The validity of indicators for assessing quality of care: a review of the European literature on hospital readmission rate

Claudia Fischer1,2,*, Helen A. Anema1,*, Niek S. Klazinga1

1 Department of Public Health, Academic Medical Center, University of Amsterdam, NL-1100 DD, Amsterdam, The Netherlands
2 Department of Public Health, Erasmus Medical Centre, NL-3000 CA, Rotterdam, The Netherlands

*These authors contributed equally to this work.

Background: Quality indicators are increasingly being implemented in Europe for policy and management purposes. Many of these indicators were initially developed and implemented in the USA. However, the suitability of directly adopting indicators that have been developed in a different health care system can be questioned. Therefore, we investigate the validity behind the readmission rate indicator in the European setting. Methods: A systematic literature study was conducted to identify the status of scientific research on the validity of this indicator (January 1999 and April 2010). Descriptive information as well as information on the data source, indicator definition, risk adjustment factors, and conclusions was assessed. Results: The majority of the 486 included studies focused on the actual use of the indicator as an outcome measure in European countries. Only 21 studies specifically addressed its validity, or important prerequisites of validity. There is little consensus over the time-frame used to calculate the indicator, the type of readmission that is included, and the case-mix adjustment applied. Conclusions: Despite the increase in Europe of the use of the readmission rate as a measure of quality of care, the amount of research performed on its validity is scarce. Those studies that report on validity replicate earlier, mainly US findings (<1999) of methodological problems and express reservations on its large-scale use. The readmission rate as an indicator should be used with care. Users should address issues related to definition, time-frame and case-mix adjustment as part of the process to enhance validity in the European settings.

Introduction

Recent transformations in European (EU) health care systems, such as the change from government-driven health care to a free market system, ask for an objective assessment of the quality of hospital care. Performance indicators, measurement tools that assess a particular health care structure, process or outcome, are generally expected to provide such objective assessments of health care quality. However, in order to do so, the structures (e.g. organization of care), processes (what has been done to a patient) and outcomes (final health status) that are measured should be interrelated, and should be informative about the same underlying construct.

An important starting point in quality of care research has its origin in the USA in the 1970s. Here, the interest in quality of care increased substantially when the Healthcare Cost and Utilization Project (HCUP) developed a set of health care quality indicators in 1989.3 In those years, a substantial amount of knowledge became available on the difficulties that arise when developing and maintaining valid and reliable health care quality indicators.4 Ashton et al., for example, aimed to resolve the (at that time current) controversy in the USA over the validity of early readmissions as a measure of quality of care. Despite the results being somewhat conflicting, the summary odds ratios indicated that the process of medical care does affect the risk of readmission within 30 days. It was suggested however to handle the reported figures cautiously, as a null effect could not be ruled out.

In many health care systems the rate of readmissions (RR) to a hospital has become a well-known indicator which measures how many patients are readmitted to the hospital after they have been discharged. The rationale behind monitoring readmissions is that a readmission is related to substandard care. Within the Donabedian model, this indicator can be considered an intermediate outcome indicator, as it is a proxy for the rate of adverse events or positive outcomes such as increased life expectancy or reduced morbidity (genuine outcome indicator). The fact that hospital readmissions are high in cost,6 a burden to patients,7 and can be easily computed from routine statistics of administrative databases,8 makes it plausible that readmission scores are monitored. Indeed, over a decade ago several studies, mostly performed in the USA, reported that readmissions were valid measures of quality as they appear to be largely caused by substandard care received during the prior hospitalization.9–12 However, other studies failed to confirm a valid relation between the quality of care and the RR.13,14 In all it seemed that the validity of the RR as a measure of health care quality was still debated.

Within the last decade, Europe followed up on the USA’s example and several countries implemented performance indicators that were originally developed there. Since local health care factors, such as the proximity of the hospital and availability of beds,15,14 influence the probability of a readmission, and European health care systems are substantially different from that of the USA (both at the governmental and hospital level) the validity of the RR might be compromised. Our aim is to explore the validity behind the RR indicator in the European setting. The validity is assessed by looking at the consistency in the use of definition of the indicator, the readmission time-frame used and the use of case-mix adjustment.

