Document Version
Early version, also known as pre-print

Link to publication record in King's Research Portal

Citation for published version (APA):
Beaussier, A-L., Demeritt, D., Rothstein, H., & Griffiths, A. (Accepted/In press). Steering by their own lights: Why regulators across Europe use different indicators to measure healthcare quality. HEALTH POLICY.

Citing this paper
Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher’s definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher’s website for any subsequent corrections.

General rights
Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the Research Portal

Take down policy
If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 25. Mar. 2020
Steering by their own lights:
How European regulators use indicators to measure healthcare quality

Anne-Laure Beaussier¹, David Demeritt²*, Alex Griffiths³ and Henry Rothstein²

1. Centre de Sociologie des Organisations (CSO), Sciences Po-CNRS, 19 rue Amélie, 75007 Paris, France
2. Department of Geography, King’s College London, Strand, London WC2R 2LS
3. Data Science Directorate, Statica Research, London, SE22 9PN

* Correspondence to david.demeritt@kcl.ac.uk

ABSTRACT

Despite widespread faith that quality indicators are key to healthcare improvement and regulation, surprisingly little is known about what is actually measured in different countries, nor how, nor why. To address that gap, this article compares the official indicator sets—comprising some 1100 quality measures—used by statutory hospital regulators in England, Germany, France, and the Netherlands. The findings demonstrate that those countries' regulators strike very different balances in: the dimensions of quality they assess (e.g. between safety, effectiveness, and patient centredness); the hospital activities they target (e.g. between clinical and non-clinical activities and management); and the 'Donabedian' measurement style of their indicators (between structure, process and outcome indicators). We argue that these contrasts reflect: i) how the distinctive problems facing each country's healthcare system create different 'demand-side' pressures on what national indicator sets measure; ii) how the configuration of national healthcare systems and governance traditions create 'supply-side' constraints on the kinds of data that regulators can use for indicator construction. Our analysis suggests fundamental differences in the meaning of quality and its measurement across countries that are likely to impede international efforts to benchmark quality and identify best practice.

HIGHLIGHTS:

- Indicator sets differ in how they define, measure, and assess healthcare quality
- National sets shaped by varying governance traditions and healthcare system configuration
- Targeting of quality dimensions and hospital activities shaped by system-specific 'demand-side' pressures
- Measurement styles shaped by 'supply-side' constraints on data access and indicator construction
- International benchmarking is easier when healthcare systems and governance traditions are similar

Forthcoming in Health Policy
Accepted 24 February 2020
INTRODUCTION

Lord Kelvin (1) famously opined that only “when you can measure what you are speaking about and express it in numbers” can you “know something about it”. No doubt he would have applauded how healthcare quality, long regarded as too ineffable to define (2), is now subject to pervasive measurement to support everything from quality assurance and improvement to patient choice and payment by results. Indeed, quantitative indicators are now central to an international “quality movement” (3–7), which has emerged over the last 25 years in response to spiralling costs, safety scandals and demands for more responsive and accessible care. In England’s National Health Service (NHS) for example, the number of performance indicators has skyrocketed from 70 in 1982 to more than 2000 today (8,9). Likewise in the US the number of healthcare quality indicators endorsed by the National Quality Forum has more than doubled over the last decade to 1078 (10,11).

Measurement may be the first step to improvement (12), but the proliferation of indicators creates its own problems. For one thing, measurement and reporting are costly (10), with one recent study estimating the burden on a major US medical centre at 1% of total revenue (13), despite efforts to rationalise and reduce excessive reporting (4,14). In turn, there are many competing ways of conceptualizing and measuring quality and of selecting, normalising, aggregating, and visualising quality indicators (15–19). This creates difficulties in benchmarking performance—both of individual providers (20–23) and entire healthcare systems (24–26)—and can lead to poor choices by patients, policymakers, and practitioners alike (27).

To improve the consistency, reliability, and validity of care quality metrics, the World Health Organization, European Commission, OECD, and Institute of Medicine have published alternative frameworks for defining key indicator sets (10,28–30). Indeed, the European Commission regards the development of comparable national indicator sets as vital to helping patients exercise their rights to accessing cross-border healthcare (2011/24/EU). However, despite efforts to create universal indicators that could help facilitate convergence and discourage countries from steering by their own lights (31,32), high-level international comparative studies have tentatively pointed to considerable unevenness in the selection and use of indicators across countries (7).

Part of the problem is that international frameworks tend to regard care quality as an objective phenomenon for which universally applicable measures can—and should—be adopted regardless of the institutional contexts and purposes for which it is being assessed. Yet, as, Pollitt et al (8), have observed in relation to the general measurement of healthcare system performance, patterns of indicator adoption and use can depend on the distinctive problems facing different countries’ healthcare systems and governance traditions. For example, Pollitt et al (8) suggest that governments within pluralist political systems face fewer institutional constraints than governments in corporatist political systems built on compromise among the social partners. Certainly, a number of high-level cross-national comparisons have pointed to distinctive national variation in the philosophies and regulatory mandates that underpin indicator use as well as the sources on which indicator sets draw and the purposes to which they are put (33–36).
Beyond high-level observation, however, little is known about international variation in what actually gets measured and how, whether there are any distinctive patterns to that variation, and what might explain such variation. In order to address those lacunae, we undertake the first indicator-by-indicator comparison of the official sets used across advanced healthcare systems, examining four neighbouring EU countries; England, Germany, France and the Netherlands. We focus on the statutory regulation of acute hospital care because that is the area of healthcare provision where international efforts to define and measure quality are most advanced. In so doing, we consider whether and how the availability, design, and selection of quality measures varies and what those patterns reveal about regulatory priorities, institutional barriers to quality monitoring, and fundamental understandings of quality itself. We conclude by reflecting on the opportunities for, and barriers to, future convergence, in ways that could enable meaningful comparison across countries.

