Data quality issues impede comparability of hospital treatment delay performance indicators

L.M. Verweij · J. Tra · J. Engel · R.A. Verheij · M.C. de Bruijne · C. Wagner

Abstract

Aim To assess the comparability of five performance indicator scores for treatment delay among patients diagnosed with ST-segment elevation myocardial infarction (STEMI) undergoing primary percutaneous coronary intervention in relation to the quality of the underlying data.

Methods Secondary analyses were performed on data from 1017 patients in seven Dutch hospitals. Data were collected using standardised forms for patients discharged in 2012. Comparability was assessed as the number of occasions the indicator threshold was reached for each hospital.

Results Hospitals recorded different time points based on different interpretations of the definitions. This led to substantial differences in indicator scores, ranging from 57 to 100% of the indicator threshold being reached. Some hospitals recorded all the required data elements for calculating the performance indicators but none of the data elements could be retrieved in a fully automated way. Moreover, recording accessibility and completeness of time points varied widely within and between hospitals.

Conclusion Hospitals use different definitions for treatment delay and vary greatly in the extent to which the necessary data are available, accessible and complete, impeding comparability between hospitals. Indicator developers, users and hospitals providing data should be aware of these issues and aim to improve data quality in order to facilitate comparability of performance indicators.

Keywords Acute coronary syndromes · Quality indicators · Data quality · Hospital information systems

Introduction

Assessment of the quality of care by means of performance indicators is an integral part of modern day health care. Performance indicators are a tool in quality improvement and provide the government, physicians, patients, scientific society and insurance companies an indication of hospital performance, which is increasingly demanded [1]. As comparing performance indicator scores between hospitals can have major consequences, including lay press ranking lists and government and insurance company sanctions, performance indicator scores need to be comparable.

There are several steps in the process that leads from an event happening in clinical practice to a performance indicator intended to measure the performance of a clinical practice regarding that event [2]. This process is illustrated in Fig. 1. Variations in any of these steps will lead to different performance indicator scores. Ideally, data recorded for performance indicators are based on sound clinical practice guidelines, in which the definitions and inclusion and exclusion criteria of the performance indicator are clear and unambiguous and then processed in a uniform way to calculate the performance indicator in a uniform way. In reality, however, definitions are far from unambiguous and data are recorded in a variety of ways, impeding comparability of indicators for external quality control [3, 4]. This means that users of performance indicators need to be aware of the possible impact of variations in definitions and quality of...
the data in terms of availability, accessibility and completeness [5, 6]. The more unambiguous the definitions and the higher the quality of the underlying data, the more likely the performance indicator scores will be accurate and consistent between hospitals [7].

For patients diagnosed with ST-segment elevation myocardial infarction (STEMI), international guidelines recommend timely invasive treatment by primary percutaneous coronary intervention (PCI), generally within 90 min of first medical contact [8, 9]. Delays in timely invasive treatment by PCI caused by, for example, residential distance rapidly decrease the benefits over alternative treatments [10, 11], while shortening delays has the potential to contribute to decreased heart failure and mortality [12, 13]. It is, however, unclear to what extent the treatment delay indicator scores are comparable between hospitals. This study therefore aims to investigate to what extent variations in definitions influence performance indicator scores. Moreover, we investigate to what extent the quality of data in terms of availability, accessibility and completeness influences performance indicator scores. We conclude by providing recommendations for improving comparability of performance indicator scores.

### Methods

**Patient data**

Secondary data were used from two university hospitals and five tertiary teaching hospitals performing PCI participating in the acute coronary syndromes (ACS) program evaluation, within the larger national safety management program: ‘VMS safety management program’ [14].

Data from these seven hospitals were collected manually by six chart abstractors using standardised case report forms. All abstractors had a background in research and received instructions for the chart review procedures by JT and JE. The chart abstractors collected data by means of retrospective review of the medical records in electronic or paper-based medical, nursing or catheterisation laboratory records of patients discharged between 1 January and 31 December 2012. Each month, eligible records of patients discharged in the preceding month were selected from the hospital billing system using the diagnosis treatment combination code. To determine the STEMI population, chart abstractors first considered all the records of patients diagnosed with ACS for inclusion. Next, the chart abstractors checked whether the discharge letter confirmed the ACS diagnosis. When the discharge diagnosis was unclear, the record was discussed with a cardiologist or other attending physician working in the field of cardiology. Charts of patients with a treatment delay not exceeding 6 h were included in the study [15]. Charts of patients without a discharge diagnosis of STEMI, those not undergoing an acute PCI, patients with secondary ACS (e.g. due to anaemia), those undergoing elective procedures, patients with missing or uninformative charts and the charts of patients under the age of 18 years were excluded from the study. Chart abstractors signed a confidentiality agreement and all data were stored on a password protected network server of the VU University Medical Centre.

