{Y} = [\mathbf{Y}_1,\ldots,\mathbf{Y}_k,\ldots,\mathbf{Y}_m]'$, MSCA decomposes $\mathbf{Y}$, according to equation (1):

$$\mathbf{Y} = \mathbf{T}_{(q)}\mathbf{A}_{(q)} + \mathbf{V}_{(v)}\mathbf{B}_{(v)} + \mathbf{E} \tag{1}$$

where $\mathbf{T}_{(q)}$ describes the PCs within groups in a reduced space of dimension q, while $\mathbf{V}_{(v)}$ are the PC between groups in a reduced space of dimension v, $\mathbf{A}_{(q)}$ and $\mathbf{B}_{(v)}$ are loadings matrices and $\mathbf{E}$ is the matrix of residuals.

The obtained scores of the between structure $\mathbf{V}_{(v)}$ express a synthetic measurement of the performance attributable to different hierarchical units of the second level (ASL), whereas the scores of the within structure $\mathbf{T}_{(q)}$ express a synthetic measurement of Hospital performance net to the clustering effect (ASL).

In this way, the overall data variability (sum of variances of $\mathbf{Y}$ columns) is decomposed in two reduced spaces reproducing the variability existing within groups ($\mathbf{T}_{(q)}\mathbf{A}_{(q)}$) and in the variability between groups ($\mathbf{V}_{(v)}\mathbf{B}_{(v)}$). The values of q and v for the within and the between components are less than the full rank dimensionality, $\Sigma_k p_k$ and m respectively.

Although Equation (1) is written in compact form, it reveals that, after having chosen v and q, each $\mathbf{Y}_k$ matrix can be decomposed into $\mathbf{T}_{k(q)}\mathbf{A}_{(q)}+\mathbf{V}_{k(v)}\mathbf{B}_{k(v)}$ plus the residual term $\mathbf{E}_k$, where $\mathbf{T}_{k(q)}$,$\mathbf{V}_{k(v)}$ and $\mathbf{B}_{k(v)}$ are the k-th submatrices composing $\mathbf{T}_{(q)}$, $\mathbf{V}_{(v)}$ and $\mathbf{B}_{(v)}$, respectively, whereas $\mathbf{A}_{(q)}$ is fixed among the m groups.

The fundamental property of MSCA is that the components $\mathbf{T}_{k(q)}$ and $\mathbf{V}_{k(v)}$, are orthogonal within each block and for all groups.

The estimation of the PCs between and PCs within (and their respective loadings) is obtained by least squares criteria, minimizing the sum of squares (SSQ) of the loss function f_{MSCA}, as shown in equation (2):

$$f_{MSCA} = \Sigma_k \, SSQ\, [\mathbf{Y}_k - (\mathbf{T}_{k(q)}\mathbf{A}_{(q)} + \mathbf{V}_{k(v)}\mathbf{B}_{k(v)})] \qquad k=1,\ldots,m \tag{2}$$

under Following specific constraints on the component scores: $\mathbf{1}_{nk}{}'\mathbf{T}_{k(q)}=\mathbf{0}$ and $\Sigma_k n_k \, \mathbf{v}_{k(v)} = \mathbf{0}$, where $\mathbf{v}_{k(v)}$ is the k-th row of the matrix $\mathbf{V}_{k(v)}$.

Considering that $\mathbf{T}_{k(q)}$ and $\mathbf{V}_{k(v)}$ must be orthogonal, the loadings parameters can be obtained separately in each block. The interesting property of MSCA is that the loadings matrix $\mathbf{A}_{(q)}$ in equation (3) is constrained to be the same for different m groups, otherwise the latent scores $\mathbf{T}_{(q)}$ obtained in different geometric subspaces would not be comparable among m groups.

As measure of fit, the explained percentage of variance extracted by latent dimensions can be used. Since MSCA consists of independent sub models, three indices of fit can be defined: the percentage of variance explained by the within model, by the between model and by the whole model (Timmermann, 2006). The MSCA procedure can be extended in many different situations or structures. For example, longitudinal data, showing a clear hierarchical structure, may be used to estimate the between component as a static component (Hospitals' performance in a period of reference) and the within component as a dynamic component (trends in time for each Hospital).

3. PLSPM-MSCA INTEGRATION

As discussed, MSCA has the suitable property of decomposing overall KPA variability into separate components referring to ASL and Hospital contribution. This justifies efforts aimed to implement MSCA within the BSC structural model.