METHODS

Our qualitative descriptive study collected and classified the indicators used by regulatory agencies in a sample of European countries to monitor the quality of hospital care in their jurisdictions.

Sample

We selected England, Germany, France, and the Netherlands; four neighbouring EU member states with advanced economies and similarly well-developed but differently structured systems of healthcare and varied governance traditions (37). Their contrasting organisation of payers and providers within their respective healthcare systems and their distinctive regulatory arrangements might be expected to offer different opportunities for indicator construction and create different demands for quality measures. At the same time, all four countries have participated in the OECD Healthcare indicators project (6,28) and initiatives by the World Health Organization (38) and EU (39) to develop standardised quality measures of international health system effectiveness. There are, therefore, good institutional reasons to expect convergence beyond the universal desire to follow best practice in quality measurement.

International comparisons of how regulators in different countries monitor healthcare quality pose considerable methodological difficulties, not least because regulators often operate within complex and nationally distinctive landscapes of state and non-state organisations, such as medical professional associations, clinical disease registries, and insurers that have developed their own healthcare quality indicators for overlapping or different purposes (40). In this paper, however, we address that problem of comparison by restricting our analysis to the official indicator sets used by the supervisory organisations charged by law with monitoring the quality of acute hospital healthcare in each country:

- **England:** Hospital care is almost entirely provided by the single-payer state-run NHS. Healthcare quality is overseen by the Care Quality Commission (CQC), which is a non-departmental public body responsible for regulating the quality of care by all health and
social care providers. As well as licencing and inspecting providers, the CQC can issue regulatory improvement notices and put the management of poor quality hospitals into ‘special measures’ (41). Its enforcement activities are based on inspection findings and analysis of the wealth of performance data routinely collected by NHS England and the Department of Health to inform the administration and financing of the NHS.

- **Germany**: Hospital care is delivered by public and private providers funded by para-public social insurance funds. Healthcare quality is overseen by the *Gemeinsamer Bundesausschuss* (G-BA), which is a Federal joint committee of medical professionals, social insurers, and healthcare providers operating independently of the Ministry of Health. The G-BA sets and monitors quality standards and determines which procedures and providers are eligible for reimbursement. The G-BA also works collaboratively to design quality indicators that draw on hospital quality assurance data collected and published by external contractors, which the G-BA uses to engage in ‘structured dialogue’ with providers if their performance deviates from pre-determined norms.

- **France**: Hospital care is likewise delivered by public and private providers largely funded by social insurance funds. Healthcare quality is overseen by the *Haute Autorité de Santé* (HAS), which is an independent administrative authority responsible for accrediting and certifying the quality of care provided by hospitals, clinics, and other health care facilities. HAS designs indicators in consultation with voluntary and independent health professionals and patients, conducts peer review visits of hospitals on a routine basis, collects data and publishes assessment results online. Enforcement is left to regional health agencies, which were given responsibility for health and social care planning, regulation and enforcement of national health policy priorities by the devolution laws of 2009.

- **The Netherlands**: Hospital care is delivered mostly by private not-for-profit foundations, which have been funded through mandatory, and strictly regulated, private insurance since 2006 (42). Healthcare quality is overseen by the *Inspectie voor de Gezondheidszorg en Jeugd* (IGJ), which is a government inspectorate within the Ministry of Health, Welfare and Sport, responsible for regulating the quality of health and social care and ensuring a level-playing field among providers. The IGJ executes those tasks by licencing, inspecting, and policing hospitals, using a pyramid of compliance tools and sanctions. The IGJ also publishes measures of the quality of care delivered by every Dutch hospital, based on hospital-reported data and indicators designed in collaboration with medical professional organizations and the hospital themselves.

**Indicator definition and data sources**

We define an indicator as a discrete variable providing some nominal, ordinal, or quantitative measure of healthcare quality. Indicators can either be a single measure, such as the number of ‘never events’ recorded in English hospitals (STEISNE in (43)), or they can
combine multiple measures into a ‘composite indicator’, such as the French indicator for the quality of discharge records for psychiatric patients (TDP2 PSY in (44)) which aggregates together 15 discrete fields of information (TDP2 PSY1-15 in (44)).

In 2016 we compiled a database of the hospital quality indicators used at that time by our four supervisory agencies (43–46) (see supplementary dataset). Where indicator lists were published in English, as they were for Germany (45) and England (43), we used those; otherwise we worked with the original listings in Dutch (46) and French (44) and drew on a corpus of 32 background interviews conducted with key informants from the four countries to clarify any uncertainties about particular indicators and help explain differences in indicator selection and use between countries. To ensure consistency and capture the variety and granularity of quality measures, we decomposed ‘composite’ indicators into their constitutive ‘sub-indicators’. In the process we excluded purely administrative measures used to facilitate data collection, e.g. the Dutch measure “Does your hospital perform colorectal surgery?” (17.2.2 in (46)), or enable cross-tabulation of patient survey results by patient condition, e.g. the French measure “did you need help with routine activities (washing, dressing, eating, ...)?” (E-SATIS20 in (44)).

In total, our dataset of disaggregated indicators comprises 1,100 indicators: England (226); Germany (431); France (260); Netherlands (183).

**Conceptual framework for indicator classification**

We used an iterative process of expert judgment to classify each indicator in our database in three different ways, which we summarise in Table 1 and describe at greater length in a supplementary methodological appendix. First, we categorised each indicator according to the Donabedian distinction between structure, process, and outcome-based approaches to measuring quality (47).

Second, we assessed the dimension of quality being measured by each indicator, using the ‘dimensions of quality’ framework first developed by the Institute of Medicine (IoM) (5) and later elaborated by the World Health Organization and OECD (28,29), which added equity to the original IoM dimensions of safety; medical effectiveness; patient-centeredness; timeliness and access; and efficiency. To these conventional dimensions of quality, our analysis led us to conceptualise two further dimensions: ‘well documented’, for indicators assessing the quality of administrative paperwork, medical records, and information handover between clinicians; and ‘trained & certified’, for indicators measuring the training and skills of the hospital workforce.

