Is there an association between the quality of hospitals’ research and their quality of care?

Joan Pons, Carolina Sais1, Carles Illa1, Raúl Méndez2, Eduard Suñen2, Mercè Casas1, Jordi Camí2
Catalan Agency for Health Technology Assessment and Research, Department of Health, Barcelona, Spain; 1IASIST, Barcelona, Spain; 2Parc de Recerca Biomèdica de Barcelona, Barcelona, Spain

Objective: It is often claimed that hospitals that are leaders in biomedical research provide higher health care quality, or vice versa. Although several studies have shown a relationship between teaching status and quality of care, none has analysed the association between research output and hospital outcomes. Our aim was to determine whether there is a relationship between bibliometric measures of research output in acute hospitals and hospital mortality for two common cardiac conditions.

Methods: A cross-sectional analysis of secondary data of in-hospital risk-adjusted mortality for congestive heart failure and acute myocardial infarction (2002–2004) and several bibliometric measures of publications (1996–2004) in cardiovascular disease. The setting was 50 acute Spanish public hospitals, voluntarily participating in an external quality initiative, with more than 30 medical cases of congestive heart failure and acute myocardial infarction per year, and more than five citable papers in the field of heart disease. Spearman’s rho non-parametric correlation coefficient was used to assess association.

Results: There was a low-to-moderate negative correlation between the risk-adjusted mortality ratio and the weighted citations ratio: −0.43 (95% CI −0.17 to −0.63) for congestive heart failure and −0.37 (−0.10 to −0.59) for acute myocardial infarction. Teaching status and the technological level of the hospital had a stronger correlation with hospital mortality.

Conclusions: Measures of research output could be considered for incorporation into comparisons of the quality of hospitals. A weighted citations ratio is the most suitable measure of research output, but more research is needed on the interplay between research and practice as complementary ways of developing medical knowledge.


Introduction

In addition to providing health care, many hospitals also undertake research and teaching activities. These activities are closely related, but have different aims and measures. In recent years performance measures have been developed to enable comparisons between hospitals. Attaining a high rank in these scores gives a hospital prestige that helps attract staff, students, researchers, patients, grants and health care insurers. Although their validity is improving, these data have limitations and even though they are publicly available in some countries, their real importance in stakeholder decision-making is unclear.

Health care quality is traditionally assessed through structure, process and outcome indicators. In the US, annual reports of hospital performance provide national benchmarks differentiated by hospital level and specialty. Neither teaching nor research activities are included in these reports which most frequently focus on risk-adjusted mortality ratios as the main outcome measure.

The relationship between hospital teaching status and health care quality has been widely analysed. Overall, teaching hospitals seem to have more favourable outcomes than non-teaching hospitals, a difference that...
remains true for a range of locations, conditions, and populations.\textsuperscript{7} Quality measures of both process and outcome of care for congestive heart failure (CHF) and acute myocardial infarction (AMI) have also been shown to be better in academic medical centres,\textsuperscript{8,9} although the strength of the relationship between process and outcomes measures as well as the influence of teaching status is debated.\textsuperscript{10,11}

Peer review is the most widely used system for assessing the quality of biomedical research though it has widespread critics in view of its subjective nature. More objective measurement of research quality is complex. Common quantitative measures include the number of research grants (input) won by individual researchers, groups or institutions, and the bibliometric analysis of scientific publications or other research output measures (e.g. patents). The relevance of scientific literature and its added value as new knowledge is difficult to assess, especially in the short-term. Citations analysis measures knowledge use and the contribution of published articles to subsequent publications, and is considered more accurate and less misleading than relying on journal impact factors.\textsuperscript{12,13}

However, it is well known that article citations have a skewed distribution over time.\textsuperscript{14} Also article duplication,\textsuperscript{15} mass media impact, and Matthew's effect\textsuperscript{16} (eminent scientists get disproportionately great credit for their contribution to science) may influence and distort citation frequency. While there is consensus on the unreliability of journal impact factors\textsuperscript{17} as a surrogate for an article's scientific contribution, there is less agreement on the most appropriate indicator of research quality.\textsuperscript{18,19}

While it would seem obvious that excellent health care should relate to similar excellence in biomedical research and 	extit{vice versa}, there has been surprisingly little analysis of the relationship between research output and health care performance. In this study, we test the hypothesis that a hospital's relative performance on health care quality indicators is related to its relative performance in some measures of biomedical research. The hypothesis was tested for cardiovascular conditions in Spanish public hospitals.

