Developing and Trialling an Independent, Scalable and Repeatable IT-benchmarking Procedure for Healthcare Organisations*

J. D. Liebe; U. Hübner
Health Informatics Research Group, University of Applied Sciences Osnabrück, Department of Business Management and Social Sciences, Osnabrück, Germany

Keywords
Hospital, IT-benchmarking, IT-governance, performance indicator, visualisation, guidance

Summary
Background: Continuous improvements of IT-performance in healthcare organisations require actionable performance indicators, regularly conducted, independent measurements and meaningful and scalable reference groups. Existing IT-benchmarking initiatives have focussed on the development of reliable and valid indicators, but less on the questions about how to implement an environment for conducting easily repeatable and scalable IT-benchmarks.

Objectives: This study aims at developing and trialling a procedure that meets the afore-mentioned requirements.

Methods: We chose a well established, regularly conducted (inter-) national IT-survey of healthcare organisations (IT-Report Healthcare) as the environment and offered the participants of the 2011 survey (CIOs of hospitals) to enter a benchmark. The 61 structural and functional performance indicators covered among others the implementation status and integration of IT-systems and functions, global user satisfaction and the resources of the IT-department. Healthcare organisations were grouped by size and ownership. The benchmark results were made available electronically and feedback on the use of these results was requested after several months.

Results: Fifty-nine hospitals participated in the benchmarking. Reference groups consisted of up to 141 members depending on the number of beds (size) and the ownership (public vs. private). A total of 122 charts showing single indicator frequency views were sent to each participant. The evaluation showed that 94.1% of the CIOs who participated in the evaluation considered this benchmarking beneficial and reported that they would enter again. Based on the feedback of the participants we developed two additional views that provide a more consolidated picture.

Conclusion: The results demonstrate that establishing an independent, easily repeatable and scalable IT-benchmarking procedure is possible and was deemed desirable. Based on these encouraging results a new benchmarking round which includes process indicators is currently conducted.

1. Introduction
1.1 Impact on Patient Care
More and more studies are aiming to provide evidence that information technology (IT) can impact patient care and they cautiously conclude that this is the case when IT is implemented and utilised properly, e.g. electronic health record systems (EHR) [1], computerised provider order entry systems (CPOE) [2, 3] and clinical decision support systems (CDSS) [4]. These systems may facilitate access to comprehensive and detailed patient data [5], can support adherence to guideline-based care [6] and may help in improving the process of care [7]. The services they offer enfold their potential in IT-environments that allow information logistics mechanisms to provide the right data of the right patient at the right time and place to the right receiver [8]. In the best case health information systems constitute such an environment.

1.2 Systematic IT-governance
In order for IT-environments and clinical applications to operate successfully they need to be managed at the strategic, tactical and operational level [9]. It is only then that they are suitable enablers of the healthcare enterprise’s strategy. On the other hand, managing IT-systems systematically, i.e. employing IT-governance methods, has a great potential of reducing the technical and social risks associated with introducing these systems [10]. IT-governance em-

* Supplementary online material published on our website www.methods-online.com
braces decision making for the right system considering their potential for optimizing the service but also their risks for harming patients [11] and the hazards of impeding the work of clinicians [12]. Eventually, decisions have to be made for balancing alternative technical options [13] and taking into account the various requirements of the users at the level of individuals, departments and organisations [14]. Information officers therefore must ask themselves how performing their systems are and what options new systems offer to optimise the current practice [15].

1.3 Measuring IT Performance

Performance indicators (PI) measure system characteristics, the degree with which these systems support the goals of the own enterprise [10, 16, 17], constitute the basis for comparing the performance of different organisations and ultimately for conducting IT-benchmarks [18]. They aim at providing advice on how to systematically improve IT-performance and encompass the following phases: 1) selection of pertinent performance indicators (PI) [19], 2) application of these indicators and identification of the weaknesses of the own organisation in comparison with the best of the group and 3) development of practical actions on how to overcome the deficiencies [20]. There are no standardised methods on how to define PIs, however, there is a wealth of different approaches within and outside healthcare. COBIT (Control Objectives for Information and Related Technology) [21] and ITIL (Information Technology Infrastructure Library) [22] are the most renowned frameworks for IT-governance and industry-independent approaches that define objective and subjective PIs relating to IT-operations and support.