Finally, we recorded the particular specialty or part of the hospital to which each indicator pertained, drawing on the list of medical specialties across EU member states set out in EU Directive 2005/36/EC (Annex V) on the recognition of professional qualifications and including non-clinical support services and management (see supplementary materials for further details).
Table 1.
Conceptual categories for classifying each indicator in terms of its Donabedian style of measurement; the dimensions of quality it assesses; and hospital activities it oversees

<table>
<thead>
<tr>
<th>Donabedian style of quality measurement</th>
<th>Dimensions of quality</th>
<th>Hospital department or activity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Outcome</strong>: the effect of care on the health status and/or satisfaction of the patient with their treatment</td>
<td><strong>Safety</strong>: preventing adverse outcomes for patients arising from care intended to help them</td>
<td>A&amp;E Anaesthesia Cardiology Gastroenterology Geriatrics Intensive care (ICU) Nephrology Neurology Obstetrics Oncology Orthopaedics Outpatient care Paediatrics Psychiatry Rehabilitation Respiratory medicine Other medical depts: clinical specialties unique to an indicator set with ≤5 indicators (e.g. dermatology) Hospital-wide: indicators for clinical activities that span the hospital e.g. nursing, infection control Non-clinical services: e.g. catering, parking Management: e.g. finance, administration, and other oversight functions</td>
</tr>
<tr>
<td><strong>Process</strong>: how healthcare is delivered</td>
<td><strong>Effectiveness</strong>: efficacy of care in benefitting those who need it while avoiding unnecessary treatment</td>
<td></td>
</tr>
<tr>
<td><strong>Structure</strong>: type and amount of financial, human, material or organisational resources used by a health care organization to deliver services</td>
<td><strong>Patient-centredness</strong>: responsiveness of care to patient values, preferences, and needs</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Timeliness</strong>: delays and other barriers in accessing appropriate care</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Efficiency</strong>: cost-effectiveness and productivity of providers in delivering care</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Equity</strong>: fairness and impartiality in healthcare distribution, delivery, and outcomes</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Well documented</strong>: accuracy, completeness, &amp; security of administrative record-keeping about patients and their care</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Trained &amp; certified</strong>: staff licencing, training and CPD up-to-date and appropriate</td>
<td></td>
</tr>
</tbody>
</table>

**Coding process**

To ensure validity and reliability, each indicator was coded in four iterative steps. First, the four authors worked through a sample of indicators from each country to develop a consistent understanding of our classification categories. Second, three of the authors worked together to code each indicator in turn according to those categories. Third, the
remaining author then repeated the coding exercise independently. Finally, for the small number of indicators where conflicts in coding arose, reconciliation was achieved through in-depth discussion by all four authors until consensus was reached.

**Limitations**

Our study has at least three limitations. First, our datasets represent a snapshot in time. Since 2016 they have continued to evolve, but as we explain in the ‘Discussion’ and ‘Conclusions’, there are good reasons to believe that further evolutionary developments are unlikely to affect the broad patterns of difference we observe between countries. Second, the process of coding involved a significant degree of subjectivity. However, that subjectivity is mitigated by our large N and our reconciliation processes, which increase the likelihood that even if our classification of any single indicator is uncertain and contestable, any individual coding errors are likely to cancel out within such a large dataset. Third, we have restricted our analysis to the official indicator sets used by the supervisory organisations charged by law with monitoring the quality of acute hospital healthcare in each country. Further research would be needed to analyse the various indicators used by other state and non-state organisations in each of our four case study jurisdictions.

**RESULTS**

Supervisory agencies in our four countries each collected data for hundreds of hospital quality indicators, but not one of 1,100 indicators in their official sets was concerned with equity, despite the emphasis given to it in international comparisons of healthcare system performance (5,10,28,29). Aside from this universal lacuna, the four national indicator sets differed substantially in their balance of Donabedian measurement styles, the dimensions of quality they considered, and the particular hospital activities they scrutinised. Those differences are shown in Table 2 below. We describe the distinctive patterns of indicator use in each country in the following sub-sections.
Table 2.
Numbers of quality indicators used in 2016 by the English CQC, the German B-GA, the French HAS and the Dutch IGJ disaggregated by Donabedian style, quality dimension, and hospital activities