Nevertheless, although PLSPM and MSCA obtain latent scores as linear combinations of their observed indicators, these techniques cannot be easily or consistently integrated. More specifically, the strategy of applying in both structures (equation 1) the PLSPM algorithm, using PCs obtained by MSCA, is not statistically satisfactory. Infact, both techniques have different criteria for extracting linear components: MSCA estimates PCs by maximizing the explained variability in each block, whereas PLSPM, maximizes regression criteria specified by the structural model.

Thus, we follow an alternative strategy that does not fully integrate both techniques, but that picks suggestions of MSCA.

In fact, under mentioned constrains on the between and within components scores, MSCA estimates PCs in both structures that are mutually orthogonal in the same dimension and for all dimensions, since these latent components are extracted from mutually orthogonal matrices $\mathbf{M}_k$ and $\mathbf{W}_k$ (see equation 3).

Hence, following the strategy of decomposing the data matrix in two orthogonal blocks taking into account the between and the within data structure, we can generalize, in a multilevel perspective, other statistical models that involve latent components (estimated as linear combinations of their observed indicators) in which projections are guided into meaningful directions by the suggested structural model.

More specifically, we can perform two separate PLSPM analyses utilizing the blocks of manifest variables, once they have been separated in the between and the within structure.

In this way, although PLSPM proceed by maximizing statistical criteria that are different from that of MSCA, since the PLSPM LVs scores are estimated as linear combinations of manifest variables that are mutually orthogonal in both structures, there is guarantee that all within-latent variables scores are orthogonal to the between-latent variables scores.

MSCA finds PCs and loading matrices by separating each $\mathbf{Y}_k$ block in two orthogonal terms:

$$\mathbf{Y}_k = \mathbf{Y}_k - n_k^{-1}\,\mathbf{I}_{nk}\mathbf{Y}_k + n_k^{-1}\,\mathbf{I}_{nk}\mathbf{Y}_k = [\mathbf{Y}_k - \mathbf{M}_k] + \mathbf{M}_k = [\mathbf{W}_k] + \mathbf{M}_k \qquad (3)$$

where $\mathbf{M}_k$ is the matrix where each column contains the means of p_k variables for the k-th group, $\mathbf{W}_k$ is the matrix of differences between the observed values and the means of p_k variables in the k-th group and $\mathbf{I}_{nk}$ is a square (n_k, n_k) matrix of unit elements. Collecting vertically all m blocks, the following decomposition $\mathbf{Y}=\mathbf{W}+\mathbf{M}$ holds, where $\mathbf{M}$ collects vertically all $\mathbf{M}_k$ matrices and $\mathbf{W}$ collects all $\mathbf{W}_k$ matrices.

In case of balanced groups, MSCA estimates PCs between groups $\mathbf{V}_{(v)}$ by simple Principal Component Analysis (PCA) on the matrix $\mathbf{M}$, whereas the PCs within groups $\mathbf{T}_{(q)}$ by PCA on the matrix $\mathbf{W}$ (Timmermann, 2006).

Utilizing the suggested decomposition of equation (3), we can directly apply PLSPM on two separate orthogonal matrices $\mathbf{M}$ and $\mathbf{W}$ to obtain latent scores and parameters (causal links between LV) in both structures.

More specifically, let $\mathbf{Y}_{HC}=(\mathbf{Y}_{HCk})$, $\mathbf{Y}_{PS}=(\mathbf{Y}_{PSk})$, $\mathbf{Y}_{PR}=(\mathbf{Y}_{PRk})$, $\mathbf{Y}_{EC}=(\mathbf{Y}_{ECk})$ be the four data matrices of observed indicators of the four KPAs (Human Capital, Patient satisfaction, Process and Economy), reorganized in m vertically staked blocks. Following equation (3), each k-th block is decomposed in the between and within structure: $\mathbf{Y}_{HCk}=[(\mathbf{W}_{HCk}, \mathbf{M}_{HCk})]$, $\mathbf{Y}_{PSk}=[(\mathbf{W}_{PSk}, \mathbf{M}_{PSk})]$,$\mathbf{Y}_{PRk}=[(\mathbf{W}_{PRk}, \mathbf{M}_{PRk})]$, $\mathbf{Y}_{Ek}=[(\mathbf{W}_{ECk}, \mathbf{M}_{ECk})]$. Next, two matrices are obtained, one for the between structure $\mathbf{M}$, vertically collecting the m matrices $\mathbf{M}_k=[\mathbf{M}_{HCk}, \mathbf{M}_{PSk}, \mathbf{M}_{PRk}, \mathbf{M}_{ECk}]$ and one for the within structure $\mathbf{W}$, vertically collecting the m matrices $\mathbf{W}_k=(\mathbf{W}_{HCk}, \mathbf{W}_{PSk}, \mathbf{W}_{PRk}, \mathbf{W}_{ECk})$.