**Methods**

Health care quality data for cardiovascular conditions were provided by IASIST, a private company that has produced annual public reports since 2000 on the performance of voluntarily participating hospitals in Spain.\textsuperscript{20} Benchmarks for cardiovascular conditions were released in 2003 and in 2006.\textsuperscript{21,22} Confidential agreements preclude the identification of participating hospitals. Only those hospitals in the top rankings are publicly known.

IASIST’s main data source is the hospital Basic Minimum Data Set (MDS). The MDS contains individual patient diagnostic and procedure details on all discharges and in-hospital deaths. Hospitals are classified into three different groups: district hospitals; referral hospitals with specialist training programs; and high-tech teaching hospitals with invasive cardiac procedures (angioplasty and/or cardiac surgery). The cardiac pathologies selected for this study were those with the highest admission rate (3.5% of all hospital admissions in patients \( \geq 17 \) years old): acute myocardial infarction (AMI) and congestive heart failure (CHF). Care for these patients is mainly provided in the cardiology and internal medicine departments of acute hospitals.

Diagnostic codes which include CHF and AMI as a main diagnosis were identified using the International Classification of Diseases (ICD-9-CM). Hospitals with more than 30 admissions annually for medical treatment of either CHF or AMI were included. The period of study was three years (2002–2004). One hundred and twenty-six hospitals (88%) of 143 centres participated.\textsuperscript{22} More than 80% of hospitals participating in the generic health care quality assessment were contracted by the Spanish health system, representing nearly 50% of public sector hospitals in Spain. Only public hospitals are included in this study.

The main outcome measure was the in-hospital risk-adjusted mortality ratio based on the quotient between observed and expected deaths. A logistic predictive model of mortality was developed taking into account the following independent variables: age; risk of death from main diagnosis; risk of death from the most severe secondary condition; risk of death of the most severe procedure; and type of centre. The adjustment systems used show a high discriminant value (area under the receiver-operator curve [ROC]) and goodness of fit (calibration) among deciles of observed and predicted mortality.\textsuperscript{21} Hospitals included in the Benchmark group in cardiovascular disease had 12% and 14% fewer deaths than predicted from their case-mix for CHD and AMI, respectively.\textsuperscript{22}

To assess research output, we aimed to identify all publications related to cardiovascular disease in which at least one of the authors was affiliated to a Spanish hospital. The data were taken from the \textit{Spanish National Citations Report (NCR)} for the period 1996–2004. The Spanish NCR database contains those documents considered citable (original articles, reviews, and proceedings), originating in Spain. It is derived from publication and citation statistics compiled by Thomson Scientific in the Web of Knowledge. Documents categorized under the journals’ heading Cardiovascular system, were selected for the analysis and identification of research groups in Spain. Since articles related to the cardiovascular system may be published in general and other journals, we undertook an additional search.
of the database using 3,000 MeSH terms identified from Pub Med’s Thesaurus. This process was also supplemented by manual screening to avoid duplications arising from the use of different names for institutions and authors.23,24 Centres which published less than six documents during the period 1996–2004 were excluded on the grounds that this represents a minimum level of continuity of research production. From the 194 health care centres identified, including those in primary care, 94 had less than five and 45 had only one citable document during the nine-year period.

The bibliometric measures of research output used and their meaning are displayed in Table 1. Self-citation was included in all cases. We also obtained an institutional h index because, in spite of suggestions that this is mainly used to assess individual scientific research output,25 there is growing experience with its application for groups and institutions.26 Finally, and according to the method described by Lewison,27 the journal research level (ranging from basic to clinical) was derived for each institution.