<table>
<thead>
<tr>
<th>Donabedian style</th>
<th>England</th>
<th>Germany</th>
<th>France</th>
<th>Netherlands</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structure</td>
<td>153</td>
<td>4</td>
<td>39</td>
<td>63</td>
</tr>
<tr>
<td>Process</td>
<td>12</td>
<td>119</td>
<td>145</td>
<td>90</td>
</tr>
<tr>
<td>Outcome</td>
<td>61</td>
<td>308</td>
<td>76</td>
<td>30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dimensions of quality</th>
<th>England</th>
<th>Germany</th>
<th>France</th>
<th>Netherlands</th>
</tr>
</thead>
<tbody>
<tr>
<td>Safety</td>
<td>97</td>
<td>214</td>
<td>55</td>
<td>27</td>
</tr>
<tr>
<td>Effectiveness</td>
<td>51</td>
<td>208</td>
<td>23</td>
<td>60</td>
</tr>
<tr>
<td>Patient centredness</td>
<td>27</td>
<td>-</td>
<td>49</td>
<td>-</td>
</tr>
<tr>
<td>Timeliness</td>
<td>14</td>
<td>3</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Efficiency</td>
<td>22</td>
<td>-</td>
<td>-</td>
<td>4</td>
</tr>
<tr>
<td>Equity</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Well documented</td>
<td>5</td>
<td>6</td>
<td>123</td>
<td>77</td>
</tr>
<tr>
<td>Well certified</td>
<td>10</td>
<td>-</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hospital department or activity</th>
<th>England</th>
<th>Germany</th>
<th>France</th>
<th>Netherlands</th>
</tr>
</thead>
<tbody>
<tr>
<td>A&amp;E</td>
<td>12</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Anaesthesia</td>
<td>1</td>
<td>-</td>
<td>17</td>
<td>6</td>
</tr>
<tr>
<td>Cardiology</td>
<td>24</td>
<td>148</td>
<td>17</td>
<td>31</td>
</tr>
<tr>
<td>Gastroenterology</td>
<td>16</td>
<td>37</td>
<td>6</td>
<td>18</td>
</tr>
<tr>
<td>Geriatrics</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>17</td>
</tr>
<tr>
<td>Intensive care (ICU)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>6</td>
</tr>
<tr>
<td>Nephrology</td>
<td>5</td>
<td>47</td>
<td>6</td>
<td>-</td>
</tr>
<tr>
<td>Neurology</td>
<td>9</td>
<td>18</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>Obstetrics</td>
<td>4</td>
<td>26</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Oncology</td>
<td>3</td>
<td>8</td>
<td>1</td>
<td>30</td>
</tr>
<tr>
<td>Orthopaedics</td>
<td>23</td>
<td>102</td>
<td>-</td>
<td>9</td>
</tr>
<tr>
<td>Outpatient care</td>
<td>-</td>
<td>-</td>
<td>16</td>
<td>-</td>
</tr>
<tr>
<td>Paediatrics</td>
<td>5</td>
<td>25</td>
<td>-</td>
<td>19</td>
</tr>
<tr>
<td>Psychiatry</td>
<td>4</td>
<td>-</td>
<td>17</td>
<td>-</td>
</tr>
<tr>
<td>Rehabilitation</td>
<td>-</td>
<td>-</td>
<td>22</td>
<td>-</td>
</tr>
<tr>
<td>Respiratory medicine</td>
<td>8</td>
<td>17</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>Other medical depts</td>
<td>26</td>
<td>-</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>Hospital-wide</td>
<td>33</td>
<td>3</td>
<td>89</td>
<td>14</td>
</tr>
<tr>
<td>Non-clinical services</td>
<td>4</td>
<td>-</td>
<td>9</td>
<td>-</td>
</tr>
<tr>
<td>Management</td>
<td>49</td>
<td>-</td>
<td>53</td>
<td>12</td>
</tr>
</tbody>
</table>

Total number of indicators      | 226     | 431     | 260    | 183         |
CQC indicators provided a synoptic overview of care quality across the hospital (Figure 1), covering 22 different hospital specialities, including 10 we combined together as ‘other medical departments’ because they were few in number and unique to England. As will become clear below, the scope of English hospital monitoring was far wider than in other countries, but also comparatively shallow. Most notably, the CQC’s outcomes-focus tended to preclude much scrutiny of the processes of delivering particular kinds of care. Just two of its 226 indicators measured compliance with best practice guidance (‘proportion of patients receiving all secondary prevention medication for which they are eligible’ (row 18); ‘proportion of cases complying with all nine standards of care set out by the National Hip Fracture Database’ (row 19). Rather than auditing clinical governance processes, the CQC focused instead on hospital management and on various hospital-wide indicators of quality, like waiting times, nosocomial infection, and re-admission rates, as well as patient satisfaction with non-clinical services, like catering and housekeeping.

Germany

In Germany, hospital quality indicators focused almost entirely on the safety and medical effectiveness of a few, largely surgical, interventions (Figure 2). The G-BA’s indicator set was composed primarily of outcome measures (71%). These were derived from mandatory hospital reporting of discrete outcomes from particular interventions, such as raw rates of mortality and inability to walk at discharge after knee replacement surgery (indicators 2277 and 2272 in (45)), rather than from administrative payment data or individual patient records, which might have been used to risk-adjust measures of hospital performance by taking account of varying patient mix. Process indicators focused largely on medical
effectiveness by compliance with best practice guidelines, such as the number of hip replacement surgeries fulfilling indication criteria (indicator 1082 in (45)). The G-BA’s handful of structure measures related to the availability of paediatricians at premature births and delays between diagnosis and surgery, reflecting the German concern with ensuring clinical excellence (48), rather than efficiency or patient centredness, which were not otherwise monitored.

Figure 2.
The Donabedian style, quality dimension, and hospital activities monitored by each quality indicator used in 2016 by the German G-BA (45)

This clinical orientation was also reflected in the focus of the German indicator set on intensively monitoring a limited range of largely surgical interventions, rather than considering quality at the broader hospital-level. Indeed, the hospital as an organisational entity hardly figured in the G-BA’s quality monitoring framework. Over a third (34%) of its indicators focused on a single specialty—cardiology—for which there were 148 indicators measuring the safety and effectiveness of particular surgical procedures, such as pacemaker implantation and heart transplants. Likewise, the focus of orthopaedic, nephrology, and gastroenterology indicators was also on surgical interventions rather than other kinds of treatment delivered by those specialities. The only non-surgical quality indicators were for obstetrics and the 17 indicators for the treatment of community-acquired pneumonia (classified as ‘respiratory medicine’ in Figure 2). Beyond these specialties, there were only 3 indicators for hospital-wide aspects of clinical care, such as nursing, and no indicators for non-clinical services or hospital management.