Methods

Search strategy and selection criteria

A systematic literature search was conducted in the electronic databases Medline and PubMed, using the following keywords in various combinations: ‘re-admission’, ‘readmission’, ‘rehospitisation’, ‘re-hospitalization’ and MEDLINE subject heading (MeSH term) ‘patient readmission’, present in either the title and/or the abstract. The search was limited to publications from Europe and to the time period from January 1999 to April 2010. An English abstract was required to be present in order to be able to include or exclude the study. After we screened the abstracts of included studies, two groups of articles that differed largely in magnitude were identified: (i) studies...
using the RR as an outcome indicator and (ii) studies testing the indicators validity.

**Procedure for classification and evaluation**

The classification and evaluation process was performed by two independent researchers, who discussed any disagreement. If necessary, a third researcher was consulted. From all included articles, the following data elements were extracted: Year of publication, country of affiliation of the corresponding author and the patient group/disease being focused on. Further, we checked whether they define the RR they are focussing at, and how they define it. To provide more insight into current opinion on the validity, we screened the articles focusing on the validity of the quality indicator in more depth using information from the title, the abstract and full text. The following data elements were abstracted: type of validity investigated, study design used, type of data sources, type of case mix adjustment and the conclusion. We distinguished between three types of validity: (i) face validity, the extent to which the measure appears to assess the construct, (ii) construct validity, the degree to which the measure reflects the construct and is related to other variables in predicted ways, (iii) criterion validity, the validity that relates to the ability of a measure to predict an outcome (criterion), to test it, this measure ideally gets evaluated against a 'gold standard'.

**Results**

The search identified 1062 publications (see figure 1 for flow diagram). After the initial screening, 552 articles were excluded on the basis of our inclusion criteria or a missing full English text. Of the 207 studies which did not provide full texts, most studies were published in Germany ($n = 55$), Spain ($n = 38$), France ($n = 29$) and Denmark ($n = 25$). The remaining 510 articles were screened on basis of the full text, 24 articles were reclassified. The resulting 486 articles were evaluated by the reviewers of which 465 articles focused on the actual use of the indicator as an outcome measurement and 21 studies somehow addressed the validity of the indicator, see for further bibliometric results Figure 2A, B and C.

**The RR is a poorly defined but increasingly used outcome measure**

Of the 465 articles which use the RR as an outcome measure in their study, we found 288 (partly) defining what they mean by RR. 177 studies just stated that they used the RR as an outcome measure, without defining it. 263 articles out of the 288 defined the timeframe they were investigating. 13 studies looked at readmissions within 14 weeks, while 80 studies used the timeframe 'within 30 days'. However, most of the articles applied 'longer than 30 days' ($n = 166$). Those who defined the type of RR (70 studies out of the 288) used most often 'unplanned/emergency readmission' ($n = 51$). Just 45 studies stated both, the timeframe and type of readmission they included.

In the following paragraph, the 21 validity studies that address the quality of the indicator will be the topic of discussion. The studies will be discussed in regards to the following aspects: definition, type of validity and adjustment on case mix.

As is described above, 21 European articles could be identified that address validity or validity aspects of the RR. Seven studies were included on basis of their abstract in which they stated to investigate the validity of the indicator by using the term validity or validation. $^{16-22}$ The rest was identified on basis of the full-text information. In that case, the authors inferred the validity topic based on whether they questioned validity related factors, such as the influence of case-mix/ hospital factors, or the influence of different definitions on the RR score.

Data sources used varied. Three validity studies used data from an administrative database. Of those three, one study used the MODCOD system (encoded reports), allowing to compute DRGs according to ICD-9-CM codes (CPHA, 1979) and the UNIDOC system, which is an integrated patient report processor. $^{23}$ One study used a hospital information database (PAS) $^{24}$ and the third used the hospital’s 'Clinicom' patient administration system with an in-built 28-day re-admission search tool. $^{25}$

Other data source examples are: data collected by statistic offices ($n = 2$), $^{16,17}$ hospital discharge data ($n = 1$), $^{18}$ hospital information systems ($n = 1$), $^{26}$ electronic coding systems ($n = 1$) $^{27}$ and cancer registries ($n = 1$). $^{19}$ Two studies did not provide details about the source of the studied data. $^{28,29}$

**Different ways to address validity**

Diverse aspects of an indicator’s validity can be addressed (such as face-, construct-, criterion validity or validity threatening factors such as the time window used, hospital and patient characteristics). The criterion validity of an indicator can be addressed in several ways, such as investigating the predictability of an outcome indicator, or directly comparing performance which is revealed by the performance indicator to a gold standard (expert judgement of care based on record review).

