BMJ Open

How do University Hospitals perform compared to General Hospitals? Some evidence from Italy

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-011426
Article Type:	Research
Date Submitted by the Author:	05-Feb-2016
Complete List of Authors:	Nuti, Sabina; Scuola Superiore Sant\'Anna, Institute of Management - Management&Health Laboratory Grillo Ruggieri, Tommaso; Scuola Superiore Sant\'Anna, Institute of Management - Management&Health Laboratory Podetti, Silvia; Scuola Superiore Sant\'Anna, Institute of Management - Management&Health Laboratory
Primary Subject Heading:	Health services research
Secondary Subject Heading:	Health policy
Keywords:	Performance, Evaluation, University, Hospital, Italy

How do University Hospitals perform compared to General Hospitals? Some evidence from Italy

Corresponding author:
Sabina Nuti
Postal Address: Laboratorio MeS, Istituto di Management - Scuola Superiore Sant'Anna,
Via San Zeno 2 – 56127, Pisa, Italy
Email address: s.nuti@sssup.it
Telephone: +39 050 88 38 86
Fax: +39 050 88 38 90

Authors:
Sabina Nuti, Full Professor, Management and Health Laboratory, Institute of Management, Scuola Superiore Sant'Anna, Via San Zeno 2 – 56127, Pisa, Italy, s.nuti@sssup.it.

Tommaso Grillo-Ruggieri, PhD candidate, Management and Health Laboratory, Institute of Management, Scuola Superiore Sant'Anna, Via San Zeno 2 – 56127, Pisa, Italy, t.grilloruggieri@sssup.it

Silvia Podetti, PhD, Management and Health Laboratory, Institute of Management, Scuola Superiore Sant'Anna, Via San Zeno 2 – 56127, Pisa, Italy, s.podetti@sssup.it

Keywords: performance, evaluation, university, hospital, Italy
Word Count: 3,461

Abstract

Objective

Aim of this research is to investigate how UHs perform compared to GHs in the Italian healthcare system.

Design and setting

27 overall performance indicators were selected and analysed for UHs and GHs in ten Italian Regions. The data refer to 2012 and 2013 and were selected from two performance evaluation systems based on administrative data on hospitalization discharge flows: the Inter-Regional Performance Evaluation System developed by the Management and Health Laboratory of the Scuola Superiore Sant'Anna of Pisa and the Italian National Outcome Evaluation Programme developed by the National Agency for Healthcare Services. The study was conducted in two stages and by combining two statistical techniques. In stage 1, a nonparametric Mann-Whitney *U* test was carried out to compare the performance of UHs and GHs on the selected set of indicators. In stage 2, a robust equal variance test between the two groups of hospitals was carried out to investigate differences in the amount of variability between them.

Results

The overall analysis showed heterogeneous results. In general, being in the UHs group rather than the GHs does not generally affect performance. Thus, Italian UHs cannot straightforwardly be associated with better results in terms of appropriateness, efficiency, patient satisfaction, and outcomes.

Conclusions

Policy-makers and managers should further encourage hospital performance evaluations in order to stimulate wider competition aimed at assigning *teaching status* to those hospitals that are able to assure requirements in terms of performance. In addition, UH facilities could be integrated with other providers in charge of community, primary and outpatient services creating a joint accountability for more patient-centred and integrated care.

Strenghts and limitations of the study

- This study provides evidence about differences in terms of performance between University Hospitals and General Hospitals that was lacking for the Italian context
- The analysis shows new results about hospital performance that can contribute to the debate on this topic
- A non parametric approach of analysis is applied first to Italian context
- The study is limited to the Italian healthcare system and its organizational structure

- There could be other indicators of performance as valuable and informative as those measures included in the analysis

1. INTRODUCTION

University Hospitals (UHs) can be considered as complex organizations as their mission includes three different objectives: patient care, education, and research [1]. UHs combine all the features of Minzberg's Professional Bureaucracy [2] embedded within both the healthcare organizations and the university context. In addition, UHs are usually referral centres for most complex care within a hub-and-spoke hospital network [3].

Given the three-fold mission of these institutions and the specific role that they play in the healthcare system, should UHs be considered as a 'cluster' with specific performance patterns?

This study investigates whether UHs behave homogenously regarding performance results with substantial differences with respect to general hospitals (GHs).

Evidence on this topic could provide important information for policy-makers and managers in defining specific policies and actions in order to improve the quality of care within the hospital regional network, where UHs play a specific and strategic role and in order to pursue their specific mission. In particular, in Italy as in other countries, UHs are in charge of the strategic role of training doctors of the future. Therefore, since health professionals are the most important assets for the healthcare organizations, policy-makers should ensure clinicians are trained and supported in their practice by institutions that can assure the appropriate requirements in terms of quality of care and research productivity. The analysis was carried out in Italy.

2. BACKGROUND

Teaching status has been already investigated from several perspectives by studying whether it affects the results of UHs compared with other hospitals in terms of outcomes, quality of care, productivity, costs, etc.

Firstly, reviews on outcomes, quality of care and adverse-event prevention reached mixed conclusions and highlighted the need for evidence on differences between UHs and GHs [4-5]. Some reviews underlined better overall results for UHs [6-7]; whereas, a systematic review highlighted no differences between UH and GH outcomes [8].

Secondly, studies on productivity and efficiency have usually applied Data Envelopment Analysis (DEA) and frequently highlighted better performance of GHs with respect to UHs [among others, 9-10].

Indeed, training resident students, carrying out research activities besides patient care and the role of referral centres for complex care have often been identified as elements that can increase costs [11-13]. This frequently drives additional financial resources to UHs (e.g., an increased mark-up in the reimbursement system for UH discharges) [6].

Research on this topic presents several differences in terms of data sources, measurement processes and methodology for data analysis [4]. This could raise potential issues regarding external validity and result generalizability [6-9]. Examples of these differences are:

- the data sources: e.g. medical records or administrative data;
- the definition of UHs and their ownership (public, private, for-profit, non-profit): for example, some studies consider only major UHs, whilst others include all the hospitals with a residency program;
- the indicators included in the analysis (usually, outcomes, quality of care or efficiency) and the different calculation criteria and risk-adjustment procedure used for the same measures (mortality rates, process measures, etc.);
- the statistical methods used to compare hospitals (parametric and non-parametric approaches and tests such as DEA, ANOVA, Kruskall-Wallis, Mann-Whitney, etc).

These differences may partially explain why research looking at different performance or outcomes in UHs or controlling for a potential effect of the *teaching status* have not led to straightforward results.

Finally, results may be also associated with the specific geographical context. For instance, in one of the most recent systematic reviews on this topic, more than three-fourths of the studies included in the analysis were conducted in the United States [8]. However, each specific geographical and health system context may play an important role in explaining results.

With reference to Italy, detailed studies are also lacking on this topic. Scholars have focused on governance issues or research evaluations [see for instance, 14-17]. There have been no systematic comparisons of performances between the two groups of hospitals and related research.

2.1. The Italian context

The national healthcare system in Italy follows a Beveridge model by providing universal coverage through general taxation. Regional governments are responsible for organizing and delivering health services and being accountable for performance. National government monitors the pursuit of the universal coverage in particular with respect to a package of essential services (Nationally defined basic health benefit package - Livelli Essenziali di Assistenza). National government allocates financial resources to the regional governments on an adjusted capitation basis. Regions then

reallocate resources to Local Health Authorities (LHAs), through a regionally-adjusted capitation formula.

The Italian UHs are identified as those public or private hospital authorities with both *teaching* and *research status* that are integrated within a public or private University School of Medicine. UHs are autonomous organizations with respect to LHAs, which manage the healthcare delivery in their own geographical area. In Italy, hospital care is also delivered by private or public autonomous hospitals (AHs), public general hospitals (GHs) directly managed by the LHAs and, for specific highly-specialized services, by private or public research hospitals (RHs).

Since UHs are autonomous authorities, they are not financed by LHA resources through capitation-based funding. Financial resources are directly allocated for the healthcare services delivered (through DRG tariffs) by Regional Governments, which may also assign additional resources to cover the added costs for education and research (usually through an increased percentage reimbursement of the DRG tariffs or through specific restricted funds). Therefore, the percentage or fixed amount of additional resources varies depending on the regional decisions [14].

Italian UHs have on average a much higher number of hospital beds with respect to GHs and are hub-referral centres for highly-complex and highly-specialized care, such as neuro-surgery, cardio-surgery, radiotherapy, most critical intensive care, paediatric highly-complex surgery, etc.

Evidence from Italy on the comparison of UH performance with respect to GHs may provide valuable information for both healthcare policy-makers and managers, at both regional and national level and not only in Italy. Indeed, if UHs behave as a specific 'cluster', new policies and focused-actions could be defined to support the specific role of these authorities within the hospital network in the regional and national contexts. Evidence of similar patterns of performance between these two groups of hospitals may highlight the need to look for other sources of variation between the two groups. As a consequence, other features from the *teaching and research status* may be relevant to inform policies on hospital governance, financing and network organization, considering the crucial role of UHs in training the future clinicians for the healthcare system.

The aim of this paper is thus to investigate how UHs perform in comparison to GHs.

3. METHODS

3.1. Data sources and hospital selection

The data used in this analysis were selected from two performance evaluation systems based on the same source of administrative data on hospitalization discharge flows:

- The *Inter-Regional Performance Evaluation System* (IRPES) developed by the Management and Health Laboratory of the Scuola Superiore Sant'Anna of Pisa (MeS-Lab) - where the authors of this

paper are researchers. This system provides a multi-dimensional evaluation of performance including efficiency, appropriateness, integration and quality of care. This system was firstly implemented by the regional government in Tuscany [18-19] and was then adopted – on a voluntary basis - by the majority of other Italian regions** [20-21]. The evaluation process measures through benchmarking and with specific risk adjustment processes the results achieved every year by all the Health Authorities (the local health authorities, the university hospitals, the research hospitals and the autonomous hospitals) located in these regions. Results are publicly reported [22].

- The *Italian National Outcome Evaluation Programme* (NOEP) developed by the National Agency for Healthcare Services on behalf of the Ministry of Health. This system measures outcomes nation-wide [23], i.e. for each Italian hospitals. On the basis of rigorous risk adjustment processes [among others, 24-25], these measures represent assessment tools to support clinical and organizational audit programs aimed at improving both outcome and equity in the National Health System.

Data refer to the hospital activity of 2012 and 2013, apart from two economic indicators related to balance sheets, which are available only for 2011 and 2012.

Two groups of hospitals differing in particular because of the *teaching status*, the organizational autonomy with respect to the LHAs and, due to the hospital network organization, to the average number of hospital discharges (in 2012, 32,472 for UHs and 17,606 for GHs) and the average DRG weight (in 2012, 1.3 for UHs and 1.06 for GHs) were considered in the analysis. The whole study included all the 16 UHs and 73 LHAs of the ten IRPES regions.

3.2. Performance indicators

For the purposes of this study, 27 performance indicators were selected, 10 from IRPES (Table 1) and 17 from NOEP (Table 2).

Table 1 – IRPES Indicators

IRPES INDICATORS	Rationale
EFFICIENCY AND APPROPRIATENESS	
Relative stay index: case-mix adjusted differential average Length of Stay - LOS days	Measure of the average difference from the standard LOS for admitted patients with adjustments for case-mix
Percentage of medical discharges with LOS > ministerial threshold for patients over 65	Measure of the hospital compliance with the Italian Ministry of Health standards for the LOS for medical inpatient activity for elderly patients. This measure is a proxy of the effective implementation of integrated pathways between home, community-based and hospital care for elderly patients
Percentage of ED green-coded patients visited within 1 hour	Measure of timely emergency care for ED patients whose treatment may be delayed without risk
Percentage of ED patients referred for hospital admission with ED length of stay <= 8 hours	Measure of overall timely emergency care

Percentage of medical inpatient discharges within 2 days (National Healthcare Agreement 2010)	Measure of hospital compliance in avoiding short ordinary hospitalizations for patients that could be treated in outpatient clinics or in other care settings, as requested by the Italian Ministry of Health standards in the National Healthcare Agreement of 2010
Percentage of day-surgery treatment for specific procedures (National Healthcare Agreement 2010)	Measure of hospital compliance with Italian Ministry of Health standards for delivering specific not-complex surgical procedures in day-surgery or in outpatient clinics rather than through ordinary hospitalizations
PATIENT SATISFACTION	
Percentage of patients leaving ED against/without medical advice	Proxy of Patient Satisfaction on ED services and waiting times
Percentage hospitalized patients leaving against medical advice	Proxy of Patient Satisfaction for the inpatient activity
ECONOMIC AND FINANCIAL EVALUATION	
Average cost per weighted case	Measure of the ratio of a hospital acute inpatient care expenses to the number of acute inpatient cases weighted for the DRG complexity. The weighting enhances comparability across hospitals
Average expenditure per Diagnostic Imaging weighted for tariff	Measure of efficiency that compares costs and the value of the delivered diagnostic activity (sum of ambulatory tariffs)

Table 2 – NOEP Indicators

NOEP INDICATORS
OUTCOME: measures of 30-day mortality or re-admissions for relevant inpatient activity
AMI: 30-day mortality
AMI without PTCA: 30-day mortality
AMI with PTCA within 2 days: 30-day mortality
AMI with PTCA after 2 days: 30-day mortality
AMI: 1-year mortality
AMI: MACCE after 1 year
Isolated Aortocoronary Bypass: 30-day mortality
Valvuloplasty or heart valve replacement: 30-day mortality
Congestive heart failure: 30-day mortality
Ischemic Stroke: 30-day mortality
Ischemic Stroke: 30-day re-admission
COPD exacerbation: 30-day mortality
COPD: 30-day re-admission
Proportion of Caesarean Section
Femur Fracture: 30-day mortality
Femur Fracture: Percentage of operations carried out within 2 days
Colon cancer surgery: 30-day mortality

Eight IRPES indicators regard efficiency and appropriateness, patient satisfaction, and economic and financial dimensions. Two indicators regard economic and financial evaluation. This selection was shared by the group of the IRPES regional representatives. This group is in charge of systematically

reviewing and discussing the measures included in the IRPES as relevant proxies for measuring performance in a multidimensional perspective in all the different settings of care [20].

For both sources of the selected indicators, the time coverage and the number of providers needed to perform the statistical test were guaranteed, thus ensuring the consistency of the comparative analysis between the two groups of hospitals in this single-country study [among others, 26-27].

The number of observations for the NOEP indicators may differ because not all the hospitals included in the analysis provide all the healthcare services linked to the included measures. However, the selection of these measures took into account the services usually provided by both LHA-GHs and UHs.

The analysis for the IRPES indicators compared the 16 UHs to the 73 LHAs. On the other hand, the analysis for the NOEP indicators was carried out at the hospital level, thus comparing the (at most) 19 facilities of the 16 UHs to the individual (at most) 191 GHs led by the 73 LHAs. (See Appendix I in the Supplementary File for the complete list of hospitals considered and the number of observations included for each indicators).

3.3. Statistical Methods

The study was conducted in two stages and by combining two statistical techniques. Data were processed using Stata software, version 12. In stage 1, a nonparametric Mann-Whitney *U* test was carried out to compare the performance of UHs and GHs on the selected set of indicators. This analysis determines whether or not UHs and GHs were drawn from the same target population. Previous studies have already applied this univariate analysis to illustrate differences between hospitals [among others, 28] because of its appropriateness with small samples [29-33]. For the purposes of this study, this test verified whether or not there were differences between UH and GH performance, or, in other words, whether or not UHs and GHs could be considered as two different clusters. In stage 2, we carried out a robust equal variance test between the two groups of hospitals [34] to investigate differences in the amount of variability between UHs and GHs. This test is usually used to verify the assumption of homogeneity of variance across groups, meaning that the internal variability of one group of hospitals is not significantly different with respect to the other one.

To be in line with the assumptions of the Mann-Whitney *U* test, we used an extension of Levene's test as suggested by Brown and Forsythe [35]. We applied the test only for those indicators in which the Mann-Whitney *U* test did not show significant differences between UH and GH performances. Indeed, in those cases where the performance between the two groups did not show significant differences, we tested whether there were specific patterns in terms of variability.

4. RESULTS

The Mann-Whitney *U* test on IRPES indicators showed that in relation to four measures of "Efficiency and Appropriateness" and "Economic and financial evaluation" dimensions, there were differences in performance between UHs and GHs. The test, in fact, was significant both in 2012 and 2013 for the "% of ED green-coded patients visited within one hour", the "% of medical inpatient discharges within two days" and the "% of day-surgery treatment for specific procedures". The test was significant also in both 2011 and 2012 for the "Average expenditure for Diagnostic Imaging weighted for tariff". For these indicators, GHs seemed to perform better than UHs.

On the other hand, with reference to the indicators "Relative stay index", "% of medical discharges with LOS > ministerial threshold for patient over 65", and "% of ED patient referred for hospital admission with ED length of stay <= 8 hours", the Mann-Whitney *U* test was rejected for both 2012 and 2013.

Moreover, no significant differences were found for patient satisfaction proxies "% of patients leaving ED against/without medical advice" and of "% of hospitalized patients leaving against medical advice". Moreover, in 2013 UHs accounted for fewer patients that left ED or who were discharged against medical advice, whilst in 2012 the GHs achieved better results. The test was also not significant for the "Average cost per weighted case".

Table 3 summarizes the results of the test and illustrates the average and the median values of the two groups of hospitals for each of the indicators.

Table 3 – Mann-Whitney *U* test for IRPES indicators

MANN – WHITNEY *U* TEST IRPES INDICATORS	2012					2013				
	Median UH	Median GH	Mean UH	Mean GH	Best Perf. MEDIAN	Median UH	Median GH	Mean UH	Mean GH	Best Perf. MEDIAN
EFFICIENCY AND APPROPRIATENESS										
Relative stay index (case-mix adjusted differential average LOS days)	-0.3	-0.1	-0.1	-0.2	UH	-0.2	-0.3	-0.2	-0.3	GH
% medical discharges with LOS > ministerial threshold for patients over 65	4.7	3.6	4.6	4	GH	3.7	3.5	4.2	3.8	GH
% ED green-coded patients visited within 1 hour	**72.7**	**79.2**	**70.7**	**77.3**	**GH***	**68**	**77.2**	**64.9**	**76.2**	**GH***
% ED patients referred for hospital admission with ED length of stay <= 8 hours	98.1	97.8	93.9	94.8	UH	97.5	97.5	92.8	94.5	UH/GH
% medical inpatient discharges within 2 days (National Healthcare Agreement 2010)	**21.3**	**14.6**	**22.1**	**14.9**	**GH***	**21.2**	**14.1**	**21.8**	**14.4**	**GH***
% day-surgery treatment for specific procedures (National Healthcare Agreement 2010)	**46.5**	**58.7**	**49.1**	**58.9**	**GH***	**49.2**	**59.1**	**50.2**	**59**	**GH***

PATIENT SATISFACTION										
% patients leaving ED against/without medical advice	3	3.2	3.4	3.1	GH	3.3	3.2	3.4	3.4	UH
% hospitalized patients leaving against medical advice	0.9	0.8	1	1	GH	0.7	0.8	0.9	0.9	UH
ECONOMIC AND FINANCIAL EVALUATION										
	2011					**2012**				
Average cost per weighted case	4,303	4,220	4,678	4,348	GH	4,469	4,516	4,659	4,651	UH
Average expenditure per Diagnostic Imaging weighted for tariff	**1.3**	**0.9**	**1.8**	**1.1**	**GH***	**1.4**	**1**	**1.6**	**1.1**	**GH***

Regarding the test for the NOEP indicators, for all the tested measures, the Mann-Whitney *U* test was not significant except for two measures that showed mixed results in 2012 and in 2013 (Table 4).