France

In contrast to England and Germany, the French indicator set consisted of mostly structure and process measures (Figure 3), which HAS constructed from hospital reporting and auditing randomly sampled patient files. There were just two clinical outcome indicators--
for post-operative pain-level and autonomy after discharge from stroke (DAN EVA and AVC9 in (44))-- and none of the mortality indicators so common in England and Germany, with the first HSMR (for myocardial infarction) still under development and not set for release until 2020 (49). Instead, outcome measures in France were almost entirely concerned with patient experiences of care, captured through survey questions about, *inter alia*, pain relief (E-SATIS29 in (44)), parking (E-SATIS1 in (44)), and the welcome provided by administrative staff (E-SATIS2 in (44)). In the absence of objectively measurable indicators of clinical outcomes, safety was assessed through structural measures of whether hospitals had appropriate protocols for managing nosocomial infections, while effectiveness was largely captured through process measures of adherence to protocols for assuring the quality of care, such as prescription of beta-blockers to heart attack patients on release from hospital (BBL in (44)). In this way, quality in France was treated as a function of hospital organisation rather than the skill of individual clinicians. Indeed, almost half of French indicators were process measures of the quality of medical record-keeping, *de facto* linking good medical practice to the paperwork needed to support the patient journey through the healthcare system.

**Figure 3.**

*The Donabedian style, quality dimension, and hospital activities monitored by each quality indicator used in 2016 by the French HAS (44)*

This organisational approach to healthcare quality in France was also reflected in the emphasis given to monitoring general hospital functions. The majority (67%) of French indicators focused on aspects of performance across the hospital, including non-clinical services like catering as well as various clinical functions, such as pain relief, patient rehab, and -most notably- nosocomial infection control, which was the subject of more than a quarter of all indicators. However, with just a few exceptions, such as psychiatry, for which France had many more indicators than any other country, much less attention was paid to monitoring individual medical specialities or interventions.
Netherlands

In the Netherlands, the IGJ drew exclusively on mandatory hospital reporting to construct its own collaboratively designed indicator set (Figure 4). The set predominantly comprised process (49%) and structure (34%) indicators, many of which, like in France, focused on the quality of documentation. For the IGJ, however, ‘well documented’ measured hospital participation in various specialty-based national registries, like the percentage of eligible operations registered with the Dutch Spine Surgery Registry (1.5.1 in (46)), rather than the quality of individual patient records, as in France. The Dutch principally assessed safety and medical effectiveness through a clinical governance focus on structure and process measures of adherence to best clinical practice, supplemented by various patient-reported outcome measures. However, unlike the CQC and G-BA, the IGJ indicator set included raw mortality indicators for just two interventions, only one of which was then risk-adjusted. Nor did the IGJ make use of patient-survey based indicators, which were instead collected by the Dutch National Health Care Institute (ZIN) to help patients choose their provider (50).

Figure 4.
The Donabedian style, quality dimension, and hospital activities monitored by each quality indicator used in 2016 by the Dutch IGJ (46)

The IGJ monitored a wide range of services across the hospital. With indicators for 13 discrete functions, including two (minimally invasive surgeries [1.7.1 in (46)], and diabetic foot ulcers [1.8.1 in (46)]) that we classified under ‘other medical departments’, it was second only to the CQC in England in the number of clinical specialities it monitored. Although there were a few indicators about hospital-wide issues like nursing care and human resource management, the focus was largely clinical. There were no indicators for non-clinical services and the vast majority (86%) of indicators focused on the quality of
particular specialties that patients might choose, like in Germany, rather than on the hospital as an organisational unit, like in France.

**DISCUSSION**

Despite the universal desire to monitor healthcare quality and substantial international efforts to identify and share best practice in measuring it, our cross-country comparison reveals striking differences in the official indicator sets used by statutory regulators to monitor the quality of hospital care in England, Germany, France, and the Netherlands.

One way in which official indicator sets differed was in their use of structure, process, and outcome indicators. Germany stood out for almost entirely eschewing structure measures in favour of outcome and process ones. By contrast, regulators in the other countries used all three indicator types more freely, with outcome indicators predominating in England, process indicators in France, and Dutch indicators evenly divided between Donabedian’s three styles of measurement.

Official indicator sets also differed in which dimensions of quality were monitored and how they were measured. Medical effectiveness and safety received universal attention, but apart from Germany – which focused almost exclusively on those two dimensions of quality – other countries had various additional quality concerns as well. Hospital record-keeping accounted for a third of Dutch indicators and half of French ones but was largely ignored in England. Patient experience was closely monitored by regulators in France and England but not by their Dutch or German counterparts. Likewise, efficiency was a concern in England and to a lesser extent in the Netherlands, but not in France or Germany.

Even when they monitored the same quality dimensions, regulators often defined and measured them in quite different ways. For example, more than 90% of all safety indicators in both England and Germany were patient-reported outcomes, which they calculated in very different ways. While the CQC overwhelmingly relied on HSMRs, the G-BA measured a broader range of adverse clinical outcomes but did not standardise them to take account of hospitals’ varying patient mix and measure relative performance. In contrast to that outcome-focus, HAS assessed hospital safety in France by checking for the existence of clinical protocols to prevent hospital-acquired infections. The Dutch IGJ measured safety in the most diverse ways, including: checking the existence of, and compliance with, hospital infection controls and other speciality-specific safety protocols; measuring patient volumes to ensure surgeons were sufficiently practiced to be safe; calculating hospital standardised emergency readmission and complication rates, but almost no mortality rates.

Official indicator sets also focused on different kinds of hospital activity. German indicators intensively monitored a small number of largely surgical interventions, almost completely ignoring other kinds of medical care or the hospital itself as an organisation. By contrast, indicators in the other three countries covered a broader range of clinical specialties and were more concerned with hospital-wide processes and management. England monitored by far the widest set of hospital activities, while France was most pre-occupied by management of hospital-wide concerns, such as infection control and catering. By contrast, the Dutch indicator set was concerned with how well hospital specialists cooperated with
various national disease-based registries to support those registries’ quality improvement activities.

These findings are consistent with comparative health policy studies that have highlighted how new policy instruments are shaped by country-specific demands and constraints of national healthcare systems and governance traditions, the interests and veto power of key actors, public preferences and the wider political system (51–53). Such factors are likely to create path dependencies in the way that quality indicators are developed and put to use in each country (8,54–56). We can go further in explaining the nationally specific character of indicator sets, however, if we differentiate between ‘demand-side’ pressures for quality indicators, and ‘supply-side’ constraints on how indicators can be constructed.