1.4 IT-benchmarking in Hospitals

Other approaches are specialised for hospitals and healthcare in general, e.g. measuring IT-sophistication in hospitals in terms of technology, functions and of integration [23] or IT-maturity [24]. Other objective performance indicators looked at selected topics, such as the number of documents created and the number and type of problems derived from trouble ticket statistics [16] and the completeness and timeliness of discharge letters and the use of patient scheduling [19]. Subjective outcome measures that serve as PIs were availability, correctness and completeness, readability, usability of information, compliance with regulations, and the time needed for information processing [25]. Other authors pursue an integrated approach of subjective and objective criteria, e.g. automation and usability scores [26], a combination of indicators developed in other studies [27] referring to 25 and 19] and system, information and service quality, use and user satisfaction aggregated to an electronic medical record composite index [18]. In addition to defining the content of PIs there is the question of measuring the PIs and comparing organisations based on these PIs. Regularity, independence, anonymity as well as a sufficiently large number of benchmark participants are characteristics stipulated in the literature that allow performance to be improved systematically [28]. Existing IT-benchmarking initiatives often only relate to individual hospitals [16, 19] or a closed group of a limited number of hospitals [14, 18].

Against this background, our study aims at developing (aim 1) and trialling (aim 2) a new approach for conducting an independent, scalable and easily repeatable IT-benchmarking procedure. This benchmarking study should focus on structural features of the organisation and functional characteristics of the IT-systems.

2. Materials and Methods

2.1 Survey-based Benchmark at National Level

In order to achieve these aims we propose to embed the benchmarking procedure within an established survey on health IT for several reasons. Using an established survey as a data source for composing reference groups entails several options. First, the benchmark is scalable in terms of the number of peers in the reference groups which can exceed the number of benchmark participants. Large reference groups help avoiding statistical flukes. Second, established surveys usually draw on a proven mechanism for developing a questionnaire, recruiting participants and analysing data. This makes them an attractive platform for benchmarks that should be easily repeated. If the survey is conducted and financed by an independent organisation the benchmark utilises data that are more likely to be unbiased with regard to potential influences exerted by the benchmark participants and third parties. Based on these presumptions we chose the IT-Report Healthcare [29, 30] as the platform and data source. The IT-Report Healthcare is a national survey of healthcare institutions in Germany that is also used in Austria and the Netherlands. It has been conducted regularly for ten years and is based on a sufficiently large number of respondents from hospitals of a different type, size and location. It is financed by public grants and publishes its survey results freely accessible on the Internet (www.it-report-healthcare.info).

The IT-benchmarking procedure described in the following related to the IT-Report Healthcare online survey of 1,368 chief information officers (CIOs) and other leading members of IT-staff in German hospitals, who represented 1,807 of Germany's 2,061 hospitals. In 254 hospitals, no such person could be identified either because they did not exist or because the position had been outsourced. The survey was conducted from March to July 2011. The invitation to the survey included the offer to take part in an anonymous IT-benchmarking.

2.2 Benchmarking Instrument and Performance Indicators

The questionnaire used in the 2011 survey was a standardized instrument that consisted of 40 main questions and 9 subordinated questions on IT-systems/functions and integration, overall satisfaction with IT-systems/functions, IT-department, management issues and on various additional variables, e.g. hospital size, ownership and location. These questions had been developed in close cooperation with other experts in this field and had been pretested by three chief information officers. The items of this questionnaire constituted 61 raw performance indicators that could be classified mainly as structural fea-
Firstly understand the “capability” of the methods inf med 4/2013 © schattauer 2013
362
Table 2 Size of the reference groups for hospital size (n = 190) and ownership (n = 180) and participant groups

<table>
<thead>
<tr>
<th>raw performance indicator</th>
<th>IT-systems/ functions and integration</th>
<th>overall satisfaction with systems/functions</th>
<th>IT-department</th>
<th>management issues</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>43</td>
<td>1</td>
<td>11</td>
<td>6</td>
</tr>
</tbody>
</table>