Table 4 - Mann-Whitney *U* test for NOEP indicators

MANN-WHITNEY *U* TEST – NOEP INDICATORS	2012					2013				
	Median UH	**Median GH**	**Mean UH**	**Mean GH**	**Best perf. MEDIAN**	**Median UH**	**Median GH**	**Mean UH**	**Mean GH**	**Best perf. MEDIAN**
OUTCOME INDICATORS										
AMI: 30-day mortality	9.6	8.8	10.0	9.3	GH	9.4	7.7	9.1	8.1	GH
AMI without PTCA: 30-day mortality	17.4	15.5	17.8	16.5	GH	17.2	15.0	18.1	15.5	GH
AMI with PTCA within 2 days: 30-day mortality	4.7	4.1	4.5	4.2	GH	4.2	3.7	4.4	3.8	GH
AMI with PTCA after 2 days: 30-day mortality	2.4	2.4	3.1	2.6	GH	2.6	2.5	2.9	2.8	GH
AMI: 1-year mortality	10.2	11.1	10.4	11.5	UH	9.7	10.6	10.1	10.9	UH
AMI: MACCE after 1 year	23.9	24.8	24.4	25.2	UH	22.3	23.2	22.8	23.6	UH
Isolated Aortocoronary Bypass: 30-day mortality	1.8	1.9	2.2	2.0	UH	1.8	2.3	2.3	2.1	UH
Valvuloplasty or heart valve replacement: 30-day mortality	2.6	3.7	2.9	3.5	UH	2.3	3.0	2.6	3.2	UH
Congestive heart failure: 30-day mortality	8.7	9.8	9.3	10.8	UH	**8.9**	**10.7**	**8.7**	**11.0**	**UH***
Ischemic Stroke: 30-day mortality	9.3	10.1	8.6	10.5	UH	9.2	9.5	9.2	10.5	UH
Ischemic Stroke: 30-day re-admission	10.7	9.4	10.4	10.3	GH	6.5	6.7	7.1	7.2	**UH**
COPD exacerbation: 30-day mortality	7.3	8.7	7.6	8.9	UH	7.9	8.1	7.8	8.7	UH
COPD: 30-day re-admission	14.3	15.6	15.0	15.4	UH	14.4	15.4	14.2	15.4	UH
Proportion of Caesarean Section	19.8	17.8	23.1	18.8	GH	20.2	18.5	22.9	19.3	GH
Femur Fracture: 30-day mortality	4.3	4.8	4.8	5.1	UH	4.4	4.7	4.7	4.8	UH
Femur Fracture: Percentage of operations carried out within 2 days	**48.6**	**54.4**	**42.4**	**53.2**	**GH***	51.6	60.3	55.1	59.8	GH
Colon cancer surgery: 30-day mortality	3.4	3.9	4.2	4.3	UH	3.0	4.2	3.7	4.6	UH

For the "Congestive heart failure: 30 day mortality" the test showed no statistical differences between UHs and GHs in 2012. However, a significantly better performance for UHs was found in 2013. Similarly, in the case of the indicator "Femur fracture: % of operations carried out within two days",

the Mann-Whitney *U* test showed significant differences between UHs and GHs in 2012, but not for 2013, with GHs having the best median performance.

In order to investigate different variations between the two groups of hospitals, the robust equal variance test [35] was carried out for a set of 23 indicators (6 IRPES indicators and 17 NOEP indicators) that rejected the Mann-Whitney *U* test.

Regarding IRPES indicators, the test was always not significant for both years included in the analysis (Table 5). UHs and GHs showed a higher standard deviation depending on the measures considered.

Table 5 – Robust Equal Variance test for IRPES Indicators

ROBUST EQUAL VARIANCE TEST - IRPES INDICATORS	2012				2013				Higher Variability In 2012	Higher Variability In 2013
	Std. Dev. UH	Std. Dev. GH	W50 - Median	Pr > F	Std. Dev. UH	Std. Dev. GH	W50 - Median	Pr > F		
EFFICIENCY AND APPROPRIATENESS										
Relative stay index (case-mix adjusted differential average LOS days)	0.9	1.4	0.3	0.6	0.9	1.2	0.6	0.4	GH	GH
% medical discharges with LOS > ministerial threshold for patients over 65	1.7	2	0.2	0.7	1.7	2.1	1	0.3	GH	GH
% ED patients referred for hospital admission with ED length of stay <= 8 hours	8.7	6.7	0.4	0.5	9.5	7.7	0.6	0.5	UH	UH
PATIENT SATISFACTION										
% patients leaving ED against/without medical advice	2	1.8	0	1	2.1	2.1	0.1	0.8	UH	UH/GH
% hospitalized patients leaving against medical advice	0.6	0.7	0.2	0.7	0.6	0.6	0	0.9	GH	UH/GH
ECONOMIC AND FINANCIAL EVALUATION										
	2011				2012				Higher Variability In 2011	Higher Variability In 2012
Average cost per weighted case	1089	775	1.2	0.3	985	850	0.5	0.5	UH	UH

For the 2012 results of NOEP indicators, the test was significant for five measures (Table 6):

- "AMI: 1-year mortality" (p-value=0.02)
- "Ischemic stroke: 30-day mortality" (p-value=0.02)
- "COPD exacerbation: 30-day mortality" (p-value=0.04)
- "Femur fracture: 30-day mortality" (p-value=0.04)
- "COPD: 30-day readmission" (p-value=0.01)

Table 6 – Robust Equal Variance test for NOEP indicators

ROBUST EQUAL VARIANCE TEST - NOEP INDICATORS	2012				2013				*Higher Variability in 2012*	*Higher Variability in 2013*
	Std. Dev. UH	Std. Dev. GH	W50 - Median	Pr > F	Std. Dev. UH	Std. Dev. GH	W50 - Median	Pr > F		
OUTCOME INDICATORS										
AMI: 30-day mortality	3.2	3.8	1.2	0.3	2.7	3.7	2.6	0.1	GH	GH
AMI without PTCA: 30-day mortality	4.7	6.2	1.4	0.2	4.7	6.6	1.5	0.2	GH	GH
AMI with PTCA within 2 days: 30-day mortality	1.4	1.9	1.2	0.3	1.8	2.1	0.9	0.4	GH	GH
AMI with PTCA after 2 days: 30-day mortality	1.6	1.4	0.3	0.6	1.1	1.4	1.0	0.3	UH	GH
AMI: 1-year mortality	**2.1**	**4.4**	**5.2**	**0.02***	3.2	3.8	0.2	0.6	**GH***	GH
AMI: MACCE after 1 year	4.0	5.3	2.5	0.1	**3.3**	**5.6**	**4.3**	**0.04***	GH	**GH***
Isolated Aortocoronary Bypass: 30-day mortality	1.4	1.6	0.0	0.9	1.6	1.4	0.0	1.0	GH	UH
Valvuloplasty or heart valve replacement: 30-day mortality	1.3	0.5	2.7	0.1	1.3	1.0	0.3	0.6	UH	UH
Congestive heart failure: 30-day mortality	3.2	5.0	2.1	0.2					GH	
Ischemic Stroke: 30-day mortality	**2.9**	**4.5**	**5.7**	**0.02***	3.9	4.6	0.6	0.4	**GH***	GH
Ischemic Stroke: 30-day re-admission	3.5	3.9	0.0	0.9	2.2	3.0	1.9	0.2	GH	GH
COPD exacerbation: 30-day mortality	**2.3**	**3.9**	**4.2**	**0.04***	2.9	4.1	1.4	0.2	**GH***	GH
COPD: 30-day re-admission	**2.3**	**4.5**	**6.7**	**0.01***	3.3	4.2	1.8	0.2	**GH***	GH
Proportion of Caesarean Section	9.0	7.1	1.0	0.3	9.4	7.1	1.9	0.2	UH	UH
Femur Fracture: 30-day mortality	**1.4**	**2.2**	**4.1**	**0.04***	2.0	2.2	0.9	0.4	**GH***	GH
Femur Fracture: Percentage of operations carried out within 2 days					16.5	17.5	0.4	0.5		UH
Colon cancer surgery: 30-day mortality	2.6	2.3	0.1	0.8	1.7	2.5	3.2	0.1	UH	GH

In 2013, the test was significant only for the indicator "AMI: MACCE after 1 year" (p-value=0.04). For these five measures, GHs frequently showed a higher Standard Deviation with respect to UHs. This was also the case for most of the other outcome measures included for both 2012 and 2013, apart from the "Proportion of Caesarean Section" and the "30-day mortality rate for Valvuloplasty or heart replacement".

5. DISCUSSION

The overall analysis showed heterogeneous results when comparing the two groups of hospitals.

Considering the IRPES indicators of appropriateness, we found a higher compliance of GHs in pursuing the Italian Ministry of Health standards on directing patients to the appropriate care settings for surgical treatments as well as in avoiding short medical hospitalizations and giving preference to outpatient clinics or day-hospital cases. This may be due to the lower complexity of general LHA-led hospitals and to a related lower complex management.

Regarding efficiency, in 2013 GHs seemed to perform better than the UHs but these results are slightly different in 2012, thus leading to ambiguous conclusions. Therefore, the three-fold mission and the greater organizational complexity of UHs seemed to lead to lower but not significantly different efficiency with respect to GHs. The more straightforward results in terms of the waiting

times in ED may be due to greater pressure in the UH emergency departments, which are usually located in city centres.

Although the differences between GHs and UHs were always not significant, in 2012 GHs accounted for higher patient satisfaction. These results changed in 2013. However, previous research focused only on the patient experience with medical staff in the hospitals in Tuscany showed a higher patient satisfaction for patients discharged by UHs with respect to patients hospitalized in GHs [see among others, 36].

In addition, the test on variability for IRPES indicators showed homogenous patterns of performance regardless of the *teaching status*. In particular, the UHs showed a larger variation in the Average cost per weighted case, which measures efficiency by comparing the average costs of inpatient cases weighted for the DRG complexity. This suggests that as a group, UHs do not generally account for higher costs, as stated by other scholars [11-13]. As individuals, UHs show highly heterogeneous results. Hence, based on our analysis, the financial and economical sustainability of UHs could be related to the individual internal organization or other factors rather than to the *teaching status*.

Finally, for the tested IRPES indicators and considering both the years considered in the analysis, a "cluster effect" linked to the *teaching status* did not seem plausible.

This is also confirmed by the analysis on the NOEP indicators, which suggested that UHs did not generally achieve better outcomes. These results contribute to the research on this topic by suggesting that there is no straightforward evidence for better outcomes associated with UHs. Interestingly, GHs performed better (although not significantly) considering indicators related to the waiting time for femur-fracture surgery and to the recourse to Caesarean sections. In most of the mortality and re-admission indicators, UHs did perform better but without a significant effect. Considering that UHs are referral centres with higher delivered volumes and patients, it is possible that these better results could also be explained by their role in the hospital network, rather than only by the *teaching status*, as suggested in other studies [see among others, 37].

In addition, GHs account for a general higher variability compared to UHs, but without significant differences. This means that, although UHs seem to be generally more concentrated around average values, the extreme values of GH results towards the maximum and minimum of the distribution do not affect the overall analysis results.

In conclusion, a straightforward evidence identifying better performance and less variability for UHs does also not seem plausible for NOEP indicators.

Summarizing these results, from a multidimensional perspective being in the UH group rather than the GHs does not generally affect performance. Instead, the results could be linked to particular

features of an individual hospital as well as managerial approaches, rather than to a specific group affiliation.

As a preliminary study on this topic, this research presents some limitations in terms of:

- the study context focused on the Italian healthcare system and its organizational structure. We believe however that the contextual factors strongly influence the results and therefore that they cannot be excluded when the research is aimed at supporting decision-making processes. This study provides evidence to enlarge the debate on this relevant topic not only in Italy but also in those countries aiming at linking *teaching status* attribution to performance evaluation.

- there could be other indicators as valuable and informative as those measures included in the analysis. However, we included the ones that regional policy-makers and healthcare managers in Italy share as valuable measure to assess and guide the system.

Further studies will investigate the relevance of individual and regional factors in affecting UH and GH results in this multidimensional perspective.

6. CONCLUSIONS

The main finding of this study is that Italian UHs cannot straightforwardly be associated with better results in terms of appropriateness, efficiency, patient satisfaction, economic and financial evaluation, and outcomes. However, this preliminary evidence may inform the debate on the future role of UHs and encourage further considerations with regard to the Italian healthcare system.

Firstly, if UHs claim to maintain their role of leading players in the hospital network and to be the main actors in charge of training clinicians of the future, hospital performance evaluations should be further encouraged in order to inform the attribution of *teaching status* based on performance results. This could stimulate wider competition between Italian hospitals aimed at assigning *teaching status* to those hospitals that achieve the best performance in specific care paths. In this respect, medical schools should base their teaching activities for both undergraduate and resident students in the hospitals that can ensure the best results and practices, since the future generation of clinicians has a crucial role in improving the quality of care.

Secondly, considering the pressure towards more population-based oriented healthcare systems, the organizational structure of Italian UHs as an independent organization could be revised towards a more integrated network with other facilities delivering community, primary and outpatient care. UH facilities could therefore be directly integrated with the other LHA-led providers also creating a joint accountability for more patient-centred care.

ENDNOTES

** The IRPES in 2014 included Basilicata, Emilia-Romagna, Friuli Venezia Giulia, Liguria, Marche, Autonomous Province of Bolzano, Autonomous Province of Trento, Toscana, Umbria, Veneto. In 2015 Lombardia, Calabria, Lazio, Puglia and Sardegna joined the network.

ACKNOWLEDGEMENTS

The authors wish to thank all the IRPES regional representatives and their staff for their invaluable suggestions and collaboration and the MeS-Lab researchers for their help during data processing and results interpretation. We also wish to thank the participants to the Wennberg International Collaborative Spring Policy Meeting 2015 held in June 2015 for their comments on the draft version of this paper.

COMPETING INTERESTS

The authors have read and understood BMJ Open policy on declaration of interests and declare that we have no competing interests. All authors have completed the Unified Competing Interest form and declare that Nuti, Grillo-Ruggieri and Podetti have support from the Network of Italian Regions that adopted the IRPES for the submitted work; Nuti, Grillo-Ruggieri and Podetti have no relationships with the Network of Italian Regions that might have an interest in the submitted work in the previous 3 years; their spouses, partners, or children have no financial relationships that may be relevant to the submitted work; and Nuti, Grillo-Ruggieri and Podetti have non-financial interests that may be relevant to the submitted work.

CONTRIBUTORS

Sabina Nuti, the lead author, led the study design. Tommaso Grillo Ruggieri and Silvia Podetti carried out the data collection and the empirical analyses. All the authors were responsible for writing the manuscript and were involved in interpreting the findings and approving the final manuscript.

TRANSPARENCY DECLARATION

The lead author affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

FUNDING

This work was financed by the Network of Regions adopting the Inter-Regional Performance Evaluation System (IRPES), coordinated by the Management and Health Laboratory of Scuola Superiore Sant'Anna of Pisa, Italy (www.meslab.sssup.it).

ROLE OF THE STUDY FUNDER

The funder was not directly involved in the study but it shared with the authors the research aim. The authors were independent from funder in designing the research and interpreting study results.

ETHICAL APPROVAL

The paper did not involve human participants and did not need an ethical approval.

DATA SHARING STATEMENT

"Data sharing: full dataset available at [/doi] with open access. No informed consent was necessary because the data used in this study are publicly reported on the following websites: http://performance.sssup.it/netval (IRPES) and http://95.110.213.190/PNEed15/index.php (NOEP)"

REFERENCES

1. Smith T, Whitchurch C. The Future of the Tripartite Mission: Re-examining the Relationship Linking Universities, Medical Schools and Health Systems. *Higher Education Management and Policy* 2002;14(2):39-52 doi: 10.1787/hemp-v14-art12-en
2. Mintzberg H. Structure in Fives: Designing Effective Organizations. Upper Saddle River, NJ: Prentice Hall 1983.
3. Palm W, Glinos IA, Garel P, et al. Building European reference networks: exploring concepts and national practices in the European Union. Copenhagen, WHO Regional Office for Europe on behalf of the European Observatory on Health Systems and Policies 2013 http://www.euro.who.int/__data/assets/pdf_file/0004/184738/e96805-final.pdf [Accessed 26 January 2016]
4. Clark J, Tugwell P. Does Academic Medicine Matter?. *PLoS Med* 2006;3(9):e340. doi:10.1371/journal.pmed.0030340
5. International Working Party to Promote and Revitalise Academic Medicine. Academic medicine: the evidence base. *BMJ* 2004;329:789-92 doi: 10.1136/bmj.329.7469.789
6. Ayanian JZ, Weissman JS. Teaching Hospitals and Quality of Care: A Review of the Literature. *Milbank Q* 2002;80(3):569–93 doi: 10.1111/1468-0009.00023
7. Kupersmith J. Quality of Care in Teaching Hospitals: A Literature Review. *Acad Med* 2005;80(5):458-66
8. Papanikolaou PN, Christidi GD, Ioannidis JP. Patient outcomes with teaching versus nonteaching healthcare: a systematic review. *PLoS Med* 2006 Sep;3(9):e341 doi: 10.1371/journal.pmed.0030341
9. Hollingsworth B. Non-parametric and parametric applications measuring efficiency in health care. *Health Care Manag Sci* 2003;6(4):203-18 doi:10.1023/A:1026255523228
10. Grosskop S, Margatitis D, Valdmanis V. Comparing Teaching and Non-teaching Hospitals: A Frontier Approach (Teaching vs. Non-teaching Hospitals). *Health Care Manag Sci* 2001;4(2):83-90 doi: 10.1023/A:1011449425940
11. Sloan F, Feldman R, Steinwald B. Effects of teaching on hospital costs. *J Health Econ* 1983;2(1):1–23 doi:10.1016/0167-6296(83)90009-7
12. Iezzoni LI, Shwartz M, Moskowitz MA, et al. Illness Severity and Costs of Admissions at Teaching and Nonteaching Hospitals". *JAMA* 1990;264(11):1426-31 doi:10.1001/jama.1990.03450110072030
13. Mechanic R, Coleman K, Dobson A. Teaching hospital costs: implications for academic missions in a competitive market. *JAMA* 1998 Sep 16;280(11):1015-9 doi:10.1001/jama.280.11.1015

14. Carbone C, Lega F, Prenestini A. La governance delle Aziende Ospedaliero-Universitarie: perché non funziona, come potrebbe migliorare. In: Cantù E. L'aziendalizzazione della sanità in Italia - Rapporto OASI 2010. Milano, IT: EGEA 2010.
15. Roversi Monaco MG. Università e Servizio Sanitario Nazionale: L'azienda ospedaliera universitaria. Padova, IT: CEDAM 2011.
16. Leardini C, Sala GA, Campedelli B. L'azienda ospedaliera universitaria integrata. Un modello di integrazione tra attività di cura, ricerca e formazione. Milano, IT: Franco Angeli 2015.
17. Barsanti S, Giovanelli S, Nuti S. Come valutare la ricerca nelle Aziende Ospedaliero-Universitarie? Metodi e proposte operative. *Mecosan* 2014;90(1):9-32 doi: 10.3280/MESA2014-090002
18. 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 doi: 10.1007/s10997-012-9218-5
19. Nuti S. La valutazione della performance in sanità. Bologna, IT: il Mulino 2008.
20. Nuti S, Vola F, Bonini A, et al. Making governance work in the healthcare sector: evidence from a "natural experiment" in Italy. *Health Econ Policy Law* 2015;11(1):17-38 doi: 10.1017/S1744133115000067
21. Nuti S, Vola F, Amat P. Il sistema di valutazione della performance dei sistemi sanitari regionali: Basilicata, Calabria, Emilia-Romagna, Friuli Venezia Giulia, Lazio, Liguria, Lombardia, Marche, P.A. Bolzano, P.A. Trento, Puglia, Sardegna, Toscana, Umbria, Veneto - REPORT 2014. Pisa, IT: ETS 2015.
22. *Inter-Regional Performance Evaluation System* (IRPES) website: http://performance.sssup.it/netval [Accessed 25 January 2016]
23. *Italian National Outcome Evaluation Programme* (NOEP) website: http://95.110.213.190/PNEed15/index.php [Accessed 25 January 2016]
24. Ash AS, Fienberg SE, Louis TA, et al. Statistical Issue in Assessing Hospital Performance. Quantitative Health Sciences Publications and Presentations. Centers for Medicare and Medicaid Services 2012. http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Downloads/Statistical-Issues-in-Assessing-Hospital-Performance.pdf [Accessed 25 January 2016]
25. Iezzoni LI. Risk Adjustment for measuring healthcare outcomes, 2nd edition. Chicago, IL: Health Administration Press 1997.

26. Campbell SM, Braspenning J, Hutchinson A, et al. Research Methods used in developing and applying quality indicators in primary care. *Qual Saf Health Care* 2002;11:358-64 doi:10.1136/qhc.11.4.358

27. Ibraihm JE, Performance indicators from all perspectives. *Int J Qual Health Care* 2001 Dec;13(6):431-2. doi: dx.doi.org/10.1093/intqhc/13.6.431 [Published Online First: 1 December 2001].