In our review, we observed in total 14 studies that investigated the validity of the RR (13 criterion validity 1 face validity, $^{30}$ no construct validity study could be identified) and five studies that did not specifically investigated or discuss valid relations between the RR and other measure of quality of care but investigated factors that might influence validity. $^{22,31-34}$ (see table 1 for summary).

The identified criterion studies are discussed below in more detail. Of the 13 criterion validity studies, 11 studies used expert judgments as a gold standard. The quality of care was merely judged by whether a re-admission was unplanned/emergency ($n = 3$), $^{17,25,24}$ unforeseen and avoidable ($n = 1$) $^{16}$ could be avoided ($n = 1$) $^{18}$ or unplanned and avoidable ($n = 3$), $^{25,26,35}$ or was caused by a complication that was likely related to the surgery or whether it was provoked by the patient himself ($n = 3$). $^{27,29}$ In two studies, the gold standard was provided by either an expert judgment about aspects of quality of care based on information from an electronic full text discharge summary, $^{26}$ or on process indicators that were derived from evidence based guidelines and were calculated from medical chart information. $^{20}$ In the latter study, the authors specifically stated that these process indicators were used as gold standard.
Finally, one study validated a National Patient Registry against corresponding data from the review of medical records serving as the gold standard.\(^{19}\) In regards to the study purpose, four studies were set-up to investigate whether the indicator is informative of insufficient care,\(^{18,20,27,29}\) or if the underlying data source is suitable for calculating the RR \((n = 9)\).\(^{16,17,19,23–26,28,35}\)

This paragraph refers to the 11 studies investigating the validity threatening factors or the feasibility of the RR.\(^{16,17,27,29,31–34}\)

Eight studies\(^{16,17,27,29,31–34}\) investigated validity threatening factors such as the time window used to determine the readmissions, and the factors that influence readmissions like patient and hospital characteristics. The other three studies merely focused on the feasibility of calculating the indicator,\(^{21,36}\) or on the validity of linking multiple admissions to one patient without the presence of unique patient identifiers.\(^{22}\)

Finally, one study merely discussed general validity issues of the RR.\(^{30}\)