2.3 Establishing Reference Groups on Validated Characteristics

All 193 hospitals that responded to the survey served as a reference data platform for the IT-benchmark. These hospitals represented organisations of all different sizes, of different ownerships and were dispersed throughout Germany. Features that were meaningful descriptors of the reference groups were identified on the basis of univariate and multiple regression analyses. Hospital size, ownership, IT-department, relationship with IT-vendors and IT-strategy proved to be significant predictors for the number of IT-systems and could be confirmed in an independent analysis [31, 32]. Among these five predictors hospital size and ownership were so-called context factors, i.e. stable factors that could not be changed within a short time frame. They were used to form the two reference groups for each hospital that was benchmarked: ownership embraced the attributes “public” and “private” and hospital size the classes “up to 199”, “200 to 399”, “400 to 599”, “600 to 799” and “800 and more” beds. The number of classes and the class width was chosen on the ground of balancing the chance of getting enough members per group and of building meaningful groups.

2.4 Measuring and Visualising the Indicators

For each benchmark participant, the 61 raw performance indicators were presented in comparison to the ones of the size and of the ownership reference group. These comparisons were visualised in the so-called single indicator frequency views, i.e. frequency charts (bar or pie charts) for each of the 61 raw performance indicators within the reference group showing the bar or pie section that belonged to the particular category of the benchmark participant in dark grey (Figures 1 to 3). The implementation status of clinical systems was measured by a 4-point-Likert scale from "no implementation" to "completely implemented in all units" (clinical systems – example see Table 1). The status of the electronic patient record system by a 5-point-Likert scale (Figure 2) and finally the status of administrative systems by a 3-point-Likert scale from "no implementation" to "completely implemented". Different Likert scales were employed for measuring the implementation status of clinical and administrative systems because we distinguished between the implementation status of clinical systems “in all clinical units” and “in at least one but not all clinical units”. This distinction did not apply to administrative systems, which could be implemented in one department only. The scale for measuring the implementation status of the electronic patient record corresponds to the scale of previous surveys of the IT-Report Healthcare and was retained to be downward compatible. The frequency data for each raw indicator were extracted from SPSS Version 19 and imported into Microsoft Excel 2010 where the charts were pro-

Table 2 Size of the reference groups for hospital size (n = 190) and ownership (n = 180) and participant groups

<table>
<thead>
<tr>
<th>hospital size (n = 190)</th>
<th>ownership (n = 180)</th>
</tr>
</thead>
<tbody>
<tr>
<td>up to 199 beds</td>
<td>25.2 % (48)</td>
</tr>
<tr>
<td>200 to 399 beds</td>
<td>31.6 % (60)</td>
</tr>
<tr>
<td>400 to 599 beds</td>
<td>14.2 % (27)</td>
</tr>
<tr>
<td>600 to 799 beds</td>
<td>9.0 % (17)</td>
</tr>
<tr>
<td>800 beds and more</td>
<td>20.0 % (38)</td>
</tr>
<tr>
<td>privat</td>
<td>21.7 % (39)</td>
</tr>
<tr>
<td>public</td>
<td>78.3 % (141)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>percentage of survey participants in reference group</th>
<th>25.2 % (48)</th>
</tr>
</thead>
<tbody>
<tr>
<td>percentage of benchmark participants in reference group</td>
<td>18.8 % (21/60)</td>
</tr>
</tbody>
</table>

*Only 190 hospitals had given information on their size and 180 hospitals on their ownership*

Methods Inf Med 4/2013 © Schattauer 2013

Downloaded from www.methods-online.com on 2018-04-14 | IP: 54.70.40.11
For personal or educational use only. No other uses without permission. All rights reserved.
duced manually. All data were quality checked by two other persons. Hundred and twenty-two of these charts were provided in electronic format to each benchmark participant.

2.5 Evaluation

All 59 benchmark participants were contacted by phone or email on the average eight months after receiving the charts to give feedback on the benchmarking procedure. In particular, we asked whether the participants considered the information useful and if yes how they could make use of them. Furthermore, we wanted to know if the participants had any comments on improving the benchmarking procedure and the information. Finally, we asked them whether they would take part in this benchmark again. We used open questions because we were not sure about what the CIOs wanted to tell us. The answers in the telephone interviews were recorded on paper and used together with the email answers to extract significant statements and to group them according to recurring themes of similar content. Two persons categorised the free text answers independent from each other, in order to ensure an objective analysis. In case of disagreement they discussed the issues to find a solution. The reason for contacting them with some delay after sending the charts was to give them enough time to fully exploit the information, e.g. use them to justify investments or show deficiencies to the board of directors.