Thus ‘demand-side’ pressures help explain how national indicator sets ended up targeting such divergent dimensions of quality and hospital activities as they have responded to the distinctive policy problems emerging in each country’s healthcare system. Thus, the indicator set for England’s NHS was synoptic in its coverage of quality dimensions and hospital activities, because the state is responsible for everything: funding and delivering healthcare as well as regulating its quality. In this context, competing public demands for safe, speedy, and yet also inexpensive care have fuelled regular political crises. In response, politicians have charged the regulator - the CQC - with an ever-expanding list of quality concerns that its indicators must somehow monitor (9).

By contrast, official indicator sets in Germany, France, and the Netherlands were less comprehensive in their coverage of quality dimensions and hospital activities because in those social- and private-insurance systems the state is less immediately accountable for healthcare and so has left some matters to the healthcare sector. In Germany, the G-BA is relatively insulated from political pressures and has, therefore, been slow to expand the narrow scope of its indicator set, which was first introduced to prevent providers from compromising the safety and effectiveness of fixed price surgical procedures (57). In France, HAS initially adopted a light touch to monitoring quality, restricting itself to patient experience surveys and assessing the quality of paperwork to guard against discontinuities in care by doctors operating in private practice within French traditions of liberal medicine (58,59). However, oversight expanded in 2006 when a public crisis over nosocomial infections (60,61) prompted the state to develop safety indicators for infection control, giving France more than three times as many such indicators as the other countries put together. In the Netherlands, IGJ indicators have focused more on the clinical effectiveness and safety of a wide range of discrete specialities, not least to ensure that the market-oriented healthcare reforms of 2006, which sought efficiency gains through managed competition, did not result in a race to the bottom on quality (62).

‘Supply-side’ explanations for indicator variety concern the way in which the configuration of national healthcare, political and regulatory systems constrains the kinds of data that regulators can use for indicator construction (8), and as such principally shape the balance between Donabedian measurement styles in each countries’ indicator set. Thus, in England’s pluralist political system where the state is not bound to secure the agreement of competing stakeholder interests, the quasi-independent CQC has been free to construct indicators from the vast quantities of administrative data on NHS structures, performance,
and patient outcomes that the state already collects routinely in discharging its responsibility for both the financing and provision of NHS healthcare. Accordingly the CQC indicator set was wide-ranging and in keeping with British commitments to ‘risk-based’ regulation (9,63), it used z-scoring techniques to highlight hospitals posing the greatest risk to quality standards, which were defined relatively in terms of statistical deviation from the mean rather than according to any absolute minimum standards (64).

However, comparable administrative data is less readily available in social and private insurance systems of Germany, France and the Netherlands, not least because payment systems vary between insurers (despite ongoing efforts to standardise according to diagnostic-related groups (57)). Consequently, regulators have had to negotiate consistent reporting standards with sometimes reluctant, and often private, providers or otherwise source their own data in the face of varying legal and institutional constraints that are deeply rooted within their particular political systems and constitutional settings.

Thus, in Germany’s fragmented healthcare system indicator construction is highly constrained by both technical data availability and political constraints on its use. Inconsistencies in patient recording systems make it difficult to take account of hospitals’ varying patient mix (57), which is one reason why mortality and other clinical outcome measures are not risk-adjusted, and hospital performance is benchmarked against absolute reference values rather than relatively as in England. Indicator development is also constrained by the federal political system which makes the 16 state (Länder) governments—and not the B-GA—responsible for organisational aspects of hospital provision, that might otherwise be served by structural indicators. Further constraints are created by the political need for corporatist consensus, and strong constitutional protections of business rights to economic activity (65), which open indicator design to legal challenge and were central to German hospitals successfully contesting minimum volumes regulation (66).

In France’s centralised pluralist political system, the government faces fewer political and constitutional constraints on data usage, but its fragmented healthcare system has forced HAS to collect much of its own indicator data, largely through auditing patient medical records, mandatory hospital reporting and patient surveys. Those data sources have favoured structure and process measures and a concern with the quality of documentation and patient experience rather than clinical outcome measures, such as mortality indicators, which France was slow in developing, because of medical professional scepticism about unadjusted mortality rates (60,61). HAS is currently developing its first HSMR, which is now possible thanks to a new database, the Système National des Données de Santé (SNDS) (67), which the state is creating to help with cost control by linking previously separate payment data with clinical in- and out-patient activity records and a national cause of death registry.

In the Netherlands, indicator selection and construction are less constrained by technical challenges of data sourcing and linkage that trouble Germany and France, not least because the 2006 health care reforms required standardisation of payment data to ensure equitable distribution of the pool of high-risk patients. Rather the main constraint has been political insofar as the Dutch corporatist governance tradition means any new indicators must gain the consent of the various medical professional and hospital associations (8). Process and
structure indicators – which account for half and a third of the indicator set respectively – are widely accepted amongst stakeholders. By contrast, outcome indicators, which are widely used by clinical registries to support quality improvement initiatives, account for only a sixth of the official indicators used by the IGJ. It explains this imbalance by noting that “the reporting burden on an outcome indicator is much greater than a structure indicator” (46), but informally it is also clear that the IGJ uses outcome measures sparingly because their wider utility for regulatory purposes is not universally accepted (68). Similarly, the indicator set does not include any patient survey-based measures because the patient experience is regarded as more relevant to informing patient choice - and thus the responsibility of ZIN - than quality assurance and regulation.