**Quality indicator definitions**

Five definitions for the treatment delay indicator were derived from literature (Table 1 and Fig. 2): (A) The Dutch ‘VMS safety management program’ guidelines [14]; (B) The adjusted Dutch ‘VMS safety management program’ evaluation [14]; (C) The mean door-to-needle time [15]; (D) The door-to-balloon time (American ACC/AHA guidelines for the management of STEMI [9, 16]); and (E) The European Society of Cardiology guidelines for the management of STEMI [8]. In these five definitions, treatment delay was defined as: (A) PCI within 90 min of first medical/paramedical contact; (B) PCI within 90 min of first electrocardiogram (ECG); (C) the mean door-to-needle time (no threshold provided); (D) PCI within 90 min of hospital arrival, and (E) PCI within 90 min after first medical contact. The B definition is an adaption of the A definition, because the time of first medical/paramedical contact was not registered consistently in all PCI centres but the time of the first ECG was. Thus, for this study, treatment delay was defined as the time from first ECG to PCI. Noteworthy is further that indicator C asks for the mean door-to-needle time, illustrating that different organisations ask hospitals to register different information. Moreover, although none of the PCI centres registered the

![Fig. 1 Comparability of data: flow from collection to interpretation.](image-url)
time of wire passage in the culprit artery, which is used by
the ESC in the last definition, we provide this definition as
an illustration because these guidelines provide the basis
for the first and second definitions. For this study, we regarded
the time from first ECG to PCI as the reference standard for
pragmatic reasons. We emphasise that this definition is not
a gold standard as there is no common gold standard for
measuring treatment delay due to national and international
differences and differences in perceptions of stakeholders.
Influence of definitions on indicator scores

To investigate the influence of the performance indicator definition on the scores, we calculated the percentage of patients for whom the treatment delay indicator was below the threshold for each hospital according to the different definitions.

Results

Patient data

Secondary data were used from two university hospitals and five tertiary teaching hospitals performing PCI. The bed capacity in these hospitals ranged between 400 to over 1100. Initially, 4471 records were reviewed for inclusion. After excluding records of patients who were not diagnosed with STEMI or excluded based on exclusion criteria (n = 3454), 1017 records were available for analyses, ranging between 112 and 236 included records per hospital.

Outcome measures

Data quality

The chart abstractors reported that some hospitals recorded all the required data elements for the calculation of the performance indicator scores. Moreover, automated access to these data was not possible in most cases. The most common ways to access the data were manual or partly automated access (four of the seven hospitals). Fully automated access was not available for any of the data elements, illustrating that data collection was time consuming and costly.
For all available and accessible data, we noted where this information was found (Table 2). For the extraction of data elements with partly automated or manual access, the chart abstractors had to review a combination of medical records, nurse records, discharge letters, electrocardiograms (ECG), procedure letters, correspondence with other health care professionals, and in paper form, scanned or in hospital information system. Table 2 illustrates that the accessibility of data did not only differ per hospital, but also per time point within hospitals.

The completeness of the available information is illustrated in Fig. 3. In 24% of patients the time of first contact was recorded, in 88% of the patients the time of ECG, in 51% of patients the time of arrival at the PCI centre, in 94% of patients the time of sheath insertion and in 64% of patients the time of first intervention was recorded. Thus, hospitals vary greatly in completeness of recording, particularly with respect to the time of first contact.

**Influence of definition on indicator scores**

Table 3 shows the percentage of patients satisfying the indicator threshold for each of the definitions and each of the hospitals. Indicator score B was reported best, with 15–50% missing data across hospitals. Missing data on indicator scores A, C and D were generally over 50% ranging from 21 to 100%. When calculable, indicator scores ranged from 57 to 100% within a given hospital, dependent on the indicator definition.

**Discussion**

This study illustrates that hospital performance indicator scores for the treatment delay performance indicator are largely incomparable, without laborious manual review.