3. Results

3.1 Participation and Raw Performance Indicators

About one-third of the survey participants (59 out of 193) took the offer of having their hospital benchmarked. They were allocated to two reference groups. As ▶Table 2 shows these groups were populated with at least 17 hospitals in the size group “600 to 799 beds” and ranging up to 141 hospitals in the ownership group “public hospitals”. As indicators related to the implementation status were measured by Likert scales, the bar charts showed the distance of the benchmarked hospital to those in the reference group who were best. ►Figure 1 and ►Figure 2 give examples of these bar charts for the implementation status of the “clinical reminder” function and the “electronic patient record system”. In both cases the hospital belonged to the group “having started implementation or providing resources for implementation” and in both cases the best hospitals of this group had implemented the IT-system/function in all clinical units. The frequency diagrams also revealed information about the positioning of the hospital within the field of hospitals of the same reference group. Figure 1 shows that the hospital was better than the statistical mode (the category with the highest frequency) in the case of the clinical reminder function, whereas in the case of the electronic patient record system the same hospital was average belonging to the group of the mode, i.e. “having started implementation” (►Figure 2).

Other performance indicators related to the position of the chief information officer (CIO) within the organisation (►Figure 3), number of full-time equivalents in the IT-
3.2 Evaluation

A total of 51 persons from the 59 benchmarking participants (86%) took part in the evaluation, 18 via email and 33 via telephone interviews. Eight persons could not be contacted neither by mail nor by phone. The large majority of the 51 persons (94.1%) reported that they could make use of the benchmarking information. Three persons (5.9%) said that they either did not have enough time to devote themselves with benchmarking or preferred other indicators (usage of systems instead of availability). The persons who considered the benchmarks beneficial gave a combination of reasons for their opinion (Table 3). Twenty persons (39.2%) made comments on how to improve the benchmarks. The majority of answers referred to the use of cost indicators. Other suggestions included the use of additional indicators or the provision of more details, guidance and follow-ups and the presentation of the results in a more consolidated manner (Table 4). Four persons explicitly mentioned and esteemed the scientific and independent approach. Ninety-four percent (94.1%) of the participants who evaluated the benchmark responded that they would enter again.

3.3 Alternative Views

Based on the wish of the participants for a consolidated perspective on the data we sought a solution to combine different types of information visually. In addition to the single indicator frequency views, which made use of
the raw indicators, we thus developed two alternative views for visualising the implementation status of different IT-systems/functions that aggregated the raw indicators and emphasised particular aspects within the data. In the first type of diagrams, the multiple indicator distance views, IT-systems/functions that belonged to the same category were depicted in a web graph together with their degree of implementation. Connecting the values for the degree of implementation yielded a characteristic picture so that the shape of the best within the reference group and that of the own institution could be compared (Figure 4). As the indicators were scaled as ordinal data discrepancies could be interpreted as distances. This view did not contain any information on frequencies.

Table 3  Types of usage of the benchmark results (n = 51)

<table>
<thead>
<tr>
<th>percentage of respondents who could make practical use of the benchmark</th>
<th>type of use (percentage of answers from all answers)</th>
</tr>
</thead>
<tbody>
<tr>
<td>learn about position in peer group</td>
<td>strategic alignment</td>
</tr>
<tr>
<td>94.1 %</td>
<td>48.5%</td>
</tr>
</tbody>
</table>

Table 4  Comments for improving the benchmarks (n = 51)

<table>
<thead>
<tr>
<th>percentage of respondents who made comments</th>
<th>type of comments (percentage of answers from all answers)</th>
</tr>
</thead>
<tbody>
<tr>
<td>use of cost data, e.g. for cost-benefit ratio</td>
<td>use other indicators in addition and provide more details</td>
</tr>
<tr>
<td>39.2 %</td>
<td>50.0%</td>
</tr>
</tbody>
</table>

Figure 4  Multiple indicator distance view for IT-systems/functions related to documentation in the size reference group 200 to 399 beds (n = 60)
4. Discussion

4.1 Feasibility

In this study, we could demonstrate that a large number of IT-performance indicators measuring structural features of a healthcare organisation and functional characteristics of the IT-system could be captured using a questionnaire, could be analysed statistically within reference groups and could be visualised in various views. The number of benchmark participants revealed the interest in this procedure and the positive feedback given in the evaluation was encouraging to continue along the way.