---

Figure 2 (A) number of publications per country in absolute numbers, (B) number of publications per patient group or disease field and (C) Number of publications per year
<table>
<thead>
<tr>
<th>Study aim</th>
<th>Design</th>
<th>Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>To discuss the validity of RR. [n.a.]</td>
<td>Opinion paper</td>
<td>RR is an unsatisfactory performance indicator since it seems difficult to select only the unplanned and avoidable RAs, current health care databases are not sufficient to detect readmissions to other hospitals, and there is no uniform time window in which RA is measured.</td>
</tr>
<tr>
<td>To compare databases to determine which information is needed to distinguish planned from unplanned readmission.</td>
<td>Retrospective</td>
<td>Electronic reports based on diagnostic and procedure codes alone are not sufficient to distinguish planned from unplanned readmission, automation of detailed (full text) clinical databases seems promising.</td>
</tr>
<tr>
<td>To investigate possible factors influencing the readmission rate: frequency of previous admission and cause of readmission.</td>
<td>Comparative control study</td>
<td>RR is unlikely to be a valid outcome indicator, until better routine data for standardization by case mix is available.</td>
</tr>
<tr>
<td>To determine relation between early unplanned readmissions of heart failure patients and suboptimal in-hospital care, or whether RA relates to clinical and demographic characteristics of the patient.</td>
<td>Case-control study</td>
<td>RR is not a valid measure of the quality of care for heart failure patients; unplanned readmissions are more strongly associated with clinical and demographic characteristics than with quality of care. Case-mix adjustment is needed.</td>
</tr>
<tr>
<td>To compare views of general practitioners and hospital staff on the reasons for unplanned readmission of elderly people.</td>
<td>Observational research</td>
<td>No specific measure for agreement was given. GP's and hospital staff judged most readmissions to be caused by a relapse or complication of index illness. Opinions differed most significantly when the reason for readmission was poor health on discharge or inadequate preparation on discharge.</td>
</tr>
<tr>
<td>To develop a computerized method that is able to detect potentially avoidable hospital readmissions on basis of routinely collected data and to develop a prediction model that adjusts the RR for case mix factors.</td>
<td>Evaluative retrospective cohort study</td>
<td>Detection of avoidable RA by computerized method was scientifically sound enough to sign quality issues. Medical risk adjusters were more important than non-medical patient characteristics. But RR is sign to gather further information from medical records, and not for public reporting.</td>
</tr>
<tr>
<td>To examine reasons for re-admission, possible errors in coding and any preventable factors in acutely readmitted patients.</td>
<td>3 month retrospective audit</td>
<td>Comparing crude rates of readmission does not quantify number of avoidable re-admissions and is only useful as a sign to conduct local studies to determine avoidable readmissions.</td>
</tr>
<tr>
<td>To measure RR and to qualify the readmissions as planned, unplanned, avoidable or unavoidable.</td>
<td>Prospective study, explorative</td>
<td>No conclusions were drawn regarding the possibility to distinguish between planned, unplanned, avoidable or unavoidable. But, RR should be monitored regularly, with a time-frame of 30 days of discharge.</td>
</tr>
<tr>
<td>To identify the frequency and characteristics of potentially avoidable RA and to compare the quality of care derived from RR to the quality of care judged by experts.</td>
<td>Cross-sectional observational study</td>
<td>Most RAs are not avoidable thus RR cannot be considered valid indicators of the quality of care for the set of specialties in a general hospital, RR depends mainly on patients clinical condition and is not uniformly defined.</td>
</tr>
<tr>
<td>To measure the validity and predictive ability of readmission.</td>
<td>Evaluative retrospective cohort study</td>
<td>RR, calculated from routinely collected data was not a valid indicator of the quality of heart failure care. Improvement of definition and measurement methods are needed as well as appropriated risk adjustments.</td>
</tr>
<tr>
<td>To evaluate the usefulness of a computerised algorithm to identify potentially avoidable readmissions on the basis of minimum bias, criterion validity (expert judgment) and measurement precision.</td>
<td>Evaluative retrospective cohort study</td>
<td>Adjusted rates of potentially avoidable readmissions are scientifically sound enough to warrant their inclusion in hospital quality surveillance and a high rate acts as a signal to hospitals to evaluate their practices.</td>
</tr>
<tr>
<td>To assess the accuracy of hospital unplanned re-admission data, and identify patterns or possible causes of unplanned general surgical re-admissions.</td>
<td>Retrospective audit of case note records</td>
<td>A hospital coded information database, may not be accurate enough for the calculation of unplanned RR as it allows inclusion of unrelated admissions. Also, factors like age, sex, history of psychiatric disease, number of drugs on discharge could not be used to predict unplanned RA.</td>
</tr>
<tr>
<td>To formulate an appropriate definition of readmission, to investigate the demographics and the predominant cause of readmitted patients (avoidable RA) and to see whether rapid throughput is leading to unacceptably high RRs.</td>
<td>Retrospective observational study</td>
<td>Older patients with more complex care needs are more likely to be readmitted, rapid throughput of patients is not associated with readmission. RR needs to be interpreted with caution; it varies with changes in the inclusion criteria. However, on basis of expert judgment it appeared that most readmissions can be avoided with better quality care.</td>
</tr>
<tr>
<td>To investigate the validity of the Danish Cholecystectomy Database (DCD) by evaluating the association between Ps calculated from this database and post operative complications.</td>
<td>Evaluative retrospective cohort study</td>
<td>The DCD is a valid method for the monitoring of quality of care in cholecystectomy, however, correct coding is important, especially with administrative data. Length of stay &gt;3days and/or RA were strongly associated with post operative complications.</td>
</tr>
<tr>
<td>Validity factors</td>
<td>Auth.</td>
<td>Study aim</td>
</tr>
<tr>
<td>------------------</td>
<td>-------</td>
<td>-----------</td>
</tr>
<tr>
<td>Heggestad, Norway</td>
<td>To demonstrate the effects of different time intervals to calculate RR. (under investigation)*</td>
<td>Modeling approach</td>
</tr>
<tr>
<td>Lyratzopoulos, UK</td>
<td>To examine the effect of patient and disease factors on the risk of emergency medical readmission.</td>
<td>Prospective observational cohort study</td>
</tr>
<tr>
<td>Spiegelhalter, UK</td>
<td>To test different options for handling over-dispersion of performance indicators.</td>
<td>Retrospective analysis</td>
</tr>
<tr>
<td>Mason, UK</td>
<td>To identify suitable outcome measures for comparing gynecology performance between hospitals.</td>
<td>Descriptive feasibility study</td>
</tr>
<tr>
<td>Tromp, The Netherlands</td>
<td>To describe an efficient, generalizable approach to validate probabilistic record linkage results and to apply this approach to validate linkage of admissions of newborns. [n.a.]</td>
<td>Validation of probabilistic record linkage</td>
</tr>
<tr>
<td>Demir, UK</td>
<td>To develop a modeling approach to tackle the issue surrounding the appropriate choice of a time window as a definition of RA. (under investigation)</td>
<td>Modeling approach</td>
</tr>
<tr>
<td>Bottle, UK</td>
<td>To apply 10 of the AHRQ indicators for use in English routine hospital admissions data as the first step in validation, and describe their rates in relation to established measures of negative outcome.</td>
<td>Descriptive feasibility study</td>
</tr>
</tbody>
</table>