Finally, the PLSPM algorithm is applied separately to the within structure $\mathbf{W}$ and to the between structure $\mathbf{M}$ in order to obtain latent scores and causal parameters consistent with the supposed structural model in two separate domains. The final product is the estimation of two structural models, one for the between and one for the within segment, possibly revealing different causal structures, which would have been otherwise masked if the nested structure of observations had been ignored.

4. APPLICATION TO THE LOMBARDY HEALTH SYSTEM

The theoretical conceptualization of BSC, specified in Figure 1, has been recently applied in the Lombardy Region Health sector in a pilot study (Lauro, 2007), with the objective of assessing the impact of health structure management and policies on business results.

This study was entrusted to the Regional Institute of Research (IReR) by the Lombardy Directorate of Health in 2007.

The collected data was drawn from Lombardy Region administrative archives in the Health sector and more specifically refers to Hospital discharge cards, archives collecting Official Regional Surveys on Patient Satisfaction (mandatory for all regional Health structures) and other administrative archives existing in the Regional Directorate of Health.

To incorporate the nested structure of Lombardy Health, we have considered presidiums of Health Agency (from here Hospitals) as the first level unit and Local Health Agencies (from here ASL) as the second level.

The choice of the first level is motivated by the consideration that indicators, especially in Economic and Process areas, are provided at this level for public structures, whereas private Health Agencies coincide *de facto* with private Presidiums (Hospitals).

With regard to the second level, the Lombardy Regional Law 31 of 1997 has revised the Lombardy Health System, separating "providers" (credited public and private Health structures), from "purchasers" (Local Health Agencies): thus, Health services are managed and pertain to ASLs.

The major effort of the pilot study of 2007 was to identify available regional information for the implementation of the Balanced Scorecard system, as a managerial tool for the Lombardy Region Directorate of Health. These efforts have provided a list of 24 Hospital indicators (Key performance index, KPI), representative of the four KPA areas, as listed in Table 1.

Human capital indicators, extracted from available administrative sources in the Lombardy Health sector, do not appear sufficient to adequately capture the quality level of human resources involved in the process of care. Patient satisfaction indicators refer to the official Regional Customer Satisfaction Questionnaire, composed by 15 items on a 0-7 ordinal scale.

Other indicators were drawn by previous specific research studies promoted by the Lombardy Directorate of Healthcare in 2004 within the "Triennial Program for the implementation of an evaluation system of Health structures and management of Health structures" (Healthcare Directorate of Lombardy Region, 2004).

Process indicators were extracted within the subproject "Study of outcomes", conducted by the inter-University centre CRISP, with the aim of extracting Health structure performance indicators from the regional Hospital discharge cards. The indicators concerning the Economic area are those proposed by Joint Commission International", responsible for the subproject "Economic Performance".

Tab. 1: BSC Key performance areas and key performance indicators.

Area (KPA)	Sub-area	Indicator (KPI)	Label
Human	*Training (Employee)*	Total days of training per employee	RUFI
Capital (HC)	*Training (Sanitary)*	Total days of training per sanitary personnel	RUFE
Process	*Outcome*	Intra-hospital mortality rate	OMIN
(PRO)	*Outcome*	Mortality rate within 30 days of demission	OM30
	Quality	% of discharges against medical advices on total discharges	QDIV
	Quality	% of transferred patients (towards other Hospitals) on total discharges	QTRS
	Quality	% of unscheduled re-admissions on total discharges	QRRP
	Health productivity	Length of stay in days (mean score)	DGM
	Health productivity	Length of stay in days per doctor (mean score)	GDM
Patient	*Organizational quality*	Quality (room, menu) of Hospitality (mean score)	QOOS
Satisfaction	*Organizational quality*	Medical and nurse Staff (mean score)	QOPE
(SAT)	*Organizational quality*	Hospital organization (mean score)	QOOS
	Clinical Quality	Information on the health status (mean score)	QCIC
	Clinical Quality	Provided care (mean score)	QCCU
	Clinical Quality	Information after Discharge (mean score)	QCID
	Overall Satisfaction	Global satisfaction (mean score)	SATT
	Overall Satisfaction	Satisfaction on the health state (mean score)	SSAL
Economy	*Economic results*	Gross operating margin	CMO
(ECO)	*Productivity*	DRG profit per day of hospitalization (mean score)	PRNG
	Productivity	DRG profit per discharge (mean score)	PRNC
	Productivity	Total turnover per employee	PRND
	Incidence of costs	% Labour cost on total costs	ICLT
	Incidence of costs	% Cost of health services on total costs	ISST
	Incidence of costs	% Cost of hotel services on total costs	ISAT