The relevant data from the two databases were combined to produce a cross-tabulation of health care performance against research output. For the statistical analysis, we used the chi-squared test for dichotomous variables and the non-parametric Kruskal-Wallis test for continuous ones. To assess the relationship between health care outcomes and research output, we used the simple, non-parametric, rank correlation coefficient Spearman’s rho. Confidence intervals (CI) were constructed assuming similar distributions of Spearman’s rho and Pearson correlation coefficient.28 The Bonferroni correction method was used to adjust the statistical significance of p values to allow for the multiple comparisons performed.

Results

After pooling of data from the two databases, a total of 50 hospitals met the inclusion criteria. There were no statistically significant differences between district, referral and high-tech hospitals, except for the number of beds and cases (Table 2). Statistically significant
differences in bibliometric measures were related to the number of citable documents and citations, the weighted citations ratio, and the institutional \( h \) index.

There was significant association between the risk-adjusted mortality ratio and the weighted citations ratio for CHF and AMI (Table 3). For CHF, the Spearman’s rho correlation coefficient was \(-0.43\) (95% CI \(-0.17\) to \(-0.63\)) between the risk-adjusted mortality ratio and weighted citations ratio. For AMI, the Spearman’s rho was \(-0.37\) (95% CI \(-0.10\) to \(-0.59\)). There was a statistically significant inverse relationship (\( p < 0.008 \)) between these two measures implying that as the weighted citations ratio increases (more citations than the average number of citations in cardiovascular disease from authors affiliated to Spanish public hospitals), the risk-adjusted mortality ratio decreases (observed mortality less than expected). The weighted citations ratio was also associated with volume and crude mortality in CHF, and average citations per publication was correlated with adjusted mortality in AMI.

These associations showed different profiles depending on the condition analysed and the hospital type (Table 4). In the case of CHF, the inverse statistically significant association between risk-adjusted mortality ratio and weighted citations ratio was present when all types of hospital were considered. The highest correlation was observed in district hospitals (\( \rho = -0.86; 95\% \text{ CI } -0.30 \) to \(-0.98; p = 0.014 \)), in spite of their small sample size (\( n = 7 \)), and in high-tech centres (\( \rho = -0.40; 95\% \text{ CI } -0.05 \) to \(-0.66; p = 0.027 \)). The association was observed irrespective of teaching status (Table 4).

For AMI, when all centres were analysed, the correlation was lower, but still statistically significant. The association was only significant for high-tech hospitals (\( \rho = -0.42; 95\% \text{ CI } -0.08 \) to \(-0.67 \)). However, for AMI, teaching status was significantly associated with the risk-adjusted mortality ratio (Table 4).

For AMI, when all centres were analysed, the correlation was lower, but still statistically significant. The association was only significant for high-tech hospitals (\( \rho = -0.42; 95\% \text{ CI } -0.08 \) to \(-0.67 \)). However, for AMI, teaching status was significantly associated with the risk-adjusted mortality ratio (Table 4).

### Discussion

**Main findings**

We found of a statistically significant, low-to-moderate negative rank correlation between in-hospital mortality for AMI and CHF and some bibliometric measures of research output in cardiovascular diseases in Spanish hospitals. Our results suggest that the weighted citations ratio, an adjusted bibliometric measure that takes into account research output of peers in the same field, could be a useful measure of research production, especially when comparing hospitals in the same country.

It is possible that hospitals willing to participate voluntarily in external performance assessments are also the most likely to be open to quality improvement. This openness to self-evaluation and to self-criticism to improve their own practice, and to translate new research findings into practice, cannot be exclusive to teaching or high-tech hospitals. Even small hospitals can do well and learn from high-performing organizations; in other words, there is room to improve health quality and research quality.