28. Garcia-Lacalle J, Martin E. Rural vs urban hospital performance in a 'competitive' public health service. *Soc Sci Med* 2010 Sep;71(6):1131-40 doi:10.1016/j.socscimed.2010.05.043 [Published Online First: 19 June 2010].

29. Mann HB, Whitney DR. On a test of whether one of 2 random variables is stochastically larger than the other. *Annals of Mathematical Statistics* 1947;18:50-60 http://projecteuclid.org/euclid.aoms/1177730491 [Accessed 26 January 2016]

30. Lehmann EL. Nonparametrics: Statistical Methods Based on Ranks. *Journal of Applied Mathematics and Mechanics* 1977;57(9):562 doi: 10.1002/zamm.19770570922

31. Zimmerman DW. Comparative power of Student t test and Mann-Whitney U test for unequal sample sizes and variances". *J Exp Educ* 1987;55:71-174.

32. Siegel S, Castellan NJJ. Nonparametric statistics for the behavioral sciences. New York, NY: McGraw-Hill 1988.

33. Hollander M, Wolfe DA. Nonparametric Statistical Methods, Solutions Manual, 2nd Edition. Chichester, UK: Wiley and Sons Ltd 1999.

34. Levene H. Robust testes for equality of variances. In: Olkin I. Contributions to Probability and Statistics. Palo Alto, CL: Stanford University Press 1966:278-292.

35. Brown MB, Forsythe AB. Robust tests for the equality of variances. *J Am Stat Assoc* 1974;69:364–67 doi:10.2307/2285659 doi: 10.1080/01621459.1974.10482955

36. Murante AM, Seghieri C, Brown A, et al. How do hospitalization experience and institutional characteristics influence inpatient satisfaction? A multilevel approach. *Int J Health Plann Manage* 2014;29(3):e247-60 doi: 10.1002/hpm.2201 [Published Online First: 1 July 2014].

37. Dimick JB, Cowan JA, Colletti LM et al. Hospital Teaching Status and Outcomes of Complex Surgical Procedures in the United States. *Arch Surg* 2004;139(2):137-141 doi:10.1001/archsurg.139.2.137

SUPPLEMENTARY FILE

APPENDIX I – Complete list of hospitals included in the analysis and number of observations included for each indicators

IRPES Indicators	UHs	GHs	Total Hospitals
2011			
Average cost per weighted case	13	42	55
Average expenditure per Diagnostic Imaging weighted for tariff	13	34	47
Total Hospitals in 2011	**26**	**77**	**103**
2012			
Relative stay index: case-mix adjusted differential average Length of Stay - LOS days	16	73	89
Percentage of medical discharges with LOS > ministerial threshold for patients over 65	16	72	88
Percentage of ED green-coded patients visited within 1 hour	16	69	85
Percentage of ED patients referred for hospital admission with ED length of stay <= 8 hours	16	70	86
Percentage of medical inpatient discharges within 2 days (National Healthcare Agreement 2010)	16	73	89
Percentage of day-surgery treatment for specific procedures (National Healthcare Agreement 2010)	16	71	87
Percentage of patients leaving ED against/without medical advice	16	70	86
Percentage hospitalized patients leaving against medical advice	16	73	89
Average cost per weighted case	15	59	74
Average expenditure per Diagnostic Imaging weighted for tariff	15	57	72
Total Hospitals in 2012	**158**	**687**	**845**
2013			
Relative stay index: case-mix adjusted differential average Length of Stay - LOS days	16	73	89
Percentage of medical discharges with LOS > ministerial threshold for patients over 65	16	72	88
Percentage of ED green-coded patients visited within 1 hour	15	71	86
Percentage of ED patients referred for hospital admission with ED length of stay <= 8 hours	16	71	87
Percentage of medical inpatient discharges within 2 days (National Healthcare Agreement 2010)	16	73	89
Percentage of day-surgery treatment for specific procedures (National Healthcare Agreement 2010)	16	71	87
Percentage of patients leaving ED against/without medical advice	16	70	86
Percentage hospitalized patients leaving against medical advice	16	72	88
Total Hospitals in 2013	**64**	**286**	**350**

NOEP Indicators	UHs	GHs	Total Hospitals
2012			
AMI: 30-day mortality	18	111	129
AMI without PTCA: 30-day mortality	15	85	100
AMI with PTCA within 2 days: 30-day mortality	16	37	53
AMI with PTCA after 2 days: 30-day mortality	14	20	34
AMI: 1-year mortality	17	113	130
AMI: MACCE after 1 year	17	113	130
Isolated Aortocoronary Bypass: 30-day mortality	13	5	18
Congestive heart failure: 30-day mortality	19	153	172
Ischemic Stroke: 30-day mortality	17	112	129
COPD exacerbation: 30-day mortality	17	144	161
Proportion of Cesarean Section	16	127	143
Femur Fracture: 30-day mortality	17	119	136
Femur Fracture: Percentage of operations carried out within 2 days	17	121	138
Colon cancer surgery: 30-day mortality	18	95	113
COPD: 30-day re-admission	17	137	154
Valvuloplasty or heart valve replacement: 30-day mortality	13	5	18
Ischemic Stroke: 30-day re-admission	17	104	121
Total Hospitals for 2012	**278**	**1,601**	**1,879**
2013			
AMI: 30-day mortality	17	111	128
AMI without PTCA: 30-day mortality	15	77	92
AMI with PTCA within 2 days: 30-day mortality	15	41	56
AMI with PTCA after 2 days: 30-day mortality	14	18	32
AMI: 1-year mortality	17	109	126
AMI: MACCE after 1 year	17	109	126
Isolated Aortocoronary Bypass: 30-day mortality	13	4	17
Congestive heart failure: 30-day mortality	19	172	191
Ischemic Stroke: 30-day mortality	16	106	122
COPD exacerbation: 30-day mortality	17	134	151
Proportion of Cesarean Section	17	124	141
Femur Fracture: 30-day mortality	16	118	134
Femur Fracture: Percentage of operations carried out within 2 days	16	118	134
Colon cancer surgery: 30-day mortality	17	97	114
COPD: 30-day re-admission	17	133	150
Valvuloplasty or heart valve replacement: 30-day mortality	14	5	19
Ischemic Stroke: 30-day re-admission	16	101	117
Total Hospitals for 2013	**273**	**1,577**	**1,850**

SUPPLEMENTARY FILE

APPENDIX I – Complete list of hospitals included in the analysis and number of observations included for each indicators

IRPES Indicators	UHs	GHs	Total Hospitals
2011			
Average cost per weighted case	13	42	55
Average expenditure per Diagnostic Imaging weighted for tariff	13	34	47
Total Hospitals in 2011	**26**	**77**	**103**
2012			
Relative stay index: case-mix adjusted differential average Length of Stay - LOS days	16	73	89
Percentage of medical discharges with LOS > ministerial threshold for patients over 65	16	72	88
Percentage of ED green-coded patients visited within 1 hour	16	69	85
Percentage of ED patients referred for hospital admission with ED length of stay <= 8 hours	16	70	86
Percentage of medical inpatient discharges within 2 days (National Healthcare Agreement 2010)	16	73	89
Percentage of day-surgery treatment for specific procedures (National Healthcare Agreement 2010)	16	71	87
Percentage of patients leaving ED against/without medical advice	16	70	86
Percentage hospitalized patients leaving against medical advice	16	73	89
Average cost per weighted case	15	59	74
Average expenditure per Diagnostic Imaging weighted for tariff	15	57	72
Total Hospitals in 2012	**158**	**687**	**845**
2013			
Relative stay index: case-mix adjusted differential average Length of Stay - LOS days	16	73	89
Percentage of medical discharges with LOS > ministerial threshold for patients over 65	16	72	88
Percentage of ED green-coded patients visited within 1 hour	15	71	86
Percentage of ED patients referred for hospital admission with ED length of stay <= 8 hours	16	71	87
Percentage of medical inpatient discharges within 2 days (National Healthcare Agreement 2010)	16	73	89
Percentage of day-surgery treatment for specific procedures (National Healthcare Agreement 2010)	16	71	87
Percentage of patients leaving ED against/without medical advice	16	70	86
Percentage hospitalized patients leaving against medical advice	16	72	88
Total Hospitals in 2013	**64**	**286**	**350**

NOEP Indicators	UHs	GHs	Total Hospitals
2012			
AMI: 30-day mortality	18	111	129
AMI without PTCA: 30-day mortality	15	85	100
AMI with PTCA within 2 days: 30-day mortality	16	37	53
AMI with PTCA after 2 days: 30-day mortality	14	20	34
AMI: 1-year mortality	17	113	130
AMI: MACCE after 1 year	17	113	130
Isolated Aortocoronary Bypass: 30-day mortality	13	5	18
Congestive heart failure: 30-day mortality	19	153	172
Ischemic Stroke: 30-day mortality	17	112	129
COPD exacerbation: 30-day mortality	17	144	161
Proportion of Cesarean Section	16	127	143
Femur Fracture: 30-day mortality	17	119	136
Femur Fracture: Percentage of operations carried out within 2 days	17	121	138
Colon cancer surgery: 30-day mortality	18	95	113
COPD: 30-day re-admission	17	137	154
Valvuloplasty or heart valve replacement: 30-day mortality	13	5	18
Ischemic Stroke: 30-day re-admission	17	104	121
Total Hospitals for 2012	**278**	**1,601**	**1,879**
2013			
AMI: 30-day mortality	17	111	128
AMI without PTCA: 30-day mortality	15	77	92
AMI with PTCA within 2 days: 30-day mortality	15	41	56
AMI with PTCA after 2 days: 30-day mortality	14	18	32
AMI: 1-year mortality	17	109	126
AMI: MACCE after 1 year	17	109	126
Isolated Aortocoronary Bypass: 30-day mortality	13	4	17
Congestive heart failure: 30-day mortality	19	172	191
Ischemic Stroke: 30-day mortality	16	106	122
COPD exacerbation: 30-day mortality	17	134	151
Proportion of Cesarean Section	17	124	141
Femur Fracture: 30-day mortality	16	118	134
Femur Fracture: Percentage of operations carried out within 2 days	16	118	134
Colon cancer surgery: 30-day mortality	17	97	114
COPD: 30-day re-admission	17	133	150
Valvuloplasty or heart valve replacement: 30-day mortality	14	5	19
Ischemic Stroke: 30-day re-admission	16	101	117
Total Hospitals for 2013	**273**	**1,577**	**1,850**

BMJ Open

Do university hospitals perform better than general hospitals? A comparative analysis among Italian regions

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-011426.R1
Article Type:	Research
Date Submitted by the Author:	22-Mar-2016
Complete List of Authors:	Nuti, Sabina; Scuola Superiore Sant\'Anna, Institute of Management - Management&Health Laboratory Grillo Ruggieri, Tommaso; Scuola Superiore Sant\'Anna, Institute of Management - Management&Health Laboratory Podetti, Silvia; Scuola Superiore Sant\'Anna, Institute of Management - Management&Health Laboratory
Primary Subject Heading:	Health services research
Secondary Subject Heading:	Health policy
Keywords:	Performance, Evaluation, University, Hospital, Italy

Do university hospitals perform better than general hospitals? A comparative analysis among Italian regions

Corresponding author:
Sabina Nuti
Postal Address: Laboratorio MeS, Istituto di Management - Scuola Superiore Sant'Anna,
Via San Zeno 2 – 56127, Pisa, Italy
Email address: s.nuti@sssup.it
Telephone: +39 050 88 38 86
Fax: +39 050 88 38 90

Authors:
Sabina Nuti, Full Professor, Management and Health Laboratory, Institute of Management, Scuola Superiore Sant'Anna, Via San Zeno 2 – 56127, Pisa, Italy, s.nuti@sssup.it.

Tommaso Grillo-Ruggieri, PhD candidate, Management and Health Laboratory, Institute of Management, Scuola Superiore Sant'Anna, Via San Zeno 2 – 56127, Pisa, Italy, t.grilloruggieri@sssup.it

Silvia Podetti, PhD, Management and Health Laboratory, Institute of Management, Scuola Superiore Sant'Anna, Via San Zeno 2 – 56127, Pisa, Italy, s.podetti@sssup.it

Keywords: performance, evaluation, university, hospital, Italy
Word Count: 3,806

Abstract

Objective

Aim of this research is to investigate how university hospitals (UHs) perform compared to general hospitals (GHs) in the Italian healthcare system.

Design and setting

27 overall performance indicators were selected and analysed for UHs and GHs in ten Italian regions. The data refer to 2012 and 2013 and were selected from two performance evaluation systems based on administrative data on hospitalization discharge flows: the Inter-Regional Performance Evaluation System developed by the Management and Health Laboratory of the Scuola Superiore Sant'Anna of Pisa and the Italian National Outcome Evaluation Programme developed by the National Agency for Healthcare Services. The study was conducted in two stages and by combining two statistical techniques. In stage 1, a nonparametric Mann-Whitney *U* test was carried out to compare the performance of UHs and GHs on the selected set of indicators. In stage 2, a robust equal variance test between the two groups of hospitals was carried out to investigate differences in the amount of variability between them.

Results

The overall analysis showed heterogeneous results. In general, being in the UHs group rather than the GHs does not generally affect performance. Thus, Italian UHs cannot straightforwardly be associated with better results in terms of appropriateness, efficiency, patient satisfaction, and outcomes.

Conclusions

Policy-makers and managers should further encourage hospital performance evaluations in order to stimulate wider competition aimed at assigning *teaching status* to those hospitals that are able to assure requirements in terms of performance. In addition, UH facilities could be integrated with other providers in charge of community, primary and outpatient services creating a joint accountability for more patient-centred and integrated care.

Strenghts and limitations of the study

- This study provides evidence about differences in terms of performance between university hospitals and general hospitals that was lacking for the Italian context
- The analysis shows new results about hospital performance that can contribute to the debate on this topic
- A non parametric approach of analysis is applied first to Italian context
- The study is limited to the Italian healthcare system and its organizational structure

- There could be other indicators of performance as valuable and informative as those measures included in the analysis

1. INTRODUCTION

University hospitals (UHs) can be considered as complex organizations as their mission includes three different objectives: patient care, education, and research [1]. UHs combine all the features of Minzberg's Professional Bureaucracy [2] embedded within both the healthcare organizations and the university context. In addition, UHs are usually referral centres for most complex care within a hub-and-spoke hospital network [3].

Given the three-fold mission of these institutions and the specific role that they play in the healthcare system, should UHs be considered as a 'cluster' with specific performance patterns?

This study investigates whether UHs behave homogenously regarding performance results with substantial differences with respect to general hospitals (GHs).

Evidence on this topic could provide important information for policy-makers and managers in defining specific policies and actions in order to improve the quality of care within the hospital regional network, where UHs play a specific and strategic role and in order to pursue their specific mission. In particular, in Italy as in other countries, UHs are in charge of the strategic role of training doctors of the future. Therefore, since health professionals are the most important assets for the healthcare organizations, policy-makers should ensure clinicians are trained and supported in their practice by institutions that can assure the appropriate requirements in terms of quality of care and research productivity. The analysis was carried out in Italy.

2. BACKGROUND

Teaching status has been already investigated from several perspectives by studying whether it affects the results of UHs compared with other hospitals in terms of outcomes, quality of care, productivity, costs, etc.

Firstly, reviews on outcomes, quality of care and adverse-event prevention reached mixed conclusions and highlighted the need for evidence on differences between UHs and GHs [4-5]. Some reviews underlined better overall results for UHs [6-7]; whereas, a systematic review highlighted no differences between UH and GH outcomes [8].

Secondly, studies on productivity and efficiency have usually applied Data Envelopment Analysis (DEA) and frequently highlighted better performance of GHs with respect to UHs [among others, 9-10].

Indeed, training resident students, carrying out research activities besides patient care and the role of referral centres for complex care have often been identified as elements that can increase costs [11-13]. This frequently drives additional financial resources to UHs (e.g., an increased mark-up in the reimbursement system for UH discharges) [6].

Research on this topic presents several differences in terms of data sources, measurement processes and methodology for data analysis [4]. This could raise potential issues regarding external validity and result generalizability [6-9]. Examples of these differences are:

- the data sources: e.g. medical records or administrative data;
- the definition of UHs and their ownership (public, private, for-profit, non-profit): for example, some studies consider only major UHs, whilst others include all the hospitals with a residency program;
- the indicators included in the analysis (usually, outcomes, quality of care or efficiency) and the different calculation criteria and risk-adjustment procedure used for the same measures (mortality rates, process measures, etc.);
- the statistical methods used to compare hospitals (parametric and non-parametric approaches and tests such as DEA, ANOVA, Kruskall-Wallis, Mann-Whitney, etc).

These differences may partially explain why research looking at different performance or outcomes in UHs or controlling for a potential effect of the *teaching status* have not led to straightforward results.

Finally, results may be also associated with the specific geographical context. For instance, in one of the most recent systematic reviews on this topic, more than three-fourths of the studies included in the analysis were conducted in the United States [8]. However, each specific geographical and health system context may play an important role in explaining results.

With reference to Italy, detailed studies are also lacking on this topic. Scholars have focused on governance issues or research evaluations [see for instance, 14-17]. There have been no systematic comparisons of performances between the two groups of hospitals and related research.

2.1. The Italian context

The national healthcare system in Italy follows a Beveridge model by providing universal coverage through general taxation. Regional governments are responsible for organizing and delivering health services and being accountable for performance. National government monitors the pursuit of the universal coverage in particular with respect to a package of essential services (Nationally defined basic health benefit package - Livelli Essenziali di Assistenza). National government allocates financial resources to the regional governments on an adjusted capitation basis. Regions then

reallocate resources to Local Health Authorities (LHAs), through a regionally-adjusted capitation formula.

In Italy, hospital care is delivered by public general hospitals (GHs) directly managed by the LHAs, private or public autonomous hospitals (AHs), private or public university hospitals (UHs) and research hospitals (RHs). AHs, UHs and RHs are autonomous organizations with respect to LHAs managing the healthcare delivery in their own geographical area.

UHs can be classified considering ownership and different institutional and organizational settings [18]. In Italy, the *teaching status* can be attributed to hospitals owned by private university medical schools, hospitals owned by public university medical schools and hospitals jointly owned by both public university medical schools and the Regional Administration. In this last case, the CEO is jointly appointed by the two institutions. Following the national laws (D.Lgs 502/92 and D.Lgs 517/99), these hospitals are identified as teaching facilities by the Ministry of Health, the Ministry of Education and the Regional Administrations. Regardless of the ownership and the organizational settings, health professionals employed by university, besides teaching and research activity, provide also patients care and receive an integrative 30% remuneration. These costs are directly sustained not by the universities but by the hospital administration.

Considering patient care activity, since UHs are autonomous authorities, they are not financed by LHA resources through capitation-based funding.

UHs can therefore be financed through a pay for service system based on DRG tariffs or through a budget-cost control system, depending on the regional healthcare organization.

As occurs in other countries [among others, 19], Regions may also assign additional resources to UHs to cover the added costs for education and research. Depending on the regional financial strategy, these resources are usually allocated through an increased percentage reimbursement of the DRG tariffs or through specific restricted funds. Therefore, the percentage or fixed amount of additional resources varies depending on the regional decisions [14]. Moreover, at the national level, inpatient services for residents of other regions are reimbursed considering DRGs tariffs increased of 7%.

Italian UHs have on average a much higher number of hospital beds with respect to GHs and are hub-referral centres for highly-complex and highly-specialized care, such as neuro-surgery, cardio-surgery, radiotherapy, most critical intensive care, paediatric highly-complex surgery, etc.

Evidence from Italy on the comparison of UH performance with respect to GHs may provide valuable information for both healthcare policy-makers and managers, at both regional and national level and not only in Italy. Indeed, if UHs behave as a specific ‘cluster’, new policies and focused-

actions could be defined to support the specific role of these authorities within the hospital network in the regional and national contexts. Evidence of similar patterns of performance between these two groups of hospitals may highlight the need to look for other sources of variation between the two groups. As a consequence, other features from the *teaching and research status* may be relevant to inform policies on hospital governance, financing and network organization, considering the crucial role of UHs in training the future clinicians for the healthcare system.