Three reasons contribute to this incomparability. First, definitions vary for treatment delay performance indicators across the literature, which leads hospitals to vary in the extent to which different time points are recorded and/or used for calculating performance indicators. These differences are also due to the low number of patients and missing data. This is partly due to the choices hospitals make regarding which times to record, but also due to the format in which organisations compel hospitals to report indicators (as percentage or mean). To compare indicator definitions among patients with all data points would be a methodologically sound method. In practice, information is not available for all the data points in any of the patients, as hospitals use different definitions for treatment delay and vary greatly in the extent to which the necessary data are available, accessible and complete. So, this leads to substantially different indicator scores, especially between definitions A and B versus D. Second, the chart abstractors reported that some hospitals had all the required data elements for calculation of the performance indicators and data could not be retrieved easily in any of the hospitals. Moreover, data accessibility not only varied between hospitals, but also between data elements within hospitals. The same hospital could therefore have a relatively low indicator score following one definition and relatively a high score following another definition. Third, we found large variations between hospitals in completeness of time records.

Previous studies on the comparability of medical data in the Netherlands and across Europe similarly showed that required data elements for performance indicators were generally poorly available, accessible and incomplete [3, 16, 17, 18]. This may partly be due to the enormous number of indicators hospitals have to report on for external quality control. In order to compare indicator scores among hospitals it is thus necessary to standardise definitions and record...
In order to prevent incomparability in the future. Hospital associations in the Netherlands are now working on these steps. Despite the lack of solutions, we feel it is important to inform practice of the critical notion that hospital performance indicator scores for the treatment delay performance indicator are largely incomparable, without laborious manual review.

Our study has several limitations. The time points extracted to calculate indicator scores per hospital may be an overestimation of data completeness compared with indicator scores calculated and supplied by hospitals themselves, because data were extracted by chart abstractors who went to great lengths to obtain data. Moreover, the data obtained by our chart abstractors may deviate from hospital data as the chart abstractors made decisions to clarify which data were necessary to calculate performance indicator scores, such as manually checking all diagnoses in the discharge letter based on the diagnosis and procedure codes. Also, the presence of researchers collecting data on site and the provision of feedback of performance may have influenced documentation of times and performance indicator scores. However, as the patient safety program for which the data were primarily collected was designed to improve guideline adherence and provide hospitals with feedback of their own performances, it would not be appropriate to withhold this information. Consequently, another limitation is the secondary use of data that were obtained for the goal of measuring guideline adherence. For example, the exclusion of uninformative charts means that data were preselected on their quality. In spite of these limitations, our results show that the comparability of indicator scores is influenced by data quality issues.

**Conclusion**

In sum, hospitals use different definitions of this one particular quality indicator and vary greatly in the extent to which the necessary data are actually available, accessible...
### Table 3

| Hospital (number) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Total (patient with missing data % and indicator reached n) |
|-------------------|---|---|---|---|---|---|---|----------------------------------------------------------|
| Definition of treatment delay | [ ] | [ ] | [ ] | [ ] | [ ] | [ ] | [ ] |----------------------------------------------------------|
| A. Dutch ‘VMS safety management program’ guideline | 82% | 93% | 86% | 44% | 69% | 99% | 71% | 85% |
| B. Adjusted Dutch ‘VMS safety management program’ guideline | 13; 23 | n/a | 10; 16 | 36; 63 | 40; 56 | n/a | 49; 69 | 156; 236 |
| C. Mean door-to-needle time (IGZ) | 91% | 83% | 100% | 87% | 86% | 83% | 93% | 50% |
| D. Door-to-balloon time (ACC/AHA) | 24% | 100% | 100% | 86% | 21% | 100% | 91% | 74% |
| Total (n of patients) | 127 | 120 | 112 | 112 | 171 | 139 | 236 | 1017 |

IGZ Dutch Health Care Inspectorate, n/a data for indicator not available or fewer than 10 cases

*Indicator asks for mean door-to-needle time.

and complete, impeding comparability between hospitals. It is important to increase awareness among developers, users and producers of performance indicators regarding the impact of variations in indicator definitions and data quality on indicator scores.

**Acknowledgements** The study was funded by the Dutch Ministry of Health, Welfare and Sport. We acknowledge Suzanne Vonk for helping to interpret the data.

**Conflict of interest** The authors declare that they have no conflict of interests.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