Hospitals from all size groups expressed the wish to take part in the benchmarks. This illustrates an interest in getting benchmarked that is independent from hospital size.

Capturing IT-performance indicators in an online survey with CIOs using a standardised and well tested questionnaire helps to fulfil two success criteria: a) data acquisition is comparatively time saving, which is a criterion that Jahn and Winter deemed desirable [27] and b) the chance to receive high quality and meaningful data is great as persons that oversee the IT-landscape of the organisation are the respondents – which is a criterion stipulated by Hübner-Bloder and Ammenwerth [17].

4.2 Statistical and Scientific Approach

The combination of conducting a regular national survey on IT-adoption in healthcare organisations and of benchmarking its participants or at least a subset of them allows a statistical approach to comparing similar organisations. The main advantage is that the size of the reference groups does not depend on the number of benchmarking participants but on the number of survey participants (scalability). Furthermore, reference groups can be established on the ground of statistical analyses that give appropriate information for building them. The feasibility of the benchmarking and the intention of the benchmark participants to enter again constitute a solid ground for its repeatability. The evaluation results show that the objective and scientific approach was appreciated, which con-

Figure 5 Multiple indicator innovation view for IT-systems/functions related to documentation in the size reference group 200 to 399 beds (n = 60)

The second type of diagrams, the multiple indicator innovation views, allowed the user to classify the own organisation in terms of its type of innovation (Figure 5). Similar to the multiple indicator distance view a group of IT-systems/functions was targeted. Each system/function was localised in two dimensions: first the implementation status and second the frequency of the specific manifestation of the implementation status in the own organisation. For example, the hospital in Figure 5 had implemented the electronic nursing record “in all units” (x-axis) like 10% of the hospitals (y-axis) in the reference group. By positioning the IT-system/function in the plane, the user could identify to which quadrant the own organisation belonged to with regard to this IT-system/function: a) the lower right corresponded with innovators/early adopters, b) the upper right with an advanced majority, c) the upper left with an unprogressive majority and d) the lower left with laggards or technology grouches. These terms were chosen referring partly to Rogers’ Diffusion of Innovation theory [33]. If a hospital was very advanced with regard to having implemented a specific system and if only few other hospitals in this group shared the same implementation status this hospital was called an early adopter (lower right). If the hospital had not implemented a specific system and there were more hospitals with a better implementation status then this hospital lagged behind (lower left). The two upper quadrants identified majorities, i.e. an implementation status that was shared by 50% or more. In the upper right the majority was advanced, and in the upper left the majority was rather behind. The hospital in Figure 5 was an early adopter within its reference group in terms of the electronic nursing record system and the order entry system (medication); it lagged behind in terms of the electronic surgery and anaesthesia record systems.
firms and corroborates the initial goal of providing independent results.

4.3 Repeatability and International Use

There is good evidence on the repeatability of the benchmark as results from a current call for participation indicate. A total of 182 hospitals out of 260 participants of the 2013 survey expressed their interest to take part in the benchmark. Forty participants of the 180 hospitals had been already benchmarked two years ago. The current procedure includes dedicated process indicators in combination with structural and functional performance indicators.

The IT-benchmarking procedure we are proposing is also suitable for other countries in our view. The standardised questionnaire has already been utilized in surveys outside Germany, i.e. in Austria in 2007 and 2009 [30, 34] and in The Netherlands in 2011 [35] and served to compare the adoption of health IT in two countries.

4.4 Validity of the Performance Indicators

The IT-performance indicators in our study were selected in accordance with the literature, mainly based on concepts as “system capability in terms of tasks” to denote “system quality” [18], “functional sophistication” and “integration sophistication” [23]. Several of the raw performance indicators, i.e. HIS architecture, full time equivalents (FTEs) per bed, IT-budget trend and data capture were identical or very similar to the ones studied by Hübner-Bloder and Ammenwerth [17]. All items were chosen – in cooperation with chief information officers – to represent the wide spectrum of functions/systems, key items reflecting integration and organisation. This hints at the content validity of our measurement instrument.