a: Inferences are made when authors do not describe topic clearly.
b: Time window larger than within 30 days.
c: Unplanned/emergency readmissions.
d: Time window within 30 days.
e: All readmissions.
f: Unplanned and avoidable.
g: Avoidable readmissions.
h: Unforeseen and avoidable.
i: Time window within 14 days. n.a. = not applicable, RR = Readmission Rate, RA = Readmission.
Validity of indicators for assessing quality of care 489

A definition that often lacks precision

The identified studies learned that the time windows and the criteria 'avoidable' and 'unplanned' were the main issues threatening the validity of this indicator. It seems that there is a large variability in the time range between discharge and readmission as it ranged from '14 days', to 'within 180 days'. The largest number of articles however, calculated readmissions within 30 days (28–30 d; n = 11). Two out of them specifically investigated the increase in risk to be readmitted for various patient dependent variables. First, Kossovsky et al. investigated heart failure patients in a case–control design and observed a significant increase of readmission risk for: previous diagnosis of heart failure, age, history of cardio revascularization. Odds ratio’s varied from 1.14 (poor readiness for discharge) to 4.1 (age, for patients >80 years). In addition the authors observed an association between readiness for discharge and a subsequent early readmission. The second study conducted a case–control design and observed the following significant predictors of all cause readmission: male sex, age, number of coded comorbidities, admission via GP referral (decrease in risk), primary diagnosis of heart failure and of chronic obstructive pulmonary disease or asthma, and higher level of deprivation. Hazard ratios varied from 0.93 (admission via GP referral) to 1.49 (>4 coded comorbidities).

Three studies investigated the possibility to apply satisfactory risk adjustment for patient-related factors, descriptively compared the rates between different medical specialties or statistically compared the rates between medical and patient characteristics. In all, there was some evidence for the influence of male gender, age, length of stay, previous hospitalization within 6 months, life threatening diseases that are prone to serious disability or complications (cancer, heart disease, high risk surgery) and disease categories such as nephrology and haematology. The last one addressed the effects of hospital factors on the variability of the indicator stays relatively behind in comparison with the total number of patients discharged alive within the reference pool. In our opinion, however, the observations reported in this review are of a robust character and it seems unlikely that inclusion of non-English literature would change the scope and conclusions. Also, as the research field of performance indicators is relatively young, sensitive search terms are not well established yet. Our experience showed that a combination of self-entered search terms and selecting relevant MeSH terms revealed the most appropriate articles. 

Different options for adjusting to patient and disease characteristics

Six studies examined to what extent patient characteristics, demographical, social and medical factors, affected the RR. Two out of them specifically investigated the increase in risk to be readmitted for various patient dependent variables. First, Kossovsky et al. investigated heart failure patients in a case–control design and observed a significant increase of readmission risk for: previous diagnosis of heart failure, age, history of cardio revascularization. Odds ratio’s varied from 1.14 (poor readiness for discharge) to 4.1 (age, for patients >80 years). In addition the authors observed an association between readiness for discharge and a subsequent early readmission. The second study conducted a case–control design and observed the following significant predictors of all cause readmission: male sex, age, number of coded comorbidities, admission via GP referral (decrease in risk), primary diagnosis of heart failure and of chronic obstructive pulmonary disease or asthma, and higher level of deprivation. Hazard ratios varied from 0.93 (admission via GP referral) to 1.49 (>4 coded comorbidities).