### Table 3  Rank correlation coefficients (Spearman’s rho) between bibliometric measures and outcome measures

<table>
<thead>
<tr>
<th></th>
<th>Number of documents</th>
<th>Number of citations</th>
<th>Average citations per publication</th>
<th>International collaboration</th>
<th>Weighted citations ratio</th>
<th>Institutional h index</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Congestive Heart Failure</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Volume per year rho</td>
<td>0.24</td>
<td>0.26</td>
<td>0.20</td>
<td>0.24</td>
<td>0.38</td>
<td>0.24</td>
</tr>
<tr>
<td>p value</td>
<td>0.091</td>
<td>0.064</td>
<td>0.158</td>
<td>0.088</td>
<td>0.007</td>
<td>0.091</td>
</tr>
<tr>
<td>Crude mortality rho</td>
<td>0.07</td>
<td>-0.03</td>
<td>-0.11</td>
<td>-0.10</td>
<td>-0.33</td>
<td>0.08</td>
</tr>
<tr>
<td>p value</td>
<td>0.653</td>
<td>0.847</td>
<td>0.458</td>
<td>0.490</td>
<td>0.018</td>
<td>0.602</td>
</tr>
<tr>
<td>Adjusted mortality rho</td>
<td>-0.08</td>
<td>-0.18</td>
<td>-0.26</td>
<td>-0.15</td>
<td>-0.43</td>
<td>-0.09</td>
</tr>
<tr>
<td>p value</td>
<td>0.599</td>
<td>0.199</td>
<td>0.067</td>
<td>0.289</td>
<td>0.002</td>
<td>0.550</td>
</tr>
<tr>
<td><strong>Acute Myocardial Infarction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Volume per year rho</td>
<td>0.15</td>
<td>0.14</td>
<td>0.02</td>
<td>0.03</td>
<td>0.06</td>
<td>0.19</td>
</tr>
<tr>
<td>p value</td>
<td>0.289</td>
<td>0.318</td>
<td>0.912</td>
<td>0.827</td>
<td>0.672</td>
<td>0.189</td>
</tr>
<tr>
<td>Crude mortality rho</td>
<td>0.25</td>
<td>0.21</td>
<td>0.06</td>
<td>0.15</td>
<td>-0.03</td>
<td>0.20</td>
</tr>
<tr>
<td>p value</td>
<td>0.077</td>
<td>0.153</td>
<td>0.665</td>
<td>0.287</td>
<td>0.863</td>
<td>0.174</td>
</tr>
<tr>
<td>Adjusted mortality rho</td>
<td>0.00</td>
<td>-0.17</td>
<td>-0.39</td>
<td>-0.10</td>
<td>-0.37</td>
<td>-0.12</td>
</tr>
<tr>
<td>p value</td>
<td>1.00</td>
<td>0.250</td>
<td>0.005</td>
<td>0.501</td>
<td>0.008</td>
<td>0.411</td>
</tr>
</tbody>
</table>

\( n = 50 \)
care quality and to produce relevant research for all
types of hospital.

The widely held assumption that good performers in
terms of health care quality also have a good quality
research record seems to be valid. However, in spite of
the statistical significance of the correlation, its relatively
low strength, coupled with the small, non-random
sample of hospitals, call for some caution in interpreting
these findings. As expected, all correlations were nega-
tive, but were generally low-to-moderate and only
reached higher values in the case of CHF in district,
non-teaching hospitals. For AMI, the highest corre-
lation coefficients appeared in teaching and high-tech
hospitals. We cannot explain the differences between
the two conditions except by hypothesizing a link
between the potential role of some other structural or
process of care variables in the management of these
conditions and their differential availability in different
hospitals. Possible variables could include thrombolytic
therapy and ‘door-to-needle’ time or the more appropri-
ate use of other diagnostic and therapeutic procedures.

Medical research and practice are two overlapping
fields, but with different enterprises, cultures, missions,
goals, and reward systems. In spite of being undertaken
under the same roof, often performed by the same
doctor or applied to the same patient, their quality is
not necessarily directly interrelated. Studies in cardiac
and cancer care support the hypothesis that patients
treated at hospitals participating in clinical trials have
better outcomes than patients treated at non-
participating hospitals.29 This could partly explain the
link between research and quality of health care.
However, better standardization of the process of care,
more trained personnel and professional involvement
in active policies to ensure continuous quality improve-
ment of practice could also be explanatory factors.