The aim of this paper is thus to investigate how UHs perform in comparison to GHs.

3. METHODS

3.1. Data sources and hospital selection

The data used in this analysis were selected from two performance evaluation systems based on the same source of administrative data on hospitalization discharge flows:

- The *Inter-Regional Performance Evaluation System* (IRPES) developed by the Management and Health Laboratory of the Scuola Superiore Sant'Anna of Pisa (MeS-Lab) - where the authors of this paper are researchers. This system provides a multi-dimensional evaluation of performance including efficiency, appropriateness, integration and quality of care. This system was firstly implemented by the regional government in Tuscany [20-21] and was then adopted – on a voluntary basis - by the majority of other Italian regions** [22-23]. The evaluation process measures through benchmarking and with specific risk adjustment processes the results achieved every year by all the Health Authorities (the local health authorities, the university hospitals, the research hospitals and the autonomous hospitals) located in these regions. Results are publicly reported [24].

- The *Italian National Outcome Evaluation Programme* (NOEP) developed by the National Agency for Healthcare Services on behalf of the Ministry of Health. This system measures outcomes nation-wide [25], i.e. for each Italian hospitals. On the basis of rigorous risk adjustment processes [among others, 26-27], these measures represent assessment tools to support clinical and organizational audit programs aimed at improving both outcome and equity in the National Health System.

Data refer to the hospital activity of 2012 and 2013, apart from two economic indicators related to balance sheets, which are available only for 2011 and 2012.

Two groups of hospitals differing in particular because of the *teaching status*, the organizational autonomy with respect to the LHAs and, due to the hospital network organization, to the average number of hospital discharges (in 2012, 32,472 for UHs and 17,606 for GHs) and the average DRG weight (in 2012, 1.3 for UHs and 1.06 for GHs) were considered in the analysis. The whole study included all the 16 UHs and 73 LHAs of the ten IRPES regions.

3.2. Performance indicators

For the purposes of this study, 27 performance indicators were selected, 10 from IRPES (Table 1) and 17 from NOEP (Table 2).

Table 1 – IRPES Indicators

IRPES INDICATORS	Rationale
EFFICIENCY AND APPROPRIATENESS	
Relative stay index: case-mix adjusted differential average Length of Stay - LOS days	Measure of the average difference from the standard LOS for admitted patients with adjustments for case-mix
Percentage of medical discharges with LOS > ministerial threshold for patients over 65	Measure of the hospital compliance with the Italian Ministry of Health standards for the LOS for medical inpatient activity for elderly patients. This measure is a proxy of the effective implementation of integrated pathways between home, community-based and hospital care for elderly patients
Percentage of ED green-coded patients visited within 1 hour	Measure of timely emergency care for ED patients whose treatment may be delayed without risk
Percentage of ED patients referred for hospital admission with ED length of stay <= 8 hours	Measure of overall timely emergency care
Percentage of medical inpatient discharges within 2 days (National Healthcare Agreement 2010)	Measure of hospital compliance in avoiding short ordinary hospitalizations for patients that could be treated in outpatient clinics or in other care settings, as requested by the Italian Ministry of Health standards in the National Healthcare Agreement of 2010
Percentage of day-surgery treatment for specific procedures (National Healthcare Agreement 2010)	Measure of hospital compliance with Italian Ministry of Health standards for delivering specific not-complex surgical procedures in day-surgery or in outpatient clinics rather than through ordinary hospitalizations
PATIENT SATISFACTION	
Percentage of patients leaving ED against/without medical advice	Proxy of Patient Satisfaction on ED services and waiting times
Percentage hospitalized patients leaving against medical advice	Proxy of Patient Satisfaction for the inpatient activity
ECONOMIC AND FINANCIAL EVALUATION	
Average cost per weighted case	Measure of the ratio of a hospital acute inpatient care expenses to the number of acute inpatient cases weighted for the DRG complexity. The weighting enhances comparability across hospitals. The measure includes the percentage cost of hospital university staff financed by the Regional Administration for their patient care activity. This allows to account for the overall hospital staff costs.
Average expenditure per Diagnostic Imaging weighted for tariff	Measure of efficiency that compares costs and the value of the delivered diagnostic activity (sum of ambulatory tariffs)

Table 2 – NOEP Indicators

NOEP INDICATORS
OUTCOME: measures of 30-day mortality or re-admissions for relevant inpatient activity
AMI: 30-day mortality
AMI without PTCA: 30-day mortality
AMI with PTCA within 2 days: 30-day mortality

AMI with PTCA after 2 days: 30-day mortality
AMI: 1-year mortality
AMI: MACCE after 1 year
Isolated Aortocoronary Bypass: 30-day mortality
Valvuloplasty or heart valve replacement: 30-day mortality
Congestive heart failure: 30-day mortality
Ischemic Stroke: 30-day mortality
Ischemic Stroke: 30-day re-admission
COPD exacerbation: 30-day mortality
COPD: 30-day re-admission
Proportion of Caesarean Section
Femur Fracture: 30-day mortality
Femur Fracture: Percentage of operations carried out within 2 days
Colon cancer surgery: 30-day mortality

Eight IRPES indicators regard efficiency and appropriateness, patient satisfaction, and economic and financial dimensions. Two indicators regard economic and financial evaluation. This selection was shared by the group of the IRPES regional representatives. This group is in charge of systematically reviewing and discussing the measures included in the IRPES as relevant proxies for measuring performance in a multidimensional perspective in all the different settings of care [22].

For both sources of the selected indicators, the time coverage and the number of providers needed to perform the statistical test were guaranteed, thus ensuring the consistency of the comparative analysis between the two groups of hospitals in this single-country study [among others, 28-29].

The number of observations for the NOEP indicators may differ because not all the hospitals included in the analysis provide all the healthcare services linked to the included measures. However, the selection of these measures took into account the services usually provided by both LHA-GHs and UHs.

The analysis for the IRPES indicators compared the 16 UHs to the 73 LHAs. On the other hand, the analysis for the NOEP indicators was carried out at the hospital level, thus comparing the (at most) 19 facilities of the 16 UHs to the individual (at most) 191 GHs led by the 73 LHAs. (See Appendix I in the Supplementary File for the complete list of hospitals considered and the number of observations included for each indicators).

3.3. Statistical Methods

The study was conducted in two stages and by combining two statistical techniques. Data were processed using Stata software, version 12. In stage 1, a nonparametric Mann-Whitney *U* test was carried out to compare the performance of UHs and GHs on the selected set of indicators. This analysis determines whether or not UHs and GHs were drawn from the same target population.

Previous studies have already applied this univariate analysis to illustrate differences between hospitals [among others, 30] because of its appropriateness with small samples [31-35]. For the purposes of this study, this test verified whether or not there were differences between UH and GH performance, or, in other words, whether or not UHs and GHs could be considered as two different clusters. In stage 2, we carried out a robust equal variance test between the two groups of hospitals [36] to investigate differences in the amount of variability between UHs and GHs. This test is usually used to verify the assumption of homogeneity of variance across groups, meaning that the internal variability of one group of hospitals is not significantly different with respect to the other one.

To be in line with the assumptions of the Mann-Whitney *U* test, we used an extension of Levene's test as suggested by Brown and Forsythe [37]. We applied the test only for those indicators in which the Mann-Whitney *U* test did not show significant differences between UH and GH performances. Indeed, in those cases where the performance between the two groups did not show significant differences, we tested whether there were specific patterns in terms of variability.

4. RESULTS

The Mann-Whitney *U* test on IRPES indicators showed that in relation to four measures of "Efficiency and Appropriateness" and "Economic and financial evaluation" dimensions, there were differences in performance between UHs and GHs. The test, in fact, was significant both in 2012 and 2013 for the "% of ED green-coded patients visited within one hour", the "% of medical inpatient discharges within two days" and the "% of day-surgery treatment for specific procedures". The test was significant also in both 2011 and 2012 for the "Average expenditure for Diagnostic Imaging weighted for tariff". For these indicators, GHs seemed to perform better than UHs.

On the other hand, with reference to the indicators "Relative stay index", "% of medical discharges with LOS > ministerial threshold for patient over 65", and "% of ED patient referred for hospital admission with ED length of stay <= 8 hours", the Mann-Whitney *U* test was rejected for both 2012 and 2013.

Moreover, no significant differences were found for patient satisfaction proxies "% of patients leaving ED against/without medical advice" and of "% of hospitalized patients leaving against medical advice". Moreover, in 2013 UHs accounted for fewer patients that left ED or who were discharged against medical advice, whilst in 2012 the GHs achieved better results. The test was also not significant for the "Average cost per weighted case" and this occurred also after deleting outliers.

Table 3 summarizes the results of the test and illustrates the average and the median values of the two groups of hospitals for each of the indicators.

Table 3 – Mann-Whitney *U* test for IRPES indicators

MANN – WHITNEY *U* TEST IRPES INDICATORS	2012					2013				
	Median UH	**Median GH**	**Mean UH**	**Mean GH**	**Best Perf. MEDIAN**	**Median UH**	**Median GH**	**Mean UH**	**Mean GH**	**Best Perf. MEDIAN**
EFFICIENCY AND APPROPRIATENESS										
Relative stay index (case-mix adjusted differential average LOS days)	-0.3	-0.1	-0.1	-0.2	UH	-0.2	-0.3	-0.2	-0.3	GH
% medical discharges with LOS > ministerial threshold for patients over 65	4.7	3.6	4.6	4	GH	3.7	3.5	4.2	3.8	GH
% ED green-coded patients visited within 1 hour	**72.7**	**79.2**	**70.7**	**77.3**	**GH***	**68**	**77.2**	**64.9**	**76.2**	**GH***
% ED patients referred for hospital admission with ED length of stay <= 8 hours	98.1	97.8	93.9	94.8	UH	97.5	97.5	92.8	94.5	UH/GH
% medical inpatient discharges within 2 days (National Healthcare Agreement 2010)	**21.3**	**14.6**	**22.1**	**14.9**	**GH***	**21.2**	**14.1**	**21.8**	**14.4**	**GH***
% day-surgery treatment for specific procedures (National Healthcare Agreement 2010)	**46.5**	**58.7**	**49.1**	**58.9**	**GH***	**49.2**	**59.1**	**50.2**	**59**	**GH***
PATIENT SATISFACTION										
% patients leaving ED against/without medical advice	3	3.2	3.4	3.1	GH	3.3	3.2	3.4	3.4	UH
% hospitalized patients leaving against medical advice	0.9	0.8	1	1	GH	0.7	0.8	0.9	0.9	UH
ECONOMIC AND FINANCIAL EVALUATION										
	2011					**2012**				
Average cost per weighted case	4,303	4,220	4,678	4,348	GH	4,469	4,516	4,659	4,651	UH
Average expenditure per Diagnostic Imaging weighted for tariff	**1.3**	**0.9**	**1.8**	**1.1**	**GH***	**1.4**	**1**	**1.6**	**1.1**	**GH***

Regarding the test for the NOEP indicators, for all the tested measures, the Mann-Whitney *U* test was not significant except for two measures that showed mixed results in 2012 and in 2013 (Table 4).

Table 4 - Mann-Whitney *U* test for NOEP indicators

MANN-WHITNEY *U* TEST – NOEP INDICATORS	2012					2013				
	Median UH	Median GH	Mean UH	Mean GH	Best perf. MEDIAN	Median UH	Median GH	Mean UH	Mean GH	Best perf. MEDIAN
OUTCOME INDICATORS										
AMI: 30-day mortality	9.6	8.8	10.0	9.3	GH	9.4	7.7	9.1	8.1	GH
AMI without PTCA: 30-day mortality	17.4	15.5	17.8	16.5	GH	17.2	15.0	18.1	15.5	GH
AMI with PTCA within 2 days: 30-day mortality	4.7	4.1	4.5	4.2	GH	4.2	3.7	4.4	3.8	GH
AMI with PTCA after 2 days: 30-day mortality	2.4	2.4	3.1	2.6	GH	2.6	2.5	2.9	2.8	GH
AMI: 1-year mortality	10.2	11.1	10.4	11.5	UH	9.7	10.6	10.1	10.9	UH
AMI: MACCE after 1 year	23.9	24.8	24.4	25.2	UH	22.3	23.2	22.8	23.6	UH
Isolated Aortocoronary Bypass: 30-day mortality	1.8	1.9	2.2	2.0	UH	1.8	2.3	2.3	2.1	UH
Valvuloplasty or heart valve replacement: 30-day mortality	2.6	3.7	2.9	3.5	UH	2.3	3.0	2.6	3.2	UH
Congestive heart failure: 30-day mortality	8.7	9.8	9.3	10.8	UH	**8.9**	**10.7**	**8.7**	**11.0**	**UH***
Ischemic Stroke: 30-day mortality	9.3	10.1	8.6	10.5	UH	9.2	9.5	9.2	10.5	UH
Ischemic Stroke: 30-day re-admission	10.7	9.4	10.4	10.3	GH	6.5	6.7	7.1	7.2	**UH**
COPD exacerbation: 30-day mortality	7.3	8.7	7.6	8.9	UH	7.9	8.1	7.8	8.7	UH
COPD: 30-day re-admission	14.3	15.6	15.0	15.4	UH	14.4	15.4	14.2	15.4	UH
Proportion of Caesarean Section	19.8	17.8	23.1	18.8	GH	20.2	18.5	22.9	19.3	GH
Femur Fracture: 30-day mortality	4.3	4.8	4.8	5.1	UH	4.4	4.7	4.7	4.8	UH
Femur Fracture: Percentage of operations carried out within 2 days	**48.6**	**54.4**	**42.4**	**53.2**	**GH***	51.6	60.3	55.1	59.8	GH
Colon cancer surgery: 30-day mortality	3.4	3.9	4.2	4.3	UH	3.0	4.2	3.7	4.6	UH

For the "Congestive heart failure: 30 day mortality" the test showed no statistical differences between UHs and GHs in 2012. However, a significantly better performance for UHs was found in 2013. Similarly, in the case of the indicator "Femur fracture: % of operations carried out within two days", the Mann-Whitney *U* test showed significant differences between UHs and GHs in 2012, but not for 2013, with GHs having the best median performance.

In order to investigate different variations between the two groups of hospitals, the robust equal variance test [37] was carried out for a set of 23 indicators (6 IRPES indicators and 17 NOEP indicators) that rejected the Mann-Whitney *U* test.

Regarding IRPES indicators, the test was always not significant for both years included in the analysis (Table 5). UHs and GHs showed a higher standard deviation depending on the measures considered.

Table 5 – Robust Equal Variance test for IRPES Indicators

ROBUST EQUAL VARIANCE TEST - IRPES INDICATORS	2012				2013				Higher Variability in 2012	Higher Variability in 2013
	Std. Dev. UH	Std. Dev. GH	W50 - Median	Pr > F	Std. Dev. UH	Std. Dev. GH	W50 - Median	Pr > F		
EFFICIENCY AND APPROPRIATENESS										
Relative stay index (case-mix adjusted differential average LOS days)	0.9	1.4	0.3	0.6	0.9	1.2	0.6	0.4	GH	GH
% medical discharges with LOS > ministerial threshold for patients over 65	1.7	2	0.2	0.7	1.7	2.1	1	0.3	GH	GH
% ED patients referred for hospital admission with ED length of stay <= 8 hours	8.7	6.7	0.4	0.5	9.5	7.7	0.6	0.5	UH	UH
PATIENT SATISFACTION										
% patients leaving ED against/without medical advice	2	1.8	0	1	2.1	2.1	0.1	0.8	UH	UH/GH
% hospitalized patients leaving against medical advice	0.6	0.7	0.2	0.7	0.6	0.6	0	0.9	GH	UH/GH
ECONOMIC AND FINANCIAL EVALUATION										
	2011				2012				Higher Variability in 2011	Higher Variability in 2012
Average cost per weighted case	1,089	775	1.2	0.3	985	850	0.5	0.5	UH	UH

For the 2012 results of NOEP indicators, the test was significant for five measures (Table 6):

- "AMI: 1-year mortality" (p-value=0.02)
- "Ischemic stroke: 30-day mortality" (p-value=0.02)
- "COPD exacerbation: 30-day mortality" (p-value=0.04)
- "Femur fracture: 30-day mortality" (p-value=0.04)
- "COPD: 30-day readmission" (p-value=0.01)

Table 6 – Robust Equal Variance test for NOEP indicators

ROBUST EQUAL VARIANCE TEST - NOEP INDICATORS	2012				2013				Higher Variability in 2012	Higher Variability in 2013
	Std. Dev. UH	Std. Dev. GH	W50 - Median	Pr > F	Std. Dev. UH	Std. Dev. GH	W50 - Median	Pr > F		
OUTCOME INDICATORS										
AMI: 30-day mortality	3.2	3.8	1.2	0.3	2.7	3.7	2.6	0.1	GH	GH
AMI without PTCA: 30-day mortality	4.7	6.2	1.4	0.2	4.7	6.6	1.5	0.2	GH	GH
AMI with PTCA within 2 days: 30-day mortality	1.4	1.9	1.2	0.3	1.8	2.1	0.9	0.4	GH	GH
AMI with PTCA after 2 days: 30-day mortality	1.6	1.4	0.3	0.6	1.1	1.4	1.0	0.3	UH	GH

AMI: 1-year mortality	**2.1**	**4.4**	**5.2**	**0.02***	3.2	3.8	0.2	0.6	**GH***	GH
AMI: MACCE after 1 year	4.0	5.3	2.5	0.1	**3.3**	**5.6**	**4.3**	**0.04***	GH	**GH***
Isolated Aortocoronary Bypass: 30-day mortality	1.4	1.6	0.0	0.9	1.6	1.4	0.0	1.0	GH	UH
Valvuloplasty or heart valve replacement: 30-day mortality	1.3	0.5	2.7	0.1	1.3	1.0	0.3	0.6	UH	UH
Congestive heart failure: 30-day mortality	3.2	5.0	2.1	0.2					GH	
Ischemic Stroke: 30-day mortality	**2.9**	**4.5**	**5.7**	**0.02***	3.9	4.6	0.6	0.4	**GH***	GH
Ischemic Stroke: 30-day re-admission	3.5	3.9	0.0	0.9	2.2	3.0	1.9	0.2	GH	GH
COPD exacerbation: 30-day mortality	**2.3**	**3.9**	**4.2**	**0.04***	2.9	4.1	1.4	0.2	**GH***	GH
COPD: 30-day re-admission	**2.3**	**4.5**	**6.7**	**0.01***	3.3	4.2	1.8	0.2	**GH***	GH
Proportion of Caesarean Section	9.0	7.1	1.0	0.3	9.4	7.1	1.9	0.2	UH	UH
Femur Fracture: 30-day mortality	**1.4**	**2.2**	**4.1**	**0.04***	2.0	2.2	0.9	0.4	**GH***	GH
Femur Fracture: Percentage of operations carried out within 2 days					16.5	17.5	0.4	0.5		UH
Colon cancer surgery: 30-day mortality	2.6	2.3	0.1	0.8	1.7	2.5	3.2	0.1	UH	GH

In 2013, the test was significant only for the indicator "AMI: MACCE after 1 year" (p-value=0.04). For these five measures, GHs frequently showed a higher Standard Deviation with respect to UHs. This was also the case for most of the other outcome measures included for both 2012 and 2013, apart from the "Proportion of Caesarean Section" and the "30-day mortality rate for Valvuloplasty or heart replacement".

5. DISCUSSION

The overall analysis showed heterogeneous results when comparing the two groups of hospitals. Considering the IRPES indicators of appropriateness, we found a higher compliance of GHs in pursuing the Italian Ministry of Health standards on directing patients to the appropriate care settings for surgical treatments as well as in avoiding short medical hospitalizations and giving preference to outpatient clinics or day-hospital cases. This may be due to the lower complexity of general LHA-led hospitals and to a related lower complex management.

Regarding efficiency, in 2013 GHs seemed to perform better than the UHs but these results are slightly different in 2012, thus leading to ambiguous conclusions. Therefore, the three-fold mission and the greater organizational complexity of UHs seemed to lead to lower but not significantly different efficiency with respect to GHs. The more straightforward results in terms of the waiting times in ED may be due to greater pressure in the UH emergency departments, which are usually located in city centres.

Although the differences between GHs and UHs were always not significant, in 2012 GHs accounted for higher patient satisfaction. These results changed in 2013. However, previous research focused only on the patient experience with medical staff in the hospitals in Tuscany showed a higher patient satisfaction for patients discharged by UHs with respect to patients hospitalized in GHs [see among others, 38].