Obviously, the chosen indicators do not cover the full picture of the Regional Health system, limiting its capacity for providing a realistic overall picture: for example Human Capital indicators concern only the training area, without considering innovation, system integration and the quality of Human resources (Pink'*et al.*, 2001), whereas the Process indicators do not cover information concerning safety, timeliness, accessibility, productivity and clinical efficiency of Health structures (*Baker* and P*ink,* 1995).

The limited number of indicators is evidence of the information gap in the Lombardy Health sector, even though it is one of most developed systems in Italy. However, despite the limitations due to the lack of available information in regional administrative archives, the limited number of proposed indicators is a consequence of experimental nature of this study which attempts to conceptualize BSC in the Italian Health sector.

The present application involves 163 public and private Lombardy Region Hospitals nested in 15 ASLs. Each indicator, calculated at Hospital level, refers to year 2007.

The elimination of the units containing large quota of missing values has determined a final presence of 129 Hospitals in the analyzed database.

In order to better compare the indicators (Tenenhaus *et al.*, 2005), they have been normalized on a 0-100 scale (100 is the score for the best performance, whereas indicators concerning costs indicate a 100 value for lowest cost).

Before showing the estimated equations of BSC in the between and within structure, Table 2 presents results of measurement models for both structures aiming at evaluating, for each LV, the unidimensionality of LVs, underlying blocks of manifest variables, mono-factorial validity and"discriminant validity. A block of formative indicators is considered one-dimensional when measures such as Cronbach's α ($C\alpha$) and Dillon-Goldstein's Rho ($DG\rho$), summarizing correlations between observed indicators of the same block, are greater than a minimal threshold (typically 0.7, see Tenenhaus *et al.*, 2005). Mono-factorial validity holds when correlations between observed indicators of a block and their LV are higher than correlations with other LVs, whereas discriminant validity (more variance is shared between an LV and its block of indicators than with another LV) of an LV holds if their squared correlations with other LVs are smaller than average communality (AVE, i.e. the average variance of the manifest indicator's set explained by the LV).

In the between structure, all blocks present highly satisfactory unidimensionality measures. Furthermore, mono-factorial validity and discriminant validity exhibit acceptable results, justifying the estimation of causal relations between LVs and their observed indicators. The Economy dimension (in which ISST and ISAT indicators have been eliminated since they presented weak correlations with other indicators) presents the lowest unidimensionality measures and weak Mono-factorial validity.

Tab. 2: Unidimensionality, Mono-factorial validity and Discriminant validity (Between-structure)

Dimension	Index and indicators	Human Capital	Process	Patient Satisfaction	Economy
		Between Structure			
...mensionality	Ca	0.932	0.742	0.980	0.735
(...indices)	DGp	0.969	0.797	0.986	0.710
Mono-factorial	Human Capital	**1**	0.054	0.038	0.020
validity	Process	0.054	**1**	0.165	0.272
...relations and	Patient Satisfaction	0.038	0.165	**1**	0.010
AVE)	Economy	0.020	0.272	0.010	**1**
	AVE	0.932	0.384	0.894	0.317
	RUFI	**0.986**	-0.212	0.224	0.175
	RUFE	**0.945**	-0.253	0.123	0.062
	SSAL	0.417	-0.450	**0.795**	-0.055
	QOOS	0.168	-0.259	**0.973**	0.178
	QOPE	0.201	-0.407	**0.974**	0.144
	QOOR	0.091	-0.292	**0.964**	0.142
	QCIC	0.116	-0.386	**0.982**	0.135
	QCCU	0.178	-0.496	**0.987**	0.082
	QCID	0.074	-0.387	**0.891**	-0.036
Discriminant	SATT	0.224	-0.403	**0.981**	0.150
validity	EGDM	-0.215	**0.807**	-0.584	0.267
(loadings)	OMIN	0.036	**0.592**	0.234	0.631
	OM30	-0.229	**0.682**	-0.366	0.357
	QDIV	-0.068	**0.415**	-0.209	0.004
	QTRS	-0.308	**0.832**	-0.406	0.315
	QRRP	0.016	**0.552**	-0.044	0.520
	EDGM	0.139	**0.230**	-0.103	0.196
	ECMO	0.145	0.478	0.172	**0.988**
	PRNC	-0.129	0.041	-0.020	**0.167**
	PRND	0.241	0.474	-0.437	**0.731**
	ICLT	-0.067	0.457	-0.191	**0.485**