CONCLUSIONS

Our research on the use of quantitative indicators by healthcare quality regulators in four neighbouring EU states shows that they define, measure, and monitor the quality of acute hospital care in starkly different ways. However, we go beyond the banal observation that countries have their own ways of doing things, much like they have different national flags. Rather we argue that contrasting indicator set designs reflect fundamental differences in national regulatory priorities, institutional configurations of payers and providers, and even understandings of quality itself. Although national indicator sets will continue to evolve, the patterns we identify here are likely to persist. That path-dependence reflects distinctive ‘demand-side’ pressures shaping the particular dimensions of quality and hospital activities targeted by national indicator sets, as well as ‘supply-side’ constraints on data availability and access shaping the Donabedian measurement styles adopted in different healthcare and regulatory systems.

Our analysis helps explain why international efforts to benchmark hospital quality and identify universal measures are so difficult (26,54). In the absence of universal agreement about the meaning of quality, countries necessarily steer by their own lights when selecting quality indicators. Nevertheless, our analysis does suggest that international benchmarking could be made more tractable by looking for families of countries with similarly structured healthcare systems and governance traditions, where supply-side constraints on, and demand-side pressures for, measuring healthcare quality are better aligned.

ACKNOWLEDGEMENTS AND COMPETING INTERESTS STATEMENT

Contributors All four authors co-conceived the study, conducted background interviews, analysed the data and contributed to writing the article.

Funding The Economic and Social Research Council (ES/K006169/1) and the Wellcome Trust (210346/Z/18/Z).

Ethics Ethical approval was obtained for the two projects under which this study was conducted from the Research Ethics Committee of King’s College London (REP(GSSHM)/13/14-5, MRA-17/18-5908).
**Competing interests** AG worked at the CQC up until 2016 and now works for Statica Research with a focus on patient feedback. DD, HR and A-LB report a grant from the ESRC (detailed above) and DD and HR report a grant from the Wellcome Trust (detailed above) during the conduct of the study

**REFERENCES**


SUPPLEMENTAL. METHODOLOGICAL APPENDIX

We classified each indicator in our database in three different ways, summarised in Table 2 (see below). First, we categorised each indicator according to the Donabedian distinction between structure, process, and outcome-based approaches to measuring quality. Structure indicators measure the type and amount of financial, human, material and organisational resources used by health care organizations to deliver services. By contrast, process indicators measure the delivery of appropriate (or inappropriate) healthcare to the relevant population, consistent with best current professional knowledge. Lastly, outcome indicators measure the effect of care on the health status of the patient as well as improvement in patient knowledge, behaviour and satisfaction with their treatment.

Though clear in theory, Donabedian’s threefold model was not always easy to apply in practice. One ambiguity concerned the classification of indicators measuring the existence and application of care protocols. Indicators assessing a hospital’s adoption of protocols, for example to deal with nosocomial infections, were classified as structure indicators, insofar as they were about the resources available to a hospital in delivering care. By contrast, indicators measuring how well protocols were followed in practice, were classified as process indicators. Another ambiguity concerned the classification of outcome indicators that were measuring outcomes for organisations rather than Donabedian’s classic definition of outcome measures as pertaining to “the effects of care on the health status of patients and populations” [1]. We classified indicators addressing organisational outcomes, such as the CQC’s use of the Monitor risk rating of the financial health and governance of an NHS trust (MONITOR01: row 138), as structure indicators insofar as they were concerned more with the structure and inputs to care than outcomes for patients.

Second, we classified each indicator in terms of the regulatory goals it was assessing, using the ‘dimensions of quality’ framework [2]. This basic approach to conceptualising quality was pioneered by the Institute of Medicine (IoM) [2], which identified five dimensions of quality—safety; effectiveness; patient-centeredness; timeliness and access; efficiency—as “definable, preferably measurable, attributes of the system that are related to its functioning to maintain, restore, or improve health” [3]. This framework was further elaborated by the World Health Organization and OECD [4,5], which added equity as another distinct dimension of healthcare quality that indicators might attempt to capture.

Applying this basic framework to our indicators required some further elaboration to define those different dimensions of quality more closely. Differentiating between safety and effectiveness could be particularly challenging. For example, should indicators about mortality among very small pre-term infants be classified as medical effectiveness, since survival is a key measure of the effectiveness of neonatal care, or of safety, insofar as the first duty of the physician is to do no harm? Faced with such borderline cases, we categorised those indicators dealing especially with the harm done to patients during an episode of care (for instance, in-hospital mortality or nosocomial infections following a procedure) as safety. In turn, we treated indicators about longer-term outcomes of care (for instance survival rates after x years for heart transplant and per cent of patients readmitted after a procedure) as effectiveness. The notions of harm, directness and immediacy were our major criteria for differentiating between these two dimensions of quality. Using a similar logic, we classified ‘volume’ indicators dealing with the number of patients treated
for a specific condition and with a specific procedure (for instance volume of bariatric surgery performed in a year) as indicators of safety or effectiveness rather than access, because professional guidelines often recommend a minimum volume of operations be performed each year in order to ensure the skill of the team and hence the safety of the procedure [6].

A number of indicators did not fit the conventional dimensions of quality described by the IoM, WHO, and OECD, which forced us to invent two additional quality dimensions. In particular, we observed that many indicators, especially in France and the Netherlands, were concerned with measuring the quality of the administrative paperwork, medical records, and information handover to relevant professional or administrative organisations. To capture this regulatory goal, we created an additional dimension of quality we termed “well documented”. Another focus that conventional dimensions of quality failed to capture was the concern, especially in England and the Netherlands, with assessing the training and skills of the hospital workforce. To capture this regulatory desideratum, we created an additional category we called “trained and certified”, giving us a total of eight dimensions of quality for classifying each indicator in our dataset (Table 1).