In addition, the test on variability for IRPES indicators showed homogenous patterns of performance regardless of the *teaching status*. In particular, the UHs showed a larger variation in the Average cost per weighted case, which measures efficiency by comparing the average costs of inpatient cases weighted for the DRG complexity. This suggests that as a group, UHs do not generally account for higher costs, as stated by other scholars [11-13]. As individuals, UHs show highly heterogeneous results. Hence, based on our analysis, the financial and economical sustainability of UHs could be related to the individual internal organization or other factors rather than to the *teaching status*.

Finally, for the tested IRPES indicators and considering both the years considered in the analysis, a "cluster effect" linked to the *teaching status* did not seem plausible.

This is also confirmed by the analysis on the NOEP indicators, which suggested that UHs did not generally achieve better outcomes. These results contribute to the research on this topic by suggesting that there is no straightforward evidence for better outcomes associated with UHs. Interestingly, GHs performed better (although not significantly) considering indicators related to the waiting time for femur-fracture surgery and to the recourse to Caesarean sections. In most of the mortality and re-admission indicators, UHs did perform better but without a significant effect. Considering that UHs are referral centres with higher delivered volumes and patients, it is possible that these better results could also be explained by their role in the hospital network, rather than only by the *teaching status*, as suggested in other studies [see among others, 39].

In addition, GHs account for a general higher variability compared to UHs, but without significant differences. This means that, although UHs seem to be generally more concentrated around average values, the extreme values of GH results towards the maximum and minimum of the distribution do not affect the overall analysis results. In conclusion, a straightforward evidence identifying better performance and less variability for UHs does also not seem plausible for NOEP indicators.

Summarizing these results, from a multidimensional perspective being in the UH group rather than the GHs does not generally affect performance. Hence, the different institutional and organizational settings between them seem not to result in significant dissimilarities. Instead, the variations in hospital performance could be linked to particular features of each individual hospital or its managerial approach. Furthermore, these variations may also be determined by the Regional Healthcare System, rather than to a specific cross-regional group affiliation.

In Italy, there is evidence that hospital performance improvement may be affected by regional strategies combining different tools [22]. This is the case of Tuscany and Basilicata regions, which applied a combination of different integrated governance tools and registered a higher performance improvement in the last years with respect to other regions.

With reference to Tuscany, these strategies and in particular the use of the IRPES for 10 years may have also reduced the gap between hospital groups. Indeed, the three Tuscany UHs account for 30% of overall regional hospitalizations and therefore the overall hospital performance improvement has been homogenously spread regardless of group affiliation [23-25; 40].

As a preliminary study on this topic, this research presents some limitations in terms of:

- the study context focused on the Italian healthcare system and its organizational structure. We believe however that the contextual factors strongly influence the results and therefore that they cannot be excluded when the research is aimed at supporting decision-making processes. This study provides evidence to enlarge the debate on this relevant topic not only in Italy but also in those countries aiming at linking *teaching status* attribution to performance evaluation.
- there could be other indicators as valuable and informative as those measures included in the analysis. However, we included the ones that regional policy-makers and healthcare managers in Italy share as valuable measure to assess and guide the system.

Further studies will investigate the relevance of individual and regional factors in affecting UH and GH results in this multidimensional perspective.

6. CONCLUSIONS

The main finding of this study is that Italian UHs cannot straightforwardly be associated with better results in terms of appropriateness, efficiency, patient satisfaction, economic and financial evaluation, and outcomes. However, this preliminary evidence may inform the debate on the future role of UHs and encourage further considerations with regard to the Italian healthcare system.

Firstly, if UHs claim to maintain their role of leading players in the hospital network and to be the main actors in charge of training clinicians of the future, hospital performance evaluations should be further encouraged in order to inform the attribution of *teaching status* based on performance results. This could stimulate wider competition between Italian hospitals aimed at assigning *teaching status* to those hospitals that achieve the best performance in specific care paths. In this respect, medical schools should base their teaching activities for both undergraduate and resident students in the hospitals that can ensure the best results and practices, since the future generation of clinicians has a crucial role in improving the quality of care.

Secondly, considering the pressure towards more population-based oriented healthcare systems, the organizational structure of Italian UHs as an independent organization could be revised towards a more integrated network with other facilities delivering community, primary and outpatient care. UH facilities could therefore be directly integrated with the other LHA-led providers also creating a joint

accountability for more patient-centred care. In this perspective, in Italy recent national legislation (Disegno di Legge n. 2111-B/2016) has allowed as a pilot experience some Special-Administrative Regions (such as Friuli Venezia Giulia) to incorporate within the LHAs the UHs.

ENDNOTES

** The IRPES in 2014 included Basilicata, Emilia-Romagna, Friuli Venezia Giulia, Liguria, Marche, Autonomous Province of Bolzano, Autonomous Province of Trento, Toscana, Umbria, Veneto. In 2015 Lombardia, Calabria, Lazio, Puglia and Sardegna joined the network.

ACKNOWLEDGEMENTS

The authors wish to thank all the IRPES regional representatives and their staff for their invaluable suggestions and collaboration and the MeS-Lab researchers for their help during data processing and results interpretation. We also wish to thank the participants to the Wennberg International Collaborative Spring Policy Meeting 2015 held in June 2015 for their comments on the draft version of this paper.

COMPETING INTERESTS

The authors have read and understood BMJ Open policy on declaration of interests and declare that we have no competing interests. All authors have completed the Unified Competing Interest form and declare that Nuti, Grillo-Ruggieri and Podetti have support from the Network of Italian regions that adopted the IRPES for the submitted work; Nuti, Grillo-Ruggieri and Podetti have no relationships with the Network of Italian regions that might have an interest in the submitted work in the previous 3 years; their spouses, partners, or children have no financial relationships that may be relevant to the submitted work; and Nuti, Grillo-Ruggieri and Podetti have non-financial interests that may be relevant to the submitted work.

CONTRIBUTORS

Sabina Nuti, the lead author, led the study design. Tommaso Grillo Ruggieri and Silvia Podetti carried out the data collection and the empirical analyses. All the authors were responsible for writing the manuscript and were involved in interpreting the findings and approving the final manuscript.

TRANSPARENCY DECLARATION

The lead author affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

FUNDING

This work was financed by the Network of regions adopting the Inter-Regional Performance Evaluation System (IRPES), coordinated by the Management and Health Laboratory of Scuola Superiore Sant'Anna of Pisa, Italy (www.meslab.sssup.it).

ROLE OF THE STUDY FUNDER

The funder was not directly involved in the study but it shared with the authors the research aim. The authors were independent from funder in designing the research and interpreting study results.

ETHICAL APPROVAL

The paper did not involve human participants and did not need an ethical approval.

DATA SHARING STATEMENT

"Data sharing: full dataset available at [/doi] with open access. No informed consent was necessary because the data used in this study are publicly reported on the following websites: http://performance.sssup.it/netval (IRPES) and http://95.110.213.190/PNEed15/index.php (NOEP)"

REFERENCES

1. Smith T, Whitchurch C. The Future of the Tripartite Mission: Re-examining the Relationship Linking Universities, Medical Schools and Health Systems. *Higher Education Management and Policy* 2002;14(2):39-52 doi: 10.1787/hemp-v14-art12-en
2. Mintzberg H. Structure in Fives: Designing Effective Organizations. Upper Saddle River, NJ: Prentice Hall 1983.
3. Palm W, Glinos IA, Garel P, et al. Building European reference networks: exploring concepts and national practices in the European Union. Copenhagen, WHO Regional Office for Europe on behalf of the European Observatory on Health Systems and Policies 2013 http://www.euro.who.int/__data/assets/pdf_file/0004/184738/e96805-final.pdf [Accessed 26 January 2016]
4. Clark J, Tugwell P. Does Academic Medicine Matter?. *PLoS Med* 2006;3(9):e340. doi:10.1371/journal.pmed.0030340
5. International Working Party to Promote and Revitalise Academic Medicine. Academic medicine: the evidence base. *BMJ* 2004;329:789-92 doi: 10.1136/bmj.329.7469.789
6. Ayanian JZ, Weissman JS. Teaching Hospitals and Quality of Care: A Review of the Literature. *Milbank Q* 2002;80(3):569–93 doi: 10.1111/1468-0009.00023
7. Kupersmith J. Quality of Care in Teaching Hospitals: A Literature Review. *Acad Med* 2005;80(5):458-66
8. Papanikolaou PN, Christidi GD, Ioannidis JP. Patient outcomes with teaching versus nonteaching healthcare: a systematic review. *PLoS Med* 2006 Sep;3(9):e341 doi: 10.1371/journal.pmed.0030341
9. Hollingsworth B. Non-parametric and parametric applications measuring efficiency in health care. *Health Care Manag Sci* 2003;6(4):203-18 doi:10.1023/A:1026255523228
10. Grosskop S, Margatitis D, Valdmanis V. Comparing Teaching and Non-teaching Hospitals: A Frontier Approach (Teaching vs. Non-teaching Hospitals). *Health Care Manag Sci* 2001;4(2):83-90 doi: 10.1023/A:1011449425940

11. Sloan F, Feldman R, Steinwald B. Effects of teaching on hospital costs. *J Health Econ* 1983;2(1):1–23 doi:10.1016/0167-6296(83)90009-7

12. Iezzoni LI, Shwartz M, Moskowitz MA, et al. Illness Severity and Costs of Admissions at Teaching and Nonteaching Hospitals". *JAMA* 1990;264(11):1426-31 doi:10.1001/jama.1990.03450110072030

13. Mechanic R, Coleman K, Dobson A. Teaching hospital costs: implications for academic missions in a competitive market. *JAMA* 1998 Sep 16;280(11):1015-9 doi:10.1001/jama.280.11.1015

14. Carbone C, Lega F, Prenestini A. La governance delle Aziende Ospedaliero-Universitarie: perché non funziona, come potrebbe migliorare. In: Cantù E. L'aziendalizzazione della sanità in Italia - Rapporto OASI 2010. Milano, IT: EGEA 2010.

15. Roversi Monaco MG. Università e Servizio Sanitario Nazionale: L'azienda ospedaliera universitaria. Padova, IT: CEDAM 2011.

16. Leardini C, Sala GA, Campedelli B. L'azienda ospedaliera universitaria integrata. Un modello di integrazione tra attività di cura, ricerca e formazione. Milano, IT: Franco Angeli 2015.

17. Barsanti S, Giovanelli S, Nuti S. Come valutare la ricerca nelle Aziende Ospedaliero-Universitarie? Metodi e proposte operative. *Mecosan* 2014;90(1):9-32 doi: 10.3280/MESA2014-090002

18. Bevan G, Rutten F. The organisation and functions of University Hospital in different countries. *Finance Accountability & Management* 1987;3(2):77-115.

19. Bevan G. The medical service increment for teaching (SIFT): A £400m anachronism for the English NHS?. *BMJ* 1999;319:908. doi: http://dx.doi.org/10.1136/bmj.319.7214.908

20. 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. doi: 10.1007/s10997-012-9218-5

21. Nuti S. La valutazione della performance in sanità. Bologna, IT: il Mulino 2008.

22. Nuti S, Vola F, Bonini A, et al. Making governance work in the healthcare sector: evidence from a "natural experiment" in Italy. *Health Econ Policy Law* 2015;11(1):17-38. doi: 10.1017/S1744133115000067

23. Nuti S, Vola F, Amat P. Il sistema di valutazione della performance dei sistemi sanitari regionali: Basilicata, Calabria, Emilia-Romagna, Friuli Venezia Giulia, Lazio, Liguria, Lombardia, Marche, P.A. Bolzano, P.A. Trento, Puglia, Sardegna, Toscana, Umbria, Veneto - REPORT 2014. Pisa, IT: ETS 2015.

24. *Inter-Regional Performance Evaluation System* (IRPES) website: http://performance.sssup.it/netval [Accessed 25 January 2016]

25. *Italian National Outcome Evaluation Programme* (NOEP) website: http://95.110.213.190/PNEed15/index.php [Accessed 25 January 2016]

26. Ash AS, Fienberg SE, Louis TA, et al. Statistical Issue in Assessing Hospital Performance. Quantitative Health Sciences Publications and Presentations. Centers for Medicare and Medicaid Services 2012. http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Downloads/Statistical-Issues-in-Assessing-Hospital-Performance.pdf [Accessed 25 January 2016]

27. Iezzoni LI. Risk Adjustment for measuring healthcare outcomes, 2nd edition. Chicago, IL: Health Administration Press 1997.

28. Campbell SM, Braspenning J, Hutchinson A, et al. Research Methods used in developing and applying quality indicators in primary care. *Qual Saf Health Care* 2002;11:358-64. doi:10.1136/qhc.11.4.358

29. Ibraihm JE, Performance indicators from all perspectives. *Int J Qual Health Care* 2001 Dec;13(6):431-2. doi: dx.doi.org/10.1093/intqhc/13.6.431 [Published Online First: 1 December 2001].

30. Garcia-Lacalle J, Martin E. Rural vs urban hospital performance in a 'competitive' public health service. *Soc Sci Med* 2010 Sep;71(6):1131-40 doi:10.1016/j.socscimed.2010.05.043 [Published Online First: 19 June 2010].

31. Mann HB, Whitney DR. On a test of whether one of 2 random variables is stochastically larger than the other. *Annals of Mathematical Statistics* 1947;18:50-60 http://projecteuclid.org/euclid.aoms/1177730491 [Accessed 26 January 2016]

32. Lehmann EL. Nonparametrics: Statistical Methods Based on Ranks. *Journal of Applied Mathematics and Mechanics* 1977;57(9):562 doi: 10.1002/zamm.19770570922

33. Zimmerman DW. Comparative power of Student t test and Mann-Whitney U test for unequal sample sizes and variances". *J Exp Educ* 1987;55:71-174.

34. Siegel S, Castellan NJJ. Nonparametric statistics for the behavioral sciences. New York, NY: McGraw-Hill 1988.

35. Hollander M, Wolfe DA. Nonparametric Statistical Methods, Solutions Manual, 2nd Edition. Chichester, UK: Wiley and Sons Ltd 1999.

36. Levene H. Robust testes for equality of variances. In: Olkin I. Contributions to Probability and Statistics. Palo Alto, CL: Stanford University Press 1966:278-292.

37. Brown MB, Forsythe AB. Robust tests for the equality of variances. *J Am Stat Assoc* 1974;69:364–67. doi:10.2307/2285659

38. Murante AM, Seghieri C, Brown A, et al. How do hospitalization experience and institutional characteristics influence inpatient satisfaction? A multilevel approach. *Int J Health Plann Manage* 2014;29(3):e247-60 doi:10.1002/hpm.2201 [Published Online First: 1 July 2014].

39. Dimick JB, Cowan JA, Colletti LM et al. Hospital Teaching Status and Outcomes of Complex Surgical Procedures in the United States. *Arch Surg* 2004;139(2):137-141. doi:10.1001/archsurg.139.2.137

40. Nuti S, Marcacci L. The Performance Evaluation System of Health Care in Tuscany – Report 2010. Calenzano, IT: Edizioni del Poligrafico Fiorentino 2012.

APPENDIX I – Complete list of hospitals included in the analysis and number of observations included for each indicators

IRPES Indicators	UHs	GHs	Total Hospitals
2011			
Average cost per weighted case	13	42	55
Average expenditure per Diagnostic Imaging weighted for tariff	13	34	47
Total Hospitals in 2011	**26**	**77**	**103**
2012			
Relative stay index: case-mix adjusted differential average Length of Stay - LOS days	16	73	89
Percentage of medical discharges with LOS > ministerial threshold for patients over 65	16	72	88
Percentage of ED green-coded patients visited within 1 hour	16	69	85
Percentage of ED patients referred for hospital admission with ED length of stay <= 8 hours	16	70	86
Percentage of medical inpatient discharges within 2 days (National Healthcare Agreement 2010)	16	73	89
Percentage of day-surgery treatment for specific procedures (National Healthcare Agreement 2010)	16	71	87
Percentage of patients leaving ED against/without medical advice	16	70	86
Percentage hospitalized patients leaving against medical advice	16	73	89
Average cost per weighted case	15	59	74
Average expenditure per Diagnostic Imaging weighted for tariff	15	57	72
Total Hospitals in 2012	**158**	**687**	**845**
2013			
Relative stay index: case-mix adjusted differential average Length of Stay - LOS days	16	73	89
Percentage of medical discharges with LOS > ministerial threshold for patients over 65	16	72	88
Percentage of ED green-coded patients visited within 1 hour	15	71	86
Percentage of ED patients referred for hospital admission with ED length of stay <= 8 hours	16	71	87
Percentage of medical inpatient discharges within 2 days (National Healthcare Agreement 2010)	16	73	89
Percentage of day-surgery treatment for specific procedures (National Healthcare Agreement 2010)	16	71	87
Percentage of patients leaving ED against/without medical advice	16	70	86
Percentage hospitalized patients leaving against medical advice	16	72	88
Total Hospitals in 2013	**64**	**286**	**350**

NOEP Indicators	UHs	GHs	Total Hospitals
2012			
AMI: 30-day mortality	18	111	129
AMI without PTCA: 30-day mortality	15	85	100
AMI with PTCA within 2 days: 30-day mortality	16	37	53
AMI with PTCA after 2 days: 30-day mortality	14	20	34
AMI: 1-year mortality	17	113	130
AMI: MACCE after 1 year	17	113	130
Isolated Aortocoronary Bypass: 30-day mortality	13	5	18
Congestive heart failure: 30-day mortality	19	153	172
Ischemic Stroke: 30-day mortality	17	112	129
COPD exacerbation: 30-day mortality	17	144	161
Proportion of Cesarean Section	16	127	143
Femur Fracture: 30-day mortality	17	119	136
Femur Fracture: Percentage of operations carried out within 2 days	17	121	138
Colon cancer surgery: 30-day mortality	18	95	113
COPD: 30-day re-admission	17	137	154
Valvuloplasty or heart valve replacement: 30-day mortality	13	5	18
Ischemic Stroke: 30-day re-admission	17	104	121
Total Hospitals for 2012	**278**	**1,601**	**1,879**
2013			
AMI: 30-day mortality	17	111	128
AMI without PTCA: 30-day mortality	15	77	92
AMI with PTCA within 2 days: 30-day mortality	15	41	56
AMI with PTCA after 2 days: 30-day mortality	14	18	32
AMI: 1-year mortality	17	109	126
AMI: MACCE after 1 year	17	109	126
Isolated Aortocoronary Bypass: 30-day mortality	13	4	17
Congestive heart failure: 30-day mortality	19	172	191
Ischemic Stroke: 30-day mortality	16	106	122
COPD exacerbation: 30-day mortality	17	134	151
Proportion of Cesarean Section	17	124	141
Femur Fracture: 30-day mortality	16	118	134
Femur Fracture: Percentage of operations carried out within 2 days	16	118	134
Colon cancer surgery: 30-day mortality	17	97	114
COPD: 30-day re-admission	17	133	150
Valvuloplasty or heart valve replacement: 30-day mortality	14	5	19
Ischemic Stroke: 30-day re-admission	16	101	117
Total Hospitals for 2013	**273**	**1,577**	**1,850**

BMJ Open

Do university hospitals perform better than general hospitals? A comparative analysis among Italian regions

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-011426.R2
Article Type:	Research
Date Submitted by the Author:	29-Apr-2016
Complete List of Authors:	Nuti, Sabina; Scuola Superiore Sant\'Anna, Institute of Management - Management&Health Laboratory Grillo Ruggieri, Tommaso; Scuola Superiore Sant\'Anna, Institute of Management - Management&Health Laboratory Podetti, Silvia; Scuola Superiore Sant\'Anna, Institute of Management - Management&Health Laboratory
Primary Subject Heading:	Health services research
Secondary Subject Heading:	Health policy
Keywords:	Performance, Evaluation, University, Hospital, Italy

Do university hospitals perform better than general hospitals? A comparative analysis among Italian regions

Corresponding author:
Sabina Nuti
Postal Address: Laboratorio MeS, Istituto di Management - Scuola Superiore Sant'Anna,
Via San Zeno 2 – 56127, Pisa, Italy
Email address: s.nuti@sssup.it
Telephone: +39 050 88 38 86
Fax: +39 050 88 38 90

Authors:
Sabina Nuti, Full Professor, Management and Health Laboratory, Institute of Management, Scuola Superiore Sant'Anna, Via San Zeno 2 – 56127, Pisa, Italy, s.nuti@sssup.it.