Tab. 2 Continued: Unidimensionality, Mono-factorial validity and Discriminant validity (Within structure)

Dimension	Within Structure				
	Index and indicators	**Human Capital**	**Process**	**Patient Satisfaction**	**Economy**
Unidimensionality (indices)	Ca	0.805	0.902	0.880	0.725
	DGρ	0.852	0.980	0.810	0.830
Mono-factorial validity (correlations and AVE)	Human Capital	**1**	0.097	0.002	0.051
	Process	0.097	**1**	0.040	0.216
	Patient Satisfaction	0.002	0.040	**1**	0.150
	Economy	0.051	0.216	0.150	**1**
	AVE	0.746	0.399	0.861	0.562
Discriminant validity (loadings)	RUFI	**0.977**	0.051	-0.341	-0.256
	RUFE	**0.734**	0.024	-0.112	-0.051
	SSAL	0.085	**0.839**	0.218	0.252
	QOOS	0.071	**0.925**	0.121	0.310
	QOPE	0.066	**0.927**	0.110	0.292
	QOOR	0.015	**0.941**	0.162	0.349
	QCIC	0.097	**0.974**	0.162	0.393
	QCCU	0.033	**0.957**	0.169	0.347
	QCID	-0.034	**0.877**	0.287	0.462
	SATT	0.055	**0.975**	0.196	0.387
	EGDM	-0.281	0.207	**0.461**	0.345
	OMIN	-0.110	0.116	**0.666**	0.289
	OM30	-0.295	-0.009	**0.744**	0.269
	QDIV	-0.272	0.175	**0.780**	0.431
	QTRS	-0.161	0.100	**0.824**	0.284
	QRRP	-0.003	0.144	**0.325**	0.099
	EDGM	-0.070	0.298	**0.439**	0.124
	ECMO	-0.168	0.419	0.387	**0.912**
	PRNC	-0.144	0.154	0.113	**0.538**
	PRND	-0.136	0.049	0.168	**0.589**
	ICLT	-0.235	0.295	0.498	**0.883**

The within-group structure confirms unidimensionality, mono-factorial validity and discriminant validity for all KPAs, demonstrating improved measures in the Economy area.

Table 3 shows the results of the structural component (causal model) of PLSPM for both structures. In both structural models, the scores of PCs between and PCs within, obtained by MSCA, were used as the initial LV scores (instead of utilizing arbitrary weights) in the first step of PLSPM algorithms.

For each endogenous LV, Table 3 shows the variance of the equation (Adjusted R-squared, AdjR^2) explained by its exogenous LVs, their specific contribution to the explained variance (%R^2), the estimated path coefficient, the empirical standard errors, t-statistics, both estimated with 500 bootstrap replications and significance values, respectively.

Tab. 3: Structural regression equations (between and within structure)

Structure	Endogenous LV	AdjR^2	Exogenous LV	%R^2	Path Coeff.	Std. error	t-value	Sign.
Between	Process	0.05	Human Capital	100%	-0.232	0.086	-2.692	0.008
	Patient Satisfaction	0.17	Human Capital	12%	0.107	0.083	1.286	0.201
			Process	88%	-0.381	0.083	-4.586	0.000
	Economy	0.43	Human Capital	7%	0.240	0.069	3.471	0.001
			Process	85%	0.718	0.074	9.686	0.001
			Patient Satisfaction	8%	0.345	0.074	4.700	0.001
Within	Process	0.05	Human Capital	100%	-0.311	0.081	-3.830	0.000
	Patient Satisfaction	0.10	Human Capital	11%	0.048	0.053	0.899	0.185
			Process	89%	0.199	0.086	2.318	0.011
	Economy	0.31	Human Capital	9%	-0.225	0.177	-1.265	0.104
			Process	52%	0.465	0.147	3.155	0.001
			Patient Satisfaction	39%	0.387	0.128	3.026	0.001

In the between structure, adjusted R-squared demonstrate, the weak predictive power of equations modeling Processes and Patient Satisfaction KPAs. Furthermore, in the first equation, the Human Capital coefficient presents a negative sign, whereas it is not statistically significant in the second equation.