Three studies investigated the possibility to apply satisfactory risk adjustment for patient-related factors, descriptively compared the rates between different medical specialties or statistically compared the rates between medical and patient characteristics. In all, there was some evidence for the influence of male gender, age, length of stay, previous hospitalization within 6 months, life threatening diseases that are prone to serious disability or complications (cancer, heart disease, high risk surgery) and disease categories such as nephrology and haematology. The last one addressed the effects of hospital factors on the variability of the indicator stays relatively behind in comparison with the total number of patients discharged alive within the reference pool. In our opinion, however, the observations reported in this review are of a robust character and it seems unlikely that inclusion of non-English literature would change the scope and conclusions. Also, as the research field of performance indicators is relatively young, sensitive search terms are not well established yet. Our experience showed that a combination of self-entered search terms and selecting relevant MeSH terms revealed the most appropriate articles. 

The time window applied needs to be in accordance with the type of disease

The reported differences in whether RR is valid or not might be explained by the various ways that were used to measure the RR (e.g. time window used). Our review showed similar variety in the use of time windows as the number of days between discharge and readmission ranged from 14 days post-discharge to 180 days. Most validity studies calculate the RR on basis of early readmissions within 30 days, whereas the majority of the articles using the indicator as an outcome measure use a timeframe >30 days. According to Heggestad, it seems that the time used to calculate the readmission is largely dependent upon the reason why the RR is measured, and whether it is important to have a sensitive or a specific measure. Also, the optimal time window is largely dependent on the type of disease the patient was originally treated for.

Readmission needs to be avoidable and unplanned

Not only the time window varied, the type of readmission that is included varied as well, as some counted all and others counted only unplanned readmissions. The importance of this distinction is shown in the conflicting results regarding the validity of the RR, in studies that included all readmissions compared to those only including unplanned readmissions. Despite the fact that unplanned readmissions are more likely to be related to substandard care than planned readmissions, our review showed that in European publications this distinction is not always made. Perhaps this is caused by the difficulty in distinguishing between planned and unplanned readmissions. Kossovsky, indeed, concludes that an administrative database that consists of (discharge) diagnosis and procedure codes alone is not sufficient for that purpose. However, even the inclusion of unplanned readmissions might not be valid as unplanned readmissions could well be caused by a new infection not related to the previous hospital stay. Instead, it is argued to focus on potentially avoidable readmissions that are unforeseen at time of discharge. In all, on the basis of their literature appraisal, Rumball-Smith and Hider recommend that the RR can be defined as following: 'the number of patients who experienced unintended, acute readmission or death within 30-days of discharge from the index admission, divided by the total number of patients discharged alive within the reference pool'.

Case-mix factors have to be taken into account

Medical and demographical patient characteristics such as the severity and complexity of the diagnosis of the readmitted patient, the length of specifically defined how the RR was measured and even a smaller part of the studies used recommended unplanned readmissions. Exploring the validity studies, however, learned that the majority cast doubt on its validity in the European setting. The studies highlight substantial problems with respect to the adjustment of factors that are beyond medical control but increase the risk to be admitted, the degree to which avoidable readmissions can be accurately detected and the time window that is used to detect relevant readmissions.

Limitations

Some limiting aspects of the methods used in our current review require further discussion. First, the search was limited to PubMed and Medline, other databases were not addressed. Further, in our methods we mentioned that papers not providing English full texts were excluded. As such, the total number of validity studies presented in this literature review is an underestimation of the total number. In our opinion, however, the observations reported in this review are of a robust character and it seems unlikely that inclusion of non-English literature would change the scope and conclusions. Also, as the research field of performance indicators is relatively young, sensitive search terms are not well established yet. Our experience showed that a combination of self-entered search terms and selecting relevant MeSH terms revealed the most appropriate articles.
stay, and age and socioeconomic status, seem to influence the validity of the RR as well. The number of factors that could potentially affect the RR seems numerous, but sex, age, number of coded comorbidities and the disease that a patient was treated for are important factors to take into account when measuring the RR.

Conclusion

With this article, we hope to increase awareness of the methodological pitfalls of performance indicators and stimulate research activities on their validity in different national set-ups in Europe. Our literature study showed that a performance indicator of US origin, such as the RR, is increasingly used in Europe. However, the amount of research performed on its validity remained scarce. Those studies that do report on its validity replicate earlier findings, mainly from the USA (<1999), of methodological problems such as the lack of a uniform definition, the impact of case mix factors and the questionable reliability of the databases used to compute the RR.