Tommaso Grillo-Ruggieri, PhD candidate, Management and Health Laboratory, Institute of Management, Scuola Superiore Sant'Anna, Via San Zeno 2 – 56127, Pisa, Italy, t.grilloruggieri@sssup.it

Silvia Podetti, PhD, Management and Health Laboratory, Institute of Management, Scuola Superiore Sant'Anna, Via San Zeno 2 – 56127, Pisa, Italy, s.podetti@sssup.it

Keywords: performance, evaluation, university, hospital, Italy
Word Count: 3,835

Abstract

Objective

The aim of this research was to investigate how university hospitals (UHs) perform compared to general hospitals (GHs) in the Italian healthcare system.

Design and setting

Twenty-seven indicators of overall performance were selected and analysed for UHs and GHs in ten Italian regions. The data refer to 2012 and 2013 and were selected from two performance evaluation systems based on administrative data on hospitalization discharge flows: the Inter-Regional Performance Evaluation System developed by the Management and Health Laboratory of the Scuola Superiore Sant'Anna of Pisa and the Italian National Outcome Evaluation Programme developed by the National Agency for Healthcare Services. The study was conducted in two stages and by combining two statistical techniques. In stage 1, a nonparametric Mann-Whitney *U* test was carried out to compare the performance of UHs and GHs on the selected set of indicators. In stage 2, a robust equal variance test between the two groups of hospitals was carried out to investigate differences in the amount of variability between them.

Results

The overall analysis gave heterogeneous results. In general, performance was not affected by being in the UH rather than the GH group. It is thus not possible to directly associate Italian UHs with better results in terms of appropriateness, efficiency, patient satisfaction, and outcomes.

Conclusions

Policy-makers and managers should further encourage hospital performance evaluations in order to stimulate wider competition aimed at assigning *teaching status* to those hospitals that are able to meet performance requirements. In addition, UH facilities could be integrated with other providers that are responsible for community, primary and outpatient services, thereby creating a joint accountability for more patient-centred and integrated care.

Strengths and limitations of the study

- This study provides evidence about differences in terms of performance between university hospitals and general hospitals that was lacking in Italy
- The analysis shows new results about hospital performance that can contribute to the debate on this topic
- For the first time a nonparametric approach of analysis was applied for this topic to the Italian context

- The study is limited to the Italian healthcare system and its organizational structure
- There could be other performance indicators that are as valuable and informative as those measures included in the analysis

1. INTRODUCTION

University hospitals (UHs) can be considered as complex organizations given that their mission includes three different objectives: patient care, education, and research [1]. UHs combine all the features of Mintzberg's Professional Bureaucracy [2] embedded within both the healthcare organizations and the university context. In addition, UHs are usually referral centres for most complex care within a hub-and-spoke hospital network [3].

Given the three-fold mission of these institutions and the specific role that they play in the healthcare system, should UHs be considered as a 'cluster' with specific performance patterns?

This study investigates whether UHs behave homogenously regarding performance results with substantial differences with respect to general hospitals (GHs).

Evidence on this topic could provide important information for policy-makers and managers in defining specific policies and actions in order to improve the quality of care within the regional network of hospitals, where UHs play a specific and strategic role, and in order to pursue their specific mission.

In particular, in Italy as in other countries, UHs are in charge of the strategic role of training doctors of the future. Therefore, since health professionals are the most important assets for the healthcare organizations, policy-makers should ensure clinicians are trained and supported by institutions that can ensure the appropriate requirements in terms of quality of care and research productivity. The analysis was carried out in Italy.

2. BACKGROUND

Teaching status has been already investigated from several perspectives by studying whether it affects the results of UHs compared with other hospitals in terms of outcomes, quality of care, productivity, costs, etc.

Firstly, reviews on outcomes, quality of care and adverse-event prevention reached mixed conclusions and highlighted the need for evidence on differences between UHs and GHs [4-5]. Some reviews underlined better overall results for UHs [6-7], whereas a systematic review highlighted no differences between UH and GH outcomes [8].

Secondly, studies on productivity and efficiency have usually applied Data Envelopment Analysis (DEA) and frequently highlighted better performance of GHs with respect to UHs [among others, 9-10].

Indeed, training resident students, carrying out research activities besides patient care and the role of referral centres for complex care have often been identified as elements that can increase costs [11-13]. This frequently drives additional financial resources to UHs (e.g., an increased mark-up in the reimbursement system for UH discharges) [6].

Research on this topic presents several differences in terms of data sources, measurement processes and methodology for data analysis [4]. This could raise potential issues regarding external validity and result generalizability [6-9]. Examples of these differences are:

- The data sources: e.g. medical records or administrative data;
- The definition of UHs and their ownership (public, private, for-profit, non-profit): for example, some studies consider only major UHs, whereas others include all the hospitals with a residency program;
- The indicators included in the analysis (usually, outcomes, quality of care or efficiency) and the different calculation criteria and risk-adjustment procedure used for the same measures (mortality rates, process measures, etc.);
- The statistical methods used to compare hospitals (parametric and nonparametric approaches and tests such as DEA, ANOVA, Kruskal-Wallis, Mann-Whitney, etc.).

These differences may partially explain why research looking at different performance or outcomes in UHs or controlling for a potential effect of the *teaching status* have not led to straightforward results.

Finally, results may be also associated with the specific geographical context. For instance, in one of the most recent systematic reviews on this topic, more than three-fourths of the studies included in the analysis were conducted in the United States [8]. However, each specific geographical and health system context may play an important role in explaining results.

With reference to Italy, detailed studies are also lacking on this topic. Scholars have focused on governance issues or research evaluations [see for instance, 14-17]. There have been no systematic comparisons of performances between the two groups of hospitals and related research.

2.1. The Italian context

The national healthcare system in Italy follows a Beveridge model by providing universal coverage through general taxation. Regional governments are responsible for organizing and delivering health services and being accountable for performance. National government monitors the pursuit of the

universal coverage in particular with respect to a package of essential services (Nationally defined basic health benefit package - Livelli Essenziali di Assistenza). National government allocates financial resources to the regional governments on an adjusted capitation basis. Regions then reallocate resources to Local Health Authorities (LHAs), through a regionally-adjusted capitation formula.

In Italy, hospital care is delivered by public general hospitals (GHs) directly managed by the LHAs, private or public autonomous hospitals (AHs), private or public university hospitals (UHs) and research hospitals (RHs). AHs, UHs and RHs are autonomous organizations with respect to LHAs managing the healthcare delivery in their own geographical area.

UHs can be classified considering ownership and different institutional and organizational settings [18]. In Italy, the *teaching status* can be attributed to hospitals owned by private university medical schools, hospitals owned by public university medical schools and hospitals jointly owned by both public university medical schools and the Regional Administration. In this last case, the CEO is jointly appointed by the two institutions. Following the national laws (D.Lgs 502/92 and D.Lgs 517/99), these hospitals are identified as teaching facilities by the Ministry of Health, the Ministry of Education and the Regional Administrations. Regardless of the ownership and the organizational settings, health professionals employed by universities, besides teaching and carrying out research, also provide patients care and receive an additional 30% remuneration. These costs are directly sustained not by the universities but by the hospital administration.

Considering patient care activity, since UHs are autonomous authorities, they are not financed through capitation-based funding as the LHAs, but through different financing mechanisms depending on regional strategies.

At the national level, UH inpatient services delivered for residents of other regions are reimbursed considering a DRG tariff increase of 7%.

At the regional level, UHs can be financed through a pay for service system based on DRG tariffs (e.g. Lombardy Region) or through a budget-cost control system. In the first case, UH DRG tariffs are increased of a certain percentage (usually the 3% circa), depending on the case-mix delivered and the regional strategy. In the second case, as well as in other countries [see among others, 19], Regions usually assign additional resources to UHs through specific funds linked to education, research and complex care delivery (e.g. in Tuscany the amount of these funds accounted for the 30% of the UH overall budget). Therefore, UHs receive an amount of additional resources with respect to GHs, but this varies depending on the regional policies [14].

Italian UHs have on average a much higher number of hospital beds with respect to GHs and are hub-referral centres for highly-complex and highly-specialized care, such as neuro-surgery, cardio-surgery, radiotherapy, most critical intensive care, paediatric highly-complex surgery, etc.

Evidence from Italy on the comparison of UH performance with respect to GHs may provide valuable information for both healthcare policy-makers and managers, at both regional and national levels and not only in Italy. Indeed, if UHs behave as a specific 'cluster', new policies and focused-actions could be defined to support the specific role of these authorities within the hospital network in the regional and national contexts. Evidence of similar patterns of performance between these two groups of hospitals may highlight the need to look for other sources of variation. Therefore, other features from the *teaching and research status* may be relevant to inform policies on hospital governance, financing and network organization, considering the crucial role of UHs in training the future clinicians for the healthcare system.

The aim of this paper is thus to investigate how UHs perform in comparison to GHs.

3. METHODS

3.1. Data sources and hospital selection

The data used in this analysis were selected from two performance evaluation systems based on the same source of administrative data on hospitalization discharge flows:

- The *Inter-Regional Performance Evaluation System* (IRPES) developed by the Management and Health Laboratory of the Scuola Superiore Sant'Anna of Pisa (MeS-Lab) - where the authors of this paper are researchers. This system provides a multi-dimensional evaluation of performance including efficiency, appropriateness, integration and quality of care. This system was firstly implemented by the regional government in Tuscany [20-21] and was then adopted – on a voluntary basis - by the majority of other Italian regions** [22-23]. The evaluation process measures through benchmarking and with specific risk adjustment processes the results achieved every year by all the Health Authorities (the local health authorities, the university hospitals, the research hospitals and the autonomous hospitals) located in these regions. Results are publicly reported [24].

- The *Italian National Outcome Evaluation Programme* (NOEP) developed by the National Agency for Healthcare Services on behalf of the Ministry of Health. This system measures outcomes nation-wide [25], i.e. for each Italian hospitals. On the basis of rigorous risk adjustment processes [among others, 26-27], these measures represent assessment tools to support clinical and organizational audit programs aimed at improving both outcome and equity in the National Health System.

Data refer to the hospital activity of 2012 and 2013, apart from 2 economic indicators related to balance sheets, which are available only for 2011 and 2012.

Two groups of hospitals were considered in the analysis. The groups differed in particular in terms of whether they had *teaching status* and in the organizational autonomy with respect to the LHAs. They also differed in terms of the average number of hospital discharges (in 2012, 32,632 for UHs and approximately 17,606 for GHs) and the average DRG weight (in 2012, 1.3 for UHs and 1.06 for GHs). The whole study included all the 15 UHs and 73 LHAs of the ten IRPES regions.

3.2. Performance indicators

For the purposes of this study, 27 performance indicators were selected, 10 from IRPES (Table 1) and 17 from NOEP (Table 2).

Table 1 – IRPES Indicators

IRPES INDICATORS	Rationale
EFFICIENCY AND APPROPRIATENESS	
Relative stay index: case-mix adjusted differential average Length of Stay - LOS days	Measure of the average difference from the standard LOS for admitted patients with adjustments for case-mix
Percentage of medical discharges with LOS > ministerial threshold for patients over 65	Measure of the hospital compliance with the Italian Ministry of Health standards for the LOS for medical inpatient activity for elderly patients. This measure is a proxy of the effective implementation of integrated pathways between home, community-based and hospital care for elderly patients
Percentage of ED green-coded patients visited within 1 hour	Measure of timely emergency care for ED patients whose treatment may be delayed without risk
Percentage of ED patients referred for hospital admission with ED length of stay <= 8 hours	Measure of overall timely emergency care
Percentage of medical inpatient discharges within 2 days (National Healthcare Agreement 2010)	Measure of hospital compliance in avoiding short ordinary hospitalizations for patients that could be treated in outpatient clinics or in other care settings, as requested by the Italian Ministry of Health standards in the National Healthcare Agreement of 2010
Percentage of day-surgery treatment for specific procedures (National Healthcare Agreement 2010)	Measure of hospital compliance with Italian Ministry of Health standards for delivering specific not-complex surgical procedures in day-surgery or in outpatient clinics rather than through ordinary hospitalizations
PATIENT SATISFACTION	
Percentage of patients leaving ED against/without medical advice	Proxy of Patient Satisfaction on ED services and waiting times
Percentage hospitalized patients leaving against medical advice	Proxy of Patient Satisfaction for the inpatient activity
ECONOMIC AND FINANCIAL EVALUATION	
Average cost per weighted case	Measure of the ratio of a hospital acute inpatient care expenses to the number of acute inpatient cases weighted for the DRG complexity. The weighting enhances comparability across hospitals. The measure includes the percentage cost of hospital university staff financed by the Regional Administration for their patient care activity. This allows to account for the overall hospital staff costs.
Average expenditure per Diagnostic Imaging	Measure of efficiency that compares costs and the value of the

weighted for tariff	delivered diagnostic activity (sum of ambulatory tariffs)

Table 2 – NOEP Indicators

NOEP INDICATORS
OUTCOME: measures of 30-day mortality or re-admissions for relevant inpatient activity
AMI: 30-day mortality
AMI without PTCA: 30-day mortality
AMI with PTCA within 2 days: 30-day mortality
AMI with PTCA after 2 days: 30-day mortality
AMI: 1-year mortality
AMI: MACCE after 1 year
Isolated Aortocoronary Bypass: 30-day mortality
Valvuloplasty or heart valve replacement: 30-day mortality
Congestive heart failure: 30-day mortality
Ischemic Stroke: 30-day mortality
Ischemic Stroke: 30-day re-admission
COPD exacerbation: 30-day mortality
COPD: 30-day re-admission
Proportion of Caesarean Section
Femur Fracture: 30-day mortality
Femur Fracture: Percentage of operations carried out within 2 days
Colon cancer surgery: 30-day mortality

Eight IRPES indicators regard efficiency and appropriateness, patient satisfaction, and economic and financial dimensions. Two indicators regard economic and financial evaluation. This selection was shared by the group of the IRPES regional representatives. This group is in charge of systematically reviewing and discussing the measures included in the IRPES as relevant proxies for measuring performance in a multidimensional perspective in all the different settings of care [22].

For both sources of the selected indicators, the time coverage and the number of providers needed to perform the statistical test were guaranteed, thus ensuring the consistency of the comparative analysis between the two groups of hospitals in this single-country study [among others, 28-29].

The number of observations for the NOEP indicators may differ because not all the hospitals included in the analysis provide all the healthcare services linked to the included measures. However, the selection of these measures took into account the services usually provided by both LHA-GHs and UHs.

The analysis for the IRPES indicators compared the 15 UHs to the 73 LHAs. On the other hand, the analysis for the NOEP indicators was carried out at the hospital level, thus comparing the (at most)

19 facilities of the 15 UHs to the individual (at most) 191 GHs led by the 73 LHAs. (See Appendix I in the Supplementary File for the complete list of hospitals considered and the number of observations included for each indicators).

3.3. Statistical Methods

The study was conducted in two stages and by combining two statistical techniques. Data were processed using Stata software, version 12. In stage 1, a nonparametric Mann-Whitney *U* test was carried out to compare the performance of UHs and GHs on the selected set of indicators. This analysis determines whether UHs and GHs were drawn from the same target population. Previous studies have already applied this univariate analysis to illustrate differences between hospitals [among others, 30] because of its appropriateness with small samples [31-35]. For the purposes of this study, this test verified whether there were differences between UH and GH performance, or, in other words, whether UHs and GHs could be considered as two different clusters. In stage 2, we carried out a robust equal variance test to investigate differences in the amount of variability between UHs and GHs [36]. This test is usually used to verify the assumption of homogeneity of variance across groups, meaning that the internal variability of one group of hospitals is not significantly different with respect to the other one.

To be in line with the assumptions of the Mann-Whitney *U* test, we used an extension of Levene's test as suggested by Brown and Forsythe [37]. We applied the test only for those indicators in which the Mann-Whitney *U* test did not show significant differences between UH and GH performances. Indeed, in those cases where the performance between the two groups did not show significant differences, we tested whether there were specific patterns in terms of variability.

4. RESULTS

The Mann-Whitney *U* test on IRPES indicators showed that in relation to four measures of "Efficiency and Appropriateness" and "Economic and financial evaluation" dimensions, there were differences in performance between UHs and GHs. The test, in fact, was significant both in 2012 and 2013 for the "% of ED green-coded patients visited within one hour", the "% of medical inpatient discharges within two days" and the "% of day-surgery treatment for specific procedures". The test was significant also in both 2011 and 2012 for the "Average expenditure for Diagnostic Imaging weighted for tariff". For these indicators, GHs seemed to perform better than UHs.

On the other hand, with reference to the indicators "Relative stay index", "% of medical discharges with LOS > ministerial threshold for patient over 65", and "% of ED patient referred for hospital

admission with ED length of stay <= 8 hours", the Mann-Whitney *U* test was rejected for both 2012 and 2013.

Moreover, no significant differences were found for patient satisfaction proxies "% of patients leaving ED against/without medical advice" and of "% of hospitalized patients leaving against medical advice". Moreover, in 2013 UHs accounted for fewer patients that were discharged against medical advice, whereas in 2012 the GHs achieved better results. The test was also not significant for the "Average cost per weighted case" and this occurred also after deleting outliers.

Table 3 summarizes the results of the test and illustrates the average and the median values of the two groups of hospitals for each of the indicators.

Table 3 – Mann-Whitney *U* test for IRPES Indicators

MANN – WHITNEY *U* TEST IRPES INDICATORS	2012					2013				
	Median UH	Median GH	Mean UH	Mean GH	Best Perf. MEDIAN	Median UH	Median GH	Mean UH	Mean GH	Best Perf. MEDIAN
EFFICIENCY AND APPROPRIATENESS										
Relative stay index (case-mix adjusted differential average LOS days)	-0.2	-0.1	0	-0.2	UH	0	-0.3	-0.1	-0.3	GH
% medical discharges with LOS > ministerial threshold for patients over 65	4.8	3.6	4.6	4	GH	3.7	3.5	4.3	3.8	GH
% ED green-coded patients visited within 1 hour	**73.1**	**79.2**	**72.7**	**77.3**	**GH***	**68.4**	**77.2**	**67.2**	**76.2**	**GH***
% ED patients referred for hospital admission with ED length of stay <= 8 hours	98.8	97.8	93.9	94.8	UH	98.2	97.5	93.2	94.5	UH
% medical inpatient discharges within 2 days (National Healthcare Agreement 2010)	**21.5**	**14.6**	**22.3**	**14.9**	**GH***	**21.8**	**14.1**	**21.9**	**14.4**	**GH***
% day-surgery treatment for specific procedures (National Healthcare Agreement 2010)	**46.2**	**58.8**	**48**	**58.9**	**GH***	**48.4**	**59.1**	**49**	**59**	**GH***
PATIENT SATISFACTION										
% patients leaving ED against/without medical advice	3.2	3.2	3,6	3.1	GH	3.5	3.2	3.6	3.4	GH
% hospitalized patients leaving against medical advice	0.9	0.8	1	1	GH	0.7	0.8	0.9	0.9	UH
ECONOMIC AND FINANCIAL EVALUATION										
	2011					**2012**				
Average cost per weighted case	4,471	4,317	4.782	4,398	GH	4,484	4,516	4,745	4,651	UH
Average expenditure per Diagnostic Imaging weighted for tariff	**1.4**	**0.9**	**1.8**	**1.1**	**GH***	**1.4**	**1**	**1.6**	**1.1**	**GH***

Regarding the test for the NOEP indicators, for all the tested measures, the Mann-Whitney *U* test was not significant except for two measures that showed mixed results in 2012 and in 2013 (Table 4) (in the Appendix II box-plots for IRPES and NOEP indicators with significant differences between UHs and GHs are showed).