By contrast, the third structural equation, explaining Economy KPA, shows discrete fit (AdjR^2 =0.43) and estimated coefficients consistent with expectations. The contribution of the Process dimension is particularly significant (contributing to 84% of the explained variance), followed by Patient Satisfaction (8%) and Human Capital (7%).

Results of the structural model in the within-structure are quite similar to those of the between structure, especially for first two equations.

In contrast to the between-structure model, the Economy structural equation presents lower fit (AdjR^2 =0.31) with a more balanced contribution of Process (52%) and Patient Satisfaction (39%) areas, while Human Capital becomes statistically not significant.

Finally, Table 4 shows the synthetic measures of fit for both structures. They refer to the structural components (the mean of R^2 over all endogenous model), to the measurement components (the mean AVE values over all LVs), and to the entire model (measuring the Absolute GoF, the geometric mean of synthetic fit measures for structural and measurement components).

Fit statistics confirm the weak performances of the specified models in both structures, due to the modest contribution of the structural component, despite the satisfactory goodness of fit of the measurement component.

Tab. 4: PLSPM goodness of fit indices (between and within structure).

Model Component	Index	Betweenstructure	Withinstructure
Measurement model	Mean AVE	0.652	0.642
Structural model	Mean R^2	0.217	0.153
Global model	Absolute GoF	0.370	0.314

These results show that the BSC theoretical framework, specifying three causal relationships between four Key performance areas, is too complex and not supported by enough empirical evidence to quantify the performance of the Lombardy Region Health System.

This may be caused by the lack of useful indicators: it is therefore not surprising that Human Capital scores, collecting only two manifest indicators, do not support the model, confirming weak significance and demonstrating unexpected signs of the relationships.

Moreover, we note that PLSPM results, proposed to the entire sample of Hospitals without taking into account the hierarchical structure of Health System organization (Hospitals nested in ASL), have revealed no significant causal relationships among the four latent dimensions (Lauro, 2007).

These considerations confirm the utility of PLSPM proposed in a version which embodies hierarchical data structure, as an appropriate statistical tool for further in-depth research on the estimation of causal relationships in the BSC framework.

5. CONCLUSION

This article has illustrated how a solid conceptual scheme such as the Balance Scorecard can be implemented in a statistical perspective to assess the causal relationships between Human Capital, Process production, Patient Satisfaction and Economy performances in health sectors.

Methodologically, we have utilized a proper statistical tool involving multi-dimensional latent constructs in the presence of hierarchical data, as typically occurs in the health sector. The PLSPM model in an MSCA perspective allows the exploration and estimation of two structural models, one for the between and one for the within structure, possibly revealing different causal structures.

As shown, the full integration of MSCA within PLSPM algorithm is still not satisfactory, essentially because two methods have different criteria. However, the orthogonality properties of MSCA scores in the between and in the within structure are respected in the PLSPM's final latent estimates.

Nevertheless, it would be opportune to conduct further research addressing the problem of PLSPM within a multilevel structure.

Recently, Multilevel Partial Least Squares (MLPLS, de Noord and Theobald, 2005) has been proposed as providing better options for analyzing such process data, but only in simplified models composed of two latent variables (PLS regression).

The results of application, focused on Hospitals in the Lombardy Region Health sector, have suggested that the performance of Process and Patient Satisfaction are poorly modeled in a Balance Scorecard framework. Moreover, the Human capital dimension is neither robust nor significant and therefore is not appropriate for investigating its impact on economic performance.

On the whole, the hypothesized model for BSC exhibits weak causal relations in the equations, despite the satisfactory goodness of fit of the measurement models (especially for the Process area, whereas Patient Satisfaction and Economy suffer from weak mono-factorial validity).

Only the Economy equation delivers results in-line with expectations, with, however, modest fit, demonstrating that efforts to improve patient satisfaction and process quality produced a minimal statistical impact on Hospitals' financial resources.

Results concerning the measurement models and the Economy structural equation suggest that a more simplified version of the theoretical model would yield more promising results, particularly in the presence of more complete data.

One possible factor that may explain the modest fit of the empirical results is a misspecification of the causal model.