It seems that a ‘best recipe’ to calculate the RR in a valid way does not exist. Clarke (2004) nicely summarizes the challenges that arise when calculating the RR. The author suggests stopping using this indicator altogether, particularly when it is calculated from routine data sources and used for the comparison of readmission rates between different hospitals on a macro (national) level. Many databases do not allow for the tracking of patients from one hospital to another, as unique patient identifiers are less frequently used within the health care context. As a consequence, patients that are readmitted to another hospital, something that particularly happens with patients that are dissatisfied with the care they received, are missed in the calculation. Walraven et al. conclude that the true proportion of potentially avoidable readmissions is simply unclear.

Implications

The validity of the measure has to be strengthened for its user’s purpose (i.e., managers, policy makers and developers). Moreover, the actual purpose of the indicator is important when studying its validity as each type of use (accountability, improvement, consumer information, pay-for-performance) may place different demands on the degree of validity. Further, it is suggested to pay attention to and enhance the indicators’ local validity.

In sum, performance indicators are not those easily obtainable measures of health care quality most users would like them to be. In fact, insufficient validity of performance indicators seems to be a more common problem as can be learned from the research field of the Hospitalized Standardized Mortality Rates.

Nevertheless, if the RR is measured as an indicator of quality of care, it is best to ensure that these readmissions are related to the index admission, are unplanned or even better, can be identified as avoidable. Secondly, the time window that is used to calculate the readmissions should be adapted to the type of care that is investigated and thirdly, the data used for calculating the indicator should have undergone reliability analysis. The data quality needs to be of such high standard that the readmissions can be accurately related to an index admission and that patient-specific information is available to adjust for patient factors. Only under these circumstances does the readmission rate provide useful information about the quality of care a hospital provides.

Funding

The research was funded by a VWS grant and as part of the European Commission’s Seventh Framework Programme (FP7/2007–2013) under grant agreement 223248.

Conflicts of interest: None declared.

References

Effectiveness of interventions to promote healthy weight in general populations of children and adults: a meta-analysis

Helene Luckner1,2, John R. Moss1, Christian A. Gericke3

1 School of Population Health and Clinical Practice, University of Adelaide, Australia
2 Department of Health Care Management, Berlin University of Technology, Germany
3 Peninsula CLAHRC, National Institute for Health Research, Peninsula Medical School, Universities of Exeter & Plymouth, UK

Correspondence: Christian A. Gericke, Deputy Director, Peninsula CLAHRC, National Institute for Health Research, Peninsula Medical School, Universities of Exeter & Plymouth, B432 Portland Square, Plymouth PL4 8AA, UK, tel: +44 1752 586 811 (ext. 757), fax: +44 1752 586788, e-mail: christian.gericke@pcmrd.ac.uk

Background: Responding to the obesity epidemic requires robust evidence to help prioritize the allocation of scarce resources to preventive interventions. The aim of this study was to evaluate interventions that promote healthy weight [defined as reduction in body mass index (BMI) or percentage body fat] in general populations (unselected by weight) using a comprehensive meta-analysis. Interventions with both single and multiple components were considered.

Methods: Studies were first identified through well-conducted systematic reviews complemented by a search for single studies in five large medical databases up to 6 November 2008. Sixty-eight controlled studies were included. For each intervention type and age group, all relevant studies were pooled in a random effects meta-analysis. Results: In children, the highest reductions in mean BMI were achieved through promoting reduced television viewing [−0.27 kg/m² (95% CI −0.4 to −0.13 kg/m²)]. Programmes combining physical activity, specifically themed or general health education and nutrition achieved a lower reduction [−0.1 kg/m²]. Other interventions had high heterogeneity or showed no statistically significant reduction in outcomes. In adults, single component interventions were found to reduce both outcome measures. Their mean percentage body fat was reduced through education by −1.22% (95% CI −1.92 to −0.52).

Conclusion: The evidence for the effectiveness of promoting healthy weight in general populations is limited, though multi-component interventions in schools and encouraging reduced children’s television viewing are promising strategies. Improving the reporting of outcomes is vital, as imputation of inadequately reported measures may have contributed to the observed heterogeneity. Longer follow-up is essential for understanding policy relevance.