### References

1. Brook RH, McGlynn EA, Cleary PD. Measuring quality of care. N Engl J Med. 1996;335:966–70.
2. Khan NA, McGilchrist M, Padmanabhan S, et al. Data quality tool. Deliverable of the TRANSFoRm project (Translational Research and Patient Safety in Europe). Utrecht: NIVEL; 2013.
3. Anema HA, Kieviit J, Fischer C, et al. Influence of hospital information systems, indicator data collection and computation on reported Dutch hospital performance indicator scores. BMC Health Serv Res. 2013;13:212–22.
4. Kringsos DS, Anema HA, ten Asbroek AHA, et al. Beperkt Zicht. Onderzoek naar de betrouwbaarheid, validiteit en bruikbaarheid van prestatie-indicatoren over de kwaliteit van de Nederlandse ziekenhuiszorg. [Research on the reliability, validity and usefulness of performance indicators on the quality of Dutch hospital care]. Amsterdam: AMC; 2012.
5. Mainz J. Defining and classifying clinical indicators for quality improvement. Int J Qual Health Care. 2003;15:523–30.
6. Williams SC, Watt A, Schmalz SP, et al. Assessing the reliability of standardized performance indicators. Int J Qual Health Care. 2006;18:246–55.
7. Wollersheim H, Hermens R, Hulscher M, et al. Clinical indicators: development and applications. Neth J Med. 2007;65:15–22.
8. Steg PG, James SK, Atar D, et al. ESC Guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation: the Task Force on the management of ST-segment elevation acute myocardial infarction of the European Society of Cardiology (ESC). Eur heart J. 2012;33:2569–619.
9. O’Gara PT, Kushner FG, Ascheim DD, et al, American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. 2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. Circulation. 2013;127:362–425.
10. Boersma E, Primary Coronary Angioplasty vs. Thrombolysis Group. Does time matter? A pooled analysis of randomized clinical trials comparing primary percutaneous coronary intervention and in-hospital fibrinolysis in acute myocardial infarction patients. Eur Heart J. 2006;27:779–88.
11. Postma S, Dambrink JH, Boer MJ de, et al. The influence of residential distance on time to treatment in ST-elevation myocardial infarction patients. Neth Heart J. 2014;22:513–9.
12. Cannon CP, Gibson CM, Lambrew CT, et al. Relationship of symptom-onset-to-ballooning time and door-to-balloon time with mortality in patients undergoing angioplasty for acute myocardial infarction. JAMA. 2000;283:2941–5.
13. Luca G De, Suryapranata H, Ottervanger JP, et al. Time delay to treatment and mortality in primary angioplasty for acute myocardial infarction: every minute of delay counts. Circulation. 2004;109:1223–5.
14. Tra J, Engel J, Wulp I van der, et al. Monitoring guideline adherence in the management of acute coronary syndrome in hospitals: design of a multicentre study. Neth Heart J. 2014;22:346–53.
15. IGZ. Basisset kwaliteitsindicatoren ziekenhuizen [Dutch Health Inspection. Quality indicators for hospitals]. 2012. http://www. igz.nl/Images/Basisset%20kwaliteitsindicatoren%20ziekenhui zen%202012–2_tcm294–305782.pdf. pp. 86–9. Accessed 26 February 2015
16. Krumholz HM, Anderson JL, Bachelder BL, et al. ACC/AHA. 2008 performance measures for adults with ST-elevation and non-ST-elevation myocardial infarction. Circulation. 2008;118:2596–648.
17. Anema HA, Veer SN van der, Kieviit J, et al. Influences of definition ambiguity on hospital performance indicator scores: examples from The Netherlands. Eur J Public Health. 2014;24:73–8.
18. Groene O, Kristensen S, Arah OA, et al, DUQuE Project Consortium. Feasibility of using administrative data to compare hospital performance in the EU. Int J Qual Health Care. 2014;26:108–15.
19. Stroetmann VN (Ed.), Kalra D, Lewalle P, et al. Semantic Interoperability for Better Health and Safer Healthcare. Deployment and Research Roadmap for Europe. European Communities, European Commission, Information Society and Media Directorate-General, Belgium 2009.
20. NFU. Kwaliteitsvaluatie door verbeterde zorg documentatie. [Netherlands Federation of University medical centers. Quality evaluation by improved care registration]. Utrecht: NFU; 2013.
21. Dentler K, Cornet R, ten Teije A, et al. Influence of data quality on computed Dutch hospital quality indicators: a case study in colorectal cancer surgery. BMC Med Inform DecMak. 2014;14:32–3.

22. Cimino JJ. Collect once, many: enabling the reuse of clinical data through controlled terminologies. J AHIMA. 2007;78:24–9.
23. Blumenthal D, Tavenner M. The 'meaningful use' regulation for electronic health records. N Engl J Med. 2010;363:501–4.
24. Herrett E, Shah AD, Boggon R, et al. Completeness and diagnostic validity of recording acute myocardial infarction events in primary care, hospital care, disease registry, and national mortality records: cohort study. BMJ. 2013;346:f2350.
25. Chung S-C, Gedeborg R, Nicholas O, et al. Acute myocardial infarction: a comparison of short-term survival in national outcome registries in Sweden and the UK. Lancet. 2014;383:1305–12.