To this end, some authors argue that the assumptions of causality relationships between the four KPAs in the BSC framework remain subjective and may change in different areas (Ittner and Larcker 1998; Morard and Stancu, 2005; Kanji and Moura, 2002; Chow *et al.*, 1998). Some empirical studies found no causality between quality and financial results (Ittner and Larcker, 1998), whereas others (Kanji and Sà, 2002; Morard and Stancu, 2005) threw doubt upon the causal structure among the KPA areas proposed by Kaplan and Norton (2001).

They argue that the financial stability of a company is based on the indispensable conditions necessary to achieve the goals of the three other perspectives, which in the case of Heath structures are: "patient" perspective (Satisfaction), "employee" perspective (Human Capital) and "quality" perspective (Process). Further, Zelman and colleagues (1999) state that heterogeneity of Health organizations may mitigate the full benefit of the BSC approach, suggesting some key modifications to take into account those unique characteristics. Chow and colleagues (1998) argue that each Health organization must engage in the full range of strategic management activities, from defining its mission to the selection of goals and strategies, in order to develop its own unique scorecard toward the selected goals.

However, the modest quantity of available indicators in Lombardy regional health archives appears to be the principal factor limiting the strategic perspective of the present study.

Much could be learned about BSC's role in developing effective strategies by investigating the information properties of the available systems. The improvement of existing regional archives, providing data availability at a fine grained level, represents an imperative priority for further research in this area.

The limitations of the study notwithstanding, the present paper provides an illustration of how a solid statistical methodology can be used to explore the implementation of BSC and how the dimensions of BSC facilitate the development of competitive strategic outcomes, helping advance understanding by identifying the areas with the most significant impact on Hospital economic performance.

REFERENCES

AHN H. 2001. Applying the balanced scorecard concept: an experience report. *Long Range Planning*. **34** (4): 441-461.

ATKINSON AA, BALAKRISHNAN R, BOOTH P, COTE J. et al., 1997. New directions in management accounting research. *Journal of Management Accounting Research* —**9**: 79-108.

BAKER GR, PINK GH. 1995. A balanced scorecard for Canadian hospitals. *Healthcare Management Forum* **8** (4): 7-21.

BENTLER PM. 1982. Linear system with multiple levels and types of latent variables. In *System* *under indirect observation: causality, structure, prediction*, Joreskog K, Wold H. (Eds.). North Holland: Amsterdam; 101-130.

BLUMENTHAL D. 1996. Quality of care—what is it? *The New England Journal of Medicine* **335** (12): 891-894.

CHIN WW. 1998. The partial least squares approach to structural equation modeling. In *Modern Methods for Business Research.*, Marcoulides GA. (Ed.), Erlbaum: New York; 295-336.

CHOW CW, GANULIN D, TEKNIKA O, HADDAD K *et al.*, 1998. The balanced scorecard: a potent tool for energizing and focusing healthcare organization management."*Journal of Healthcare Management* 43 (3): 263-280.

COX DR, FITZPATRICK R, FLETCHER AE, GORE *et al.*, 1992. Quality of life assessment: can we keep it simple? *Journal of the Royal Statistical Society*, Series A, **155**: 353-393.

DE NOORD OE, THEOBALD EH, 2005. Multilevel component analysis and multilevel PLS of chemical process data. *Journal of Chemometric*, **19:** 301–307.

GELPER S. CROUX C. 2007. The predictive power of the European Economic Sentiment Indicator (ESI).*FETEW Research report.* Katholieke Universiteit Leuven, KBI_0707.

Healthcare Directorate of Lombardy Region. 2004. *Triennial program for the implementation of an evaluation system and management of health structures* Lombardy Region press. Milano.

HULLAND J. 1999. Use of partial least squares (PLS) in strategic management research: a review of four recent studies. *Strategic Management Journal* **20** (1): 195-204.

INAMDAR N, KAPLAN RS, BOWER M. 2002. Applying the balanced scorecard in healthcare provider organizations. *Journal of Healthcare Management*: 179-195.

ITTNER CD, LARCKER D.F. 1998. Innovations in performance measurement: trends and research implications. *Journal of Management Accounting Research* **10**: 205-238.

KANJI GK, SÁ PME. 2002. Kanji's business scorecard. *Total Quality Management* **13** (1): 13-27.

KAPLAN RS, NORTON DP. 1992. The balanced scorecard. The measures that drive performance. *Harvard Business Review* **70** (1): 171-179.

KAPLAN RS, NORTON DP. 2001. *The strategy-focused organization: how balanced scorecard companies thrive in the new business environment.* Harvard Business School Press. Boston.

KMENTA J. 1991. Latent variables in econometrics. *Statistica Neederlandica* **45:** 73-84.