<table>
<thead>
<tr>
<th>Safety</th>
<th>preventing adverse outcomes for patients arising from care intended to help them</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medical effectiveness</td>
<td>efficacy of care in benefitting those who need it while avoiding unnecessary treatment</td>
</tr>
<tr>
<td>Patient-centeredness</td>
<td>responsiveness of care to patient values, preferences, and needs</td>
</tr>
<tr>
<td>Timeliness</td>
<td>delays and other barriers in accessing appropriate care</td>
</tr>
<tr>
<td>Efficiency</td>
<td>cost-effectiveness and productivity of providers in delivering care</td>
</tr>
<tr>
<td>Equity</td>
<td>fairness and impartiality in healthcare distribution, delivery, and outcomes</td>
</tr>
<tr>
<td>Well documented</td>
<td>accuracy, completeness, &amp; security of administrative record-keeping patient care and associated clinical processes</td>
</tr>
<tr>
<td>Well-trained and certified</td>
<td>staff licensing, training and continuing professional development up-to-date and appropriate</td>
</tr>
</tbody>
</table>

Table 1. Dimensions of quality used in our classification

Third, we sought to identify the particular specialty or hospital activity to which a given indicator pertained with a field we term “hospital department or activity”. This classification exercise was complex, as there are many different ways to classify hospital activities, so we adopted an iterative approach to refining this field. We began by considering the series of occupation codes describing divisions of clinical work in the British NHS, which may be defined by body systems (e.g. dermatology), patient demography (e.g. paediatrics), clinical technology (e.g. nuclear medicine), clinical function (e.g. rheumatology), disease type (e.g. oncology) or combinations of these factors. This list of recognised specialities in the NHS is tied to the EU directive 2005/36/EC on the recognition of professional qualifications, whose
Annex V lists the equivalences for various medical specialties in different EU member states, on which we based our classification.

We further refined that very long list of specialities by comparing it against the major section headings in each indicator set to identify some commonly used designations like ‘oncology’, ‘cardiology’, and ‘psychiatry’. Specialities for which only one country maintained an indicator and the total number of such indicators was less than or equal to five were consolidated into a category we termed ‘other medical depts.’ We also folded together various hospital-wide clinical functions like nursing, imaging, and pathology into a category we called ‘hospital-wide’ for indicators about clinical activities that span the hospital. However we maintained a separate category rehabilitation services, for which France maintains 22 indicators, on the grounds that in French hospitals it is a discrete function. In looking through the indicators it also became clear that there were many targeting various non-clinical ‘hospitality’ functions of a hospital, like catering and housekeeping. To represent them we created an additional category: ‘Non-clinical services’. We also added a category called ‘Management’ to refer to any indicators measuring the quality of administrative functions such as finance, human resources, ICT systems, hospital records. In total we had 20 discrete ways of classifying the particular hospital activities that a given indicator is assessing (Table 2).

Table 2. Hospital activity categories used to classify quality indicators

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A&amp;E</td>
<td>Indicators targeting the quality of care in emergency departments</td>
</tr>
<tr>
<td>Anaesthesia</td>
<td>Indicators assessing administration and control of pain relief</td>
</tr>
<tr>
<td>Cardiology</td>
<td>Indicators for the treatment of heart surgery and other cardio-vascular conditions, including, but excluding carotid surgery and stroke care</td>
</tr>
<tr>
<td>Gastroenterology</td>
<td>Indicators relating to treatments of the digestive system, including liver and pancreatic surgeries</td>
</tr>
<tr>
<td>Geriatrics</td>
<td>Indicators focused on the care of elderly and therefore potentially vulnerable patients</td>
</tr>
<tr>
<td>Intensive care (ICU)</td>
<td>Indicators targeting care in intensive care units</td>
</tr>
<tr>
<td>Nephrology</td>
<td>Indicators for kidney and pancreas treatments, except for cancer</td>
</tr>
<tr>
<td>Neurology</td>
<td>Indicators for neurological conditions, including carotid surgery and stroke care</td>
</tr>
<tr>
<td>Obstetrics</td>
<td>Indicators relating to the treatment of both mother and baby in childbirth, including neonatal care</td>
</tr>
<tr>
<td>Oncology</td>
<td>Indicators for cancer care</td>
</tr>
<tr>
<td>Orthopaedics</td>
<td>Indicators concerned with treatment of musculoskeletal conditions</td>
</tr>
<tr>
<td>Outpatient care</td>
<td>Indicators focused on the quality of care delivered by hospitals to patients ‘at home’ or otherwise not admitted to wards</td>
</tr>
<tr>
<td>Paediatrics</td>
<td>Indicators focused on the care of children, but excluding neo-natal care</td>
</tr>
<tr>
<td>Psychiatry</td>
<td>Indicators for mental health services</td>
</tr>
<tr>
<td>Rehabilitation</td>
<td>Indicators for specific rehabilitation services to restore function and prepare patients for release from hospital</td>
</tr>
<tr>
<td>Respiratory medicine</td>
<td>Indicators for the care of community acquired pneumonia and other respiratory conditions, but excluding lung cancer</td>
</tr>
</tbody>
</table>
Other medical depts: indicators for clinical specialties that were unique to a particular indicator set and numbering less than 5 in total were lumped together under this category.

Hospital-wide: Indicators for clinical activities that span multiple hospital departments or the entire hospital: e.g. nursing, infection control, pathology, imaging, etc.

Non-clinical services: Indicators assessing the quality of patient-focused but non-clinical functions and services like catering, parking, visiting hours, and other ‘hospitality’ functions.

Management: Indicators targeting hospital administration, finances and governance.

Applying this classification framework could be challenging when a particular indicator targeted an activity at the intersection of two specialties. For indicators about neonatal care, we classified those pertaining to the birth as obstetrics and the rest as paediatrics. We classified indicators pertaining to stroke care and carotid surgery as ‘neurology’ rather than ‘cardiology.’ For cancer, we defaulted to oncology, classifying the Dutch indicators about lung cancer (9.3 Lung carcinoma: rows 139-42) under that category rather than ‘respiratory medicine.’ The complete set of indicators and our classification of them is available in the supplementary materials for this paper.

REFERENCES

2 Institute of Medicine, Services B on HC, Delivery C on the NQR on HC. Envisioning the National Health Care Quality Report. National Academies Press 2001.