Limitations

We were restricted in the availability of outcome data to
the two most common cardiac diseases (AMI and CHF),
whereas our bibliometric analysis included research
covering the whole field of cardiovascular medicine.
We cannot exclude the possibility that a more accurate
match between conditions treated and condition-
specific research output would provide different
results. The approach to link a paper to a hospital
through the affiliation of, at least, one of the authors
also has limitations not only when authors have multiple
affiliations, but also when the link of an author to the
institution is tenuous.

The main weakness stems from the data sources
which were collected for other purposes, over different
time periods. In addition, neither the discharge
dataset from which IASIST derived its data, nor the
Reuter Thomson Scientific (Web of Knowledge) databases
are exhaustive. Any risk-adjustment model, in order to
be parsimonious, cannot include all patients’ diagnoses
and procedures. However, statistical analysis has shown
a good discrimination and calibration of the mortality
predictive model used by IASIST.20 Furthermore,
several US studies demonstrate that these popular hos-

dial profiling systems (The Top 20, The Best Hospitals),
in spite of their limitations, produce results in accord-
ance with more elaborate data sources used to assess
health care performance in AMI.30 Our small sample
size (around 21% of public acute hospitals in Spain)
and the non-random hospital selection further limit
the scope for inferences and generalizations.

Equally one should bear in mind that health sciences
research production in Spain is broader than what is
collected in Reuters Thomson Scientific database. For
instance, from the 320 medical journals included in the
Spanish Medical Index, only 44 are in the Medline
data-base, and 14 in the Reuters Thomson Scientific Citations
Index.31 However, only the latter database, and its
product, the Spanish National Citation Report, was used
when research groups were identified.23 Therefore,
some research is missing. We cannot exclude the possi-

bility that the use of other bibliometric databases for
citation analysis (e.g. Scopus, Google Scholar) would
produce different results, given their different journal
coverage, time-frame, and accuracy.32

Implications

As well as not being able to identify the mediating vari-
ables between health care quality and research quality,
we cannot identify the direction of causality which may
be uni-directional or in both directions simultaneously.
It is possible that clinical judgment is shaped and goes
on being refined by the continued interaction between
experience (practice) and research (new knowledge).33,34

Further research is necessary to see whether these find-
ings can be generalized to the link between other
measures of hospital health care quality and bibliometric
research measures, and to other settings. There are few
studies in this field. A Spanish study of the relationship
between reputation of hospitals and scientific production
in four specialties showed a positive correlation. However,
health care quality was based in the subjective opinion of a
small sample of members of the Spanish scientific societies
involved.35

An adjusted bibliometric measure of research pro-
duction could plausibly be incorporated as a com-
plementary indicator in the comparative evaluation of
quality between Spanish hospitals.

Acknowledgements

We thank Elinor Thompson for her comments and editing; Marta
Millaret for searching and managing references; and Joan Rodés,
Professor Emeritus, University of Barcelona, who inspired the study.
References

1 Anderson G, Steinberg E, Heyssel R. The pivotal role of the academic health center. Health Aff (Millwood) 1994;13:146–58
11 Seglen PO. Why the impact factor of journals should not be used for evaluating research. BMJ 1997;314:498–502
13 Garfield E. The history and the meaning of the journal impact factor. JAMA 2006;295:90–3
24 Hirsch J. An index to quantify an individual’s scientific research output. Proc Natl Acad Sci USA. 2005;102:16569–72
25 Van Raan AFJ. Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgement for 147 chemistry research groups. Scientometrics 2006;67:491–502
26 Lewison G, Paraje G. The classification of biomedical journals by research level. Scientometrics 2004;60:145–57
30 Figueredo-Gasparti E. Valoración curricular de las publicaciones científicas. Med Clin (Barc) 2005;125:661–5
33 Woolf SH. The meaning of translational research and why it matters. JAMA 2008;299:211–13
34 Asenjo MA, Bertrán MJ, Guinovart C, Llach M, Prat A, Trilla A. Análisis de la reputación de los hospitales españoles: relación con su producción científica en cuatro especialidades. Med Clin (Barc) 2006;126:768–70