LAURO NC. 2007. Valutazione dell'impatto delle politiche. Realizzazione di un sistema pilota in ambito sanitario. *IReR publication* 2007 **A023**. Milano.

MORAND B., STANCU A., 2005. Structural equation modeling in a rationalization tentative of balanced scorecard. *Cahiers de Recherche* **09**. Université de Genève. Genève.

MUTHÉN B. 1994. Multilevel covariance structure analysis. *Sociological Methods & Research.* **22**: 376-398.

NORTHCOTT D., FRANCE N. 2005. The balanced scorecard in New Zealand health sector performance management: dissemination to diffusion. *Australian Accounting Review* **15** (37): 34-46.

OLIVEIRA J. 2001. *The balanced scorecard: an integrate* approach to performance evaluation. *Healthcare Financial Management* **55** (5): 42-46.

ØVRETVEIT J, GUSTAFSON D. 2003. Using research to inform quality programmes. *British Medical Journal,* **326** (5): 759-61.

PINK GH. MCKILLOP I, SCHRAA EG, PREYRA *et al.*, 2001. *Creating a balanced scorecard for a hospital system. Journal of Healthcare Finance* **27** (3): 1-20.

SAHNEY VK. *1998*. Balanced scorecard as a framework for driving performance in *managed care* organizations. *Managed Care Quarterly* **6** (2): 1-8.

STELZL I. 1986. Changing the causal hypothesis without changing the fit: some rules for generating equivalent path models. *Multivariate Behavioral Research* **21:** 309-331.

TENENHAUS M, VINZI V, CHATELIN YM, LAURO NC. 2005. Partial least squares. *Computational Statistics and Data Analysis* **48** (1): 159-205.

TIMMERMAN ME. 2006. Multilevel component analysis, *British Journal of Mathematical and Statistical Psychology* **59**: 301-320.

TUOMELA TS. 2005. The Interplay of different levers of control: a case study of introducing a new performance measurement system. *Management Accounting Research* **16** (3): 293-320.

VITTADINI G. 1989. Indeterminacy problems in the LISREL model. *Multivariate Behavioral Research* **24** (4): 397-414.

WOLD H. 1982. Soft modelling: the basic design and some extensions. In *Systems under indirect observation: causality, structure, prediction,* Jöreskog K, Wold H. (Eds.). North Holland: 1-54.

ZELMAN WN, PINK GH, MATTHIAS CB. 2003. Use of the balanced scorecard in health care. *Journal of Healthcare Finance* **29** (4): 1-16.

ZELMAN WN, BLAZER D, GOWER JM, BUMGARNER PO. *et al.,* 1999. Issues for academic health centers to consider before implementing a balanced-scorecard effort. *Academic Medicine* **74** (12): 1269-1277.

IL BALANCED SCORECARD E LA VALUTAZIONE DI SISTEMI SANITARI: UN APPROCCIO CON VARIABILI LATENTI

Riassunto

Il Balanced Scorecard è una tecnica utilizzata nell'ambito del management strategico, per l'implementazione di strategie aziendali all'interno un complesso sistema di relazioni tra misure sintetiche (KPA, Key Performance Area) aziendali.

Il presente articolo descrive un approccio per l'implementazione del BSC al fine di misurare e valutare le performance sanitarie ed economiche delle strutture di cura, suggerendo, agli stakeholders regionali, ottiche interpretative e di fattibilità per la costruzione di strategie di valutazione e di benchmarking tra erogatori. Il contributo originale consiste nell'esplorazione di basi dati e indicatori esistenti a livello regionale e nella proposta di metodologie statistiche che possono facilitare la diffusione e l'utilizzo del BSC come schema concettuale per una valutazione di sistema della sanità lombarda.

Dal punto di vista metodologico, la struttura di relazioni causali tra le KPA specificate nel BSC è stata indagata utilizzando l'approccio Partial Least Square Path Modelling in un'ottica multilevel, modo da tener conto della struttura gerarchica delle osservazioni disponibili (176 Ospedali, valutati nel 2007, appartenenti alle ASL lombarde).

I risultati empirici hanno evidenziato due differenti strutture causali nella componente "between" e in quella "within", e solidi modelli di misura per la stima dei punteggi delle KPA. Tuttavia, le relazioni causali tra le dimensioni specificate nel BSC sono statisticamente deboli, suggerendo possibili modificazioni della struttura di relazioni di causalità tra KPA, contenute nella specificazione originale del BSC.