4.5 Performance Indicators and Visualisation

We utilised a very large number of performance indicators, which resulted in a large number of single indicator frequency views per hospital. These views were not only very detailed with regard to the many raw indicators but also with regard to providing frequency information derived from the reference group. Based on suggestions made by participants we developed more aggregated views in which similar indicators were grouped: one view focuses on the distance to the best of the group (multiple indicator distance view), and the other on the identification of systems and functions in which the participant belongs to the leading edge and those where they do not (multiple indicator innovation view). These views will have to be fully evaluated in the next series of benchmarks.

The performance indicators used in our trial primarily embrace structural features of the organisation and functional characteristics of the IT-systems. Knowing that the structure is not the only constituent of quality [36] it also becomes clear that other features need to be measured – first and foremost the degree with which IT supports and optimises clinical processes and eventually how IT contributes to improving clinical outcomes.

In order for indicators and views to be actionable in very practical terms, the challenge will be to design them carrying meaningful and comprehensive messages while still keeping them simple enough.

4.6 Guidance

Measuring performance indicators and benchmarking institutions serves the ultimate goal of knowing what to do in order to become better. Our benchmarks yield exact details about the where, i.e. technical, human and financial resources, in which there is the greatest need for action. Statistical benchmarks, however, also allow for advice on the how. The multiple regression analyses employed to constituting the reference groups via context factors (“hard” factors) gave additional information on “soft” factors, i.e. circumstances that can be changed in a given time. Those hospitals that had implemented the largest number of IT-systems had IT-departments, were reference customers (priority customer), i.e. entertained a good relationship with their main IT-vendor, and had a strategic IT-plan [32]. Drawing on these results the global advice for hospitals that want to change is to check their organisation in terms of these features. Mutual exchange between the IT-department of the organisation and the IT-vendor lays the foundation of an innovation partnership: New products are made available easily to the hospital, and can be tested and adjusted according to the customer’s input. If there is a strategic IT-plan implementing new systems does not happen ad hoc but is embedded into a portfolio of new projects and into a clear understanding of how new systems contribute to reaching overall strategic goals of the organisation. It also may be assumed that IT-investments can be pushed through more effectively in the case of a strategic IT-plan.

4.7 Limitations

The main limitation of this benchmark is its focus on structure and function. We are, therefore, currently developing indicators for the quality of IT accompanying clinical processes. Based on this additional information we will measure the structural, functional and workflow parameters together. They will serve as input for the calculation of a workflow composite score (WCS) that is scrutinized in terms of its consistency, criterion and construct validity. Although asking CIOs warrants high quality data on technical issues CIOs and other members of the IT-staff can give no evidence on how IT is perceived by clinicians and if they value its contribution to patient care. More comprehensive benchmarks will have to include their perspective as well, in particular when it comes to scrutinising clinical information and processes. Their view would definitely enhance the credibility and validity of the information base. A complete evaluation will be necessary that ideally includes all benchmark participants.

In general, statistical benchmarking is based on the quality of the entire sample and a bias due to self-selection remains a challenge.

Statistical benchmarks of the type we are proposing will probably also always have problems taking into account the idiosyncrasies of individual organisations. If a benchmark needs to achieve an in-depth analysis of differences between the
peers then other types of benchmarks seem more appropriate, i.e. with a small number of participants and semi-structured interviews. We therefore argue the case for multiple approaches to benchmarking depending on the goals, the environment and the resources available.

5. Conclusion
The proposed IT-benchmarking procedure meets the requirements of independency, scalability and repeatability, and adds a new approach to existing procedures. However, processual performance indicators need to complement the structural indicators. Furthermore the various views will have to undergo rigorous testing and new views will have to be designed. More multivariate statistical analyses should be employed to identify further combinations of factors that are associated with IT-performance and allow a systematic and evidence based improvement of current practice. If an organisation, however, wishes to change its structure this benchmark already yields valuable information on detailed targets, e.g. implementation status of particular IT-systems/functions, and on global measures, e.g. establishing a strategic IT-plan.

References