Table 4 - Mann-Whitney *U* test for NOEP Indicators

MANN-WHITNEY *U* TEST – NOEP INDICATORS	2012					2013				
	Median UH	Median GH	Mean UH	Mean GH	Best perf. MEDIAN	Median UH	Median GH	Mean UH	Mean GH	Best perf. MEDIAN
OUTCOME INDICATORS										
AMI: 30-day mortality	9.8	8.8	10.1	9.3	GH	9.1	7.7	8.9	8.1	GH
AMI without PTCA: 30-day mortality	17.4	15.5	17.7	16.5	GH	16.8	15.0	17.5	15.5	GH
AMI with PTCA within 2 days: 30-day mortality	4.8	4.1	4.6	4.2	GH	4.1	3.7	4.4	3.8	GH
AMI with PTCA after 2 days: 30-day mortality	2.7	2.4	3.2	2.6	GH	2.6	2.5	2.9	2.8	GH
AMI: 1-year mortality	10.4	11.1	10.6	11.5	UH	9.8	10.6	10.2	10.9	UH
AMI: MACCE after 1 year	24	24.8	24.5	25.2	UH	22.4	23.2	23.1	23.6	UH
Isolated Aortocoronary Bypass: 30-day mortality	1.8	1.9	2.2	2.0	UH	2	2.3	2.4	2.1	UH
Valvuloplasty or heart valve replacement: 30-day mortality	2.6	3.7	2.9	3.5	UH	2.3	3.0	2.8	3.2	UH
Congestive heart failure: 30-day mortality	8.4	9.8	9.3	10.8	UH	**8.8**	**10.7**	**8.7**	**11.0**	**UH***
Ischemic Stroke: 30-day mortality	9.4	10.1	8.8	10.5	UH	9.2	9.5	9.3	10.5	UH
Ischemic Stroke: 30-day re-admission	11.1	9.4	10.5	10.3	GH	6.7	6.7	7.2	7.2	UH
COPD exacerbation: 30-day mortality	7.2	8.7	7.6	8.9	UH	7.2	8.1	7.7	8.7	UH
COPD: 30-day re-admission	14.2	15.6	15.0	15.4	UH	14.2	15.4	14.2	15.4	UH
Proportion of Caesarean Section	19.9	17.8	23.6	18.8	GH	20.2	18.5	22.5	19.3	GH
Femur Fracture: 30-day mortality	4.2	4.8	4.7	5.1	UH	4.4	4.7	4.7	4.8	UH
Femur Fracture: Percentage of operations carried out within 2 days	**48.4**	**54.4**	**41.5**	**53.2**	**GH***	50.6	60.3	54.2	59.8	GH
Colon cancer surgery: 30-day mortality	3.4	3.9	4.4	4.3	UH	3.0	4.2	3.7	4.6	UH

For the "Congestive heart failure: 30 day mortality" the test showed no statistical differences between UHs and GHs in 2012. However, a significantly better performance for UHs was found in 2013. Similarly, in the case of the indicator "Femur fracture: % of operations carried out within two days", the Mann-Whitney *U* test showed significant differences between UHs and GHs in 2012, but not for 2013, with GHs having the best median performance.

In order to investigate different variations between the two groups of hospitals, the robust equal variance test [37] was carried out for a set of 23 indicators (6 IRPES indicators and 17 NOEP indicators) that rejected the Mann-Whitney *U* test.

Regarding IRPES indicators, the test was always not significant for both years included in the analysis (Table 5). UHs and GHs showed a higher standard deviation depending on the measures considered.

Table 5 – Robust Equal Variance test for IRPES Indicators

ROBUST EQUAL VARIANCE TEST - IRPES INDICATORS	2012				2013				Higher Variability in 2012	Higher Variability in 2013
	Std. Dev. UH	Std. Dev. GH	W50 - Median	Pr > F	Std. Dev. UH	Std. Dev. GH	W50 - Median	Pr > F		
EFFICIENCY AND APPROPRIATENESS										
Relative stay index (case-mix adjusted differential average LOS days)	0.9	1.4	0.2	0.6	0.8	1.2	0.7	0.4	GH	GH
% medical discharges with LOS > ministerial threshold for patients over 65	1.7	2	0.1	0.8	1.7	2.1	0.7	0.4	GH	GH
% ED patients referred for hospital admission with ED length of stay <= 8 hours	9	6.7	0.5	0.5	9.7	7.7	0.3	0.6	UH	UH
PATIENT SATISFACTION										
% patients leaving ED against/without medical advice	1.9	1.8	0.1	0.8	2	2.1	0	1	UH	GH
% hospitalized patients leaving against medical advice	0.7	0.7	0.1	0.8	0.6	0.6	0	1	GH	GH
ECONOMIC AND FINANCIAL EVALUATION										
	2011				2012				Higher Variability in 2011	Higher Variability in 2012
Average cost per weighted case	1,068	785	1.1	0.3	962	850	0.8	0.4	UH	UH

For the 2012 results of NOEP indicators, the test was significant for four measures (Table 6):

- "AMI: 1-year mortality" (p-value=0.02)
- "Ischemic stroke: 30-day mortality" (p-value=0.02)
- "Femur fracture: 30-day mortality" (p-value=0.02)
- "COPD: 30-day readmission" (p-value=0.02)

Table 6 – Robust Equal Variance test for NOEP indicators

ROBUST EQUAL VARIANCE TEST - NOEP INDICATORS	2012				2013				Higher Variability in 2012	Higher Variability in 2013
	Std. Dev. UH	Std. Dev. GH	W50 - Median	Pr > F	Std. Dev. UH	Std. Dev. GH	W50 - Median	Pr > F		
OUTCOME INDICATORS										
AMI: 30-day mortality	3.3	3.8	0.8	0.4	2.6	3.7	2.8	0.1	GH	GH
AMI without PTCA: 30-day mortality	4.8	6.2	1.1	0.3	4.4	6.6	2.3	0.1	GH	GH
AMI with PTCA within 2 days: 30-day mortality	1.4	1.9	1.4	0.2	1.8	2.1	0.4	0.5	GH	GH
AMI with PTCA after 2 days: 30-day mortality	1.6	1.4	0.5	0.5	1.2	1.4	0.9	0.3	UH	GH
AMI: 1-year mortality	**1.9**	**4.4**	**5.6**	**0.02***	3.3	3.8	0.2	0.7	**GH***	GH
AMI: MACCE after 1 year	4.1	5.3	2.1	0.2	**3.2**	**5.6**	**4.4**	**0.04***	GH	**GH***

Isolated Aortocoronary Bypass: 30-day mortality	1.4	1.6	0.0	0.9	1.6	1.4	0.0	0.9	GH	UH
Valvuloplasty or heart valve replacement: 30-day mortality	1.3	0.5	2.7	0.1	1.2	1.0	0.2	0.6	UH	UH
Congestive heart failure: 30-day mortality	3.3	5.0	1.8	0.2					GH	
Ischemic Stroke: 30-day mortality	**2.9**	**4.5**	**5.8**	**0.02***	4	4.6	0.5	0.5	**GH***	GH
Ischemic Stroke: 30-day re-admission	3.6	3.9	0.0	0.9	2.2	3.0	1.8	0.2	GH	GH
COPD exacerbation: 30-day mortality	2.3	3.9	3.7	0.1	2.9	4.1	1.1	0.3	GH	GH
COPD: 30-day re-admission	**2.4**	**4.5**	**5.9**	**0.02***	3.4	4.2	1.1	0.3	**GH***	GH
Proportion of Caesarean Section	9.1	7.1	1.2	0.3	9.2	7.1	1.1	0.3	UH	UH
Femur Fracture: 30-day mortality	**1.3**	**2.2**	**5.2**	**0.02***	2.1	2.2	0.6	0.5	**GH***	GH
Femur Fracture: Percentage of operations carried out within 2 days					16.7	17.5	0.6	0.4		GH
Colon cancer surgery: 30-day mortality	2.7	2.3	0	0.9	1.7	2.5	2.7	0.1	UH	GH

In 2013, the test was significant only for the indicator "AMI: MACCE after 1 year" (p-value=0.04). For these measures, GHs showed a higher Standard Deviation with respect to UHs. This was also the case for most of the other outcome measures included for both 2012 and 2013, apart from the "Proportion of Caesarean Section" and the "30-day mortality rate for Valvuloplasty or heart valve replacement".

5. DISCUSSION

The overall analysis showed heterogeneous results when comparing the two groups of hospitals. Considering the IRPES indicators of appropriateness, we found a higher compliance of GHs in pursuing the Italian Ministry of Health standards on directing patients to the appropriate care settings for surgical treatments as well as in avoiding short medical hospitalizations and giving preference to outpatient clinics or day-hospital cases. This may be due to the lower complexity of general LHA-led hospitals and to a related lower complex management.

Regarding efficiency, in 2013 GHs seemed to perform better than the UHs but these results are slightly different in 2012, thus leading to ambiguous conclusions. Therefore, the three-fold mission and the greater organizational complexity of UHs seemed to lead to lower but not significantly different efficiency with respect to GHs. The more straightforward results in terms of the waiting times in ED may be due to greater pressure in the UH emergency departments, which are usually located in city centres.

Although the differences between GHs and UHs were always not significant, in 2012 GHs accounted for higher patient satisfaction. These results changed in 2013. However, previous research focused only on the patient experience with hospital medical staff in Tuscany showed a higher patient satisfaction for patients discharged by UHs with respect to patients hospitalized in GHs [see among others, 38].

In addition, the test on variability for IRPES indicators showed homogenous patterns of performance regardless of the *teaching status*. In particular, the UHs showed a larger variation in the Average cost per weighted case, which measures efficiency by comparing the average costs of inpatient cases weighted for the DRG complexity. This suggests that, as a group, UHs do not generally account for higher costs, contrary to what has been stated by other scholars [11-13]. UHs, as individuals, show highly heterogeneous results. Hence, based on our analysis, the financial and economical sustainability of UHs could be related to the individual internal organization or other factors rather than to the *teaching status*.

Finally, for the tested IRPES indicators and considering both the years considered in the analysis, a "cluster effect" linked to the *teaching status* did not seem plausible.

This is also confirmed by the analysis on the NOEP indicators, which suggested that UHs did not generally achieve better outcomes. These results contribute to the research on this topic by suggesting that there is no straightforward evidence for better outcomes associated with UHs. Interestingly, GHs performed better (although not significantly) considering indicators related to the waiting time for femur-fracture surgery and to the recourse to Caesarean sections. In most of the mortality and re-admission indicators, UHs did perform better but without a significant effect. Considering that UHs are referral centres with higher delivered volumes and patients, it is possible that these better results could also be explained by their role in the hospital network, rather than only by the *teaching status*, as suggested in other studies [see among others, 39].

In addition, GHs account for a general higher variability compared to UHs, but without significant differences. This means that, although UHs seem to be generally more concentrated around average values, the extreme values of GH results towards the maximum and minimum of the distribution do not affect the overall analysis results. In conclusion, a straightforward evidence identifying better performance and less variability for UHs does also not seem plausible for NOEP indicators.

Summarizing these results, from a multidimensional perspective being in the UH rather than the GH group does not generally affect performance. Hence, the different institutional and organizational settings between them do not seem to result in significant dissimilarities. Instead, the variations in hospital performance could be linked to particular features of each individual hospital or its managerial approach. Furthermore, these variations may also be determined by the Regional Healthcare System, rather than to a specific cross-regional group affiliation.

In Italy, there is evidence that hospital performance improvement may be affected by regional strategies combining different tools [22]. This is the case of Tuscany and Basilicata regions, which applied a combination of different integrated governance tools and registered a higher performance improvement in the last years with respect to other regions.

In fact, with reference to Tuscany, the regional UHs generally achieve a higher performance with respect to the UHs of the other IRPES regions [23-25; 40]. Nevertheless, the analysis of the impact of these regional strategies on performance of UHs needs to be furtherly investigated.

As a preliminary study on this topic, this research presents some limitations. Firstly, the study context focused on the Italian healthcare system and its organizational structure. We believe however that the contextual factors strongly influence the results. Therefore, these factors cannot be excluded when the research is aimed at supporting decision-making processes. This study provides evidence to enlarge the debate on this relevant topic not only in Italy but also in those countries aiming at linking *teaching status* attribution to performance evaluation. Secondly, there could be other indicators as valuable and informative as those measures included in the analysis. However, we included the ones that regional policy-makers and healthcare managers in Italy share as valuable measure to assess and guide the system.

Further studies will investigate the relevance of individual and regional factors in affecting UH and GH results in this multidimensional perspective.

6. CONCLUSIONS

The main finding of this study is that Italian UHs cannot straightforwardly be associated with better results in terms of appropriateness, efficiency, patient satisfaction, economic and financial evaluation, and outcomes. However, this preliminary evidence may inform the debate on the future role of UHs and encourage further considerations with regard to the Italian healthcare system.

Firstly, if UHs wish to maintain their role of leading players in the hospital network and to be the main actors in charge of training clinicians of the future, hospital performance evaluations should be further encouraged in order to inform the attribution of *teaching status* based on performance results. This could stimulate wider competition between Italian hospitals aimed at assigning *teaching status* to those hospitals that achieve the best performance in specific care paths. In this respect, medical schools should base their teaching activities for both undergraduate and resident students in the hospitals that can ensure the best results and practices, since the future generation of clinicians has a crucial role in improving the quality of care.

Secondly, considering the pressure towards more population-based oriented healthcare systems, the organizational structure of Italian UHs as an independent organization could be revised towards a more integrated network with other facilities delivering community, primary and outpatient care. UH facilities could therefore be directly integrated with the other LHA-led providers also creating a joint accountability for more patient-centred care. In this perspective, in Italy recent national legislation

(Disegno di Legge n. 2111-B/2016) has allowed as a pilot experience the Special-Administrative Regions (such as Friuli Venezia Giulia) to incorporate the UHs within the LHAs.

In conclusion, further studies on this topic will investigate whether performance of Italian UHs may be affected by regional strategies and systems of governance, such as the use of a transparent performance evaluation system.

ENDNOTES

** The IRPES in 2014 included Basilicata, Emilia-Romagna, Friuli Venezia Giulia, Liguria, Marche, Autonomous Province of Bolzano, Autonomous Province of Trento, Toscana, Umbria, Veneto. In 2015 Lombardia, Calabria, Lazio, Puglia and Sardegna joined the network.

ACKNOWLEDGEMENTS

The authors wish to thank all the IRPES regional representatives and their staff for their invaluable suggestions and collaboration and the MeS-Lab researchers for their help during data processing and results interpretation. We also wish to thank the participants to the Wennberg International Collaborative Spring Policy Meeting 2015 held in June 2015 for their comments on the draft version of this paper.

COMPETING INTERESTS

The authors have read and understood BMJ Open policy on declaration of interests and declare that we have no competing interests. All authors have completed the Unified Competing Interest form and declare that Nuti, Grillo-Ruggieri and Podetti have support from the Network of Italian regions that adopted the IRPES for the submitted work; Nuti, Grillo-Ruggieri and Podetti have no relationships with the Network of Italian regions that might have an interest in the submitted work in the previous 3 years; their spouses, partners, or children have no financial relationships that may be relevant to the submitted work; and Nuti, Grillo-Ruggieri and Podetti have non-financial interests that may be relevant to the submitted work.

CONTRIBUTORS

Sabina Nuti, the lead author, led the study design. Tommaso Grillo Ruggieri and Silvia Podetti carried out the data collection and the empirical analyses. All the authors were responsible for writing the manuscript and were involved in interpreting the findings and approving the final manuscript.

TRANSPARENCY DECLARATION

The lead author affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

FUNDING

This work was financed by the Network of regions adopting the Inter-Regional Performance Evaluation System (IRPES), coordinated by the Management and Health Laboratory of Scuola Superiore Sant'Anna of Pisa, Italy (www.meslab.sssup.it).

ROLE OF THE STUDY FUNDER

The funder was not directly involved in the study but it shared with the authors the research aim. The authors were independent from funder in designing the research and interpreting study results.

ETHICAL APPROVAL

The paper did not involve human participants and did not need an ethical approval.

DATA SHARING STATEMENT

"Data sharing: full dataset available at [/doi] with open access. No informed consent was necessary because the data used in this study are publicly reported on the following websites: http://performance.sssup.it/netval (IRPES) and http://95.110.213.190/PNEed15/index.php (NOEP)"

REFERENCES

1. Smith T, Whitchurch C. The Future of the Tripartite Mission: Re-examining the Relationship Linking Universities, Medical Schools and Health Systems. *Higher Education Management and Policy* 2002;14(2):39-52 doi: 10.1787/hemp-v14-art12-en
2. Mintzberg H. Structure in Fives: Designing Effective Organizations. Upper Saddle River, NJ: Prentice Hall 1983.
3. Palm W, Glinos IA, Garel P, et al. Building European reference networks: exploring concepts and national practices in the European Union. Copenhagen, WHO Regional Office for Europe on behalf of the European Observatory on Health Systems and Policies 2013 http://www.euro.who.int/__data/assets/pdf_file/0004/184738/e96805-final.pdf [Accessed 26 January 2016]
4. Clark J, Tugwell P. Does Academic Medicine Matter?. *PLoS Med* 2006;3(9):e340. doi:10.1371/journal.pmed.0030340
5. International Working Party to Promote and Revitalise Academic Medicine. Academic medicine: the evidence base. *BMJ* 2004;329:789-92 doi: 10.1136/bmj.329.7469.789
6. Ayanian JZ, Weissman JS. Teaching Hospitals and Quality of Care: A Review of the Literature. *Milbank Q* 2002;80(3):569–93 doi: 10.1111/1468-0009.00023
7. Kupersmith J. Quality of Care in Teaching Hospitals: A Literature Review. *Acad Med* 2005;80(5):458-66
8. Papanikolaou PN, Christidi GD, Ioannidis JP. Patient outcomes with teaching versus nonteaching healthcare: a systematic review. *PLoS Med* 2006 Sep;3(9):e341 doi: 10.1371/journal.pmed.0030341

9. Hollingsworth B. Non-parametric and parametric applications measuring efficiency in health care. *Health Care Manag Sci* 2003;6(4):203-18 doi:10.1023/A:1026255523228

10. Grosskop S, Margatitis D, Valdmanis V. Comparing Teaching and Non-teaching Hospitals: A Frontier Approach (Teaching vs. Non-teaching Hospitals). *Health Care Manag Sci* 2001;4(2):83-90 doi: 10.1023/A:1011449425940

11. Sloan F, Feldman R, Steinwald B. Effects of teaching on hospital costs. *J Health Econ* 1983;2(1):1–23 doi:10.1016/0167-6296(83)90009-7

12. Iezzoni LI, Shwartz M, Moskowitz MA, et al. Illness Severity and Costs of Admissions at Teaching and Nonteaching Hospitals". *JAMA* 1990;264(11):1426-31 doi:10.1001/jama.1990.03450110072030

13. Mechanic R, Coleman K, Dobson A. Teaching hospital costs: implications for academic missions in a competitive market. *JAMA* 1998 Sep 16;280(11):1015-9 doi:10.1001/jama.280.11.1015

14. Carbone C, Lega F, Prenestini A. La governance delle Aziende Ospedaliero-Universitarie: perché non funziona, come potrebbe migliorare. In: Cantù E. L'aziendalizzazione della sanità in Italia - Rapporto OASI 2010. Milano, IT: EGEA 2010.

15. Roversi Monaco MG. Università e Servizio Sanitario Nazionale: L'azienda ospedaliera universitaria. Padova, IT: CEDAM 2011.

16. Leardini C, Sala GA, Campedelli B. L'azienda ospedaliera universitaria integrata. Un modello di integrazione tra attività di cura, ricerca e formazione. Milano, IT: Franco Angeli 2015.

17. Barsanti S, Giovanelli S, Nuti S. Come valutare la ricerca nelle Aziende Ospedaliero-Universitarie? Metodi e proposte operative. *Mecosan* 2014;90(1):9-32 doi: 10.3280/MESA2014-090002

18. Bevan G, Rutten F. The organisation and functions of University Hospital in different countries. *Finance Accountability & Management* 1987;3(2):77-115.

19. Bevan G. The medical service increment for teaching (SIFT): A £400m anachronism for the English NHS?. *BMJ* 1999;319:908. doi: http://dx.doi.org/10.1136/bmj.319.7214.908

20. 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. doi: 10.1007/s10997-012-9218-5

21. Nuti S. La valutazione della performance in sanità. Bologna, IT: il Mulino 2008.

22. Nuti S, Vola F, Bonini A, et al. Making governance work in the healthcare sector: evidence from a "natural experiment" in Italy. *Health Econ Policy Law* 2015;11(1):17-38. doi: 10.1017/S1744133115000067

23. Nuti S, Vola F, Amat P. Evaluating the Network Healthcare System Performance: A.P. Bolzano, A.P. Trento, Apulia, Basilicata, Calabria, Emilia-Romagna, Friuli Venezia Giulia, Lazio, Liguria, Lombardy, Marche, Sardinia, Tuscany, Umbria, Veneto. 2014 Results of the Italian Regional Collaborative. Pisa, IT: ETS 2016.
24. *Inter-Regional Performance Evaluation System* (IRPES) website: http://performance.sssup.it/netval [Accessed 25 January 2016]
25. *Italian National Outcome Evaluation Programme* (NOEP) website: http://95.110.213.190/PNEed15/index.php [Accessed 25 January 2016]
26. Ash AS, Fienberg SE, Louis TA, et al. Statistical Issue in Assessing Hospital Performance. Quantitative Health Sciences Publications and Presentations. Centers for Medicare and Medicaid Services 2012. http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Downloads/Statistical-Issues-in-Assessing-Hospital-Performance.pdf [Accessed 25 January 2016]
27. Iezzoni LI. Risk Adjustment for measuring healthcare outcomes, 2nd edition. Chicago, IL: Health Administration Press 1997.
28. Campbell SM, Braspenning J, Hutchinson A, et al. Research Methods used in developing and applying quality indicators in primary care. *Qual Saf Health Care* 2002;11:358-64. doi:10.1136/qhc.11.4.358
29. Ibraihm JE, Performance indicators from all perspectives. *Int J Qual Health Care* 2001 Dec;13(6):431-2. doi: dx.doi.org/10.1093/intqhc/13.6.431 [Published Online First: 1 December 2001].
30. Garcia-Lacalle J, Martin E. Rural vs urban hospital performance in a 'competitive' public health service. *Soc Sci Med* 2010 Sep;71(6):1131-40 doi:10.1016/j.socscimed.2010.05.043 [Published Online First: 19 June 2010].
31. Mann HB, Whitney DR. On a test of whether one of 2 random variables is stochastically larger than the other. *Annals of Mathematical Statistics* 1947;18:50-60 http://projecteuclid.org/euclid.aoms/1177730491 [Accessed 26 January 2016]
32. Lehmann EL. Nonparametrics: Statistical Methods Based on Ranks. *Journal of Applied Mathematics and Mechanics* 1977;57(9):562 doi: 10.1002/zamm.19770570922
33. Zimmerman DW. Comparative power of Student t test and Mann-Whitney U test for unequal sample sizes and variances". *J Exp Educ* 1987;55:71-174.
34. Siegel S, Castellan NJJ. Nonparametric statistics for the behavioral sciences. New York, NY: McGraw-Hill 1988.

35. Hollander M, Wolfe DA. Nonparametric Statistical Methods, Solutions Manual, 2nd Edition. Chichester, UK: Wiley and Sons Ltd 1999.

36. Levene H. Robust testes for equality of variances. In: Olkin I. Contributions to Probability and Statistics. Palo Alto, CL: Stanford University Press 1966:278-292.

37. Brown MB, Forsythe AB. Robust tests for the equality of variances. *J Am Stat Assoc* 1974;69:364–67. doi:10.2307/2285659

38. Murante AM, Seghieri C, Brown A, et al. How do hospitalization experience and institutional characteristics influence inpatient satisfaction? A multilevel approach. *Int J Health Plann Manage* 2014;29(3):e247-60 doi:10.1002/hpm.2201 [Published Online First: 1 July 2014].

39. Dimick JB, Cowan JA, Colletti LM et al. Hospital Teaching Status and Outcomes of Complex Surgical Procedures in the United States. *Arch Surg* 2004;139(2):137-141. doi:10.1001/archsurg.139.2.137

40. Nuti S, Marcacci L. The Performance Evaluation System of Health Care in Tuscany – Report 2010. Calenzano, IT: Edizioni del Poligrafico Fiorentino 2012.

SUPPLEMENTARY FILE

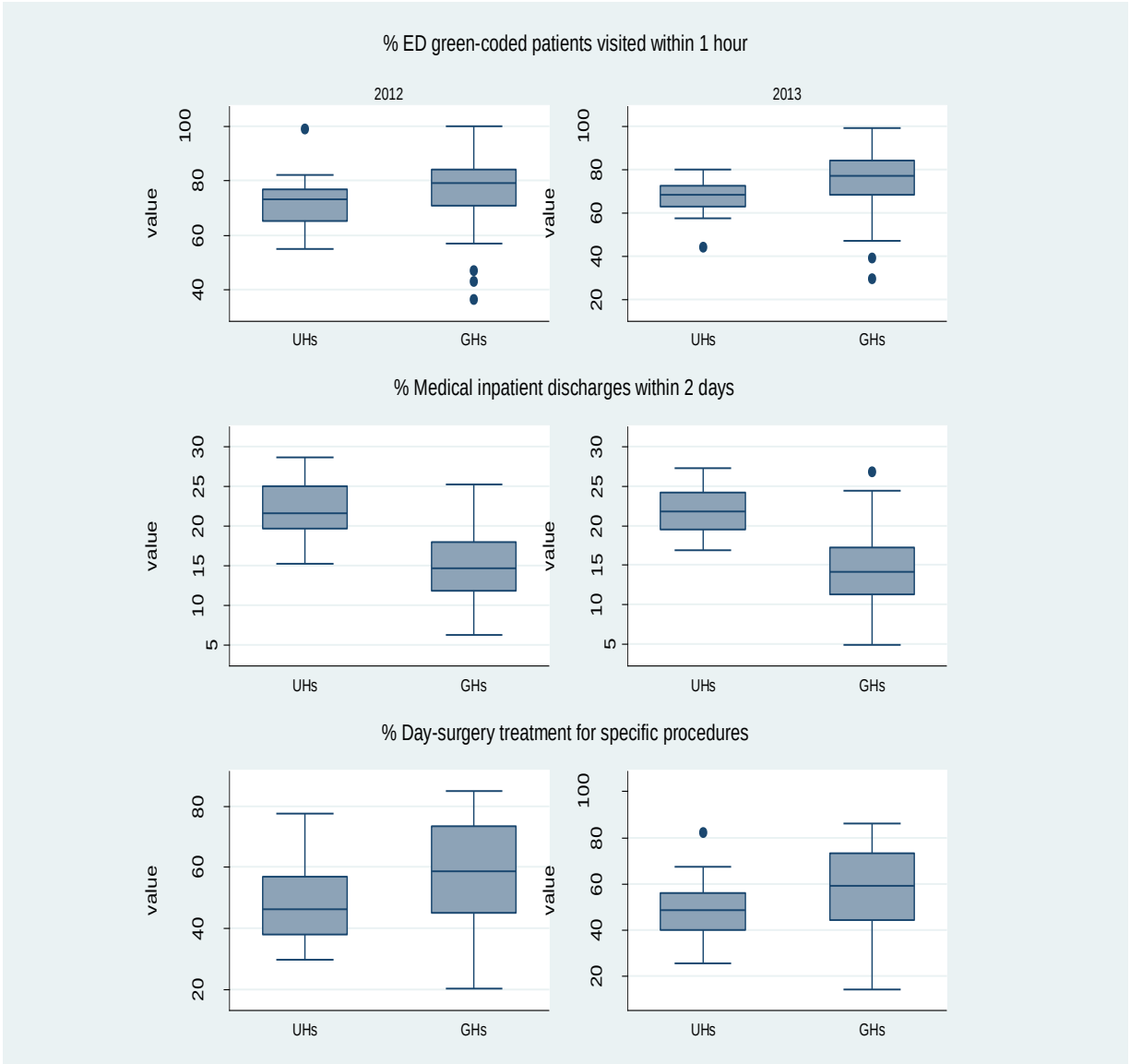
APPENDIX I – Complete list of hospitals included in the analysis and number of observations included for each indicators

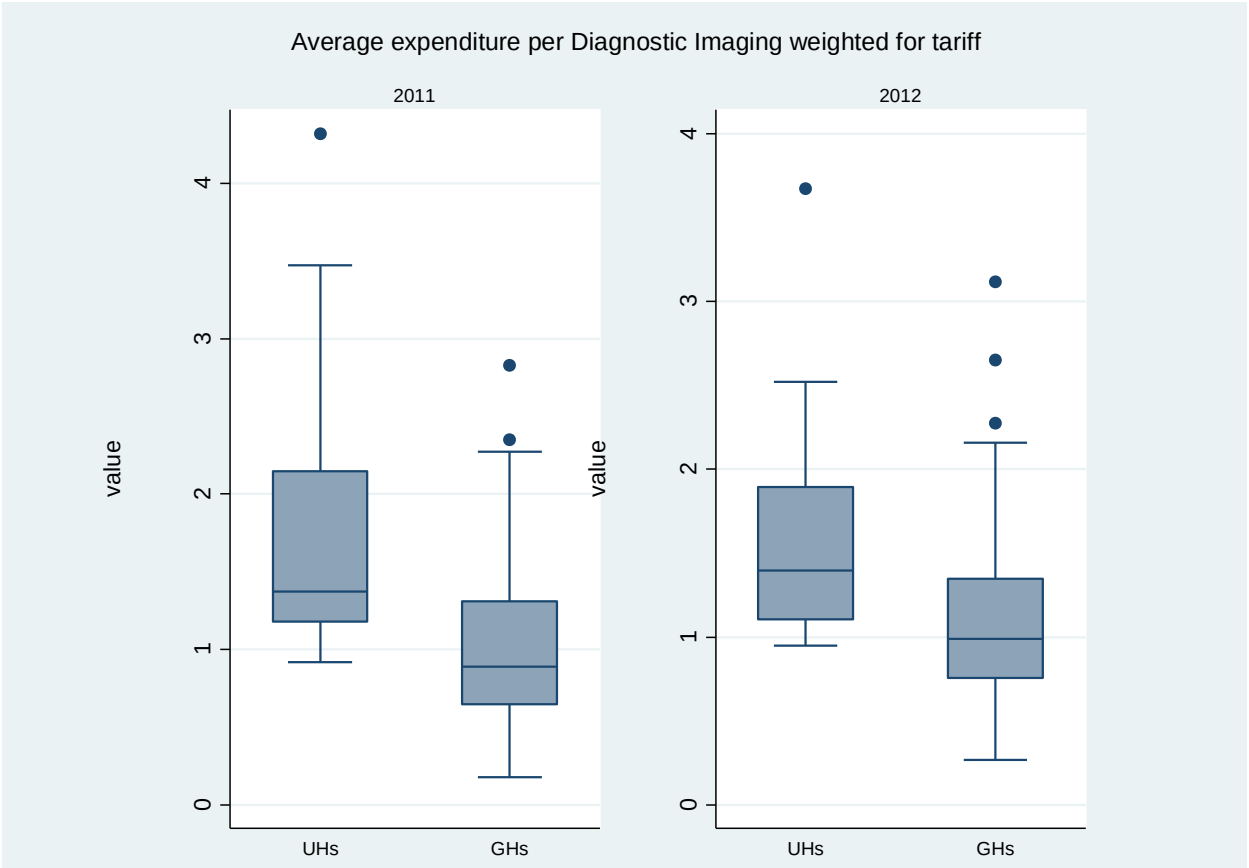
IRPES Indicators	UHs	GHs	Total Hospitals
2011			
Average cost per weighted case	12	38	50
Average expenditure per Diagnostic Imaging weighted for tariff	12	34	46
2012			
Relative stay index: case-mix adjusted differential average Length of Stay - LOS days	15	73	88
Percentage of medical discharges with LOS > ministerial threshold for patients over 65	15	72	87
Percentage of ED green-coded patients visited within 1 hour	15	69	84
Percentage of ED patients referred for hospital admission with ED length of stay <= 8 hours	15	70	85
Percentage of medical inpatient discharges within 2 days (National Healthcare Agreement 2010)	15	73	88
Percentage of day-surgery treatment for specific procedures (National Healthcare Agreement 2010)	15	71	86
Percentage of patients leaving ED against/without medical advice	15	70	85
Percentage hospitalized patients leaving against medical advice	15	73	88
Average cost per weighted case	14	59	73
Average expenditure per Diagnostic Imaging weighted for tariff	14	57	71
2013			
Relative stay index: case-mix adjusted differential average Length of Stay - LOS days	15	73	88
Percentage of medical discharges with LOS > ministerial threshold for patients over 65	15	72	87
Percentage of ED green-coded patients visited within 1 hour	14	71	85
Percentage of ED patients referred for hospital admission with ED length of stay <= 8 hours	15	71	86
Percentage of medical inpatient discharges within 2 days (National Healthcare Agreement 2010)	15	73	88
Percentage of day-surgery treatment for specific procedures (National Healthcare Agreement 2010)	15	71	86
Percentage of patients leaving ED against/without medical advice	15	70	85
Percentage hospitalized patients leaving against medical advice	15	72	87

NOEP Indicators	UHs	GHs	Total Hospitals
2012			
1.AMI: 30-day mortality	17	111	128
3.AMI without PTCA: 30-day mortality	14	85	99
4.AMI with PTCA within 2 days: 30-day mortality	15	37	52
5.AMI with PTCA after 2 days: 30-day mortality	13	20	33
88.AMI: 1-year mortality	16	113	129
89.AMI: MACCE after 1 year	16	113	129
14. Isolated Aortocoronary Bypass: 30-day mortality	13	5	18
15. Congestive heart failure: 30-day mortality	18	153	171
18. Ischemic Stroke: 30-day mortality	16	112	128
21. COPD exacerbation: 30-day mortality	16	144	160
37. Proportion of Caesarean Section	15	127	142
38. Femur Fracture: 30-day mortality	16	119	135
42. Femur Fracture: Percentage of operations carried out within 2 days	16	121	137
83. Colon cancer surgery: 30-day mortality	16	95	111
22. COPD: 30-day re-admission	16	137	153
35. Valvuloplasty or heart valve replacement: 30-day mortality	13	5	18
19. Ischemic Stroke: 30-day re-admission	16	104	120
2013			
AMI: 30-day mortality	16	111	127
AMI without PTCA: 30-day mortality	14	77	91
AMI with PTCA within 2 days: 30-day mortality	14	41	52
AMI with PTCA after 2 days: 30-day mortality	13	18	31
AMI: 1-year mortality	16	109	125
AMI: MACCE after 1 year	16	109	125
Isolated Aortocoronary Bypass: 30-day mortality	12	4	16
Congestive heart failure: 30-day mortality	18	172	190
Ischemic Stroke: 30-day mortality	15	106	121
COPD exacerbation: 30-day mortality	16	134	150
Proportion of Caesarean Section	15	124	139
Femur Fracture: 30-day mortality	15	118	133
Femur Fracture: Percentage of operations carried out within 2 days	15	118	133
Colon cancer surgery: 30-day mortality	16	97	113
COPD: 30-day re-admission	16	133	149
Valvuloplasty or heart valve replacement: 30-day mortality	13	5	18
Ischemic Stroke: 30-day re-admission	15	101	116

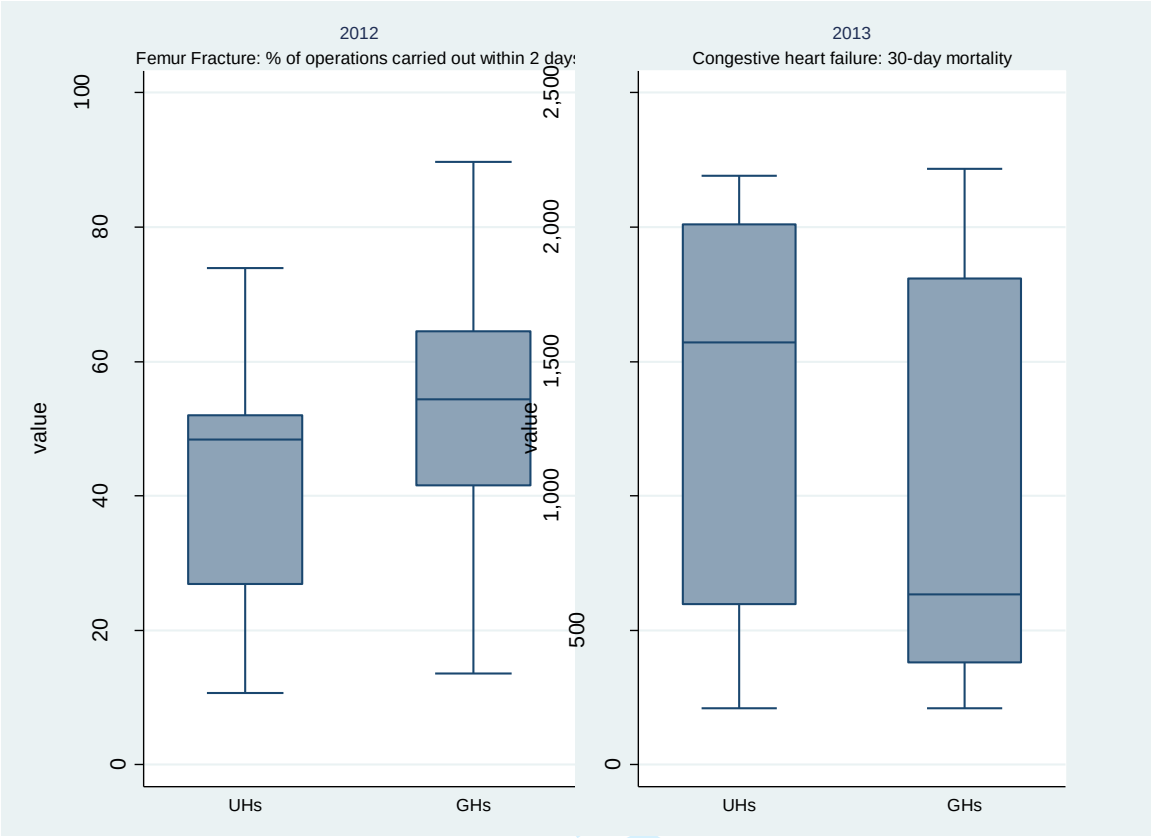
APPENDIX II – Box plots for IRPES and NOEP indicators with significant differences between UHs and GHs at the Mann-Whitney *U* test

A] IRPES indicators




Average expenditure per Diagnostic Imaging weighted for tariff
2011
2012
value
value
0
1
2
3
4
UHs
GHs

B] NOEP indicators



STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	Recommendation
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract **[Within the title, page 1]**
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found **[Within the abstract, page 2]**
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported **[within the Introduction, page 3 and the Background, pages 3-4]**
Objectives	3	State specific objectives, including any prespecified hypotheses **[within the Introduction, page 3]**
Methods		
Study design	4	Present key elements of study design early in the paper **[within the paragraph 3.3, pages 8-9]**
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection **[within the paragraph 3.1, page 6]**
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants **[within the paragraph 3.1, page 6]**
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable **[within the paragraph 3.2, pages 7-8]**
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group **[within the paragraph 3.1, page 6]**
Bias	9	Describe any efforts to address potential sources of bias **[within the paragraph 3.2, page 8]**
Study size	10	Explain how the study size was arrived at **[within the paragraph 3.2, page 8, and Appendix I]**
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why **[within the paragraph 3.3, pages 8-9, and paragraph 3.1, page 6]**
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding **[within the paragraph 3.3, pages 8-9]**
		(*b*) Describe any methods used to examine subgroups and interactions **[]**
		(*c*) Explain how missing data were addressed **[]**
		(*d*) If applicable, describe analytical methods taking account of sampling strategy **[within the paragraph 3.3, pages 8-9]**
		(*e*) Describe any sensitivity analyses **[]**
Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed **[]**
		(b) Give reasons for non-participation at each stage **[]**
		(c) Consider use of a flow diagram **[]**
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders **[]**

		(b) Indicate number of participants with missing data for each variable of interest **[]**
Outcome data	15*	Report numbers of outcome events or summary measures **[Tables 3,4,5,6, pages 10-12]**
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included **[]**
		(*b*) Report category boundaries when continuous variables were categorized **[]**
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period **[]**
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses **[]**
Discussion		
Key results	18	Summarise key results with reference to study objectives **[within Discussion, pages 13-15]**
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias **[within Discussion, pages 14-15]**
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence **[within Discussion, pages 13-15]**
Generalisability	21	Discuss the generalisability (external validity) of the study results **[within Discussion, pages 14-15]**
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based **[within Funding, page 17]**

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.